PolyPublie
Polytechnique Montréal

POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

Titre: Title:	Predicting Time to Next Better Offer for a Kidney Transplant Candidate Based on Marked Point Process
Auteur: Author:	Hugo-Maxime Tocco
Date:	2020
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Tocco, H.-M. (2020). Predicting Time to Next Better Offer for a Kidney Transplant Candidate Based on Marked Point Process [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/5526/

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/5526/
Directeurs de recherche: Advisors:	Andrea Lodi, & Jonathan Jalbert
Programme: Program:	Maîtrise recherche en mathématiques appliquées

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal

https://publications.polymtl.ca

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Predicting Time to Next Better Offer for a Kidney Transplant Candidate Based on Marked Point Process

HUGO-MAXIME TOCCO
Département de mathématiques et de génie industriel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Mathématiques appliquées

Décembre 2020

© Hugo-Maxime Tocco, 2020.

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Ce mémoire intitulé :

Predicting Time to Next Better Offer for a Kidney Transplant Candidate Based on Marked Point Process

présenté par **Hugo-Maxime TOCCO**
en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Louis-Martin ROUSSEAU, président
Andrea LODI, membre et directeur de recherche
Jonathan JALBERT, membre et codirecteur de recherche
Aurélie LABBE, membre

ACKNOWLEDGEMENTS

First of all, I want to thank my research directors Andrea Lodi and Jonathan Jalbert. They both supervised me all along my thesis, and they both taught me as much scientific knowledge as perseverance and work ethic. They always managed to give me relevant guidance throughout my work, and helped me to achieve my master thesis.

I want to thank all the CERC team, with a special mention to Mehdi Taobane, the lab manager who always efficiently managed the calendar, always made some time when needed, followed closely the progression of the master, and also helped me to progress during my master.

I thank all my lab colleagues and friends with who I shared precious moments, inside André-Aseinstadt building as well as outside in everyday life. I am sure I will miss the rich international, multilingual, serious and fun at the same time stimulating environment that we all managed together to develop in our lab.

Finally, I thank my family and my friends who always have been there for me and gave me support, even with an ocean distance and a six hours time difference for some of them.

RÉSUMÉ

Les patients souffrant d'insuffisance rénale terminale ont besoin de recourir à une thérapie de substitution rénale pour pallier au dysfonctionnement de leur reins, en effet, ceux-ci n'assurent plus correctement leur fonction vitale de purification du sang. En comparaison avec l'option de dialyse à laquelle un patient a recours par défaut, la transplantation rénale est une solution plus viable à long terme, tant en termes de contraintes quotidiennes qu'en termes de chances de survie. Toutefois, un patient recevant une proposition de greffe de rein est confronté à un dilemme, dans lequel il doit choisir entre accepter l'offre, ou bien la refuser et patienter jusqu'à une prochaine offre tout en restant sous dialyse. Au Québec, la décision résulte d'un commun accord entre le patient et son néphrologue au terme de la Prise de Décision Participative (PDP).

Le nombre de patients atteints d'insuffisance rénale terminale au Canada était de presque 42 000 en 2013, et a augmenté de 35% entre 2009 et 2018 ce qui soulève une réelle question médicale, mais aussi économique étant donné que la dialyse est une solution bien plus dispendieuse que la transplantation rénale. Par ailleurs, la donation d'organe après décès est la pratique la plus répandue, en comparaison à la donation du vivant, et cette tendance est en hausse.

L'attribution de reins en provenance de donneurs décédés est généralement assurée par des institutions gouvernementales de santé publique. Transplant Quebec (TQ) est l'institution provinciale qui en est reponsable au Québec. Notre travail se concentre sur le système d'attribution général de TQ (excluant les listes d'attente prioritaires) qui concerne la majorité des patients. Une fois qu'un patient reçoit une offre de greffe, il peut soit l'accepter, soit la refuser au terme de la PDP. D'un côté, un rein de moindre qualité (qui dépend de l'âge du donneur, des antécédents médicaux, etc.) pourrait motiver un refus de l'offre et l'attente d'une prochaine proposition. De l'autre, les temps d'attente peuvent être longs, et les perspectives d'une prochaine offre, et en particulier d'une meilleure offre, réduites.

Par conséquent, la décision peut être difficile à prendre, d'autant plus que le patient manque d'informations compréhensibles pour la documenter. En effet, les outils d'aide à la décision actuels requièrent un certain savoir médical pour être correctement utilisés. De plus, ces outils ne fournissent pas d'informations personnalisées pour un patient en particulier, mais plutôt des résultats généraux à l'échelle d'un système d'attribution entier. Par ailleurs, les travaux actuels visent souvent à répondre aux enjeux de l'allocateur d'organes, comme l'optimisation des coûts et des ressources, plutôt qu'à faire des enjeux d'un patient pris individuellement une

priorité. Ainsi, le développement d'outils appropriés d'aide à la décision pourrait grandement aider les patients lors de la PDP.

Nous considérons la situation dans laquelle un patient reçoit une offre de greffe. Au cours de sa réflexion lors de la PDP, il peut procéder à une simple disjonction de cas. D'une part, qu'arrive-t-il s'il accepte l'offre ? D'autre part, qu'arrive-t-il s'il la refuse ? Notre objectif est de fournir des informations pertinentes et compréhensibles afin d'aider le patient à se représenter la situation dans laquelle il refuserait l'offre.

Premièrement, l'outil d'aide à la décision développé doit être en mesure de donner des prédictions personnalisées de temps d'attentes pour le patient, qui seront facilement interprétables. Le premier temps à prédire est le temps d'attente avant qu'une prochaine offre soit faite au patient. Ce temps est appelé Temps avant Prochaine Offre (TPO). Aussi, l'outil doit fournir le temps d'attente avant qu'un rein de meilleure qualité que le rein actuel soit proposé au patient. On appelle ce temps Temps avant Prochaine Meilleure Offre (TPMO). Le TPMO est plus utile au patient que le TPO étant donné que la principale motivation pour décliner l'offre actuelle est d'en espérer une meilleure prochainement.

Deuxièmement, nous proposons de modéliser l'arrivée des reins dans le système d'attribution par un Processus Ponctuel Marqué (PPM), que nous caractérisons à l'aide de sa Fonction d'Intensité Conditionnelle (FIC). Ce cadre mathématique permet de modéliser à la fois le temps d'occurence et la marque (qualité du rein) des événements d'intérêt (arrivée d'un rein). Dans notre solution, l'hypothèse de stationnarité est formulée concernant les temps d'arrivée des reins ainsi que leur qualité, afin de concevoir un modèle simple et interprétable. La fonction d'intensité du processus d'arrivée des reins est supposée constante (comme dans un processus de Poisson homogène) et la distribution de leur qualité est supposée stationnaire. Les paramètres du modèle sont estimés en utilisant une reconstitution de l'historique des offres passées faites au patient avant l'offre initiale, ce qui permet une grande personnalisation des prédictions du modèle. Le PPM fournit des prédictions des temps d'attentes à l'échelle d'un patient, notamment une borne supérieure du temps d'attente au bout duquel le patient aura reçu une offre avec 95% de certitude. La modélisation de la qualité dans le PPM est utilisée pour procéder à une filtration du processus, afin de considérer uniquement les offres de meilleure qualité que l'offre actuelle. De cette manière, nous avons accès aux mêmes prédictions pour le TPMO que pour le TPO.

Enfin, la solution développée est testée et ajustée sur des données réelles provenant du système d'attribution de TQ. Les résultats montrent que notre modèle donne de meilleurs résultats en termes de pouvoir de prédiction que des solutions concurrentes ayant les mêmes objectifs, à la fois pour le TPO et pour le TPMO. Par exemple, le Pourcentage d'Erreur Absolu

Moyen (PEAM) sur l'ensemble test avec notre modèle est inférieur de 25.5% à celui obtenu avec un processus de Poisson non marqué pour les prédictions du TPO, et de 22.3% pour les prédictions du TPMO. De la même manière, nous observons une amélioration du PEAM de respectivement 37.2% et 70.7% avec notre modèle par rapport à une méthode de prédiction basique (qui consiste à donner un temps d'attente moyen calculé sur l'ensemble des patients du jeu de données). En outre, notre méthode a montré la cohérence de ses performances d'un ensemble de données à un autre, entre l'ensemble de validation et l'ensemble de test.

Au delà des résultats obtenus, l'approche du PPM qui a été développée et appliquée pour répondre à notre problématique présente une grande adaptabilité ainsi que des possiblités d'amélioration. Un travail futur pourrait par exemple considérer un modèle de PPM plus avancé qui serait défini par une FIC constante par morceaux.

ABSTRACT

End-Stage Kidney Disease (ESKD) is a medical condition in which kidneys can no longer ensure their vital blood purification function, forcing patients to resort to renal replacement therapy. Compared to the default dialysis option, kidney transplantation is a more viable long-term solution both in terms of daily constraints and survival outcomes. Nonetheless, a patient receiving a kidney transplant offer is facing a dilemma, where he needs to decide to accept it or not with the help of his nephrologist during the Shared Decision Making (SDM).

The number of ESKD patients across Canada was nearly 42,000 by the end of 2013, and has increased by 35% between 2009 and 2018, accounting for a real medical and economical issue as the dialysis treatment is far more expensive than the transplantation. In parallel, Deceased Donor Kidney Transplant (DDKT) is the most common kidney transplant option, compared to the Living Donor Kidney Transplant (LDKT) alternative, and the tendency is amplifying.

The deceased donor kidneys distribution is generally ensured by governmental health care institutions. Transplant Quebec (TQ) is the provincial institution responsible for this attribution to ESKD patients on waiting lists in Quebec. This work focuses on the general attribution list that includes the majority of patients. Once a patient is proposed a kidney, he can either accept it or refuse it at the end of the SDM. On the one hand, a low-quality kidney (age of donor, medical conditions, etc.) could motivate the patient to refuse an offer and wait for a better one. On the other hand, waiting times can be long and the perspectives of a future offer can be poor, especially a better one.

The decision can be difficult to make and patients lack understandable information to inform their reflection. Indeed, existing decision-aid tools involve medical knowledge and health education to be used correctly, which poses the problem of interpretation for the patient. Furthermore, they do not provide personalized information but give results that refer to an entire attribution system instead. Also, the existing works often stand for the allocator's stakes, like minimizing costs or optimizing resources, rather than making the individual patient the main priority. Thus, the development of appropriate decision-aid tools could help patients during SDM.

We consider the situation where a given patient on the general waiting list is being proposed an initial kidney offer. To help the patient during the SDM, a simple case disjunction can be done. On the one hand, what happens if he accepts the kidney? On the other hand, what happens if he declines it? We aim to provide relevant and helpful elements to help the

patient picture the situation if he was to say no to the kidney proposal.

First, the objective of the designed decision-aid tool is to provide highly understandable and personalized expected waiting times to the patient. The tool has to predict the time before another kidney offer is proposed to the patient. This waiting time is referred to as Time before Next Offer (TNO). Furthermore, the tool has to give the time the patient will have to wait before an offer with a better kidney quality than the current one. This waiting time is referred to as Time before Next Better Offer (TNBO). TNBO is more relevant than TNO since the main reason that motivates declining the current offer is the expectation of a better one.

Second, we propose the Marked Point Process (MPP) approach to model the arrival of kidneys in the attribution system, and we characterize the process by defining its Conditional Intensity Function (CIF). This stochastic framework enables to model both times and marks (kidney qualities) of events (arrivals of kidneys). In this work, the stationarity assumption is made about the times of arrivals of kidneys and their qualities, in order to devise a simple model. The rate of kidney arrivals is constant (like in a homogeneous Poisson process) and the associated quality distribution is stationary. Parameters of the model are inferred using a reconstruction of the history of the patient's past offers before the initial offer, which enables highly personalized predictions. The MPP model gives access to a one-patient scaled expectation of TNO, including an upper bound of the time by which an offer would have occurred with a 95% confidence (the confidence level is customizable). The quality aspect of the MPP is used to carry out a process thinning, which only considers kidney arrivals with a better quality than the current one. This way, we can access the same type of predictions for TNBO and for TNO.

Finally, the developed solution is tuned and tested on real attribution system records furnished by TQ. The results demonstrate that the proposed model outperforms competing solutions with same objectives in terms of predictive power, both on TNO and TNBO issues. Based on the Mean Absolute Percentage Error (MAPE) score on the test set, our MPP model outperforms an unmarked Poisson process method by 25.5% for TNO problem and by 22.3% for TNBO problem. Moreover, we respectively observe a MAPE decrease of 37.2% and 70.7% when using our model instead of a baseline policy (a basic statistic that averages waiting times of all patients in the dataset). Additionally, our method proves to be consistent in terms of performances between validation set and test set.

Beyond the computational results, the MPP framework that has been set to address this problem provides high flexibility and possibilities of model enhancement. A piecewise-constant defined CIF could be considered to achieve a more advanced MPP model in future research.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

RÉSUMÉ iv

ABSTRACT vii

TABLE OF CONTENTS ix

LIST OF TABLES xii

LIST OF FIGURES xiii

LIST OF SYMBOLS AND ACRONYMS xv

LIST OF APPENDICES xvi

CHAPTER 1 INTRODUCTION 1
- 1.1 Renal replacement therapy 1
- 1.2 Kidney offer and patient's decision 2
- 1.3 Research objectives and methodology 3
- 1.4 Thesis outline 4

CHAPTER 2 LITERATURE REVIEW 5
- 2.1 Decision-Aid 5
- 2.2 Marked Point Process in health care systems 7
- 2.3 Predicting next kidney offer for a kidney transplant candidate declining current one 8

CHAPTER 3 MATHEMATICAL GROUNDS 10
- 3.1 Poisson process 10
- 3.2 Marked Point Process 13
  - 3.2.1 The MPP and the History 13
  - 3.2.2 The Conditional Intensity Function approach 14
  - 3.2.3 Inference 16
  - 3.2.4 Predictions 18
  - 3.2.5 MPP methodology recap 22

CHAPTER 4 GOING BEYOND [1] 23
4.1 Original Research 23
4.1.1 Predict the Time before Next Offer 23
4.1.2 Predicting the Time before Next Better Offer 26
4.1.3 Remarks 27
4.1.4 Gap left to fill 27
4.2 Predicting waiting times based on Marked Point Process 28
4.2.1 Marked Point Process applied to kidney offer 28
4.2.2 Addressing limitations 31
4.2.3 Computational results 32

CHAPTER 5 MPP MODELING FOR TIME BEFORE NEXT BETTER OFFER . 34
5.1 Marked point process based kidney offer problem 34
5.1.1 Problem components modeling and notations 34
5.1.2 Building the history of the process 37
5.1.3 Ground Intensity Function: the rate of the process 38
5.1.4 Mark Conditional Density Function: the quality of kidney 39
5.1.5 Inference 45
5.2 Marked Point Process predictions 53
5.2.1 Exact prediction: Time before Next Offer 53
5.2.2 Exact prediction: Time before Next Better Offer 54
5.2.3 Simulation 58

CHAPTER 6 EXPERIMENTS AND RESULTS 66
6.1 Data from attribution system records 66
6.1.1 The data 66
6.1.2 One first experiment to handle data 69
6.1.3 Method adjustment to tackle hard to fit patients challenge 72
6.2 Verification Methodology 74
6.2.1 Notations 74
6.2.2 Data particularity 74
6.2.3 Validation tools 79
6.2.4 Baseline Policy 80
6.2.5 Data partition 81
6.3 Define and estimate the GIF: the rate λ 82
6.3.1 Estimating the process rate with $\widehat{\lambda}_1$ 82
6.3.2 Addressing limitations of $\widehat{\lambda}_1$ with $\widehat{\lambda}_2$ 86

6.3.3 Estimating the process rate with $\widehat{\lambda}_2$ 92
6.4 Define and estimate the mark distribution: the quality Q 100
6.4.1 Selection procedure of the quality distribution 100
6.4.2 Testing Lognormal distribution 102
6.4.3 Method adjustment to tackle hard to fit patients challenge 104
6.5 Combining rate and quality: the answer to Time before Next Better Offer . . 106
6.5.1 Result on original dataset 106
6.5.2 Rebuilding dataset 107
6.5.3 Q-Q plots an P-P plots of normalized process 108
6.5.4 Performances of MPP Method - Comparison 111
6.6 Results on the test set 112
6.6.1 Results for TNO 112
6.6.2 Results for TNBO 115

CHAPTER 7 CONCLUSION AND RECOMMENDATIONS 117
7.1 Summary of work 117
7.2 Limitations 118
7.3 Future research directions 119

REFERENCES 121

APPENDICES 127

LIST OF TABLES

Table 4.1	Methods performances comparison (on test set)	33
Table 5.1	Thinning process: example for one patient	58
Table 5.2	Exact predictions and simulation predictions	64
Table 6.1	Features in the donor file	67
Table 6.2	Features in the patient file	68
Table 6.3	Representative 20 patients sample $\tilde{\mathcal{S}}_{20}$	71
Table 6.4	Three types of challenge in quality fit	71
Table 6.5	Dataset partition	81
Table 6.6	Methods performances comparison	85
Table 6.7	Solving clustering and discretization with Method 2: estimators and predictions	89
Table 6.8	Removed outliers proportions - Validation set	94
Table 6.9	Methods performances comparison	99
Table 6.10	Best density fit selected over BIC	102
Table 6.11	Best density fit selected over BIC - Adjustment	104
Table 6.12	Methods performances comparison - on original dataset	107
Table 6.13	Methods performances comparison for TNBO	111
Table 6.14	TNO - Methods performances comparison	112
Table 6.15	Methods performances comparison for TNBO	115

LIST OF FIGURES

Figure 3.1 MPP and History: earthquakes (graph from [2]) . . . 14
Figure 3.2 Steps of MPP methodology using CIF . . . 22
Figure 4.1 History Building (figures from [1]) . . . 24
Figure 4.2 Thin the history . . . 27
Figure 4.3 Build the new targets . . . 29
Figure 4.4 Methodology steps . . . 31
Figure 5.1 PDFs of Gamma and Weibull for different set of parameters . . . 42
Figure 5.2 Comparison of Gamma and Weibull PDFs characteristics . . . 43
Figure 5.3 PDFs of Lognormal for different set of parameters. . . . 44
Figure 5.4 A patient history $\mathcal{H}$ . . . 48
Figure 5.5 Quality distribution and thinning factor . . . 58
Figure 5.6 Waiting time for original and thinned Poisson process . . . 59
Figure 5.7 Time transformation of the inverse algorithm . . . 64
Figure 5.8 Example of a simulated history $\mathcal{H}^{s,k}$ . . . 65
Figure 5.9 Histogram of the simulated sample $(Z_k)_k$ . . . 65
Figure 6.1 Steps to create the synthetic dataset . . . 69
Figure 6.2 Histogram of N_{offer} and $\tilde{\mathcal{S}}_{20}$ patients selection . . . 70
Figure 6.3 Zoomed histogram of N_{offer} and $\tilde{\mathcal{S}}_{20}$ patients selection . . . 70
Figure 6.4 Quality fit on $\tilde{\mathcal{S}}_{20}$ patients for Weibull distribution . . . 72
Figure 6.5 Challenging quality fit for patient with low N_{offer} . . . 73
Figure 6.6 Histogram of observed sample . . . 75
Figure 6.7 True PDF of the waiting time T . . . 76
Figure 6.8 Two examples of possible fitted distributions . . . 77
Figure 6.9 Uncensored and censored observations (figures from [1]) . . . 78
Figure 6.10 Normalized Q-Q plot: no outliers removed . . . 83
Figure 6.11 Normalized Q-Q plot: 5 outliers removed . . . 84
Figure 6.12 Example 1: Low number of eligible offers . . . 87
Figure 6.13 Example 2: Recent activity . . . 88
Figure 6.14 Solving clustering and discretization with Method 2: Examples of histories . . . 90
Figure 6.15 Correlation between $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ . . . 91
Figure 6.16 Normalized Q-Q plot for Method 2 . . . 93
Figure 6.17 P-P Plot for different number of removed outliers . . . 96

Figure 6.18 P-P Plot for different number of removed outliers 97
Figure 6.19 Exponential Fit on Normalized data 98
Figure 6.20 Weibull vs Gamma selection over BIC criteria, in terms of N_{offer} . . . 101
Figure 6.21 Weibull/Gamma/Lognormal: Best density fit selected over BIC . . . 103
Figure 6.22 Weibull/Gamma/Lognormal: Best density fit selected over BIC - Adjustment 104
Figure 6.23 Data proportions in terms of their nature 106
Figure 6.24 Build the new targets 108
Figure 6.25 Normalized Q-Q plot for TNBO 109
Figure 6.26 P-P Plot for different number of removed outliers 110
Figure 6.27 Q-Q and P-P Plot for different number of removed outliers 113
Figure 6.28 Exponential Fit on Normalized data - Test Set 114
Figure 6.29 Q-Q and P-P Plot for different number of removed outliers 116

LIST OF SYMBOLS AND ACRONYMS

AIC Akaike Information Criterion . . . 100
BIC Bayesian Information Criterion . . . 100
CDF Cumulative Distribution Function . . . 12
CoDF Conditional Density Function . . . 15
C-index Concordance Index . . . 32
CIF Conditional Intensity Function . . . 7
DDKT Deceased Donor Kidney Transplant . . . 1
ESKD End-Stage Kidney Disease . . . 1
GIF Ground Intensity Function . . . 15
ICU Intensive Care Unit . . . 7
iGIF integrated Ground Intensity Function . . . 17
KDRI Kidney Donor Risk Index . . . 2
LDKT Living Donor Kidney Transplant . . . 1
MAPE Mean Absolute Percentage Error . . . 32
MLE Maximum Likelihood Estimation . . . 17
MPP Marked Point Process . . . 3
MSE Mean Square Error . . . 49
PDF Probability Density Function . . . 21
P-P plot Probability-Probability plot . . . 92
Q-Q plot Quantile-Quantile plot . . . 82
SDM Shared Decision Making . . . 2
TNBO Time before Next Better Offer . . . 3
TNO Time before Next Offer . . . 3
TQ Transplant Quebec . . . 2

LIST OF APPENDICES

Appendix A MPP model proposal 127

CHAPTER 1 INTRODUCTION

1.1 Renal replacement therapy

Human kidneys are vital organs that purify blood. In the number of two, several different reasons can cause one or both to start being dysfunctional. When their blood purification function ceases to be efficient enough, patients need to resort to a renal replacement therapy.

The Canadian Blood Services refer to this stage of disease as End-Stage Kidney Disease (ESKD), and define it as "a condition in which the kidneys are permanently impaired and can no longer function normally to maintain life" [3]. According to [3], nearly 42,000 Canadians were affected by this condition at the end of 2013. Moreover, according to a more recent study from the Canadian Organ Replacement Register, the number of Canadians living with ESKD has increased by 35% between 2009 and 2018 [4]. The possible treatment of this condition is renal replacement therapy, which is divided into transplantation and dialysis.

Dialysis is not a long-term solution, since it involves permanent, regular and long sessions to purify the patient's blood with the help of a dialysis machine. This solution poses heavy constraints on the daily life of the patient. Furthermore, the other renal replacement therapy option, transplantation, generally has better patient outcomes than dialysis. Indeed, statistics presented in the 2019 report [4] claim that 50.6% of patients on peritoneal dialysis survived at least 5 years, against 81.3% for patients who received a kidney from a deceased donor. The second option of kidney transplantation is the field of interest of this research.

Kidney transplants can be of two different types. On the one hand, the kidney donor can be alive, so-called Living Donor Kidney Transplant (LDKT). This type of transplant mostly comes from patient's relatives or friends willing to give their organ. Nonetheless, donor and recipient may be not compatible. In this case, exchange programs give the possibility to two willing donors to exchange their recipient, so that two compatible pairs can be created. The optimization of exchange programs has been addressed in several works, like [5] or [6].

On the other hand, the kidney donor can be deceased, so-called Deceased Donor Kidney Transplant (DDKT). In Canada, deceased donation is more frequent than alive donation, as it accounts for 60% of total transplants [3]. Additionally, this tendency is amplifying as a 44% increase in the number of deceased donors from 2005 to 2014 has been reported by [7], while the number of living donors has only raised by 10%. DDKT can occur in case of cardiac death of the donor, which is Donation after Cardiac Death, or in case of neurological death, which is Donation after Neurological Death. In both cases, the kidney is retrieved and enters

a distribution system that aims to propose it to ESKD patients waiting for a transplant.

Kidney distribution systems are generally part of a governmental health care system, often at a regional scale. Policies that define the attribution process differ from one system to another. Also, medical considerations have to be taken into account in the attribution process (donor-recipient compatibility, blood-type, etc.). This work will focus on the distribution system in Canada, and especially in Quebec. Transplant Quebec (TQ) is the provincial health care organism responsible for the attribution of organs from deceased donors. The whole attribution system is composed of several priority layers in terms of patient condition (renal emergency, pediatric patients, etc.). Patients on the general attribution list have the lowest priority, and can wait for a long time before receiving an offer. The attribution system for the general waiting list is based on a score ranking function that attributes a score for each patient and ranks them in order of their estimated priority. The scoring function used by TQ takes into account two criteria: utility and justice. The simultaneous objectives are to be fair (minimizing waiting time) and efficient (maximizing survival).

1.2 Kidney offer and patient's decision

When the attribution system finally makes a kidney proposal to a patient, the immediate question for the patient is either to accept it or decline it. This decision can pose a dilemma. Indeed, all kidneys are not of the same quality, and the acceptance of a quick offer with a low-quality kidney means increasing risks that complications occur after the graft. Clinical risks indices, like Kidney Donor Risk Index (KDRI) [8], aim to provide an estimation of a kidney quality based on several characteristics of both donor and recipient. In some cases, waiting for a better kidney coming from a younger and healthier donor could be beneficial for the patient. This is the reason why a patient on the waiting list is always offered the possibility to refuse a kidney proposal. In fact, the decision to accept or refuse the offer is the result of a systematic Shared Decision Making (SDM), involving both the patient and his nephrologist. During this process, the patient is expected to reflect on the opportunities and possible outcomes, and the doctor is there to help him but not to impose a choice, he has to provide details and knowledge to inform the patient's reflection. SDM involves a greater participation and responsibility on the patient side, which improves both outcomes of graft and satisfaction of the patient according to [9, 10].

However, in order to decide, patients lack understandable information. Indeed, existing decision-aid tools involve medical knowledge and health education to be interpreted correctly. Furthermore, they do not provide personalized information but give results that are scaled to an entire attribution system instead. Also, the existing works often stand for the allocator's

stakes, like minimizing costs or optimizing resources, rather than making the individual patient the main priority.

To help the patient during the SDM, a simple case disjunction can be done. On the one hand, what happens if he accepts the kidney? On the other hand, what happens if he declines it? In line with [1], this thesis aims to provide relevant elements to help the patient picture the situation if he was to say no to the kidney proposal.

1.3 Research objectives and methodology

We consider the situation where a given patient on the general scoring waiting list is being proposed an initial kidney offer. We want to inform him about his perspectives if he was to decline it.

The first objective is to predict the time before another kidney offer is proposed to the patient. This waiting time is referred to as Time before Next Offer (TNO).

Assume a measure that evaluates the quality of a kidney for a given receiving patient (the KDRI quality index will be used during the thesis). The higher the quality is, the higher survival perspectives are for the recipient.

The second objective is to predict the time the patient will have to wait before an offer with a better kidney quality than the current one. This waiting time is referred to as Time before Next Better Offer (TNBO). This waiting time is more relevant than TNO since the main reason that motivates declining the current offer is the expectation of a better one.

The development of a tool capable of predicting those waiting times implies mathematical modeling of the times of kidney arrivals in the attribution system, as well as their qualities. We apply the Marked Point Process (MPP) approach for modeling, since MPPs are stochastic processes providing a convenient framework that enables to model both times and marks (kidney qualities) of events (arrivals of kidneys). We make the stationarity assumption about the times of arrival of kidneys and their qualities in order to devise a simple model. It means that the rate of kidney arrivals is constant (as for a homogeneous Poisson process) and the associated quality distribution is stationary. We infer parameters of the model on a reconstruction of the history of patient's past offers before the initial offer. Finally, the developed model is adjusted, validated and tested on real attribution system records provided by TQ.

It is worth precising that even if the main objective of the tool is to inform a patient when he receives a kidney offer, the solution can be used at any other time as soon as the patient is on the waiting list. Hence, the decision-aid tool can also be used to inform patient about

possible coming offers when he enlists, or to re evaluate his perspectives in between offers.

1.4 Thesis outline

This thesis starts with a literature review about kidney transplantation and decision-aid in Chapter 2. Examples in the literature using MPP models in health care will also be discussed. Then, Chapter 3 introduces the mathematical grounds required to develop our solution. Next, Chapter 4 is dedicated to explaining the links between this thesis and the work in [1]. Chapter 5 presents the mathematical modeling of the problem we solve, and how we apply the MPP approach. Chapter 6 describes experiments we carried out to build the model according to the theory of the previous chapter, and how we validated it. It also includes the final results on test sets. Finally, we conclude the thesis in Chapter 7 with a summary of the work, and by discussing its limitations and some future research directions.

CHAPTER 2 LITERATURE REVIEW

Kidney transplantation is an important part of renal replacement therapy that raises a lot of issues and presents various research possibilities. The work conducted in this thesis is part of a larger research that aims at addressing the kidney transplant problem in a larger extent by providing useful instruments to inform the SDM. The research has been split into two main questions that a patient and his nephrologist would have to answer in order to make a decision. On the one hand, there is the *Yes question* which mainly deals with survival analysis, since the objective is to predict what will happen if the patient accepts the graft. On the other hand, the *No question* consists in predicting what happens if the patient declines the graft. Then, we refer to the project that embraces both issues as the *Yes-No question*. This thesis focuses on the second question.

Many works addressing the *Yes question* can be cited, from [8] which is widely used clinically and designs a risk index based on a traditional Cox regression, to the research conducted in [11, 12] that uses a more cutting-edge approach of artificial neural networks. Despite survival analysis is not the subject of the current thesis, the reader can refer to the literature review conducted in [1] to learn more about this first issue. It might be relevant to understand more globally the stakes of the *Yes-No question*.

Regarding the *No question*, several approaches can be considered to inform the SDM. The solution we develop is a decision-aid tool that aims to provide relevant elements to help the patient picture the situation if he was to say no to the kidney proposal. The intent we pursue is to inform him about his future opportunities of graft proposal, by providing in particular the expected time to a next offer, or to a next better offer.

Answering the *No question* is basically a decision-making issue. Decision-aid applied to clinical and health care system is an abundant topic that has been studied many times in the literature.

2.1 Decision-Aid

Decision-aid is a major issue in health care, and several issues concerning organ transplant are the subject of research works in the literature. In different research fields, patients are included at different degrees in the organ attribution process, and the stakes and objectives are not the same.

A first field considers the problem from the allocator's side. The goal is to properly optimize

the attribution system, so that the patients have the best offers. Many different studies use Markov decision processes for liver transplantation, like [13], [14] [15] and [16]. In [16] for instance, partially observable Markov decision processes are employed to predict the patient's decision. In this type of problem, the strategy adopted is often to predict the behavior of patients in order to find the best allocation policy. In this thesis, we adopt a completely different perspective, as we predict the waiting list behavior to inform the patient about his future perspectives. However, designing a device capable of correctly predicting patients' preferences is a mean to learn what is important to patients and help to develop a decision-aid tool. Moreover, finding best allocation policies is not always equivalent to reducing costs and optimizing efficiency. The work [17] develops a discrete model of the kidney allocation process in the U.S. to prove the need of modifying the policy. Indeed, it demonstrates that the initial policy leads to risks of increasing waiting lists, resulting in future incapacity to treat patients.

A second field in decision-aid research aims at forecasting the effect of a policy modification on the size of the waiting list. The article [18] addresses the problem of predicting the behavior of the renal transplant waiting list in the Valencian community[1], with the help of a discrete event simulation model that quantifies fluctuations of the waiting list. The objective here is to evaluate the consequences of the policy modification on a current system as a whole, but predicting an individual patient's waiting time is not considered. The point of view adopted is still on the policy-maker side.

A third field of research in decision-aid aims at optimizing the decisions of the patients. In the existing literature, the so-called secretary problem has been adapted to find the solution to what type of kidney a patient should accept after a given waiting time. This problem has been originally presented in [19]. It involves a series of candidates that arrive at random times to get different positions in a firm, and different rewards are given to candidates in terms of their position. Improvements of this problem have been achieved by [20] and [21]. The former paper finds the optimal strategy for the secretary problem using dynamic programming, and discusses the modeling of the horizons with uniform and Erlang distributions. The latter paper adapts the secretary problem to the kidney transplant to figure out how demanding regarding the kidney quality a patient should be as time goes. According to the results, the longer his waiting time is, the less a patient should be demanding. However, this work only gives average estimations that could be useful for an entire allocation system, instead of individual advice.

Finally, the fourth field of research aims at forecasting an informed decision of the patient.

[1] Autonomous region of Spain with its own health care system.

The work of this thesis belongs to this field, with the objective of providing individual decision-aid tool to a patient. The authors of [22] design an algorithm that provides useful information that can help a patient in his decision process, including the probability of a 3-year survival if he accepts the current kidney graft, and the probability of a 3-year survival if he declines it. One of the interesting aspects of the solution is that, to compute final outcome probabilities, it takes into account the potential happening of events while the patient is in the waiting list (death, removal of waiting list, etc.). The decision-aid tool tends to suggest the acceptance of relatively high-risk kidneys for some patients, which could be seen as an interesting unusual advice. Nonetheless, the information provided to the patient in [22] still poses the question of interpretation of probabilities for a patient. Also, even if kidney qualities are taken into account in the model, this aspect does not appear explicitly in the information provided to the patient.

In conclusion, even if the decision-aid topic applied to organ transplant is abundant in the literature, only a few works take the patient's side and provide an individual personalized instrument to help the patient before other stakeholders.

2.2 Marked Point Process in health care systems

The solution we propose in this thesis is based on a MPP modeling. Indeed, this stochastic process characterized by the Conditional Intensity Function (CIF) approach turns out to provide a framework with powerful modeling capabilities. Plenty of applications of this methodology can be found in the literature. We focus on research studies applied to the health care system.

The article [23] proposes a medical tool to help assess the severity of illness of patients in Intensive Care Unit (ICU). The study deals with the electronic health records (physiological signals, lab test results, procedural events, etc.) as a temporal stream of events with a MPP model to capture temporal dependencies among all these different types of data. The CIF used for the MPP model is a piecewise-constant conditional intensity, which means the rate of the process (which influences events arrival frequency) is piecewise constant. The developed model improves the hospitality mortality prediction over traditional ICU scoring systems.

Still with the objective of exploiting clinical data to reveal their patterns and behaviors, a functional MPP model for lupus data is proposed in [24]. The work incorporates the use of functional data analysis in a joint estimation of both the intensity function of the MPP and the intensity of the marks. The model is applied to data from 22 lupus patients consisting of times of flares in symptom severity combined with a quantitative assessment

of the severity. The methodology helps demonstrate that a rapid decrease in drug dose is significantly associated with a decrease in flare frequency. In this example, the MPP approach reveals to doctors a behavior of the disease, which could have been difficult to detect from a human perspective.

An advanced MPP model is proposed in [25] to tackle limitations of simpler point process models like Cox or Hawkes processes, which are commonly used for risk prediction in health care. The article introduces wavelet reconstruction networks, which is a multivariate point process with a sparse wavelet reconstruction kernel to model rate functions from marked, timestamped data. The method shows the ability to capture quasi-periodic events that could be used to measure adherence and forecast risk of complications for diabetes patients. Moreover, this advanced MPP model outperforms the competing aforementioned classical models. This work demonstrates that modifying an existing MPP model to a more advanced one can lead to a solution enhancement.

Another example of MPP approach applied to behavior modeling in health area is given in [26]. The method aims to relate neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. To take into account the dependencies in the model, the logarithm of the CIF is defined as a linear combination of functions that depend of the covariates. Tests have proved that the developed model could capture the simultaneous effects of multiple covariates, as well as assess their relative importance. We can see from this example that the flexibility of the MPP framework allows to formulate elaborate models that are able to take into account several covariates to predict the behavior of one dependent variable.

Consequently, the MPP approach is suitable to many problems applied in the health care system, but it has not been applied to the kidney transplant problem yet. In this thesis, we use this methodology to model the arrival of kidneys in the attribution system of TQ. More precisely, we provide a one-patient scaled personalized model of it. For a given patient, we learn the MPP model that describes the arrival of kidney offers that are proposed to him, including the modeling of kidney qualities (which is achieved by using the marked aspect of the MPP).

2.3 Predicting next kidney offer for a kidney transplant candidate declining current one

Finally, this thesis is mainly built on the groundwork set by the work of Weller [1], entitled *'Predicting next kidney offer for a kidney transplant candidate declining current one'*, and

that also aims to answer the *No question*.

Our objective is to pursue the research by focusing on fixing and improving some limitations of the solution in [1]. As a consequence, Chapter 4 is dedicated to explaining the links between both works, summarizing both solutions, and explaining the contribution of the current work. Chapter 4 is an extended literature review of thesis [1], a summary of the current thesis, and can be seen as giving a good global overview of the current thesis.

CHAPTER 3 MATHEMATICAL GROUNDS

This chapter introduces all the mathematical grounds that will be used throughout the thesis. For this, we first introduce a simple type of stochastic process, the Poisson process. Then, we present a more general type of stochastic process that also includes marks in the arrival process, the MPP. MPPs are described using the CIF approach. The reader can feel free to read this chapter at his convenience, and come later pick the information he needs.

3.1 Poisson process

This section gives the definition of a Poisson process and reminds some of major properties of this type of stochastic process. The reader can find here everything he needs to understand what is been developed in the thesis. Nevertheless, further details can be found in reference books [27–29].

Definition 3.1 (Counting process)**.** Let $\tau_0, \tau_1, \tau_2 \ldots$ be some positive random variables. Let $T_n = \sum_{j=0}^{n-1} \tau_j$ for $n \geq 1$ and fix $T_0 = 0$. Finally, let

$$N(t) = \max\{n \geq 0 : T_n \leq t\}$$

for $t \geq 0$. In other words,

$$N(t) = \begin{cases} 0 & \text{if } T_0 \leq t < T_1 \\ 1 & \text{if } T_1 \leq t < T_2 \\ 2 & \text{if } T_2 \leq t < T_3 \\ & \vdots \\ n & \text{if } T_n \leq t < T_{n+1} \\ & \vdots \end{cases}$$

Concerning the meaning of the variables,

- ▷ The random variables $T_1, T_2, \ldots$ represent the time of occurrence of the events.
- ▷ The random variables $\tau_0, \tau_1, \ldots$ represent the time between each occurrences (or the waiting times, or also the interevent times), *i.e.* τ_n stands for the time the counting process remained in its n^{th} state (n events occurred).
- ▷ The integer-valued random variable $N(t)$ represents the number of occurrences during the interval $[0, t]$.
- ▷ Finally, the stochastic process $\{N(t), t \geq 0\}$ is called a *counting process*.

Definition 3.2 (Poisson process)**.** A Poisson process is a counting process with independent and identically distributed waiting times according to the exponential distribution $\mathcal{E}xp(\lambda)$. The parameter $\lambda > 0$ is the rate of the Poisson process.

Remark 3.1. The last definition is one among three possible definitions of the Poisson process. □

Remark 3.2. It is possible to switch from the times of occurrences $T_1, T_2, \ldots$ to the interevent times $\tau_0, \tau_1, \ldots$ and reciprocally. Indeed, on the one hand

$$T_n = \sum_{j=0}^{n-1} \tau_j, \quad \forall n \geq 1.$$

And on the other hand,

$$\tau_n = T_{n+1} - T_n, \quad \forall n \in \mathbb{N}.$$

□

Proposition 3.1. *If* $\{N(t), t \geq 0\}$ *is a Poisson process with exponential distribution of rate* $\lambda > 0$*, then*

$$N(t) \sim \mathcal{P}oisson(\lambda t).$$

Proof. Refer to [27] (p.302) for a proof. ■

The following property aims to give a formula to infer the rate of the Poisson process (Méléard [30])

Proposition 3.2 (Inference)**.** *The rate* λ *of the Poisson process* $\{N(t), t \geq 0\}$ *is equal to*

$$\lambda = \lim_{t \to +\infty}^{a.s} \frac{N(t)}{t}, \tag{3.1}$$

where a.s is the almost sure convergence. As explained in Definition 3.1, $N(t)$ *counts the total number of events up to time* t*.*

The Equation (3.1) gives an asymptotic result. In practice, if we observe the process for a finite time horizon denoted as $\Delta T > 0$, then we get the estimator $\widehat{\lambda}$ of the rate

$$\widehat{\lambda} = \frac{N(\Delta T)}{\Delta T}. \tag{3.2}$$

The next theorem supports that a Poisson process can be decomposed into two sub-processes in terms of a classification of the events it counts.

Theorem 3.1 (Poisson process decomposition)**.** *Let $\{N(t), t \geq 0\}$ be a Poisson process, with rate $\lambda > 0$. Let be the probabilities $p_1, p_2 \in]0,1[$ such that $p_1 + p_2 = 1$. Let assume*

1. *Each event of the process can be classified as type 1 or as type 2.*
2. *At each occurrence, there is a probability p_1 that the event is of type 1, and there is a probability p_2 that the event is of type 2.*
3. *Event classifications are mutually independent, and independent of the time.*

Then, we have

$$\begin{cases} \{N_1(t), t \geq 0\} \text{ is a homogeneous Poisson process with rate } \lambda_1 = p_1\lambda \\ \{N_2(t), t \geq 0\} \text{ is a homogeneous Poisson process with rate } \lambda_2 = p_2\lambda \end{cases},$$

where $\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$ count the events of type 1 and type 2, respectively. Moreover, the two processes are independent.

Proof. This theorem is presented in Lefebvre [27] (p.307), where a detailed proof can be found. ■

The following property supports that any Poisson process can undergo a normalization to be turned into a unit rate Poisson process.

Proposition 3.3 (Normalized Poisson process)**.** *Let $\{N(t), t \geq 0\}$ be a Poisson process with rate $\lambda > 0$. Let T be the random variable that stands for the waiting time between two consecutive events.*
The point process defined with the interevent time variable $Z = \lambda\, T$ is a unit rate Poisson process, i.e. has a rate equal to 1. It involves in particular that $Z \sim \mathcal{E}xp(1)$.

Proof. Assume the hypotheses of the property. We define the random variable $Z = \lambda T$, we will find its distribution by computing its Cumulative Distribution Function (CDF). First, the CDF of the exponentially distributed variable T is

$$\forall t \geq 0, \quad F_T(t) = \mathbb{P}(T \leq t) = \int_0^t f_T(u)\, du = \int_0^t \lambda e^{-\lambda u}\, du = 1 - e^{-\lambda t}.$$

Then, for $z \geq 0$, the CDF of the variable Z is

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(\lambda T \leq z) = \mathbb{P}(T \leq z/\lambda) = F_T(z/\lambda) = 1 - e^{-z}.$$

Hence, we are able to identify that $F_Z : z \mapsto 1 - e^{-1 \times z}$. We recognize the CDF of an exponential variable with unit rate $\lambda_Z = 1$. Therefore, by uniqueness of the CDF of a distribution

$$Z = \lambda T \sim \mathcal{E}xp(\lambda_Z = 1).$$

Consider a new point process that we define by choosing Z as the unique interevent time variable, which means all interevent time variables are distributed like Z. Then, since the interevent time variables are exponentially and equally distributed, we know that this process is necessarily a Poisson process [27]. In this case, the rate of the process is the rate of the distribution of Z, that is to say $\lambda_Z = 1$. In conclusion, the new process we defined is a Poisson process with rate 1. ■

3.2 Marked Point Process

A temporal point process is basically a list of times of events. Many real phenomena produce data that can be represented as a temporal point pattern. For instance, occurrences of earthquakes, traffic accidents, and arrivals at an ATM, or storms. What these examples all have in common is that the number of events as well as their time of occurrence are random, which makes them inherently nondeterministic. Such phenomenon can be modeled mathematically by a stochastic process, a *point process*, which is a mathematical tool that enables us to study the phenomena, and predict future events for instance. We use the term *point* to mention that an event is considered to be instantaneous, and thus can be represented as a point on the time line.

In such phenomena, there is often more information associated to an event than just its time of occurrence. We call them marks. For an earthquake, the mark could be defined as its magnitude (a real positive number). Taking into account a relevant mark into the model makes it possible to have a complete modeling that deals with all the aspects of the phenomenon we are interested in. Such a model is called a MPP.

3.2.1 The MPP and the History

To illustrate the principle of MPP and explain how we can store its past activity in a list, we take the example of the earthquakes presented in Figure 3.1. For each event, an earthquake, there is an associated mark, the magnitude of it. If we record the events for a while until the n^{th} one, we get the history up to the present time t, denoted as $\mathcal{H}_t = \{(t_j, \kappa_j)\}_{1 \leq j \leq n}$, with chronologically ordered times $t_1 < t_2 < ... < t_n < t$ (2 events cannot happen simultaneously).

The Figure 3.1 shows how the activity of the MPP can be stored in a simple list of ordered

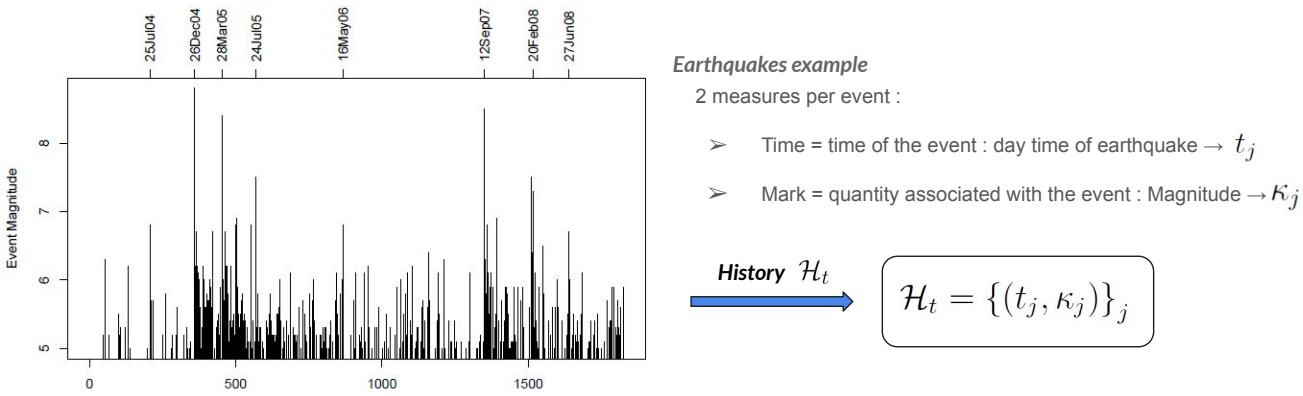


Figure 3.1 MPP and History: earthquakes (graph from [2])

pairs that includes times and marks. This is called the history of the process. Times are strictly positive numbers $t_j > 0$ that represent the time elapsed since the origin of times (usually taken as the beginning of the history). The mark space, denoted by $\mathbb{M}$, can be many different types of spaces, depending on the nature of the mark we consider. In this example, the magnitude is a strictly positive number, $\mathbb{M} = \mathbb{R}_+^*$.

Remark 3.3 (History $\mathcal{H}_t$ and $\mathcal{H}_{t_n}$). Despite that for some MPP, the distinction between $\mathcal{H}_t$ and $\mathcal{H}_{t_n}$ can be useful, we are not considering it in this work. Consider a given time $t > 0$ and the n events that happened so far at the times $t_1 < t_2 < ... < t_n < t$, the history $\mathcal{H}_t$ and $\mathcal{H}_{t_n}$ are equal. Indeed, since no events happened since the n^{th} one at t_n, the history $\mathcal{H}_t$ up to time t records the same events as the history $\mathcal{H}_{t_n}$ that is only up to time t_n.
To simplify notation, we prefer to use the notation $\mathcal{H}_t$ when it is possible, to spare defining the number n of recorded events and the time t_n of the last event.

3.2.2 The Conditional Intensity Function approach

There are several ways of dealing with MPPs. In this work, we select the CIF approach, which is based on the so-called function that defines entirely the MPP. This method offers many advantages, including its simple applicability, its high flexibility to define a model, and the convenient possibility of doing simulations. The CIF characterizes both times and marks of the events.

Definition 3.3 (Conditional Intensity Function)**.** The CIF function, which is denoted by λ^*, is the product of two functions

$$\lambda^* : \mathbb{R}_+ \times \mathbb{M} \to \mathbb{R}_+ \ ; \ (t, \kappa) \mapsto \lambda^*(t, \kappa) = \lambda_G^*(t) f^*(\kappa | t), \tag{3.3}$$

where

- ▷ $\lambda_G^* : \mathbb{R}_+ \to \mathbb{R}_+ \; ; \; t \mapsto \lambda_G^*(t)$ is the Ground Intensity Function (GIF) of the process. It is the instantaneous rate of the process, proportional to the probability of an event to occur around t, given the history $\mathcal{H}_t$ of the events. Indeed, for an infinitesimal time interval dt, we have
$$\lambda_G^*(t)dt = \mathbb{E}\left[N([t, t+dt]) \,|\, \mathcal{H}_t\right],$$
where $N(A)$ denotes the number of points occuring in an interval A.
The GIF is the factor of the CIF that describes the rate at which events occur over time, it can be seen as the mean number of events in a small region conditional on the past. The unit of the GIF is in number of events per time unit, that is to say [Time Unit]$^{-1}$.

- ▷ $f^* : \mathbb{R}_+ \times \mathbb{M} \to \mathbb{R}_+ \; ; \; (t, \kappa) \mapsto f^*(\kappa|t) = f^*(\kappa \,|\, t, \mathcal{H}_t)$ is the Conditional Density Function (CoDF) of the mark κ associated with the point t. It specifies the distribution of the mark κ given t and the history $\mathcal{H}_t$.

Remark 3.4 (Star notation). For the CIF λ^*, the GIF λ_G^* and the mark CoDF f^*, we are using the star exponent notation $*$ as in [31] and [32]. This notation reminds that these functions are conditional on the past history $\mathcal{H}_t$. Also, this notation is useful to distinguish these MPP functions from others functions or parameters. For instance, we will not confound the CIF function λ^* with the parameter denoted as λ that will be frequently used in this memoir. □

The GIF and the CoDF must be chosen carefully so that they reflect properly the behavior and properties of the phenomenon we are modeling. The choice often results in an extended analysis of the data, as well as tests to define the CIF correctly.

For the GIF, we can either decide to define directly a function especially designed for our purpose, or we can also choose a particular interevent time distribution by specifying the CoDF of the time of the next event $f(t|\mathcal{H}_t)$ and its corresponding CDF $F(t|\mathcal{H}_t)$. Indeed, we can switch from one to the other using the following relation that defines the GIF.

Definition 3.4 (Ground Intensity Function (GIF) definition)**.** For a given $t > 0$, we consider the history $\mathcal{H}_t$ up to the time t. The GIF function is defined as

$$\lambda_G^*(t) = \frac{f(t|\mathcal{H}_t)}{1 - F(t|\mathcal{H}_t)}, \tag{3.4}$$

where $f(t|\mathcal{H}_t)$ is the CoDF of the time before next event, and $F(t|\mathcal{H}_t)$ the corresponding CDF.

Remark 3.5 (Time of the next event distribution). For a given time $t > 0$, we denote as $f(t|\mathcal{H}_t)$ the CoDF of the time of the next event, and as $F(t|\mathcal{H}_t)$ its corresponding CDF. Like the CIF, these functions are conditional to the history $\mathcal{H}_t$ of the process up to the considered time t. Although, it must be noted that these functions are not densities of the absolute time t but instead characterize the next interevent time τ_n elapsed since last event at time t_n. Indeed,

- ▷ $f(t|\mathcal{H}_t) = f_{\tau_n}(t - t_n|\mathcal{H}_{t_n})$, it is the CoDF of the next interevent time τ_n.
- ▷ $F(t|\mathcal{H}_t) = F_{\tau_n}(t - t_n|\mathcal{H}_{t_n})$, it is the CDF of the next interevent time τ_n.

This notation is useful to avoid manipulating simultaneously the absolute times t_j and the interevent times τ_j, but also spare the definition of the last n^{th} event at t_n (similarly to remark 3.3 about history notation). □

Example 3.1 (GIF for a homogeneous Poisson process)**.** We consider a simple example of an unmarked process. In this case, only times of occurrence of the events are taken into account. Moreover, the CIF degenerates into the GIF since no mark CoDF is used.
The GIF of a homogeneous Poisson process with rate $\lambda > 0$ can be found using the Definition 3.4. For this type of process, we know that the interevent variables all follow an exponential distribution $\mathcal{E}xp(\lambda)$. We can then compute the resulting GIF using the relation (3.4). For a given $t > 0$, we consider the history $\mathcal{H}_{t_n}$ up to the last event at $t_n < t$, then

$$\lambda^*_G(t) = \frac{f(t|\mathcal{H}_{t_n})}{1 - F(t|\mathcal{H}_{t_n})} = \frac{\lambda e^{-\lambda(t-t_n)}}{1 - \{1 - e^{-\lambda(t-t_n)}\}} = \lambda. \tag{3.5}$$

The GIF of the homogeneous Poisson process is constant, and is in fact the rate of the process. It is understandable since λ is exactly the parameter than dictates the frequency of arrival of the events. □

The GIF and the mark CoDF we chose depend on parameters, like every probabilistic distribution, that we will have to estimate during the resolving process. We denote by Θ_G the parameters of the GIF (where 'G' stands for 'Ground'), and Θ_M the parameters of the mark CoDF (where 'M' stands for 'Mark'). We refer to all the model parameters using $\Theta = (\Theta_G, \Theta_M)$.

3.2.3 Inference

Once the model is defined, the aim is to infer its parameters $\Theta = (\Theta_G, \Theta_M)$. To do this, there are a lot of different possibilities. In the CIF approach, the likelihood function has a

quite simple expression. It usually leads to choose convenient inference methods that use this expression, such as the Maximum Likelihood Estimation (MLE) or the Bayesian inference.

3.2.3.1 Likelihood function

In order to find the expression of the likelihood that involves the CIF, we need to introduce first the integrated Ground Intensity Function (iGIF).

Definition 3.5 (integrated Ground Intensity Function (iGIF))**.** For $t > 0$, the iGIF is, by its inherent definition, given by

$$\Lambda^*(t) = \int_0^t \lambda_G^*(s)ds. \tag{3.6}$$

Then, the likelihood function can be expressed in terms of the CIF and of the iGIF we just defined.

Proposition 3.4 (Likelihood expression with CIF)**.** *Given a marked point pattern* $\mathcal{H}_T = ((t_1, \kappa_1), ..., (t_n, \kappa_n))$ *on* $[0, T) \times \mathbb{M}$*, the likelihood function is given by*

$$L = \left(\prod_{j=1}^{n} \lambda^*(t_j, \kappa_j) \right) \exp(-\Lambda^*(T)). \tag{3.7}$$

In the case of an unmarked point process, the CIF degenerates to the GIF. Then, for a given point pattern $\mathcal{H}_T = (t_1, \ldots, t_n,)$ *on* $[0, T)$*, the likelihood function is*

$$L = \left(\prod_{j=1}^{n} \lambda_G^*(t_j) \right) \exp(-\Lambda^*(T)). \tag{3.8}$$

Proof. The reader can refer to [33] for a proof. ■

3.2.3.2 Estimation

Based on the expression of the likelihood function, we can obtain an estimation of the process parameters using, for instance, the MLE method. However, it is rare to be able to find the maximum analytically, by canceling the derivative, as soon as the likelihood function becomes too complex. One of the situation where the expression of the likelihood function L is simple enough to do so is for the homogeneous Poisson process.

Example 3.2 (Homogeneous Poisson process)**.** In this unmarked case, the intensity function is constant: $\lambda_G^*(t) = \lambda > 0$, where λ is called the rate of the process (computed in Example 3.1). The unit of this rate is the number of events per time unit (for instance, number of

events per day). For $T > 0$ and the corresponding observation interval $[0, T)$ where n events happened, we use the expression in Equation (3.8) to calculate the likelihood

$$L = \left(\prod_{j=1}^{n} \lambda_G^*(t_j) \right) \exp(-\Lambda^*(T)) = \left(\prod_{j=1}^{n} \lambda \right) \exp(-\int_0^T \lambda ds) = \lambda^n \exp(-\lambda T).$$

This expression can easily be differentiated in terms of the only parameter λ, and canceled to find the maximum

$$\frac{\partial}{\partial \lambda} L = n\lambda^{n-1} \exp(-\lambda T) - \lambda^n T \exp(-\lambda T) = (n - \lambda T)\, \lambda^{n-1} \exp(-\lambda T).$$

So, since all the other factors are non null, we have

$$\frac{\partial}{\partial \lambda} L = 0 \iff n - \lambda T = 0.$$

Then, we get the MLE estimator

$$\widehat{\lambda}^{MLE} = \frac{n}{T}. \tag{3.9}$$

It is worth noting it proves that the MLE estimator matches the estimation of the rate proposed in Equation 3.2 (since $n = N(T)$). □

More generally, we will have to cancel the score equation to estimate the model parameters when using the MLE method. It means canceling the partial derivatives of the likelihood function in terms of all the different parameters. In the case of a MPP, we have to solve the system

$$\begin{cases} \frac{\partial}{\partial \Theta_G} L & = 0 \\ \frac{\partial}{\partial \Theta_M} L & = 0 \end{cases} \tag{3.10}$$

in order to find the MLE estimators for the parameters: $\widehat{\Theta_G}^{MLE}$ and $\widehat{\Theta_M}^{MLE}$. If we cannot find analytically the estimates, we can use numerical methods to obtain them. We can quote Newton-Raphson in the case of maximizing the likelihood, or even Markov Chain Monte Carlo for approximating the posterior distribution in a Bayesian approach.

3.2.4 Predictions

Once we got the MPP defined, including the parameters, we can use this model to study the natural phenomenon it describes. The possibilities it offers are various. Not exhaustively, we can be interested in particular values that characterize the phenomenon, called summary statistics, as well as in running simulations for a given particular environment, in order to

observe possible scenarios. We will see that we can predict the information we need in two main ways. On the one hand, explicit formulations that give exact predictions, and on the other hand, simulations that give empirical predictions.

3.2.4.1 Exact prediction

Many quantities can be calculated explicitly from the CIF, such as the probability of getting no events in a future time interval, or the mean waiting time to the next event. For example in the unmarked case, by choosing a specifically designed GIF, we can define an entirely new model of process. In this case, the distribution of interevent times will probably not be a known distribution, and will probably not have a simple analytical expression. Then, to predict the probability of a certain event to happen, we have to use the GIF directly. The following proposition links the expression of the GIF, with, on the one hand, the CoDF of the time of next event given the history, and, on the other hand, the CDF of the time of next event given the history.

Proposition 3.5 (GIF reverse relation)**.** *The reverse relation of the GIF definition* (3.4) *is given by*

$$f(t|\mathcal{H}_{t_n}) = \lambda_G^*(t) \exp\left(-\int_{t_n}^{t} \lambda_G^*(s)\, ds\right) \tag{3.11}$$

or

$$F(t|\mathcal{H}_{t_n}) = 1 - \exp\left(-\int_{t_n}^{t} \lambda_G^*(s)\, ds\right), \tag{3.12}$$

where t_n is the time of the last event before time t.

Proof. A proof of the property is provided in [33]. ■

Example 3.3 (Homogeneous Poisson process)**.** Let be a homogeneous Poisson process with rate $\lambda > 0$. Assume that we want to know the probability that no event happens for a time period $\widetilde{T} > 0$. In this case, there is no event since the beginning of the history at the initial time $t_0 = 0$, then the history $\mathcal{H}_{t_0}$ is empty. For the time of first event $t_1 > 0$, we want to know the probability of the event $\left\{t_1 > \widetilde{T}\right\}$.
Using the GIF inverse relation of proposition 3.5, we have that

$$\begin{aligned}
\mathbb{P}(t_1 > \widetilde{T}) &= \overline{\mathbb{P}(t_1 \leq \widetilde{T})} = 1 - F(\widetilde{T}|\mathcal{H}_{t_0}) \underset{(3.12)}{=} 1 - \left(1 - \exp\left(-\int_{t_0=0}^{\widetilde{T}} \lambda_G^*(s)\, ds\right)\right) \\
&= \exp\left(-\int_0^{\widetilde{T}} \lambda\, ds\right) = \exp\left(-\lambda \widetilde{T}\right).
\end{aligned}$$

We find the same result using the property of a Poisson process. The random number of

events $N(\widetilde{T})$ during a time $\widetilde{T}$ follows a Poisson distribution $\mathcal{P}oi(\lambda\widetilde{T})$ (refer to Property 3.1). We remind the Poisson distribution for a random variable $X \sim \mathcal{P}oi(\widetilde{\lambda})$,

$$\mathbb{P}(X = k) = \frac{\widetilde{\lambda}^k}{k!}\exp(-\widetilde{\lambda}),\ \forall k \in \mathbb{N}.$$

Then, we can compute that

$$\mathbb{P}(t_1 > \widetilde{T}) = \mathbb{P}(N(\widetilde{T}) = 0) = \frac{(\lambda\widetilde{T})^0}{0!}\exp(-\lambda\widetilde{T}) = \exp(-\lambda\widetilde{T}).$$

□

3.2.4.2 Simulation

Simulations of point processes can be useful in many ways, as explained in the following points related by [33]:

- What a point pattern looks like: We can simulate several scenarios for a given model and set of parameters. This will provide helpful information about the pattern of the point process.
- Prediction: We can simulate the future from a specific situation. Taking into account the history in the CIF makes it easy to specify a particular situation and generate a possible future from it.
- Model checking: We can use a subset of the data to fit the model, for instance the first half of data, and then simulate the second half and compare it to the real data we have left.
- Summary statistics: We can retrieve empirical summary statistics from simulation. It is certainly the most interesting possibility it offers. We give more details on this point in what comes next.

Indeed, simulation is a possible way to obtain characteristic values of the process, which can be really useful in some cases. Even if exact explicit results can be found by calculus, it may not be possible all the time. In the case either the CIF or the summary statistic wanted is particularly complex, it may not be possible to get a closed form of the statistic. The solution is to approximate it by running simulation.

For example, the mean number of events in a given time interval may not be available on closed form for a complicated model, but we can then approximate it by the average number

of points in a large number of simulations N_{sim}.

Simulation turns out to be fairly easy when the CIF is specified, which is always the case in the CIF approach we are developing. The CIF offers two major different approaches for simulating a point process; the inverse method and Ogata's modified thinning algorithm. Both are generalizations of similar methods for simulation of inhomogeneous Poisson processes.

Simulation procedure. To simulate a process means in concrete terms to simulate its events, throughout time, starting from the time origin 0, in an incremental way. The first event to be simulated (t_1, κ_1) is composed of its time t_1 and its corresponding mark κ_1.
Assume that we already have the time t_1, then simulating the corresponding mark κ_1 is not difficult. Indeed, the mark CoDF is then totally defined

$$\forall \kappa \in \mathbb{M},\ f^*(\kappa|t_1) = f^*(\kappa \,|\, t_1, \mathcal{H}_{t_1}), \tag{3.13}$$

where both t_1 and the history $\mathcal{H}_{t_1}$ up to the current time t_1 are completely known.
The random conditioning elements of the density function in Equation (3.13) being observed, we can simply draw the mark κ_1 from it, at random. This is clear how this reasoning can be extended for the simulation of any later mark κ_j.

The less obvious part of simulation is how to simulate the time t_1 of the event. It is fair to wonder how to choose it. The GIF is not of the same nature as the mark CoDF, the former is not a Probability Density Function (PDF) while the later is, and consequently, we cannot just draw a time from it. In Example 3.2 for instance, the GIF is a real number: $\lambda^*_G(t) = \lambda > 0$.

As a result, the real difficulty when performing simulation is to find a way to simulate correctly the times of the events. Both the aforementioned methods, namely the inverse method and Ogata's modified thinning algorithm, are especially designed to answer that issue. In this section, we explain the general idea and the main property of the inverse method. To learn more about the Ogata's modified thinning algorithm, refer to [34].

A simulation method: the inverse method. To understand how this method works, we need to introduce first the fundamental property the method is based on, presented in [33].

Proposition 3.6 (Inverse method)**.** *If $(s_i)_{i\in\mathbb{Z}}$ is a unit rate Poisson process on $\mathbb{R}$, and $t_i = \Lambda^{*-1}(s_i)$, then $(t_i)_{i\in\mathbb{Z}}$ is a point process with GIF $\lambda^*_G(t_i)$.*

According to proposition 3.6, the basic idea in the inverse method is simulating a unit rate Poisson process (this is just a series of independent exponentially distributed random vari-

ables with mean one) and transforming it into the desired point process using the iGIF (see Definition 3.5). The name of the method comes from the essential use of the inverse Λ^{*-1} of the iGIF.

Indeed, to perform the transformation, we need either to have access to the closed form of Λ^{*-1}, or find each t_i by searching for a $\tilde{t}$ such that $\Lambda(\tilde{t}) = s_i$. If the first option is not possible, the second one can always be achieved in practice since the iGIF is a continuous and monotonically increasing function (the result is guaranteed by the intermediate value theorem).

3.2.5 MPP methodology recap

As a recap, all the steps to follow in order to conduct the whole MPP methodology are summarized in Figure 3.2.

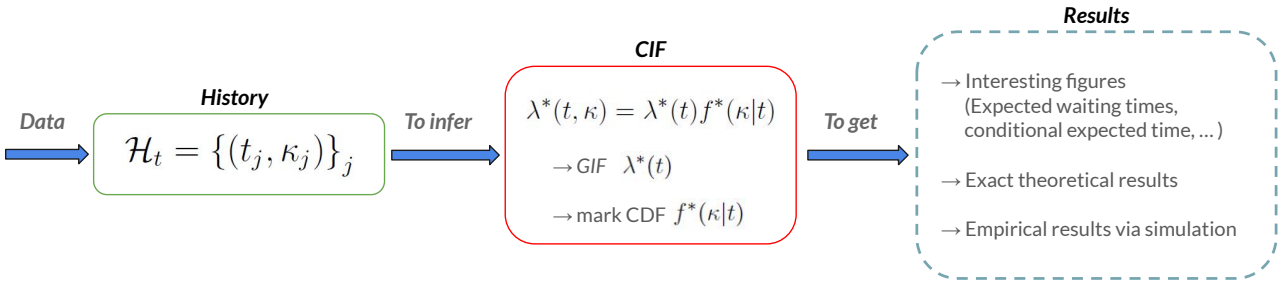


Figure 3.2 Steps of MPP methodology using CIF

CHAPTER 4 GOING BEYOND [1]

In this section, we explain how the current work is based on previous work by Weller [1], and in which extent it improves some of its limitations. We adopt a high-level vision of both works in order to have a good overall understanding, without entering too much into mathematical details. This chapter can be seen as an extended literature review focusing on the links between the current thesis and the research conducted in [1].

4.1 Original Research

In the decision-making that results from a kidney offer, the developed solution aims at answering two main emerging issues.

- TNO problem: to predict the time before another kidney is proposed to the patient. Independently of the quality of the offer.
- TNBO problem: to predict the time before a better kidney (than the current proposed one) is proposed to the patient.

If the TNO issue is non negligible, and is somehow necessary to lay the groundwork for the general mathematical modeling of the problem, the most relevant issue is the TNBO one. Indeed, for a patient who has been waiting for a kidney transplant for months, the only reason to decline the current kidney offer is the expectation of a better offer soon. We explain to what extent work conducted in [1] answers both issues.

4.1.1 Predict the Time before Next Offer

We consider the problem for one patient x_0, who is proposed a kidney from donor y_0 at time t_0 with a kidney quality of q_0 (quality given by the KDRI measure). We refer to this offer as the "initial offer" (x_0, y_0, t_0, q_0).

Building the History The main idea is to look at the past activity of the patient to infer the behavior (frequency, quality, etc.) of kidney proposals he receives. To achieve it, we would need the real patient history, for over one year at least, with all the proposed kidney offers. However, not all the information we need is available due to the nature of the data (recently enlisted patients for instance, or incomplete information). Then, the idea is to create

a pseudo-history $\mathcal{H}$ from the available data we have, which would tend to represent the real history during a $\Delta T > 0$ period of time before the initial offer at time t_0. For this purpose, we retrieve from data all kidney arrivals and corresponding waiting lists chronologically. The j^{th} waiting list on the history is denoted as w_j. For each waiting list w_j, we denote as

- ▷ y_j the corresponding donor.
- ▷ t_j the time of the offer.
- ▷ q_j the quality of the offer for our initial patient x_0 (this is computed by a function that takes as inputs y_j and x_0).

Note that in a waiting list w_j, patients are ordered according to their score, computed with a scoring function that is supposed to order them fairly in terms of different factors (medical priority, time on the waiting list, etc.). The first patient on the waiting list is the one with the highest priority, as a result he is the first one the kidney is proposed to. If he declines the offer, the kidney is proposed to the second patient on the waiting list, and so on.

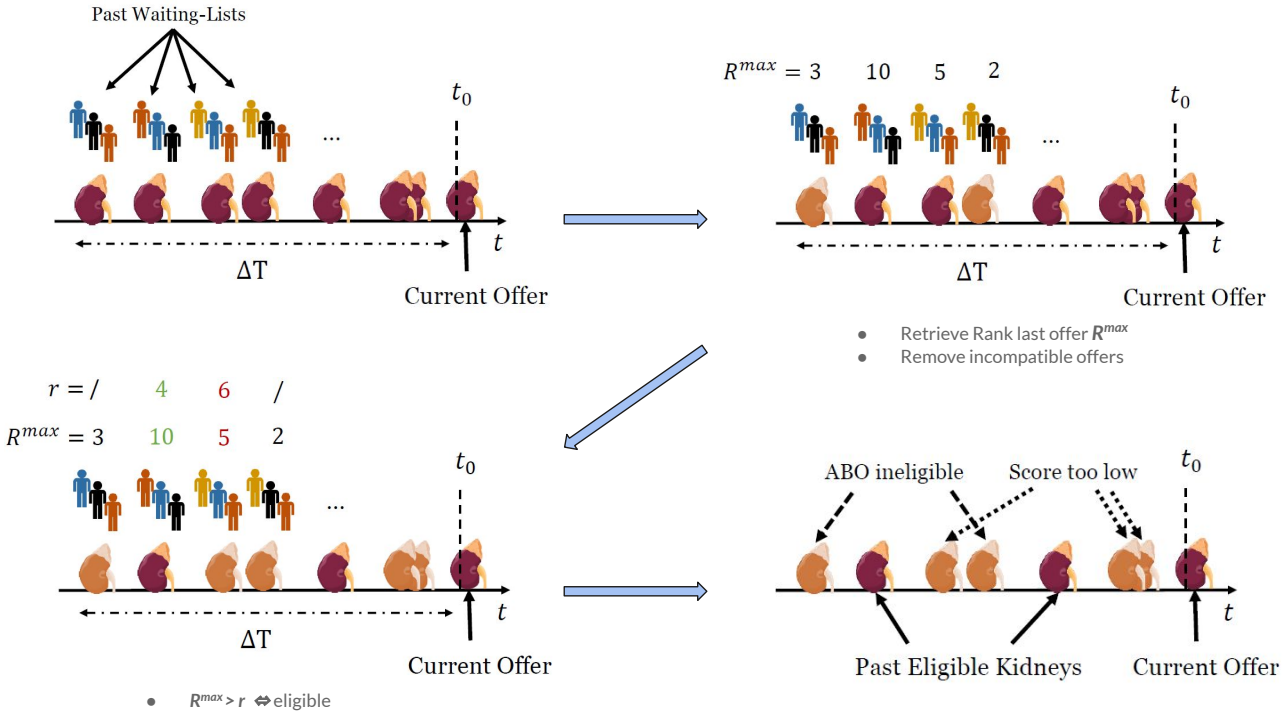


Figure 4.1 History Building (figures from [1])

Once we retrieved all the waiting lists, we need to determine for which ones the patients x_0 would have received a transplant offer. That is to say, for each waiting list w_j that occurred

at time t_j, we need to determine if the initial patient x_0 would have had received the offer or not, in the situation he would have been active in the attribution system at that time. The history building procedure is depicted in Figure 4.1. For each kidney arrival,

- We remove incompatible/non eligible offers for x_0. Every removed kidney turns orange in Figure 4.1.
- We compute the rank of last offer R_j^{max} that corresponds to the highest rank for which a patient on w_j got an offer. This is the rank of the last patient who got the offer.
- We compute, using the scoring function, the rank r_j of x_0 if he was in w_j.
- We remove offers when $r_j > R_j^{max}$, since that means the position of x_0 in the waiting list is too low to be offered the kidney.

After this procedure, we got the history $\mathcal{H} = \{(t_j, q_j)\}_j$ for the initial patient x_0.

Remark 4.1. According to previous notations from Section 3.2, we should write $\mathcal{H}_{t_0}$ for the history. However, the history is quite always up to the initial time t_0 in this problem. Then, we omit the subscript and simply use the notation $\mathcal{H}$ to stand for $\mathcal{H}_{t_0}$. □

Fitting the Poisson process Concerning the theoretical aspect, what we want to use is a stochastic process that models the arrival of the eligible kidney offers for the patient x_0. The work achieved in [1] brings explanations, proofs and justifications to support the use of the Poisson process. In such a process, the random variables that represent waiting times between two consecutive events (arrival of kidney) are all independent and equally distributed (refer to Section 3.1). We denote as τ the random variable corresponding to one of the interevent time (one is enough since they all are equally distributed). Then, $\tau \sim \mathcal{E}xp(\lambda)$, where $\lambda > 0$ is called the rate of the process. The expected value is $\mathbb{E}[\tau] = 1/\lambda$. The higher the rate is, the smaller the expectation is, *i.e.* the more rapidly an event occurs on average.

The Poisson process is entirely defined by the knowledge of its rate λ, which is the quantity we want to evaluate. We use the approximation formula from Equation (3.1) applied on the history $\mathcal{H}$ of the patient that has a finite time horizon ΔT

$$\lambda = \frac{N_{\text{offer}}}{\Delta T}, \tag{4.1}$$

where $N_{\text{offer}} = \text{Card}(\mathcal{H})$ is the number of events (or offers) on $\mathcal{H}$.

Predicting waiting times Once we have the rate of the process, we can predict characteristic values of the process that are relevant to solve our problem, like waiting times.

Former work [1] brings theoretical support and explicitly evaluates these quantities. The process we fit represents the arrival of kidneys, regardless of their qualities, thus the values we can predict are almost exclusively concerning the next arrival of a kidney, without any condition on its quality. Two important figures among them are

- Expected waiting time before next offer: $\mathbb{E}[\tau] = 1/\lambda$.
- Customized level confidence interval for the estimated time before next offer. In other words, for a confidence level $\alpha \in [0, 1]$ the solution provides the time t_α before which there is a α probability of getting an offer.

4.1.2 Predicting the Time before Next Better Offer

The issue here is to predict the time before the next better offer, *i.e.*, an offer with quality $q^+ > q_0$. The proposed TNBO solution does not differ a lot from the TNO solution. Indeed, only one simple step, the thinning of the history, is added to the existing TNO solution. Then, the steps of the TNBO solution are in this order

1. Building the history.
2. **Thinning the history**. (This is the additional step that we detail below).
3. Fitting the Poisson process.
4. Predicting waiting times.

Thinning the history This step consists in taking the freshly created history to simply remove all the offers for which $q_j \leq q_0$, *i.e.* offers for which the kidney quality is less than the desired quality. An example of thinning with 10 offers on the history is displayed in Figure 4.2. The original history before thinning is on the left, and the thinned one is represented on the right. In the end, there are 4 offers left out of 10 originally.

After this, we run the next steps exactly as before, but on the new thinned history $\widetilde{\mathcal{H}}$ instead of $\mathcal{H}$. The rationale behind this is that we kept only the better offers than the initial one, so we now have a history corresponding to the arrival of the better offers, and the corresponding smaller Poisson process rate. According to this, the predictions we are able to do now (in step 4. 'Predicting waiting times') could be used to answer TNBO problem.

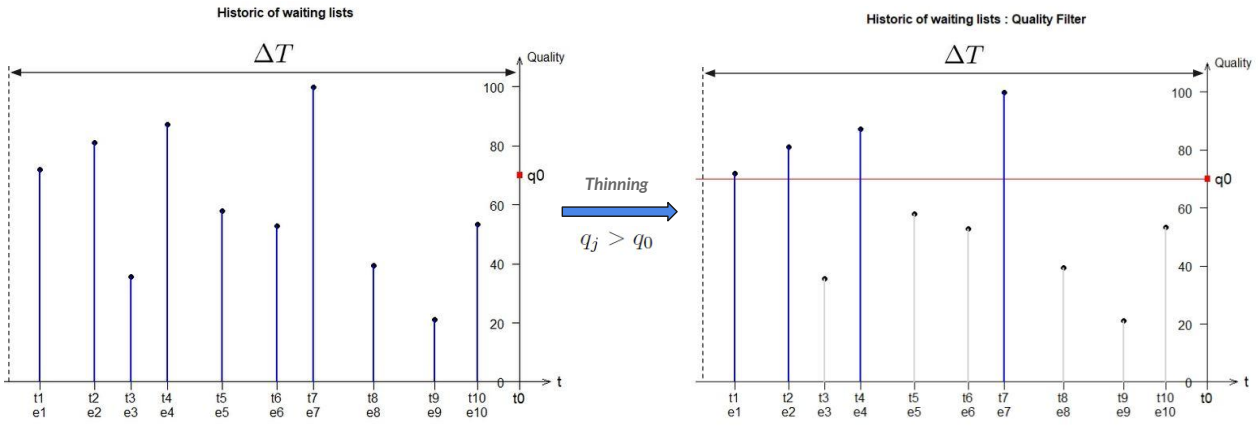


Figure 4.2 Thin the history

4.1.3 Remarks

For simplification purpose, we spared many details about the method from [1]. Further explanations can be found in the thesis of Weller [1]. Although, we give some important remarks below

- Building the pseudo history $\mathcal{H}$ from the available data is not quite obvious in reality. It involves knowledge of the attribution system used by TQ (refer to Chapter 3 in [1]), but also making assumptions and choices. Indeed, most of the times we do not have all the information we need to reconstruct a one hundred percent faithful history.
- Different methods to build the history had been developed in [1], namely: *Past waiting list*, *Current waiting list* and *Eligibility relaxation*. As best performances were obtained by the Past waiting list method, we chose this one to illustrate the history building.
- Notation: We use λ to refer to the process rate instead of the μ notation from [1]. Indeed, μ is commonly used to refer to a variable expectation, and it could lead to confusion. For instance, for an exponentially distributed random variable $X \sim \mathcal{E}xp(\lambda)$, $\mathbb{E}[X] = \mu = 1/\lambda$.

4.1.4 Gap left to fill

The work conducted in [1] constitutes a solid groundwork that addresses the *No question* issue. It provides theory, methodology as well as relevant and usable results that enables to device the desired decision-aid tool to inform a patient on his perspectives if he was to decline the current offer.

While TNO issue is completely addressed, limitations still remain concerning the TNBO solution. The current thesis focuses on the TNBO issue since the main reason that motivates a patient's refusal of a current kidney offer is the expectation of a better one.

4.2 Predicting waiting times based on Marked Point Process

In this section, we present from a high level point of view the new solution we develop in this work. The details of the mathematical modeling are presented in the dedicated Chapter 5.

This thesis uses the groundwork laid by [1] as a foundation, and the main objective is to answer the TNBO issue. The mathematical frame we use is the MPP methodology that is presented in Section 3.2. In this new approach, we do not design a different specific solution for each problem (TNO/TNBO). Instead, we provide a theoretical frame that includes quality aspect from the beginning, and that makes it possible to answer both issues at the same time.

4.2.1 Marked Point Process applied to kidney offer

In this section, the MPP methodology presented in Section 3.2 is applied to the kidney offer problem. This is the new solution developed in this thesis. As in Section 4.1.1, the reasoning is developed with one initial patient x_0 and the corresponding initial offer (x_0, y_0, t_0, q_0).

Building new targets We need to build new targets from available data, in order to retrieve times before next better offers. Those times are the real times we can observe in the data, which enables us to check the model validity. In the previous work [1], datasets only include time before next offer, which does not give the possibility to verify model performances for TNBO predictions.

The Figure 4.3 represents an example of the problem we had with former dataset targets. *'Data 1'* and *'Data 2'* refer to one row in the former research dataset. Each data consists of two direct consecutive offers (for the same patient), which does not take into account for instance the case a better offer is not the consecutive one but the one after this. We remind that in this figure, the quality q is the KDRI index quality, which is a risk index: a lower value of q means a better kidney quality for the patient.

Building the history For all the data processing part that is required to build the history, previous work from [1] was kept as is: we used the *Past waiting list* algorithm detailed in Section 4.1.1. The procedure enables to build the history $\mathcal{H} = \{(t_j, q_j)\}_{1 \leq j \leq N_{\text{offer}}}$ for the initial patient x_0.

CAN_ID	DON_ID	CENSORED	UPDATE_TM	TIME_NEXT_OFFER	KDRI	KDRI_NEXT
694	1204	1	2015-04-20	4	1.5151059	2.3535731
694	1209	1	2015-04-24	2	2.3535731	1.4091363
694	1211	1	2015-04-26	13	1.4091363	1.1232409

Real Next_Better_Offer

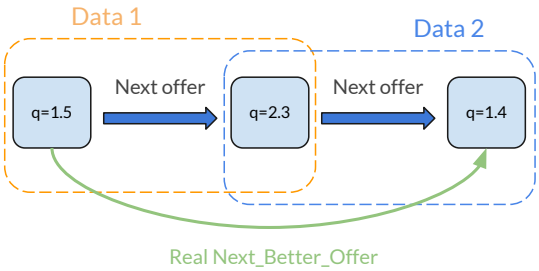


Figure 4.3 Build the new targets

Fitting the MPP To solve the problem, we apply the MPP methodology presented in Section 3.2. The main point is to treat simultaneously times and marks of the event, which are respectively the time t_j of arrival of an eligible kidney and its quality q_j. For this reason, the marks are now denoted as q_j to refer to quality (instead of the generic notation κ_j used so far).

As explained earlier, an essential point of the method is the relevant choice of the GIF and the mark CoDF. Concerning the GIF $\lambda_G^*(t)$, which specifies the arrival of the kidney, we use the theoretical groundwork developed in [1]. Then, we also consider a Poisson process with rate $\lambda > 0$. In this case, the GIF of the model is equal to

$$\lambda_G^*(t) = \lambda. \tag{4.2}$$

Regarding the mark CoDF $f^*(q|t)$, an analysis is carried out on the quality marks on the histories of patients to choose the best probability distribution for the model. The assumptions of our model imply the considered distributions to be independent of the history $\mathcal{H}$, and more especially stationary. After a comparative study, the distribution that proved to be a relevant choice is the Weibull distribution $\mathcal{W}(\theta_1, \theta_2)$, with parameters $\theta_1, \theta_2 > 0$. The corresponding mark CoDF is then, for a given $t > 0$,

$$\forall q \geq 0, \quad f^*(q|t) = f_Q(q) = \frac{\theta_1}{\theta_2}\left(\frac{q}{\theta_2}\right)^{\theta_1 - 1} \exp\left\{-\left(\frac{q}{\theta_2}\right)^{\theta_1}\right\}. \tag{4.3}$$

The next step is to estimate the model parameters. To infer the GIF parameter λ, the first

estimator to be tested was $\widehat{\lambda}_1 = \frac{N_{\text{offer}}}{\Delta T}$, the same as in [1] and which was presented earlier in Equation (4.1). However, after experiments and tests, a more efficient estimator giving better results was designed as

$$\widehat{\lambda}_2 = \frac{N_{\text{offer}}}{\sum_{j=1}^{N_{\text{offer}}} \tau_j}, \tag{4.4}$$

where the sample $\{\tau_j\}_{1 \leq j \leq N_{\text{offer}}}$ is the observed waiting time between two consecutive offers on the history $\mathcal{H}$ of the patient.
Fitting the mark CoDF parameters (θ_1, θ_2) is achieved with the help of classical inference methods carried out on the history of past qualities offers $\{q_j\}_{1 \leq j \leq N_{\text{offer}}}$, as for instance the MLE method.

Predicting waiting times When all the MPP is well-defined, we can deduce from it unknown characteristic figures we are interested in to answer the kidney problem. With the theoretical frame we set and associated computation, we can retrieve expected waiting time before next offer as well as expected waiting times before next better offer. Since the CIF includes the quality aspect, it is possible to set conditions concerning quality, as for example setting a minimum threshold for the future kidney quality to come.
To answer the TNBO issue, we want to predict the time t^+ before a next offer better than the initial one, *i.e.* an offer such that $q^+ > q_0$. Mathematical results support that the original MPP can be thinned to become a new MPP which new process rate equals to

$$\lambda^+ = \mathbb{P}(Q > q_0)\, \lambda. \tag{4.5}$$

The new process, which counts the arrival of better kidney than the current one, is a thinned process derived from the original one. The thinning factor is $\mathbb{P}(Q > q_0) = \int_{q_0}^{+\infty} f_Q(q)\, dq$, which is computed using the mark CoDF f_Q that was fitted earlier. We can notice that the higher the kidney quality has to be, the more important the thinning is, and the smaller the resulting rate λ^+ is.

Steps recovered from previous work The different steps of the implemented MPP methodology are shown in Figure 4.4. Difference is made between steps taken from Weller [1], and new developed steps. As explained earlier, the processing step to build history is retrieved from the past research work, whereas the mathematical steps and framework that come next are not.

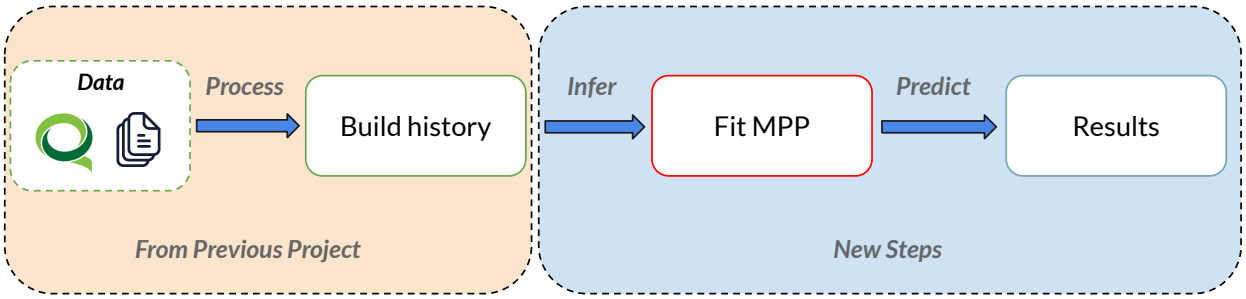


Figure 4.4 Methodology steps

4.2.2 Addressing limitations

The current research handles the kidney problem with a new perspective, which enables to tackle limitations that were encountered before. We draw up a list of improvements. From a global point of view, we can say that

- Concerning the TNO problem: We selected and retrieved from former work some relevant, correct and well justified parts. This constitutes a solid groundwork for building the new method.
- For the TNBO problem: The new solution is oriented towards answering this issue since this is the most relevant problem to solve for a patient.

Concerning more precise points to address limitations in TNBO issue,

- The current thesis provides a rigorous theoretical frame that supports the method we develop.
- The quality aspect is fully integrated in the probabilistic model, and has its own distribution, the mark CoDF. The quality aspect does not longer summarize to descriptive statistics.
- With MPP method, all the information about quality is included into the quality distribution parameters, the ones we fit using the history. The method is compact, as working with the full history is no longer necessary.
- With the fitted quality distribution, it is possible to modify the minimum quality threshold easily and as many times as wanted. The method is flexible and adaptable. Simulation on higher quality expectancy can be performed.
- No history thinning is performed when solving TNBO problem, but instead a direct thinning of the process by using the CIF. In this case, the number of eligible offers

stays the same on histories of patients. The method reduces existing problems with low priority patients for who no offers would have remain otherwise after the history thinning.

- The fact that the history remains in its entirety, so it can all be used to fit the CIF, does not cause any information loss.
- A work on data processing and dataset building has been done in order to retrieve the true next better offers. Hence, the method validation is possible.
- The method is flexible for potential future modifications: For instance, it is possible to create a more complex process GIF as well as to change the mark CoDF (*e.g.* adding dependence between events and marks). When the process CIF is not an exact distribution, empirical results can still be obtained easily from simulation. With the MPP framework, various modifications are possible.

4.2.3 Computational results

Methodologically, the MPP method we develop enables to address some limitations left unsolved in [1]. We give in this section a quick summary of main computational results to show that improvements have been made in practice too. Experiments using both the previous method from [1] and the new one have been run in order to compare performances when predicting waiting times. Extended details about experiments setups and results are given later in the dedicated Chapter 6. The new MPP method is referred to as Method 2 and competing method from [1] is referred to as Method 1.

Concerning the TNO problem, the new rate estimator $\widehat{\lambda}_2$ gives better results than the one $\widehat{\lambda}_1$ used in Method 1. For histories with horizon $\Delta T = 1000$ days ($\simeq 3$ years), the discriminative power, measured with Concordance Index (C-index) (refer to Definition 6.2), is virtually the same from one method to another. However, we note a significant 34.2% Mean Absolute Percentage Error (MAPE) (error score, refer to Definition 6.1 for details) increase when using the previous estimator $\widehat{\lambda}_1$ instead of the new estimator $\widehat{\lambda}_2$, meaning new predictions are closer to real observed values. Equivalently, to go from Method 1 to Method 2 represents a 25.5% drop in the error score.

For the TNBO problem, the method described in Section 4.1.2 was implemented in order to be compared with the new MPP method. Concerning C-index (meant to evaluate predictive power of an estimator, refer to Definition 6.2 for details), Method 1 and Method 2 give almost the same percentage of good ordering for the patients (based on the observed time), we note a slightly better performance with MPP Method, more precisely a 2.4% increase of classifying rate. We have respectively C-index of 63.8 % against 65.3%. Nonetheless, for

the error indicator, we can also note a relevant MAPE drop of 22.3% when using Method 2 instead of Method 1. The predictions of Method 2 are better than those of Method 1.

We recap figures from the comparison of the main computational results in Table 4.1.

Table 4.1 Methods performances comparison (on test set)

Method	MAPE	C-index
TNO - Method 1 $\rightarrow$ Method 2	- 25.5%	+0.6%
TNBO - Method 1 $\rightarrow$ Method 2 MPP	- 22.3%	+2.4%

In conclusion, in line with the same objectives as [1], the MPP methodology we developed provides a significant improvement for the prediction of waiting times for kidney transplant patients.

CHAPTER 5 MPP MODELING FOR TIME BEFORE NEXT BETTER OFFER

This chapter details how the MPP methodology presented in Section 3.2 applies to the kidney transplant problem, and what the next steps are in order to have predictions for our problem.

Remark 5.1 (Prior reading). From this point, global knowledge of the problem, its stakes and the kidney attribution system of TQ are assumed. To make sure the reader have this knowledge, references, including references to previous work [1], have been provided throughout Chapters 1, 2 and 4. Most importantly, the Chapter 4 consists of a summary of both previous thesis [1] and current work methods. The summary is meant to give an overall understanding and cannot go into details. For this reason, only final solutions are introduced in Chapter 4. The details of the approach carried out to obtain them are presented thoroughly in this chapter. □

5.1 Marked point process based kidney offer problem

In this section, we explain how we applied the previously detailed MPP method to our problem. We first introduce the main elements of the problem we try to solve and set up a model frame with mathematical notation. In a second phase, we apply the chosen approach to tackle the problem, which uses the CIF. This involves making decisions about the formulation of this function and its properties, in particular choosing the GIF and the mark CoDF.

5.1.1 Problem components modeling and notations

We need to set variable and parameter notations that stand for essentials elements of the problem that we already partially introduced in Chapter 4.

- Sets: $\mathcal{P}$ is the set of patients, $\mathcal{D}$ stands for the set of possible donors. A patient is denoted as $x \in \mathcal{P}$, and a donor as $y \in \mathcal{D}$.
- Times are real numbers in $\mathbb{R}$.
- Kidney qualities are denoted as $q \in \Omega_Q$, where Ω_Q is the definition domain of the quality ($\Omega_Q = \mathbb{R}_+$ in the case of the KDRI quality).

5.1.1.1 General modeling of attribution process

We explain first the general modeling of the kidney arrival process, without focusing on a particular initial patient. The variables we introduce are the same for all patients. For that reason, we add the exponent "*0*" to all variable notations (*e.g.*, T_n^0), to remind they are the earliest general variables.

- Let $(T_n^0)_{n\in\mathbb{N}}$ be the random process in $\mathbb{R}_+$ describing the random times of arrival of kidneys in the attribution system from an initial time $T_0^0 = 0$.
- Let $(Y_n^0)_{n\in\mathbb{N}}$ be the sequence of random variables in $\mathcal{D}$ describing the random donors arriving at times T_n^0, for $n \in \mathbb{N}$.

We have also global model assumptions and their resulting properties. Namely,

Assumption 5.1 (Poisson arrival of donors)**.** *We assume kidney donors are arriving on the market following a homogeneous Poisson point process of rate parameter $\lambda^0 > 0$ at times $(T_n^0)_{n\in\mathbb{N}}$, with $T_0^0 = 0$.*

This assumption means (refer to Section 3.1):

- The interevent times $\tau_n^0 = T_{n+1}^0 - T_n^0$ are *i.i.d.*, $\forall n \in \mathbb{N}$. We have a stationary process with identically distributed interevent times. It means the distribution of the waiting times does not change over time.
- Each interevent time $\tau_n^0 = T_{n+1}^0 - T_n^0$ follows an exponential distribution $\mathcal{E}xp(\lambda^0)$, $\forall n \in \mathbb{N}$.

Others assumptions are made in the model, concerning the arrival of donors:

Assumption 5.2. *The distribution of the type of incoming kidneys is independent of time: the Y_n^0 are i.i.d., $\forall n \in \mathbb{N}$.*

Assumption 5.3. *The type of incoming donor is independent of its time of arrival: Y_n^0 and T_n^0 are independent, $\forall n \in \mathbb{N}$.*

Remark 5.2 (Model assumptions). To build this general framework, we are using the same ground theory than the previous work in [1], with the objective to build something on top of it. Then, we are passing on extended details and justifications about this general framework and its assumptions, which can be found in [1]. To give an example, the validity of Assumption 5.1 is verified in Section 5.3 of [1]. □

5.1.1.2 Modeling next eligible offer

In this work, the aim is to answer a particular question: predict waiting times from a given patient and initial offer. Thus, from now on, we consider the problem for one patient $x_0 \in \mathcal{P}$, who is proposed a kidney from the donor $y_0 \in \mathcal{D}$, at the time $t_0 > 0$, with a kidney quality of $q_0 \in \Omega_Q$ (quality given by the KDRI indicator). We refer to this offer as the "initial offer" (x_0, y_0, t_0, q_0).
When we study the problem for the given patient x_0, the donors arrival process is not the same as the general process of last Section 5.1.1.1, since, for example, not all donors are compatible with the aforementioned patient x_0. We give below the notation for the resulting variables and processes, the same ones as before but without the exponent notation "*0*".

- Let $(T_n)_{n\in\mathbb{N}}$ be the random process in $\mathbb{R}_+$ describing the random times of arrival of eligible donors on the transplant system from an initial time $T_0 = 0$.
- Let $(Y_n)_{n\in\mathbb{N}}$ be the sequence of random variables in $\mathcal{D}$ describing the random eligible donors arriving at times T_n, for $n \in \mathbb{N}$.
- Since we have the characteristics of the patient x_0, we can also now introduce the quality of the kidney.
  Let $(Q_n)_{n\in\mathbb{N}}$ be the random process in Ω_Q describing the random kidney qualities arriving at times T_n, for $n \in \mathbb{N}$. For $n \in \mathbb{N}$, the quality Q_n is the quality of the donor's Y_n kidney, given our initial patient x_0.
- We denote the interevent times as $\tau_n = T_{n+1} - T_n, \forall n \in \mathbb{N}$. For $n \in \mathbb{N}$, the interevent time τ_n represents the time the MPP spent on its n^{th} state, that is to say the time elapsed with n events in total in the history. We can speak of the state of the process because the MPP we use is also a counting process; the events do not happen simultaneously so we can count, over the course of time, the total number of events that have occurred since the beginning.

Basically, studying the problem for an initial offer results in the thinning of the general arrival process. Only a certain proportion of original donors are eligible for x_0. The new process is still a Poisson process, and if we denote as $\mathcal{D}^*$ the subset of $\mathcal{D}$ that stands for the eligible donors, then the new rate can be expressed as

$$\lambda = \lambda^0 \times \mathbb{P}(Y_1^0 \in \mathcal{D}^*). \tag{5.1}$$

For an extended explanation and proof of the process thinning, we refer to [1] (See Chapter 4, p.29-31). Though, these theoretical specifications are meant to give the high level steps of the modeling, and also justify where the results come from. In this memoir, we directly start with the thinned process of eligible donors arrival $(T_n)_{n\in\mathbb{N}}$ without considering the earliest process $(T_n^0)_{n\in\mathbb{N}}$ anymore. Then, for the process $(T_n)_{n\in\mathbb{N}}$, we were able to give the following interesting key property.

Proposition 5.1 (Poisson arrival of eligible donors)**.** *The arrival of eligible kidney donors is following a homogeneous Poisson point process of rate parameter $\lambda > 0$ at times $(T_n)_{n\in\mathbb{N}}$, with $T_0 = 0$.*

And since we have a Poisson process, it means naturally it has the same properties we stated before for the earliest Poisson process. Namely,

- The interevent times $\tau_n = T_{n+1} - T_n$ are *i.i.d.*, $\forall n \in \mathbb{N}$. We have a stationary process with identically distributed interevent times. It means the distribution of the waiting times does not depend on the time being.
- Each interevent time $\tau_n = T_{n+1} - T_n$ follows an exponential distribution $\mathcal{E}xp(\lambda)$, $\forall n \in \mathbb{N}$.

In the problem of kidney graft, we want to know the distribution of the waiting time before the arrival of the next eligible donor, and the characteristics of this donor, particularly in order to compute the corresponding quality of the kidney. Hence, we need to introduce relevant notations:

- We denote as $T = T_1$ the time of the next event (starting to count after the initial event at t_0, the initial offer for the concerned patient x_0). Moreover, the time of the first event equals the first interevent time, $T = T_1 = T_1 - 0 = T_1 - T_0 = \tau_0$.
- We denote as $Y = Y_1$ the first eligible donor to arrive.
- We denote as $Q = Q_1$ the kidney quality of the donor Y_1.
- We denote as $\{(T_n, Q_n)\}_{n\in\mathbb{N}}$ the MPP describing arrival of events and associated quality.

5.1.2 Building the history of the process

We explain the way we build the pseudo-history of the MPP by retrieving past waiting lists from data. We already explained the process in Chapter 4 (Section 4.1.1). We still summarize quickly the steps to follow, to remind the reader of the notations we use.

The main idea is to look at the past activity of the patient to infer the behavior (frequency, quality, etc.) of the kidney proposals he receives. In order to do so, we build the history of

the process for our patient x_0. To have all the methodology steps detailed, please refer to Section 4.1.1. We look at a ΔT time horizon in the past to build the history for the time period $[t_0 - \Delta T, t_0]$. We get all waiting lists, chronologically ordered, and we denote the j^{th} one as w_j. Each one corresponds to the arrival of a donor (a kidney). For each waiting list w_j, we denote as

- $y_j \in \mathcal{D}$ the corresponding donor.
- $t_j > 0$ the time of the offer.
- $q_j \in \Omega_Q$ the quality of the offer for our initial patient x_0 (this is a function whose inputs are y_j and x_0).

Then, we filter the history to only keep eligible offers for our patient. After this procedure, detailed in Section 4.1.1 and depicted in Figure 4.1, we get the history $\mathcal{H} = \{(t_j, q_j)\}_{1 \leq j \leq N_{\text{offer}}}$ that is made of N_{offer} events in total. $\mathcal{H}$ represents the history of the MPP for the initial patient x_0. (See also *Remark* 4.1 for the notation $\mathcal{H}$ instead of $\mathcal{H}_{t_0}$).

Remark 5.3 (Notation). The history is composed of past events and marks. Those observations are denoted with small letters by opposition to capital letters notation used for random variables. The capital letters notation can be used for upcoming events (after time t_0), or for theoretical manipulation of a MPP object (in proofs).
Moreover, the history $\mathcal{H}$ represents the past of the MPP we are interested in, which is the arrival of donors and qualities $\{(T_n, Q_n)\}_{n \in \mathbb{N}}$. When we solve the problem, at time t_0, we already know the MPP events up to this moment, so we consider random variables (and their capital letter notations) only from the initial time $T_0 = t_0$. As for instance the time T. □

Remark 5.4 (Time origin). We generally consider t_0 as the origin of times for the random MPP to begin with. In this case, $T_0 = t_0 = 0$ and all times T_n represent elapsed times since t_0. □

5.1.3 Ground Intensity Function: the rate of the process

Concerning the GIF $\lambda_G^*(t)$ that specifies the arrival times of the kidneys, we use the results discussed in Section 5.1.1. Then, we consider a Poisson process of rate $\lambda > 0$, which is a simple process with interesting properties. The GIF of this particular type of process has been calculated in Example 3.1. Equation (3.5) gives

$$\forall t > 0, \quad \lambda_G^*(t) = \lambda. \tag{5.2}$$

Then, the model's GIF is constant and matches the rate of the Poisson process λ. It corresponds to the model parameter $\Theta_G = \lambda$ discussed earlier in Section 3.2.2.

The case of the Poisson process is a special one, since it implies first memorylessness. Indeed, exponentially distributed interevent times are memoryless, since the exponential distribution is. In addition to this independence between the event times, it also means we do not have to take account of the past in the waiting time distribution. Stochastically speaking, the process of arrival restarts from zero at each new event [27]. Then, the previous distribution notation given the history (in Definition 3.4 and Remark 3.5) can be simplified as follows

$$\forall t > t_n,\ f(t|\mathcal{H}_{t_n}) = f_{\tau_n}(t), \tag{5.3}$$

where τ_n is the waiting time before the next event after the n previous ones at $t_1, ..., t_n$.
Moreover, the Poisson process also implies that the distributions of all the waiting times are equal (the times are equally distributed). Thus, we can simply use the first interevent distribution f_{τ_0} to stand for all the interevent times distributions. $T = T_1 = \tau_0$ being the first interevent time, we get

$$\forall j \geq 1,\ f_{\tau_j} = f_{\tau_0} = f_T,\ \text{with}\ \forall\, t > 0,\ f_T(t) = \lambda e^{-\lambda t}. \tag{5.4}$$

The knowledge of the distribution f_T of the interevent times is a major convenience for all the predictions we want to make from the model. It results in simplified calculus and closed form predictions, as we will see in Section 5.2

5.1.4 Mark Conditional Density Function: the quality of kidney

For each event at t_j, there is an associated mark κ_j. In our problem, the mark is the quality q_j of the kidney: $\kappa_j = q_j$. For this reason, the marks are now denoted as q_j to refer to quality (instead of the generic notation κ_j used so far), which is more explicit.
Marks are drawn from the mark CoDF. This function $f^*(q \,|\, t) = f^*(q \,|\, t, \mathcal{H}_t)$ can theoretically depend on the time of the event (t), but also on the past events and marks ($\mathcal{H}_t$). We need to specify these dependence relations. We can deduce from the model we built that in our case, there is no dependency between times and marks.

Proposition 5.2. *The quality of incoming kidney is independent of its time of arrival: Q_n and T_n are independent, $\forall n \in \mathbb{N}$.*

Proof. For $n \in \mathbb{N}$, the quality Q_n is the quality of the donor's kidney Y_n, given our initial patient x_0. Then, Q_n is a measurable function of the random variable Y_n. In the specific case of the KDRI quality we can compute using its function f_{KDRI}, we can write: $Q_n = f_{KDRI}(Y_n, x_0)$. Moreover, Y_n and T_n are independent (according to assumption 5.3). Then,

we can use the result about independence and function composition given in [27] (p.28) to conclude in this case Q_n and T_n are also independent. ■

The immediate consequence is the independence between the time of an event t and the associated quality q. Then, the mark CoDF can be simplified in this way

$$f^*(q \,|\, t) = f^*(q \,|\, \cancel{t}, \mathcal{H}_t) = f^*(q \,|\, \mathcal{H}_t).$$

But more importantly, we can prove that the incoming quality has no dependence on the history at all.

Proposition 5.3. *The quality of incoming kidney is independent of the past times and marks, i.e. does not depend on the history.*

Proof. Let $n, j \in \mathbb{N}$, so that $n > j$. On the one hand, according to assumption 5.2, Y_n and Y_j are independent. The corresponding kidney qualities can be written as a functions of the donors. In the case of the KDRI quality: $Q_n = f_{KDRI}(Y_n, x_0)$ and $Q_j = f_{KDRI}(Y_j, x_0)$. Then, in the same way as in the proof of proposition 5.2, we can conclude Q_n and Q_j are independent.

On the other hand, the model assumption 5.3 means that the donor Y_n does not depend on the time T_n of its arrival. Consequently, it seems very reasonable to assume that Y_n does not depend neither on any past time of arrival of previous donors. Indeed, to consider this dependence before the dependence between Y_n and T_n would be somehow illogical, one can say the first assumption is stronger than the second. Then, the assumption 5.3 involves also that Y_n and T_j are independent. As a result, using again the result of independence from [27] (p.28), we can conclude that Q_n and T_j are independent. ■

Consequently, the function f^* we use in the model can be simplified, to not depend on the past, so it can be written depending only on the value q of the mark:

$$f^*(q \,|\, t) = f^*(q \,|\, \cancel{\mathcal{H}_t}) = f^*(q).$$

In this case, the mark CoDF is characterized as a stationary distribution. Therefore, we can choose a regular distribution that matches well the distribution of data. Taking into account all the simplifications, we will mostly use the more explicit notation f_Q to refer to the distribution of the quality, $f_Q(q) = f^*(q)$, except sometimes in theoretical calculus.

5.1.4.1 Candidate quality distributions

In order to choose the best law for the model, we did an analysis on the quality data points of patients histories. We consider three candidates distributions in total. The distribution should fit well the marks, so it should match the basic characteristics of the marks. In the work, we use the KDRI index as the quality marks, which is a positive quantity. Then, the distribution should correspond to a positive variable. As a first step, we try the Gamma and the Weibull distributions.

Gamma distribution. We try Gamma distribution $\Gamma(\theta_1, \theta_2)$ for different reasons:

- Histograms show that quality distributions throughout different histories are wave shaped and present a maximum.
- The curve is not symmetric.
- The position of the maximum can vary from the beginning of the curve to the half of the curve. The spread and length of the tail can deviate too.
- Due to its high flexibility, Gamma law can be adapted to various curve shapes from one patient to another.

Weibull distribution. Weibull distribution $\mathcal{W}(\theta_1, \theta_2)$ is also considered for the quality distribution:

- The Weibull law looks a lot like Gamma distribution, so it is also coherent to test it in light of previously pinpointed elements that justify the test of the Gamma distribution.
- The difference in effect between the 2 laws can be revealed by looking at their PDFs. Ignoring all the normalizing constants:

$$\text{For Gamma,} \quad \forall x \geq 0, \quad f(x) \propto x^{\theta_1 - 1} \exp\left(-\frac{x}{\theta_2}\right). \tag{5.5}$$

$$\text{For Weibull,} \quad \forall x \geq 0, \quad f(x) \propto x^{\theta_1 - 1} \exp\left(-\left(\frac{x}{\theta_2}\right)^{\theta_1}\right). \tag{5.6}$$

  As we can see from this, the probability density function for the Weibull distribution has the possibility to drop off much more quickly (for $\theta_1 > 1$) or slowly (for $\theta_1 < 1$) than the Gamma distribution. In the case where $\theta_1 = 1$, both distributions reduce to the exponential distribution, with parameter $\theta_2 = \mathbb{E}[X]$, if X is the associated random variable.

We plot the PDF of the two laws for different sets of parameters in Figure 5.1, to show flexibility and possibilities they represent in general.

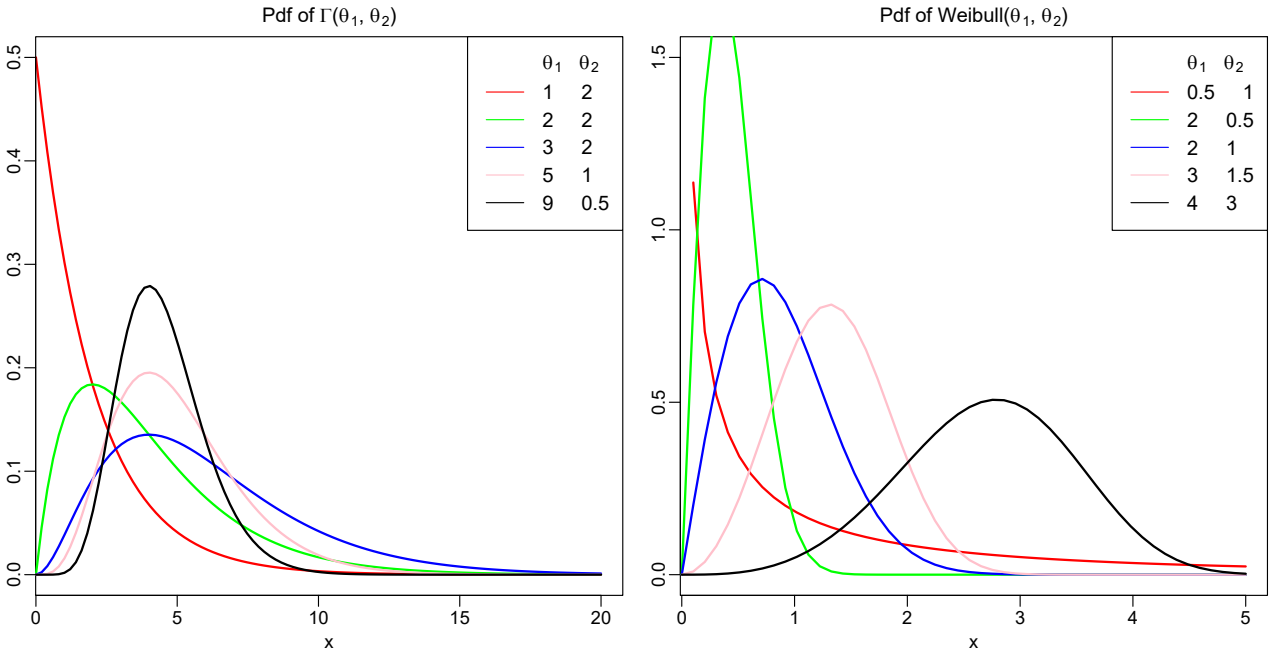


Figure 5.1 PDFs of Gamma and Weibull for different set of parameters

To better illustrate the difference between Gamma and Weibull distributions, we plot the PDF of both distributions with the same set of parameters. In Figure 5.2, for a given row, we fixed the scale parameter θ_2, and changed the shape parameter θ_1 to observe the behavior of the curves. We kept the exact same scale for the coordinate system to compare plots more easily.

Weibull law seems more capable of reaching a maximum anywhere, and seems to present more diversified forms. In particular, we can observe the possibilities offered by the θ_1 exponent inside the exponential in Weibull PDF. This exponent allows Weibull distribution to reach a more intense and earlier maximum than Gamma distribution, which presents a more flattened curve.

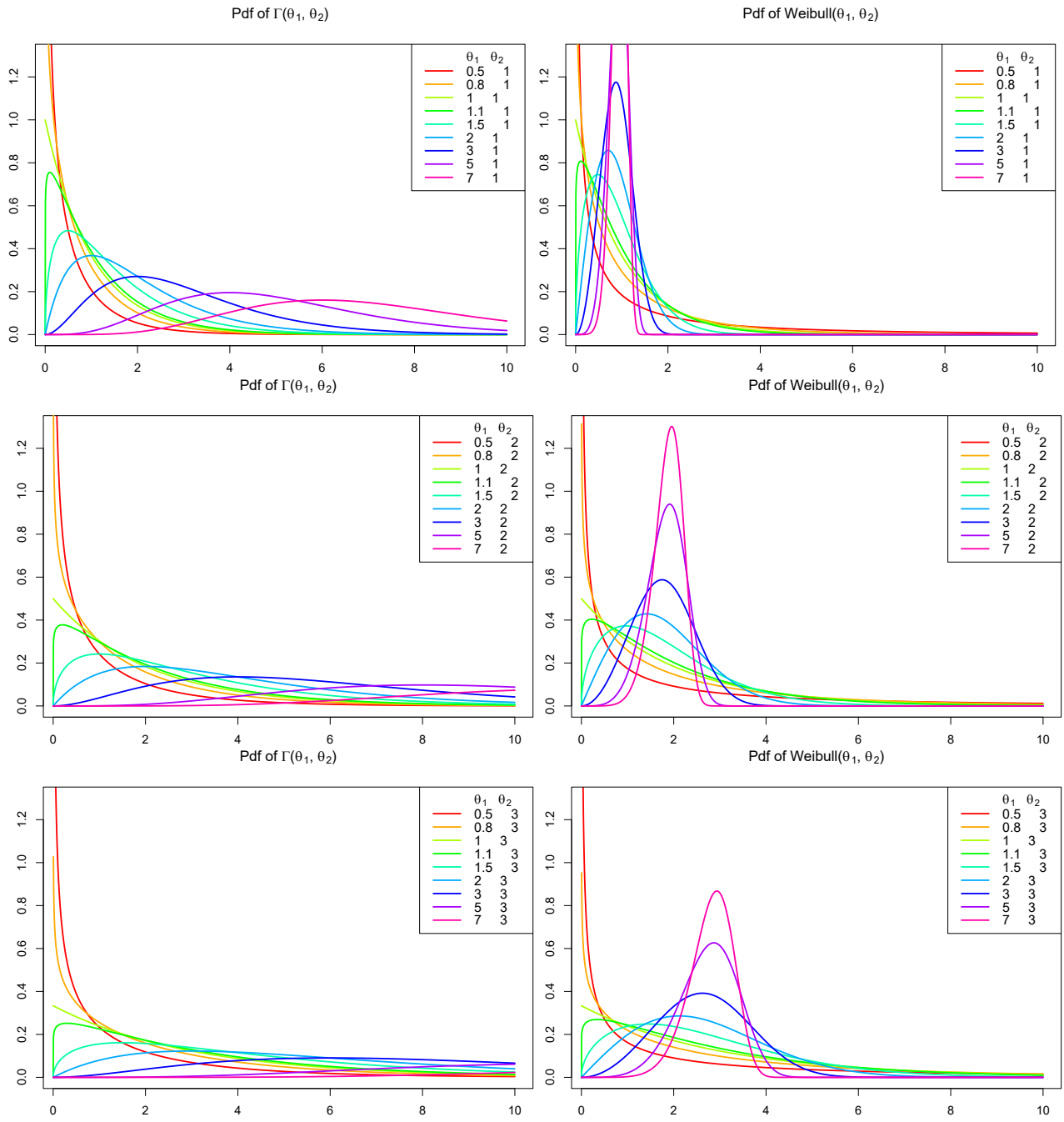


Figure 5.2 Comparison of Gamma and Weibull PDFs characteristics

Lognormal distribution. We consider also the Lognormal distribution that is denoted as Lognormal(θ_1, θ_2), with parameters $\theta_1 \in \mathbb{R}$ and $\theta_2 > 0$. This probability law is the distribution of a continuous variable whose logarithm is normally distributed. That is to say,

if $X \sim \text{Lognormal}(\theta_1, \theta_2)$, then $Y = \ln(X) \sim \mathcal{N}(\theta_1, \theta_2)$. The corresponding PDF is

$$\forall x \geq 0, \quad f(x) = \frac{1}{x\sqrt{2\pi\theta_2}} \exp\left\{-\frac{(\ln x - \theta_1)^2}{2\theta_2}\right\}. \tag{5.7}$$

We can simplify the expression by ignoring all the normalizing constants, as we did for Gamma and Weibull distributions.

$$\begin{aligned}
\forall x \geq 0,\ f(x) &\propto x^{-1} \exp\left\{-\frac{(\ln x)^2 - 2\theta_1 \ln x}{2\theta_2}\right\} \propto x^{-1} \exp\left\{-\frac{(\ln x)^2}{2\theta_2} + \frac{\theta_1}{\theta_2} \ln x\right\} \\
&\propto x^{-1} \exp\left\{-\left(\frac{\ln x}{\sqrt{2\theta_2}}\right)^2\right\} \exp\left(\frac{\theta_1}{\theta_2} \ln x\right) \\
&\propto x^{-1} \exp\left\{-\left(\frac{\ln x}{\sqrt{2\theta_2}}\right)^2\right\} x^{\frac{\theta_1}{\theta_2}}.
\end{aligned}$$

In the end, we get the non-normalized expression

$$\forall x \geq 0, \quad f(x) \propto x^{\frac{\theta_1}{\theta_2} - 1} \exp\left\{-\left(\frac{\ln x}{\sqrt{2\theta_2}}\right)^2\right\}. \tag{5.8}$$

The variable inside the exponential is a power of $\ln x$, instead of a power of x as in the two other distributions. The factor in front of the exponential is a power of x, but it depends on both distribution parameters θ_1 and θ_2, unlike for the two other distributions, where the exponent depends only on their shape parameter θ_1.

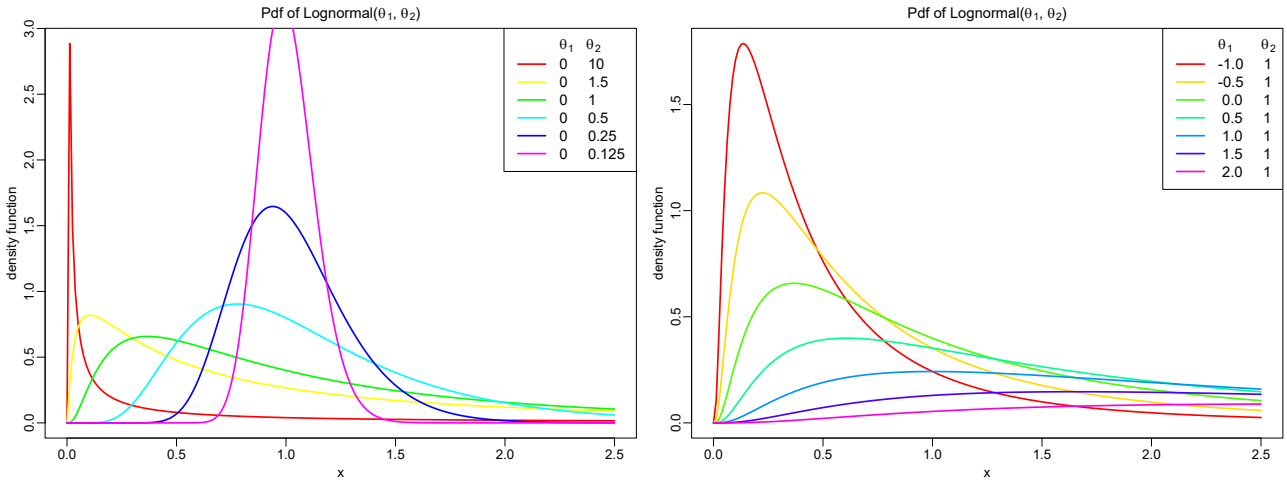


Figure 5.3 PDFs of Lognormal for different set of parameters.

This new candidate distribution can offer flexibility as well, as we show in Figure 5.3, when varying the parameters setting. The kind of shape we can obtain is quite diversified and seems as various as for Weibull distribution.

5.1.4.2 Selected quality distribution

After carrying out the selection procedures and experiments that are detailed in Section 6.4, we kept the Weibull distribution $\mathcal{W}(\theta_1, \theta_2)$, with parameters $\theta_1, \theta_2 > 0$. The corresponding mark CoDF of the MPP is then

$$f_Q(q) = f^*(q|t) = \frac{\theta_1}{\theta_2}\left(\frac{q}{\theta_2}\right)^{\theta_1 - 1} \exp\left\{-\left(\frac{q}{\theta_2}\right)^{\theta_1}\right\}. \tag{5.9}$$

The parameters of the Weibull distribution are the mark parameters of the model: $\Theta_M = (\theta_1, \theta_2)$.

5.1.5 Inference

The objective now is to infer the parameters of the model from the data. We have to estimate the parameter vector $\Theta = (\Theta_G, \Theta_M) = (\lambda, \{\theta_1, \theta_2\})$. In order to do so, we need to compute the likelihood of the process, given its history $\mathcal{H} = \{(t_j, q_j)\}_{1 \leq j \leq N_{\text{offer}}}$ on the time window $[t_0 - \Delta T, t_0]$. It is important to clarify the fact that when we handle the process described by the history, we consider the origin of time to be $t_0 - \Delta T$, so in the new time coordinates the history is recorded on $[0, \Delta T]$. This is the usual way to deal with it, since in this way both computations and notations are easier.
For all ongoing Section 5.1.5, we define $N = N_{\text{offer}}$ to simplify notation in calculations.

5.1.5.1 Likelihood

We start with the expression of the likelihood given by the proposition 3.4

$$L = \left(\prod_{j=1}^{N} \lambda^*(t_j, q_j)\right) \exp\left\{-\Lambda^*(\Delta T)\right\}.$$

We explicit the decomposition of the CIF,

$$L = \left(\prod_{j=1}^{N} \lambda_G^*(t_j) f^*(q_j|t_j)\right) \exp\left(-\int_0^{\Delta T} \lambda_G^*(s) ds\right).$$

Then, we replace the theoretical generic notations by the corresponding expressions of the model

$$L = \left(\prod_{j=1}^{N} \lambda f_Q(q_j)\right) \exp\left(-\int_0^{\Delta T} \lambda ds\right) = \lambda^N \left(\prod_{j=1}^{N} f_Q(q_j)\right) \exp(-\lambda \Delta T).$$

We can rearrange the expression so we have a product of two decorrelated factors L_{Θ_G} and L_{Θ_M}, that respectively depend on the GIF parameter $\Theta_G = \lambda$ and on the mark CoDF parameters $\Theta_M = (\theta_1, \theta_2)$

$$L = \underbrace{\lambda^N \exp(-\lambda \Delta T)}_{L_{\Theta_G}} \underbrace{\left(\prod_{j=1}^{N} f_Q(q_j)\right)}_{L_{\Theta_M}} = L_{\Theta_G}(\Theta_G | t_1, ..., t_N)\ L_{\Theta_M}(\Theta_M | q_1, ..., q_N). \tag{5.10}$$

5.1.5.2 Maximum Likelihood Estimation

Once we get the expression of the likelihood, we can find the MLE estimator by searching the maximum of the function. We have to solve the system of score equations

$$(\mathcal{S}) : \begin{cases} \frac{\partial}{\partial \Theta_G} L & = 0 \\ \frac{\partial}{\partial \Theta_M} L & = 0 \end{cases}$$

We replace L by its expression (5.10) we just found, the system becomes

$$(\mathcal{S}) \iff \begin{cases} L_{\Theta_M} \frac{\partial}{\partial \Theta_G} L_{\Theta_G}(\Theta_G | t_1, ..., t_N) & = 0 \\ L_{\Theta_G} \frac{\partial}{\partial \Theta_M} L_{\Theta_M}(\Theta_M | q_1, ..., q_N) & = 0 \end{cases}$$

Since neither the likelihood L_{Θ_G} nor L_{Θ_M} are null, we have

$$(\mathcal{S}) \iff \begin{cases} \frac{\partial}{\partial \Theta_G} L_{\Theta_G}(\Theta_G | t_1, ..., t_N) & = 0 \qquad (\mathcal{E}_G) \\ \frac{\partial}{\partial \Theta_M} L_{\Theta_M}(\Theta_M | q_1, ..., q_N) & = 0 \qquad (\mathcal{E}_M) \end{cases}$$

Solving the system is equivalent to solving both equations independently: $(\mathcal{E}_G)$ to find the maximum of L_{Θ_G}, and $(\mathcal{E}_M)$ to find the maximum of L_{Θ_M}.

For the first equation $(\mathcal{E}_G)$, we can rewrite it

$$(\mathcal{E}_G) : \frac{\partial}{\partial \Theta_G} L_{\Theta_G}(\Theta_G | t_1, ..., t_N) = \frac{\partial}{\partial \lambda} \left[\lambda^N \exp(-\lambda \Delta T)\right] = 0.$$

We recognize the equation we solved previously for the homogeneous Poisson process in

Example 3.2. Then, we get the same result as in Equation (3.9)

$$\widehat{\Theta_G}^{MLE} = \widehat{\lambda}^{MLE} = \frac{N}{\Delta T}.$$

To finish to solve the system $(\mathcal{S})$, we still have to solve the equation $(\mathcal{E}_M)$. We start by identifying that $L_{\Theta_M} = \prod_{j=1}^{N} f_Q(q_j)$ is the likelihood function of the sample $(q_1, .., q_N)$ of the qualities, which are *i.i.d.* and follow the distribution $\mathcal{W}(\theta_1, \theta_2)$.
At the same time, we can rewrite the left-hand side of equation $(\mathcal{E}_M)$ as follows

$$\frac{\partial}{\partial \Theta_M} L_{\Theta_M}(\Theta_M|q_1, ..., q_N) = \nabla_{\Theta_M}\left[L_{\Theta_M}(\Theta_M|q_1, ..., q_N)\right] = \begin{bmatrix} \frac{\partial}{\partial \theta_1} L_{\Theta_M(\Theta_M|q_1,...,q_N)} \\ \frac{\partial}{\partial \theta_2} L_{\Theta_M(\Theta_M|q_1,...,q_N)} \end{bmatrix}.$$

Therefore, to solve the equation $(\mathcal{E}_M)$ is equivalent to solve the system $(\mathcal{S}_M)$ we define below

$$(\mathcal{E}_M) : \frac{\partial}{\partial \Theta_M} L_{\Theta_M}(\Theta_M|q_1, ..., q_N) = 0 \iff (\mathcal{S}_M) : \begin{cases} \frac{\partial}{\partial \theta_1} L_{\Theta_M(\Theta_M|q_1,...,q_N)} = 0 \\ \frac{\partial}{\partial \theta_2} L_{\Theta_M(\Theta_M|q_1,...,q_N)} = 0 \end{cases}$$

We notice that the system $(\mathcal{S}_M)$ leads to maximizing the likelihood in terms of all its parameters, which is exactly as same as performing the MLE method.
Knowing this, and the fact that L_{Θ_M} is the likelihood of a Weibull $\mathcal{W}(\theta_1, \theta_2)$ sample, we can conclude that solving $(\mathcal{E}_M)$ is equivalent to finding the MLE estimators of the parameters of the Weibull distribution $\mathcal{W}(\theta_1, \theta_2)$.
It is possible to find the expression of the estimated parameters in the literature, but in practice we directly use packages already implemented in R/Python language [35] and [36], which include MLE method for Weibull distribution.

In conclusion, we proved that the estimator $\widehat{\Theta}^{MLE}$ is composed of

- The MLE estimator of L_{Θ_G}, for the rate of the process: $\widehat{\Theta_G}^{MLE} = \widehat{\lambda}^{MLE}$.
- The MLE estimator of L_{Θ_M}, for the quality distribution: $\widehat{\Theta_M}^{MLE} = \widehat{(\theta_1, \theta_2)}^{MLE}$.

Remark 5.5 (Notation). From now on, we use the notation $\widehat{\lambda}_1 = \widehat{\lambda}^{MLE}$ to refer to the estimator of the parameter $\Theta_G = \lambda$ of the process GIF, also known as rate of the process. This estimator is the same one as the one used in prior work [1]. □

5.1.5.3 Introducing another rate estimator $\widehat{\lambda_2}$

Throughout our experiments, we noticed that the estimator $\widehat{\lambda_1}$, the MLE, presented some limitations. In order to change the way we estimate the rate λ of the process, we propose a

new estimation method, referred to as Method 2, that is based on the estimator $\widehat{\lambda_2}$.

Definition 5.1 (Method 2 estimator $\widehat{\lambda}_2$)**.** The Method 2 estimator of the process rate is

$$\widehat{\lambda}_2 = \frac{1}{\frac{1}{N}\sum_{j=1}^{N} \tau_j}, \tag{5.11}$$

where

- ▷ The $\{\tau_j\}_{0\leq j\leq N}$ are the observed interevent waiting times between 2 consecutive offers on the history $\mathcal{H}$ of the patient, $\forall j \in 0, ..., N$, $\tau_j = t_{j+1} - t_j$. (The first interevent time τ_0 does not appear in the expression (5.11) of $\widehat{\lambda}_2$).
- ▷ The number of eligible offers on the history is $N = N_{\text{offer}}$.
- ▷ We take as a convention the time $t_{N+1} = \Delta T$ as the end of the history, which corresponds to the time of initial offer, so that $\tau_N = t_{N+1} - t_N = \Delta T - t_N$ does represent the last interevent waiting time.

So far, the estimator was $\widehat{\lambda}_1 = \frac{N_{\text{offer}}}{\Delta T}$. It is a temporal mean, similar to a frequency, composed of the number of events divided by the timespan during which we count the events. In the new method we consider, we use instead the mean of the observed interevent waiting times in the history: $\frac{1}{N}\sum_{j=1}^{N} \tau_j$. The estimated rate $\widehat{\lambda}_2$ is the reciprocal of this mean, which gives the right unit: [Time Unit]$^{-1}$.

We give an illustration of a patient history in Figure 5.4 to understand better the meaning of all time notations.

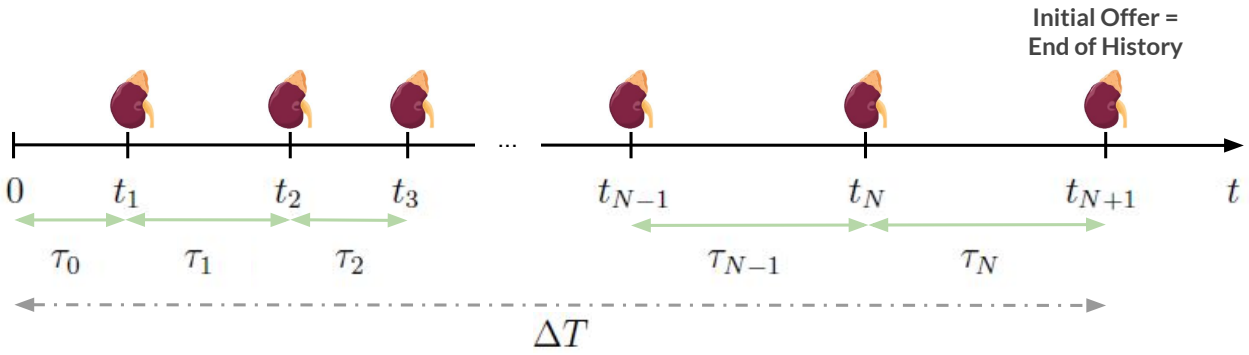


Figure 5.4 A patient history $\mathcal{H}$

5.1.5.4 Intuitive explanation

We are reasoning for one initial offer in all the calculation part, especially for mathematical proofs. For the initial patient, we have the history $\mathcal{H}$ we built, with the eligible offers that were proposed to the patient. We retrieve all the times $\{\tau_j\}_{1\leq j\leq N}$ between 2 consecutive offers. The rationale that supports the coherence of the new estimator is that the mean $\bar{\tau}_N = \frac{1}{N}\sum_{j=1}^{N}\tau_j$ of these times is supposed to tend (for a large sample $\{\tau_j\}_{1\leq j\leq N}$, *i.e.* for a large number N of eligible offers in the history) to the expected waiting time before the next offer $\mathbb{E}[T]$. Indeed, it is well known that for a sample mean such as $\bar{\tau}_N$ and an *i.i.d.* sample $\{\tau_j\}_{1\leq j\leq N}$,

$$\mathbb{E}[\bar{\tau}_N] = \mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N}\tau_j\right] = \frac{1}{N}\sum_{j=1}^{N}\mathbb{E}[\tau_j] \underset{i.i.d.}{=} \frac{1}{N}\sum_{j=1}^{N}\mathbb{E}[\tau_1] = \mathbb{E}[\tau_1].$$

Since the interevent time T before next event is also distributed like the time τ_1 (both are interevent time variables of the process), we get

$$\mathbb{E}[\bar{\tau}_N] = \mathbb{E}[\tau_1] = \mathbb{E}[T] = \frac{1}{\lambda},$$

where the last equality comes from the exponential distribution of interevent waiting times of a Poisson process. This result is one of the consequences of the fundamental law of large number [37] (p.104), which also gives the almost sure convergence of $\bar{\tau}_N$ to its expected value $\frac{1}{\lambda}$. Based on that, it seems fair to assume that the inverse of this quantity, the estimator $\widehat{\lambda}_2 = \frac{1}{\bar{\tau}_N}$, converges to the inverse of the limit, which is exactly $\widehat{\lambda}_2 \underset{N\to\infty}{\longrightarrow} \lambda$.

However, if this intuition is helpful to find the expression of $\widehat{\lambda}_2$ in the first place, it needs to be proven rigorously. Beyond the fact that the point convergence is a required property for an estimator, it is necessary to check other properties to conclude to its relevance.

5.1.5.5 Mathematical validity

In this section, we prove the mathematical validity of this new estimation method for the process rate λ as below

- We first determine the bias of the estimator. We prove the estimator to be a little biased, but asymptotically unbiased.
- We compute the variance and the Mean Square Error (MSE) of the estimator. We prove that the new estimator is consistent.

For all the following mathematical proofs, we may refer to $\widehat{\lambda}_2$ simply as $\widehat{\lambda}$ in calculus, in order to simplify notations.

Proposition 5.4 (Bias of of the estimator $\widehat{\lambda}_2$)**.**
The bias of the estimator $\widehat{\lambda}_2$ is

$$Bias\left[\widehat{\lambda}_2\right] = \frac{\lambda}{N-1}. \tag{5.12}$$

Which means $\widehat{\lambda}_2$ is asymptotically unbiased.

Proof. We use $\widehat{\lambda}$ to refer to the estimator $\widehat{\lambda}_2$. We suppose that the size of the sample is $N \geq 2$. Then,

$$\mathbb{E}\left[\widehat{\lambda}\right] = \mathbb{E}\left[\frac{1}{\frac{1}{N}\sum_{j=1}^{N}\tau_j}\right] = N\ \mathbb{E}\left[\frac{1}{\sum_{j=1}^{N}\tau_j}\right] = N\ \mathbb{E}\left[g\left(\sum_{j=1}^{N}\tau_j\right)\right],$$

where, $g(x) = \frac{1}{x}\,,\ \forall x > 0$. Moreover, the $\tau_1, ..., \tau_N$ are independent and equally distributed since $\forall j,\ \tau_j \sim \mathcal{E}xp(\lambda)$. Hence, by property (See [27] p.145), we have that

$$\sum_{j=1}^{N}\tau_j \sim \Gamma(N, \lambda),$$

where $\Gamma(N, \lambda)$ stands for the Gamma distribution with shape parameter N and scale parameter λ. Also, we have that for a variable $Z \sim \Gamma(N, \lambda)$, the variable $g(Z) = 1/Z \sim$ Inv-Gamma(N, λ) according to [38]. In this case, we can calculate the composed expectation

$$\mathbb{E}\left[g(Z)\right] = \frac{\lambda}{N-1}.$$

Using this intermediate result, we finally get the expected value of the estimator

$$\mathbb{E}\left[\widehat{\lambda}\right] = N\,\mathbb{E}\left[g(Z)\right] = \frac{N}{N-1}\lambda.$$

As a consequence, we can determine the bias of the estimator

$$Bias\left[\widehat{\lambda}\right] = \mathbb{E}\left[\widehat{\lambda}\right] - \lambda = \frac{N}{N-1}\lambda - \lambda = \frac{1}{N-1}\lambda.$$

The estimator has a positive bias, proportional to the true value of λ, so its expected value is greater than λ. But the bias decreases with the size N of the sample, and can rapidly be insignificant. Indeed,

$$Bias\left[\widehat{\lambda}\right] = \frac{1}{N-1}\lambda \underset{N\to\infty}{\longrightarrow} 0.$$

The estimator is said to be asymptotically unbiased. ■

In practice, as soon as we get a sample of size $N \geq 21$, we have a good upper bound on the relative error

$$\text{Relative Error} = \frac{\mathbb{E}\left[\widehat{\lambda}\right] - \lambda}{\lambda} = \frac{Bias\left[\widehat{\lambda}\right]}{\lambda} = \frac{\frac{1}{N-1}\lambda}{\lambda} = \frac{1}{N-1} \leq \frac{1}{21-1} = 5\%.$$

Proposition 5.5 (Variance of the estimator $\widehat{\lambda}_2$)**.**
The variance of the estimator $\widehat{\lambda}_2$ is

$$\mathbb{V}ar\left[\widehat{\lambda}_2\right] = \frac{N^2}{(N-1)^2(N-2)}\lambda^2. \tag{5.13}$$

Proof. We use $\widehat{\lambda}$ to refer to the estimator $\widehat{\lambda}_2$. We suppose that the size of the sample is $N \geq 3$. The variance of any random variable can be decomposed in this way

$$\mathbb{V}ar\left[\widehat{\lambda}\right] = \mathbb{E}\left[\widehat{\lambda}^{\,2}\right] - \mathbb{E}\left[\widehat{\lambda}\right]^2.$$

We compute the two terms that compose the variance.

▷ We already computed $\mathbb{E}\left[\widehat{\lambda}\right]$ earlier, so the second term equals $\mathbb{E}\left[\widehat{\lambda}\right]^2 = (\frac{N}{N-1})^2\lambda^2$.

▷ For the second momentum, $\mathbb{E}\left[\widehat{\lambda}^{\,2}\right] = N^2\ \mathbb{E}\left[\left(\frac{1}{\sum_{j=1}^{N}\tau_j}\right)^2\right] = N^2\ \mathbb{E}\left[g\left(\sum_{j=1}^{N}\tau_j\right)\right]$,
where, $g(x) = \frac{1}{x^2}\,,\ \forall x > 0$. Moreover, the $\tau_1, ..., \tau_N$ are independent and equally distributed since $\forall j,\ \tau_j \sim \mathcal{E}xp(\lambda)$. Hence, by property (See [27] p.145), we have that

$$\sum_{j=1}^{N}\tau_j \sim \Gamma(N, \lambda).$$

Also, we have that for a variable $Z \sim \Gamma(N, \lambda)$, the variable $1/Z \sim$ Inv-Gamma(N, λ). In this case, we can calculate the moment of second order of $1/Z$ using the formula given in [38]

$$\mathbb{E}\left[g(Z)\right] = \mathbb{E}\left[\frac{1}{Z^2}\right] = \frac{\lambda^2}{(N-1)(N-2)}.$$

Hence, we get the the second momentum $\mathbb{E}\left[\widehat{\lambda}^{\,2}\right] = \frac{N^2}{(N-1)(N-2)}\lambda^2$.

Using both transitional results, we finally get

$$\begin{aligned}\mathbb{V}ar\left[\widehat{\lambda}\right] &= \frac{N^2}{(N-1)(N-2)}\lambda^2 - \left(\frac{N}{N-1}\right)^2\lambda^2 \\ &= \frac{N^2\left[(N-1)-(N-2)\right]}{(N-1)^2(N-2)} = \frac{N^2}{(N-1)^2(N-2)}\lambda^2.\end{aligned}$$

■

Proposition 5.6 (MSE and consistency of the estimator $\widehat{\lambda}_2$)**.**
The MSE of the estimator $\widehat{\lambda}_2$ is

$$MSE(\widehat{\lambda}_2) = \frac{N+2}{(N-1)(N-2)}\lambda^2. \tag{5.14}$$

Furthermore, $MSE(\widehat{\lambda}_2) \underset{N\to\infty}{\longrightarrow} 0$ which means that $\widehat{\lambda}_2$ is a consistent estimator of the process rate λ.

Proof. We use $\widehat{\lambda}$ to refer to the estimator $\widehat{\lambda}_2$. Using the values of the bias and the variance we determined, we can compute the MSE of our estimator as

$$\begin{aligned}MSE(\widehat{\lambda}) &= \left(Bias\left[\widehat{\lambda}\right]\right)^2 + \mathbb{V}ar\left[\widehat{\lambda}\right] = \left(\frac{1}{N-1}\lambda\right)^2 + \frac{N^2}{(N-1)^2(N-2)}\lambda^2 \\ &= \frac{N^2+N-2}{(N-1)^2(N-2)}\lambda^2 = \frac{(N-1)(N+2)}{(N-1)^2(N-2)}\lambda^2 = \frac{N+2}{(N-1)(N-2)}\lambda^2 \underset{N\to\infty}{\longrightarrow} 0.\end{aligned}$$

A MSE that tends to 0 is equivalent to the consistency of the estimator (it can be proved with the squeeze theorem using Markov inequality for the upper bound, the lower bound being 0 [39]). Consequently, we can say that $\widehat{\lambda}$ is a consistent estimator of λ. ■

In conclusion, we proved that this estimator is appropriate to estimate the rate of the process. Furthermore, we will show in the experiments Section 6.3.2 that its expression is really coherent when compared to Method 1 estimator. More importantly, we will also illustrate how it addresses some problems we could have encountered with Method 1.

5.2 Marked Point Process predictions

Once we got the MPP defined, including the estimations of the parameters, we can use it to study properties and characteristic values of the process. In our case, we are interested in summary statistics like the expected waiting time before next offer $\mathbb{E}[T]$, the expected waiting time before next better offer $\mathbb{E}[T|Q > q_0]$, and also confidence intervals for both these values.
Firstly, we will show how to get exact results from the MPP we fit. And secondly, we will explain how to use simulation to get empirical results and in what situation it can be relevant.

5.2.1 Exact prediction: Time before Next Offer

In this part, we are only interested in the time before the next event happens, T. As explained previously in Chapter 4, we do not need the quality part of the MPP method to answer that question, referred to as TNO. The expected TNO is the expectation of $T \sim \mathcal{E}xp(\lambda)$. For an exponentially distributed variable, the expected value denoted as μ is given by (cf. [40])

$$\mu = \mathbb{E}[T] = \frac{1}{\lambda}. \tag{5.15}$$

To estimate this expectation, we use the estimator $\widehat{\lambda}_1$ of the rate we found in Section 5.1.5.2. This gives us the prediction $\widehat{\mu}$ of the expected TNO

$$\widehat{\mu} = \widehat{\mathbb{E}[T]} = \frac{1}{\widehat{\lambda}_1} = \frac{\Delta T}{N_{\text{offer}}}, \tag{5.16}$$

with $N_{\text{offer}} = \text{Card}(\mathcal{H})$, the number of offers in the history.

Concerning confidence intervals for the prediction, for $\alpha \in]0, 1[$ a level of confidence, we want to find an upper bound for the waiting time T. Then, we search for an interval with form $CI_\alpha = \left[0, \widetilde{t}_\alpha\right]$, so that we have $\mathbb{P}(T \in CI_\alpha) = \alpha$. Consequently, we can write

$$\alpha = \mathbb{P}(T \in CI_\alpha) = \mathbb{P}(0 \leq T \leq \widetilde{t}_\alpha) = F_T(\widetilde{t}_\alpha).$$

We can conclude that $\widetilde{t}_\alpha$ is the α-quantile of the exponential distribution $\mathcal{E}xp(\lambda)$ of T, which is denoted as $q^\alpha_{\mathcal{E}(\lambda)}$. Consequently,

$$CI_\alpha = \left[0, q^\alpha_{\mathcal{E}(\lambda)}\right]. \tag{5.17}$$

To compute the quantile in practice, we simply use $\widehat{\lambda}_1$ as an estimation of λ.

5.2.2 Exact prediction: Time before Next Better Offer

In this section, the objective is to determine the time before the next better offer happens. This involves using both the rate of the process and the quality aspect of the MPP method to answer the question, referred to as TNBO.
Let T^+ be the random variable corresponding to $T|Q > q_0$, that is to say, the time before a next better offer. By using the definition of an event-conditioned density function, we can write the expected waiting time before next better offer as

$$\mathbb{E}\left[T^+\right] = \mathbb{E}\left[T|Q > q_0\right] = \int_0^{+\infty} t\ f_{T|Q>q_0}(t)\,dt = \int_0^{+\infty} t\ \frac{f_T(t)}{\mathbb{P}(Q > q_0)}\,dt \tag{5.18}$$

$$= \frac{1}{\mathbb{P}(Q > q_0)} \underbrace{\int_0^{+\infty} t\ f_T(t)\,dt}_{\mathbb{E}[T]} = \frac{\mathbb{E}\left[T\right]}{\mathbb{P}(Q > q_0)}. \tag{5.19}$$

For the exponentially distributed time T^+, we have the relation $\mathbb{E}\left[T^+\right] = 1/\lambda^+$ that links the expected value to the exponential parameter $\lambda^+ > 0$ (see relation (5.15)). As a consequence, the rate λ^+ of the new process defined by the interevent time T^+ is

$$\lambda^+ = \frac{1}{\mathbb{E}\left[T^+\right]} = \mathbb{P}(Q > q_0)\frac{1}{\mathbb{E}\left[T\right]} = \mathbb{P}(Q > q_0)\,\lambda. \tag{5.20}$$

We can see how the new rate λ^+ is proportional to the original rate λ. Then, the new process defined by T^+, the one that counts the arrival of better kidney than the current one, is a thinned process derived from the original one. The thinning factor is $\lambda^+/\lambda = \mathbb{P}(Q > q_0)$. As a result, the higher the quality threshold q_0 is for the next offer to come, the more important the thinning of the process is.
The Equation (5.20) shows that in order to approximate λ^+, it is necessary to have an estimation for each one of the two following quantities

- ▷ The original rate λ.
- ▷ The thinning factor $\mathbb{P}(Q > q_0)$. Let us denote it as $\rho^+ = \mathbb{P}(Q > q_0)$, so we can rewrite the thinning relation (5.20) as

$$\lambda^+ = \rho^+\lambda. \tag{5.21}$$

Concerning the approximation of the original rate λ, we use the MLE estimator $\widehat{\lambda}_1$ we found in Section 5.1.5.2

$$\widehat{\lambda}_1 = \frac{N_{\text{offer}}}{\Delta T}. \tag{5.22}$$

Concerning the thinning factor ρ^+, we can first rewrite it in terms of the quality CDF F_Q

$$\rho^+ = \mathbb{P}(Q > q_0) = \int_{q_0}^{+\infty} f_Q(q)\, dq = 1 - F_Q(q_0). \tag{5.23}$$

Hence, we can estimate ρ^+ by the following estimator $\widehat{\rho^+}$

$$\widehat{\rho^+} = 1 - F_Q^{\approx}(q_0), \tag{5.24}$$

where the approximation $F_Q^{\approx}$ (of the real F_Q) is the CDF corresponding to the quality distribution of the MPP we fitted earlier in Section 5.1.5.2. That is to say, $F_Q^{\approx}$ is the CDF of the Weibull distribution with parameters $\widehat{\Theta_M}^{MLE} = \widehat{(\theta_1, \theta_2)}^{MLE}$.

Finally, combining the two estimations from Equation (5.22) and Equation (5.24), we can rewrite Equation (5.21) to get the estimator $\widehat{\lambda^+}$ of the rate λ^+ of the thinned process as

$$\widehat{\lambda^+} = \widehat{\rho^+}\, \widehat{\lambda}_1 = \left(1 - F_Q^{\approx}(q_0)\right)\ \frac{N_{\text{offer}}}{\Delta T}. \tag{5.25}$$

In the same way we did in equation 5.16 for the prediction $\widehat{\mu}$ of the TNO, the predicted waiting TNBO is given by the following estimator

$$\widehat{\mu^+} = \widehat{\mathbb{E}\left[T^+\right]} = \frac{1}{\widehat{\lambda^+}} = \frac{1}{\widehat{\rho^+}\, \widehat{\lambda}_1} = \frac{1}{1 - F_Q^{\approx}(q_0)}\ \frac{\Delta T}{N_{\text{offer}}}. \tag{5.26}$$

For the confidence intervals of the prediction, this is exactly the same situation as for Equation (5.17), but with the thinned rate λ^+ instead

$$CI_\alpha = \left[0, q^\alpha_{\mathcal{E}(\lambda^+)}\right], \tag{5.27}$$

with $q^\alpha_{\mathcal{E}(\lambda^+)}$ the α-quantile of the exponential distribution $\mathcal{E}xp(\lambda^+)$. To compute it in practice, we simply replace λ^+ by its estimation $\widehat{\lambda^+}$.

Remark 5.6 (KDRI quality). We passed on this detail in this section, but in the specific case of the KDRI indicator being the quality, the smaller the KDRI is, the better the quality is. Then, to have a better offer, we need to filter so that $q^{+,\,KDRI} < q_0^{KDRI}$. Hence, the thinning factor is in reality

$$\mathbb{P}(Q > q_0) = \mathbb{P}(Q^{KDRI} < q_0^{KDRI}) = \int_0^{q_0^{KDRI}} f_{Q^{KDRI}}(q)\, dq = F_{Q^{KDRI}}(q_0^{KDRI}).$$

However, to avoid any confusion, we keep using the more meaningful former notation, in

which a better quality is a higher quality: $q^+ > q_0$.
To be more specific, when detailing a calculus for the particular case of KDRI quality for instance, we will add the corresponding exponent notation: Q^{KDRI}. □

5.2.2.1 Mathematical proof of thinning via process decomposition

The proof we brought previously in Section 5.2.2 is sufficient if we are just interested in the next event at T^+, since we found its distribution. However, this does not stand for the next interevent waiting times that come after T^+, and neither for the nature of the new process we create.
Then, we want to prove rigorously that thinning the original process as we did really results in a Poisson process, with new rate λ^+, and with the interevent waiting time T^+. To do so, we apply the Poisson process decomposition property detailed in Theorem 3.1.

Proposition 5.7 (Poisson process decomposition applied to kidney offer problem)**.** *The point process denoted as* $\{N^+(t), t \geq 0\}$ *that counts only the events such that* $Q > q_0$*, is a homogeneous Poisson process of rate* $\lambda^+ = \rho^+\lambda$*. Its interevent waiting time is the variable* $T^+ = T|Q > q_0$ *and follows the exponential distribution* $\mathcal{E}xp(\lambda^+)$*.*

Proof. Let $\{N(t), t \geq 0\}$ be the general Poisson process of rate $\lambda > 0$ that represents the arrivals of kidney offers. The marked events it counts are the arrivals of eligible donors for patient x_0, that is to say each offer w_j. We define the types introduced in proposition 3.1 as follows:

- Type 1 corresponds to the events for which $Q > q_0$.
- Type 2 corresponds to the other events, the ones for which $Q \leq q_0$.

Because of propositions 5.2 and 5.3, the quality of the incoming kidney is independent of the time being, as well as of the past times and marks. Then, events are classified, independently of the other events, and independently of the time, as follows:

- Type 1 with a probability $p_1 = \mathbb{P}(Q > q_0)$.
- Type 2 with a probability $p_2 = \mathbb{P}\left(\overline{Q > q_0}\right) = 1 - p_1$.

As a consequence, the hypotheses of the proposition 3.1 are fulfilled and we get the result: $\{N_1(t), t \geq 0\}$ is a homogeneous Poisson process with rate $\lambda_1 = p_1\lambda$, which counts the events for which $Q > q_0$.
In fact, the type-1 events process $\{N_1(t), t \geq 0\}$ matches the definition of $\{N^+(t), t \geq 0\}$ given in the property statement. They both count the same events, so they are the same process in reality. In parallel, we remind of the notation $\rho^+ = \mathbb{P}(Q > q_0) = p_1$. Then, we

have the first part of the result: $\{N^+(t), t \geq 0\}$ is a homogeneous Poisson process of rate $\lambda^+ = \lambda_1 = p_1\lambda = \rho^+\lambda$.

In addition, we have (by definition) that the time before next better offer $T^+ = T|Q > q_0$ is exactly the first interevent waiting time of the process $\{N^+(t), t \geq 0\}$. Since in a Poisson process all interevent waiting times are *i.i.d.*, this proves the second part of the result: the interevent waiting time of $\{N^+(t), t \geq 0\}$ is the variable $T^+ = T|Q > q_0 \sim \mathcal{E}xp(\lambda^+)$. ■

In conclusion, the proposition 5.7 proves we can still use a Poisson process to model the arrivals of next better offers. Furthermore, it proves also the validity of previous Section 5.2.2 results. Namely, we can estimate the rate of the thinned process $\{N^+(t), t \geq 0\}$ with the relation (5.25)

$$\widehat{\lambda^+} = \widehat{\rho^+}\,\widehat{\lambda}_1 = \left(1 - F^{\approx}_Q(q_0)\right)\ \frac{N_{\text{offer}}}{\Delta T}.$$

And consequently, it is possible to give as a prediction of the waiting TNBO the estimator from Equation (5.26)

$$\widehat{\mu^+} = \widehat{\mathbb{E}\left[T^+\right]} = \frac{1}{1 - F^{\approx}_Q(q_0)}\ \frac{\Delta T}{N_{\text{offer}}}.$$

Remark 5.7. In this section, we clearly made a difference between the real value of a quantity and the estimator we use to approximate it. For instance, the true value of the process rate λ and the estimator $\widehat{\lambda}_1$. This was necessary to distinguish one from the other in this theoretical section where mathematical results are proven. However, in others section for which this distinction is not essential, we will directly use the notation that refers to the true quantity, *e.g.* λ to refer to the rate of the process, even if in practice we do not know its value and use instead the corresponding estimator $\widehat{\lambda}_1$. This remark particularly applies in the following Section 5.2.2.2. □

5.2.2.2 Illustration: example of thinned Poisson process

To illustrate the principle of the thinning, we apply it for one of the patients. In this example, the rate of the original process is $\lambda = 0.059$, and the thinning factor is $\rho^+ = \mathbb{P}(Q > q_0) = 0.351$. Hence, we can compute the new rate $\lambda^+ = 0.059 \times 0.351 = 0.0207$. We recap the figures in Table 5.1, where we add the corresponding expected waiting time before next event, for both original and thinned processes.

The Figure 5.5 displays the distribution of the quality, and represents the thinning factor as the area under the curve on the left side of the red vertical line (remember we chose KDRI quality, the smaller the better, as reminded in remark 5.6). The quality distribution is the Weibull law $\mathcal{W}(\theta_1, \theta_2)$ with parameter vector $(\theta_1, \theta_2) = (3.62, 1.70)$.

Table 5.1 Thinning process: example for one patient

CAN_ID	λ	$\rho^+ = \mathbb{P}(Q > q_0)$	λ^+	$\mathbb{E}[T]$	$\mathbb{E}[T^+]$
255	0.059	0.351	0.0207	17	48

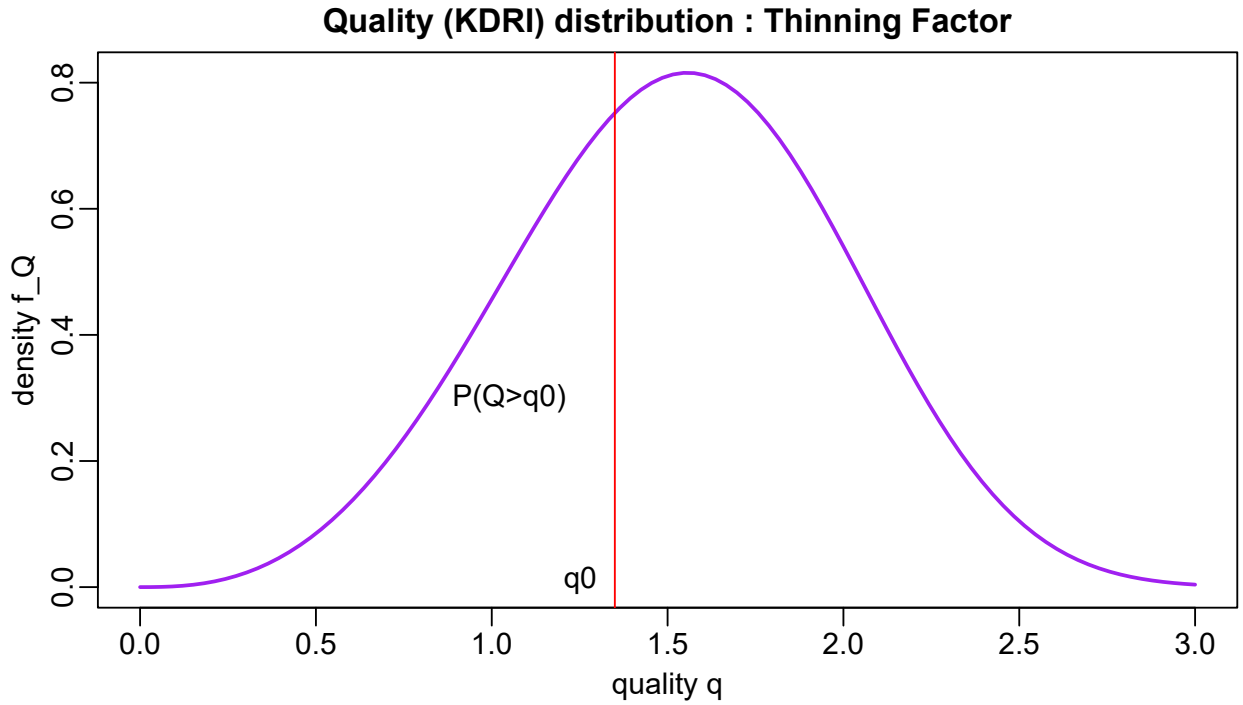


Figure 5.5 Quality distribution and thinning factor

The distribution of the waiting time T before next offer (before thinning), and the resulting distribution of the waiting time T^+ before next better offer (after thinning) are shown in Figure 5.6. The two vertical dashed lines represent the respective expected values $\mathbb{E}[T]$ and $\mathbb{E}[T^+]$ of the waiting times. Naturally, we can see that the one from the thinned process is larger than the original one. For later use, we indicate also that the 95% confidence interval for the time T (defined in Equation (5.17)) is $CI_{95\%} = [0, 50.7]$.

5.2.3 Simulation

With the model we developed we do not necessarily need simulation, since the model already enables us to find closed form results directly through theoretical calculus. Also, since in our case we have access to the exact distribution of the next interevent times (all *i.i.d.* for a Poisson process), we could simulate each event through a simple draw from the exponential

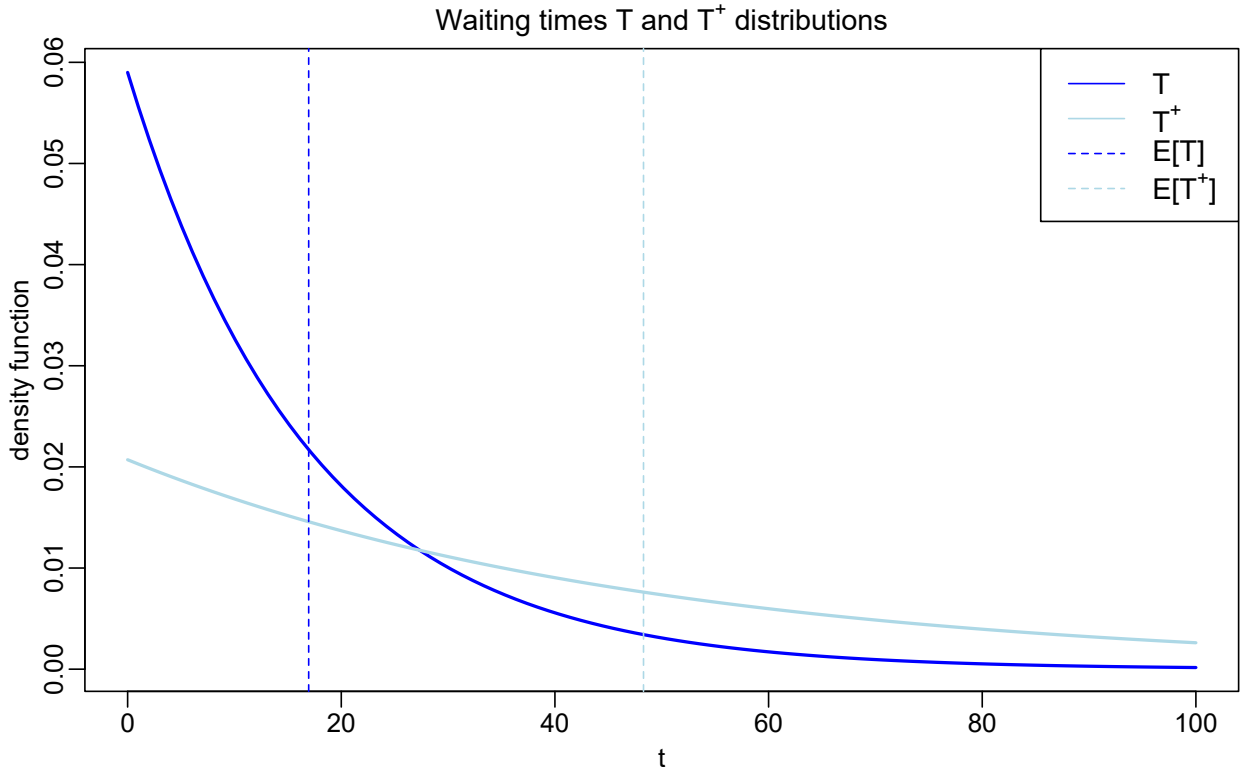


Figure 5.6 Waiting time for original and thinned Poisson process

distribution $\mathcal{E}xp(\lambda)$ followed by the draw of the corresponding quality from the mark CoDF, as explained in Section 3.2.4.2.

However, these options are not always possible in other models. Simulation is useful in many others less special cases, including if we were to modify the model's GIF or mark CoDF, which could happen if we wanted to bring improvements to the current model. For that reason, we see in this section how we can find empirically quite the same results as with closed form calculus.

5.2.3.1 Application of the inverse method

To perform simulation, we resort to the inverse method presented in Section 3.2.4.2. To use the proposition 3.6 and be able to perform the time transformation, we need to find the inverse Λ^{*-1} of the iGIF. With the current MPP model, we can have access to its closed form. First, for $\tilde{t} > 0$, the expression of the iGIF is

$$\Lambda^*(\tilde{t}) = \int_0^{\tilde{t}} \lambda_G^*(s)ds = \int_0^{\tilde{t}} \lambda \, ds = \lambda \, \tilde{t}. \tag{5.28}$$

The expression of Λ^* and of its inverse Λ^{*-1} are linked with the following fundamental relation

$$\forall \widetilde{t}, \widetilde{s} \in \mathbb{R}_+ , \qquad \Lambda^*(\widetilde{t}) = \widetilde{s} \iff \widetilde{t} = \Lambda^{*-1}(\widetilde{s}). \tag{5.29}$$

Then, to find the inverse function of Λ^*, we work on its expression until we isolate the variable $\widetilde{t}$ and express it in terms of a function of $\widetilde{s}$. Let $\widetilde{t}, \widetilde{s} \in \mathbb{R}_+$, using the expression in Equation (5.28)

$$\Lambda^*(\widetilde{t}) = \widetilde{s} \iff \lambda \widetilde{t} = \widetilde{s} \iff \widetilde{t} = \frac{\widetilde{s}}{\lambda}. \tag{5.30}$$

Hence, we are able to identify the inverse GIF as the function

$$\Lambda^{*-1} : \mathbb{R}_+ \to \mathbb{R}_+ \; ; \; \widetilde{s} \mapsto \frac{\widetilde{s}}{\lambda}. \tag{5.31}$$

Once we have the function Λ^{*-1}, we are able to apply the inverse method. The algorithm 1 describes the simulation procedure step by step.

Algorithm 1: Simulation of a MPP by inverse method

Input: Time limit $\widetilde{T}$ of the simulation.

1. **Initialization:** Set $t = 0$, $s_0 = 0$ and $j = 1$.
2. **While** $t < \widetilde{T}$ **do**
   (a) Generate $s_j = s_{j-1} + \nu_{j-1}$, where $\nu_{j-1} \sim \mathcal{E}xp(1)$.
   (b) Calculate corresponding time: $t = \Lambda^{*-1}(s_j) = s_j/\lambda$.
   (c) **If** $t < \widetilde{T}$ **then**
   - Time of the event: $t_j = t$
   - Mark of the event: generate $q_j \sim \mathcal{W}(\theta_1, \theta_2)$.
   - $j = j + 1$
3. **Output:** the simulated history $\mathcal{H}^s = \{(t_1, q_1), (t_2, q_2), \ldots\}$.

Some points to consider about the procedure:

- At some point, the simulation has to stop. Hence, there is a time limit $\widetilde{T} > 0$ to set. As soon as one of the simulated times t_j reaches it, the simulation stops. As a result, the duration of the history is fixed, but not the total number of events in it. This number is random and differs from one simulation to another.
- We remind the times s_j and t_j are absolute ones, it means they count elapsed time since the origin of the history, by opposition to interevent times. To simulate the time s_j of

the j^{th} event (see step 2.(a) of the algorithm 1), we draw first the interevent time ν_{j-1} from an $\mathcal{E}xp(1)$ distribution, and then compute the corresponding time $s_j = s_{j-1} + \nu_{j-1}$.

- The output of the algorithm is the history we just simulated, denoted as $\mathcal{H}^s$. As we just mentioned, its total length differs from one simulation to another.
- The value of the model parameters, namely $\Theta_G = \lambda$ and $\Theta_M = (\theta_1, \theta_2)$, are approximated with their respective estimators, the ones we fitted using $\mathcal{H}$ during the inference part (refer to Section 5.1.5.2).

We repeat the simulation procedure of the algorithm 1 to simulate a large given number N_{sim} of histories, resulting in a final sample of histories $\left\{\mathcal{H}^{s,k}\right\}_{1 \leq k \leq N_{sim}}$.

The idea now is to compute an average value, over all the histories, of an unknown characteristic value of the MPP we are interested in, like the mean quality of the kidney offers. We proceed as follows

- Choose an unknown parameter/characteristic value of the MPP we want to estimate. The real value of the quantity we want is denoted as z. In our example, we take $z = \mathbb{E}[Q]$.
- Choose an estimator of the unknown quantity z, we denote it as Z. Like all summary statistics, Z is a function of the data we have access to, for a given history $\mathcal{H}$, $Z = function(\mathcal{H})$. In our example, we can chose Z to be the basic sample mean

$$Z = \frac{1}{N} \sum_{j=1}^{N} q_j, \tag{5.32}$$

  with $N = \text{Card}(\mathcal{H})$ the length of the history.
- We compute the statistic Z for all the simulated histories. For $k \in 1, \ldots, N_{sim}$, let Z_k be the corresponding summary statistic computed on the simulated history $\mathcal{H}^{s,k}$. In our example, the Equation (5.32) becomes

$$Z_k = \frac{1}{N_k} \sum_{j=1}^{N_k} q_j^k, \tag{5.33}$$

  with $N_k = \text{Card}(\mathcal{H}^{s,k})$ the length of the history, and q_j^k the j^{th} quality of the history.
- Then, we average the statistics over all the histories by computing the sample mean

$$\bar{Z} = \frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} Z_k. \tag{5.34}$$

With an adequate number of simulations N_{sim}, the statistic $\bar{Z}$ is a good approximation of the unknown characteristic value z we want to infer. In the case of the example we took, it means $\bar{Z}$ is a good approximation of the quality expectation $z = \mathbb{E}[Q]$.

Remark 5.8 (Proof outline: validity of estimator $\bar{Z}$). The mathematical validity of the estimator $\bar{Z}$ could be proved with the law of large number for instance (law of large number in the case of different distributions, [37] p.104). Indeed, the $(Z_k)_k$ are always mutually independent, since they each come from a different simulation k. The result in [37] states:
If the series of the variances $\sum_{k=1}^{+\infty} \mathbb{V}ar[Z_k]/k^2$ is convergent, then

$$\frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} (Z_k - \mathbb{E}[Z_k]) \underset{a.s.}{\longrightarrow} 0, \tag{5.35}$$

where $a.s.$ stands for the almost sure convergence (a type of convergence used for random variables).

We will not prove the general validity of the estimator $\bar{Z}$ for any summary statistic Z, but show its validity in usual cases. Indeed, the hypothesis of the series convergence is true almost all the time with an usual statistic Z. Let us take the example of Z being the mean quality: Z_k are defined in Equation (5.33). Then, knowing that all the qualities q_j^k are independent and equally distributed as the variable Q

$$\begin{aligned} \forall k \in 1, \ldots, N_{sim}, \quad \mathbb{V}ar[Z_k] &= \frac{1}{N_k^2} \mathbb{V}ar\left[\sum_{j=1}^{N_k} q_j^k\right] \underset{i.i.d.}{=} \frac{1}{N_k^2} \sum_{j=1}^{N_k} \mathbb{V}ar\left[q_j^k\right] \\ &\underset{i.i.d.}{=} \frac{1}{N_k^2} \sum_{j=1}^{N_k} \mathbb{V}ar[Q] = \frac{\mathbb{V}ar[Q]}{N_k}. \end{aligned}$$

So, the series of variances has as a general term

$$\frac{\mathbb{V}ar[Z_k]}{k^2} = \frac{\mathbb{V}ar[Q]}{N_k} \frac{1}{k^2} \leq \mathbb{V}ar[Q] \frac{1}{k^2} = \mathcal{O}\left(\frac{1}{k^2}\right).$$

That proves the convergence of the series, by asymptotic comparison with the general term $1/k^2$ of a convergent Riemann series (exponent $2 > 1$). In parallel,

$$\forall k \in 1, \ldots, N_{sim}, \quad \mathbb{E}[Z_k] = \frac{1}{N_k} \sum_{j=1}^{N_k} \mathbb{E}\left[q_j^k\right] = \frac{1}{N_k} \sum_{j=1}^{N_k} \mathbb{E}[Q] = \mathbb{E}[Q].$$

Hence, the result of Equation (5.35) is rewritten as

$$\frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} (Z_k - \mathbb{E}[Q]) = \bar{Z} - \mathbb{E}[Q] \xrightarrow[a.s.]{} 0.$$

In conclusion, this last result proves that $\bar{Z}$ is a relevant estimator, since it converges to the expected value we are searching for. Moreover, the almost sure convergence means the estimator is strongly consistent. This type of consistency involves in particular the estimator $\bar{Z}$ to be asymptotically unbiased. In this particular case, the estimator $\bar{Z}$ is not only asymptotically unbiased, but it is unbiased for any value of N_{sim}, as we prove as follows

$$\mathbb{E}\left[\bar{Z}\right] = \frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} \mathbb{E}[Z_k] = \frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} \mathbb{E}[Q] = \mathbb{E}[Q] \Longleftrightarrow Bias\left[\bar{Z}\right] = \mathbb{E}\left[\bar{Z}\right] - \mathbb{E}[Q] = 0.$$

□

5.2.3.2 An example of simulation

We take again the example presented in Section 5.2.2.2. In this example, we fitted the MPP and then accessed exact predictions since, with the current MPP model, closed form calculus is possible. However, for illustration purpose, we consider in this section the simulation approach to get empirical results.

We apply the simulation procedure we explained so far, in the simple case we want to estimate the expected value $\mathbb{E}[T]$ of the Time before Next Offer T. Using the notation introduced earlier,

- The unknown characteristic value we want is $z = \mathbb{E}[T]$.
- The statistic we use to estimate it is simply the time before the first event: $Z = t_1$.
- The corresponding statistic for the k^{th} simulated history is $Z_k = t_1^k$.

For $N_{sim} = 2000$, we simulate the sample of histories using the algorithm 1. We take as the time limit, $\widetilde{T} = 350$ days ($\simeq 1$ year, which is more than enough since we only need the first event to happen). A time transformation is performed in the inverse algorithm, which turns times we simulated from an unit rate Poisson process to times that could have been generated from the GIF of the model. This time transformation is represented for one given simulated history $\mathcal{H}^{s,k}$ in Figure 5.7, for the 20 first events.

The resulting simulated history $\mathcal{H}^{s,k}$, which includes generated marks, is represented in Figure 5.8. Each point corresponds to an offer in the simulated history. The times on the x-axis

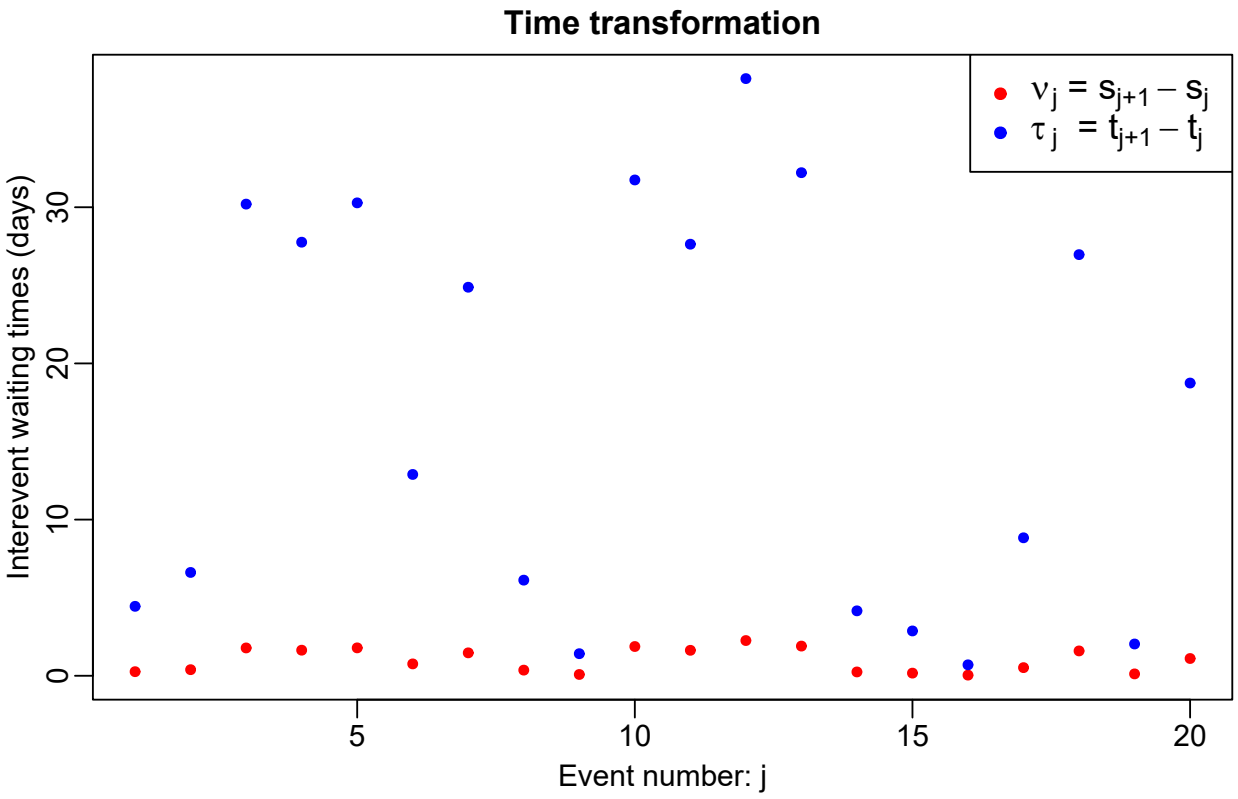


Figure 5.7 Time transformation of the inverse algorithm

are the absolute times from the beginning of the history (the cumulative sum of interevent waiting times: $t_j = \sum_{k=0}^{j-1} \tau_k$). The simulated marks q_j (KDRI quality) are on the y-axis.

Then, we compute the statistics $(Z_k)_k$ and the final statistic $\bar{Z}$. The histogram of the distribution of the $(Z_k)_k$ is given in Figure 5.9. We use it to represent the point estimation $\bar{Z}$ of the expected time $\mathbb{E}[T]$ we are searching for. Also, we can access to empirical confidence interval of this quantity, as shown in the figure. In conclusion, we get the estimation $\bar{Z} = 16.3$ and the estimated 95% confidence interval $\widehat{CI}_{95\%} = [0, 49.3]$. These values are practically the exact same results as the ones we found in Section 5.2.2.2 using closed form calculus, as we can compare in Table 5.2.

Table 5.2 Exact predictions and simulation predictions

	Exact predictions	Simulation predictions
$\mathbb{E}[T]$	17	16.3
$\widehat{CI}_{95\%}$	$[0, 50.7]$	$[0, 49.3]$

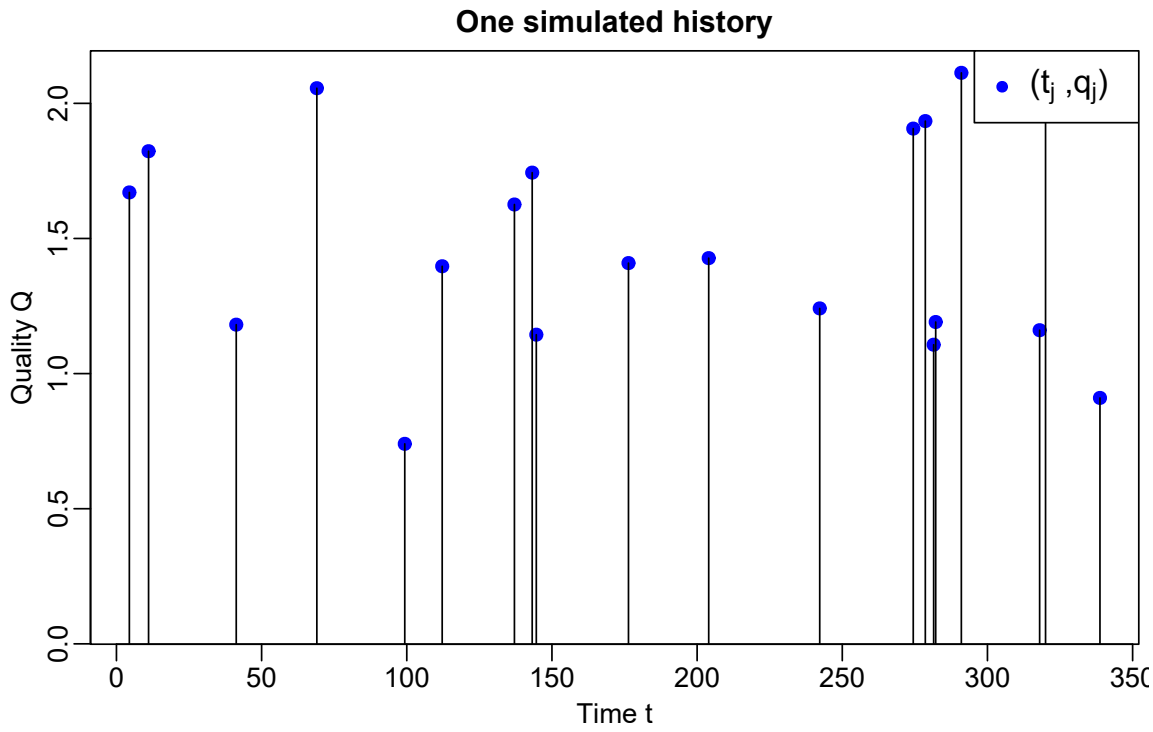


Figure 5.8 Example of a simulated history $\mathcal{H}^{s,k}$

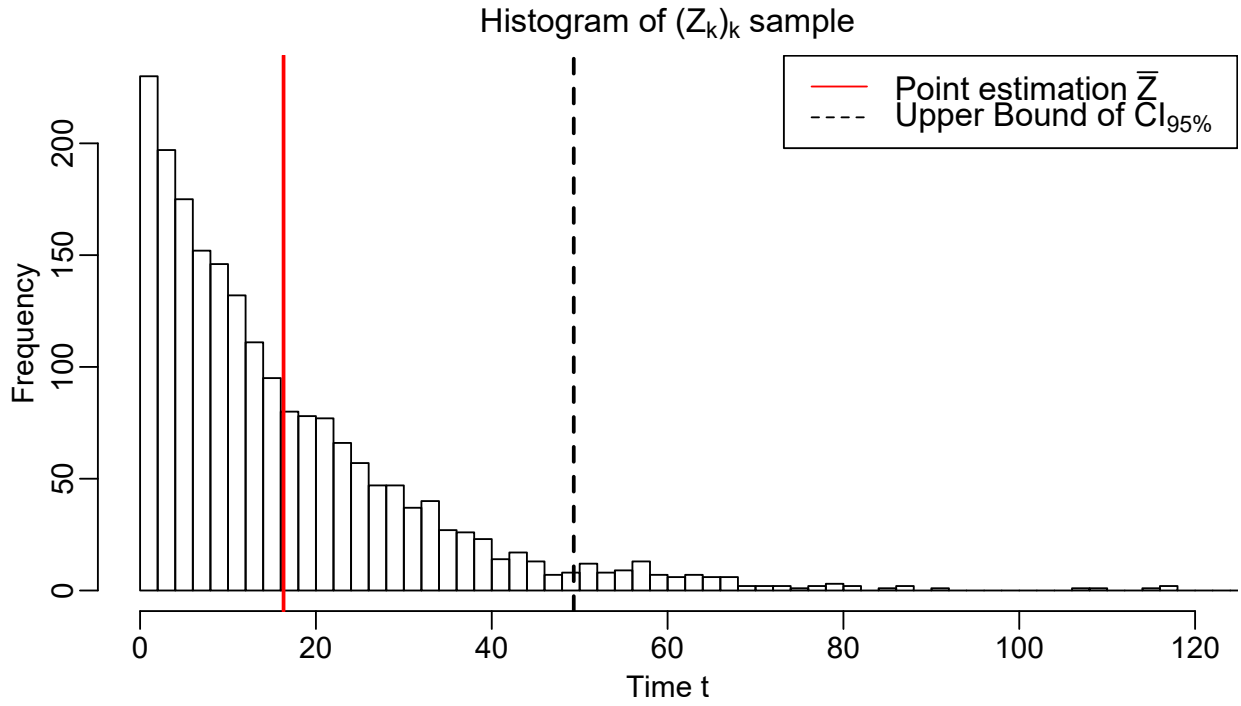


Figure 5.9 Histogram of the simulated sample $(Z_k)_k$

CHAPTER 6 EXPERIMENTS AND RESULTS

In this chapter, we conduct the experiments to define the CIF, obtain the best estimators for the parameters, and bring corrections to the model. We start by introducing the transplant attribution system records we use as data. Then, we also present the verification methodology for validating the developed MPP method. Finally, the performances on the test set are illustrated in Section 6.6.

6.1 Data from attribution system records

In this work, we applied our MPP model to real attribution system records furnished by TQ, the public institution in charge of organs attribution to patients waiting for a graft in the province of Québec. We used the same data as in the former work by Weller [1], that include records between 2012-03-29 and 2017-12-13. In this previous research, all the raw data were formatted, preprocessed and cleaned in order to make them usable to our purposes. For this reason, we will not go deeper into the details but instead focus here on the essential points for our purposes. All the details, including quantitative and qualitative analysis of the data can be found in Weller [1], Chapter 5.

6.1.1 The data

After the formatting, the records of the attribution system are contained into two main data frames. Namely,

- A donor file that contains several lines for each donor. Each line is related to a kidney offer with the corresponding patient identification. The features of this file are detailed in Table 6.1.
- A patient file that contains several lines for each patient. Each line is related to a change in the patient's status (temporarily or permanently removed from waiting-list, transplanted). The features of this file are detailed in Table 6.2.

To summarize the methodology we develop, the idea is to take an initial offer (x_0, y_0, t_0), and to fit the corresponding MPP process that describes the arrival of kidneys and their qualities. After the MPP is fitted, the model can be used to predict the time before the next offer. Consequently, the structure of the data points we need should be a pair composed of an initial offer and the corresponding time before the next offer in order to validate our prediction.

Table 6.1 Features in the donor file

Feature	Signification	Type
DON_ID	Donor identification number	Int
DON_BTH_DT	Donor birth date	YYYY-MM-DD
DON_STATUS	Donor status (e.g. DND or DCD)	Int
DON_DEATH_TM	Donor date of death	YYYY-MM-DD
DON_AGE	Donor age in years	Int
DON_GENDER	Donor gender: Male or Female	1, 2
DON_DIAB	Donor history of diabetes	0, 1, 2
DON_COCAINE	Donor history of cocaine	0, 1, 2
DON_CIGARETTE	Donor history of cigarette	0, 1, 2
DON_CORONARY	Donor coronary disease	0, 1
DON_VASC	Donor vascular disease	0, 1
DON_HTN	Donor history of hypertension	0, 1, 2
DON_CREAT	Donor creatinine ($\mu mol.L^{-1}$)	Int
DON_WGT_KG	Donor weight in kg	Int
DON_HGT_CM	Donor height in cm	Int
DON_RACE	Donor race	Int
DON_COD	Donor cause of death	Int
DON_EXC	Donor exceptional distribution	0, 1
DON_ABO	Donor blood-type	Char
DON_RH	Donor rhesus	0, 1
DON_ANTI_HCV	Donor hepatitis C serology	Int
DON_A/B/BW/CW/ DQ/DR/DRW_1/2	Donor HLA Allele A, B, BW, CW, DQ, DR or DRW at locus 1 or 2	Int
DON_ORG	Donor kidney removed	0, 1, 2
DON_RCV	Donor kidney recovered	0, 1
DON_WHY_NOT_RCV	Donor kidney why not recovered	Int
DON_WHY_REFUSED	Donor why refused	Int
DON_WHY_FAMILY_REFUSED	Donor why family refused	Int
DON_WHY_NOT_TX	Donor why not transplanted	Int
DON_WHAT_IF_NO_TX	What happened to the organ if not transplanted	Int
DON_CAN_SCORE	Donor-Candidate score if relevant	Float
CAN_RANK	Rank of the candidate for this offer	Int
CAN_ID	Candidate ID number for this offer	Int
CAN_DECISION	Candidate decision: Rejection, acceptance	0, 1
CAN_STATUS	Candidate status for this offer (Transplanted or not)	Null, 2
CAN_WHY_NO	Why candidate refused offer	Int
CTR_NO_FOR_ALL	Center refused for all candidates	0, 1
CAN_CTR_ID	Candidate center ID number	Int

Table 6.2 Features in the patient file

Feature	Signification	Type
CAN_ID	Candidate ID Number	Int
ORG_TY	Organ to which the candidate is applying	Int
CAN_BTH_DT	Candidate birth date	YYYY-MM-DD
CAN_GENDER	Candidate gender: Male, Female	1,2
CAN_WGT_KG	Candidate weigth in kg	Int
CAN_HGT_CM	Candidate height in cm	Int
CAN_ABO	Candidate blood-type	Char
CAN_RH	Candidate Rhesus	0,1
CAN_AGHBS	Candidate hepatitis B serology	Int
CAN_ANTI_HCV	Candidate hepatitis C serology	Int
CAN_ANTI_HIV	Candidate HIV serology	Int
CAN_A/B/BW/CW/ DQ/DR/DRW_1/2	Candidate HLA allele A, B, BW, CW, DQ, DR or DRW at locus 1 or 2	Int
CAN_CPRA	Candidate latest cRPA	Int $\in [0, 100]$
CAN_CPRA_DT_TM	Latest date of cPRA measurement	YYYY-MM-DD
CAN_LISTING_DT	Candidate latest date of enlisting	YYYY-MM-DD
CAN_DIAL_DT	Candidate latest date of first dialysis	YYYY-MM-DD
CAN_NB_TX	Number of transplant the candidate underwent	Int
CAN_STATUS	Candidate status on waiting list: deceased, inactive, active, DDKT	-1,0,1,2
UPDATE_TM	Date of status update	YYYY-MM-DD
CAN_WHY_RMV	Why candidate was removed from the waiting list if relevant	Int
CAN_DGN	Candidate initial diagnosis	Int
CAN_DGN2	Candidate secondary diagnosis	Int
CAN_WTG_DT	Starting date for the waiting chronometer	YYYY-MM-DD

For these reasons, in this work the data are used in two main different ways.

- Building the history: for a given initial offer an algorithm creates the history of the corresponding MPP. To do this, it retrieves all kidneys (from the donors file) arrived in the time window $[t_0 - \Delta T, t_0]$, and then it creates all the corresponding waiting lists by using the data of the patients who were enlisted at that time (from the patients file).

- Pairing initial offer and next offer: the elementary data points we use are pairs of initial offer and corresponding next offer. When we mention datasets or data points, we refer to a number of those pairs.

We summarize the data processing and the way we use it in the Figure 6.1 below.

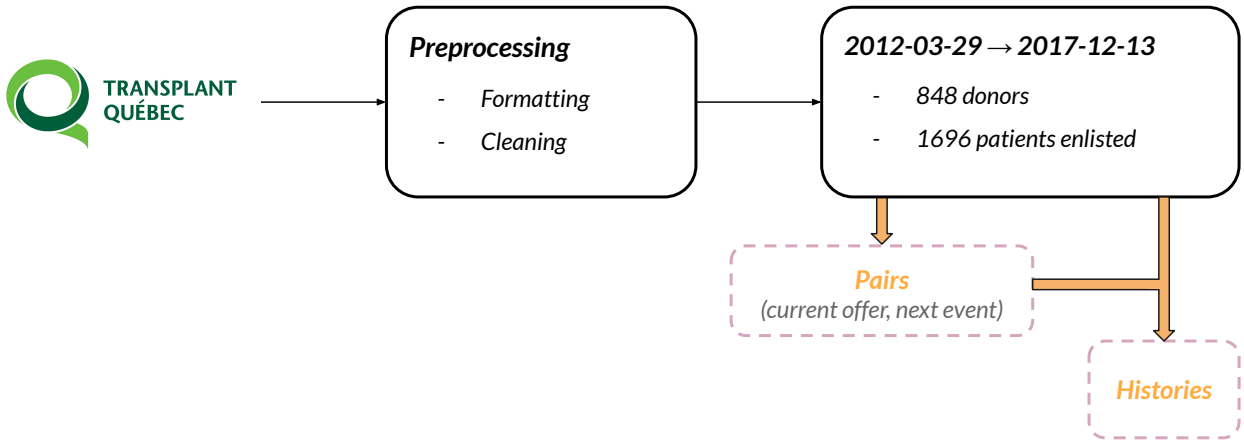


Figure 6.1 Steps to create the synthetic dataset

6.1.2 One first experiment to handle data

In this section, the objective is mainly to illustrate the kind of experiments we are performing when manipulating the data in order to investigate. Also, this section will give a better idea of what type of experiments we carry out, in contrast to Chapter 5 that was only theoretical. To implement progressively and test the method, without the computations over the entire dataset, we have to select a subset that represent well the whole set of patients. We select 20 patients and denote the representative sample as $\tilde{\mathcal{S}}_{20}$.

Since one main characteristic in our method differentiating one patient from another is the number of offers on a patient history, we use the value of N_{offer} to select the patients that will constitute $\tilde{\mathcal{S}}_{20}$. We draw the histogram of the distribution of N_{offer} among the patients in Figure 6.2.

As we can see in Figure 6.2, N_{offer} is widely ranged from high values (189) to the minimum value of 0. For most of the data, the numbers of offers N_{offer} is more than enough to infer comfortably the parameters of the quality distribution and have a good fit. However, there are problematic patients *i.e.,* the ones with small N_{offer}. It is hard to fit the quality distribution for these patients with few examples of offers in their history.

Moreover, for data with $N_{\text{offer}} < 2$, we simply cannot infer the parameters of the quality distribution. We need at least 2 offers in the history (hence 2 quality data points) to infer the 2 parameters (θ_1, θ_2) of the Weibull distribution.

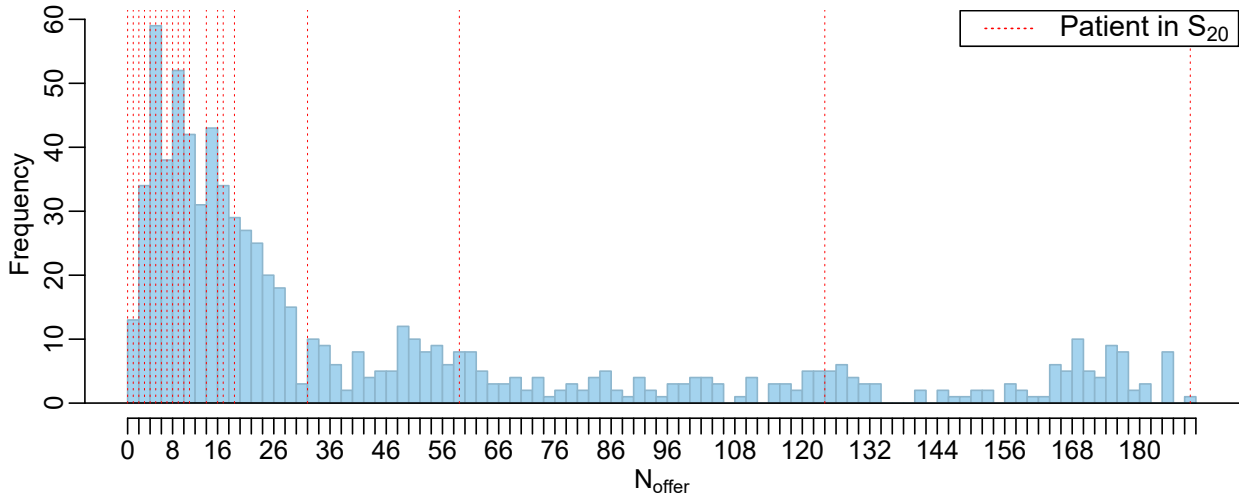


Figure 6.2 Histogram of N_{offer} and $\tilde{\mathcal{S}}_{20}$ patients selection

As a consequence, there is no point in having many patients with high number of offers in $\tilde{\mathcal{S}}_{20}$, since there will be no challenge to fit the quality distribution. Instead, we focus on potentially problematic patients with small N_{offer}. We zoom on the histogram in Figure 6.3 to better observe the distribution for low N_{offer}, and we represent the selected patients in $\tilde{\mathcal{S}}_{20}$ with red dotted lines on both Figure 6.2 and Figure 6.3.

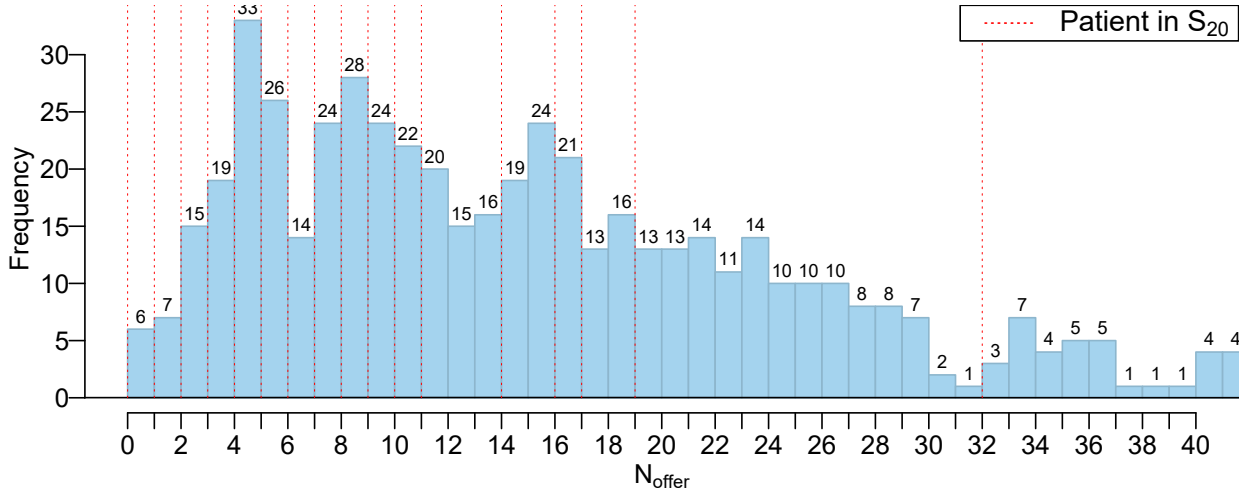


Figure 6.3 Zoomed histogram of N_{offer} and $\tilde{\mathcal{S}}_{20}$ patients selection

Then, we apply the MPP method for each patient in $\tilde{\mathcal{S}}_{20}$. We compute the history according to the past waiting list method (refer to Section 5.1.2), and we infer the parameters of the MPP model. We compute the estimator $\widehat{\lambda}_1$ of the GIF of the MPP, and we fit the Weibull quality distribution on the quality marks of the history. We give all the MPP parameters we estimated for each patient in Table 6.3. Also, we show the summary plots for the quality distribution fit for patient $n°15$ and $n°20$ in Figure 6.4.

Table 6.3 Representative 20 patients sample $\tilde{\mathcal{S}}_{20}$

Patient $n°$	1	2	3	4	5	6	7	8	9	10
N_{offer}	0	1	2	3	4	5	6	7	8	9
$\widehat{\lambda}_1$	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
$\widehat{\theta_1}$	$\varnothing$	$\varnothing$	4.92	3.08	2.84	6.00	10.40	6.17	6.42	3.80
$\widehat{\theta_2}$	$\varnothing$	$\varnothing$	2.39	1.58	1.94	1.35	1.80	1.93	1.78	1.41
Patient $n°$	11	12	13	14	15	16	17	18	19	20
N_{offer}	10	11	14	16	17	19	32	59	124	189
$\widehat{\lambda}_1$	0.010	0.011	0.014	0.016	0.017	0.019	0.032	0.059	0.124	0.189
$\widehat{\theta_1}$	2.97	3.28	5.95	2.64	4.52	2.90	2.33	3.62	2.81	2.91
$\widehat{\theta_2}$	1.70	2.04	2.29	1.62	2.09	1.99	1.62	1.70	1.65	1.67

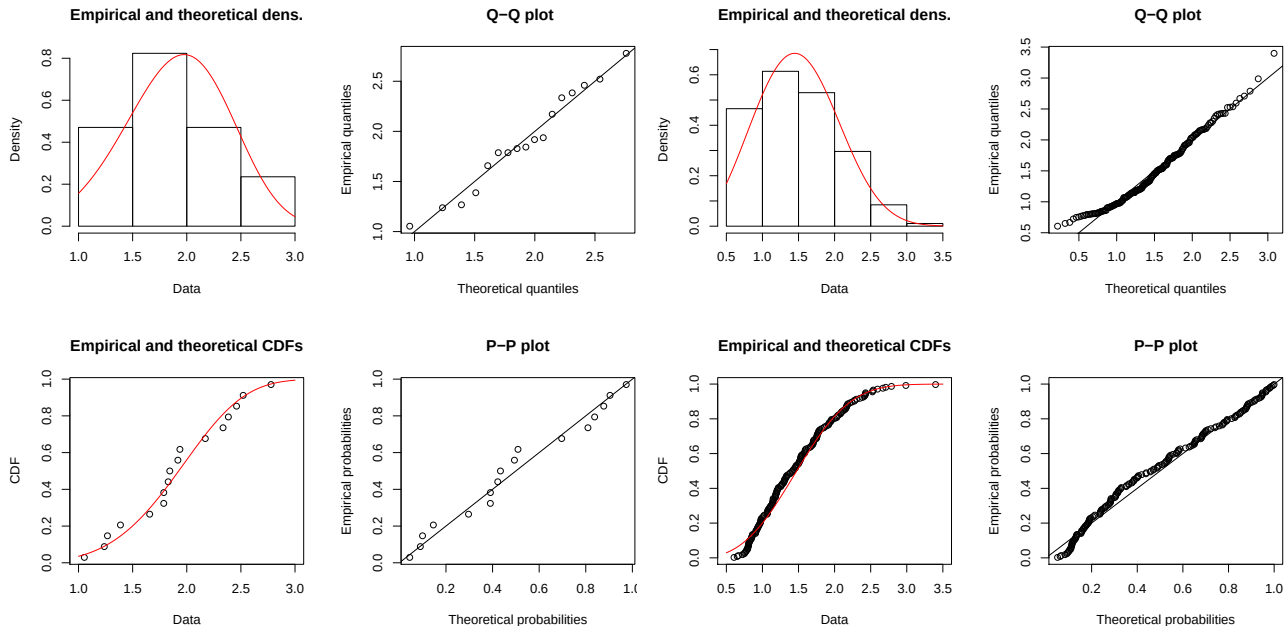
According to the distribution of N_{offer} among the patients, we face three different types of challenges in terms of the difficulty to fit the quality distribution. First, there is not enough data points ($N_{\text{offer}} < 2$). Then, between $2 \leq N_{\text{offer}} \leq 8$, the fit is challenging (the upper bound is given as an indicative basis, it is not the result of an extended analysis). Finally, there is the rest of the patients with a sufficiently large N_{offer} so it does not cause any difficulty to fit the 2-parameters quality distribution. The proportions of these different categories are compiled in Table 6.4.

Table 6.4 Three types of challenge in quality fit

	$N_{\text{offer}} < 2$	$2 \leq N_{\text{offer}} \leq 8$	$9 \leq N_{\text{offer}}$	Total
Occurrences	13	159	608	780
Proportion of dataset (%)	1.67	20.38	77.95	100

On the one hand, in Figure 6.4 for instance, the quality fit is really good for popular patients that receive a lot of offers in their history (*e.g.* patient $n°20$), or even a sufficient number of offers (*e.g.* patient $n°15$). On the other hand, the diagnosis plots for patients with a low

N_{offer} reveal some challenges, as it can be observed in Figure 6.5. The low number of quality points in the history makes it hard to infer the parameters correctly. Indeed, only a few points do not represent well a whole distribution. As a consequence, the histograms that can be observed are far from the fitted Weibull distribution. This experiment enables us to spot a weakness in the method. To tackle this difficulty and make a serious improvement of the model, we propose a solution in the next section.



(a) Patient $n°15$ (b) Patient $n°20$

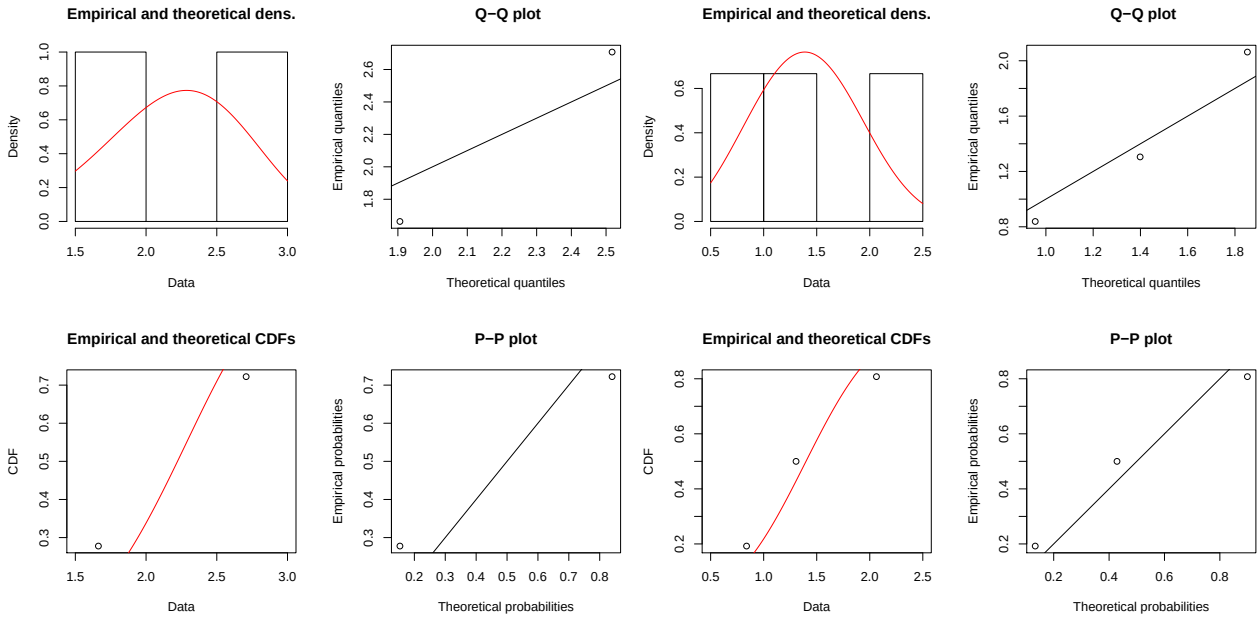
Figure 6.4 Quality fit on $\tilde{\mathcal{S}}_{20}$ patients for Weibull distribution

6.1.3 Method adjustment to tackle hard to fit patients challenge

As we just saw in the previous section, patients who have a low number of offers N_{offer} in their history pose a challenge for fitting the mark distribution (the quality of kidney offers).

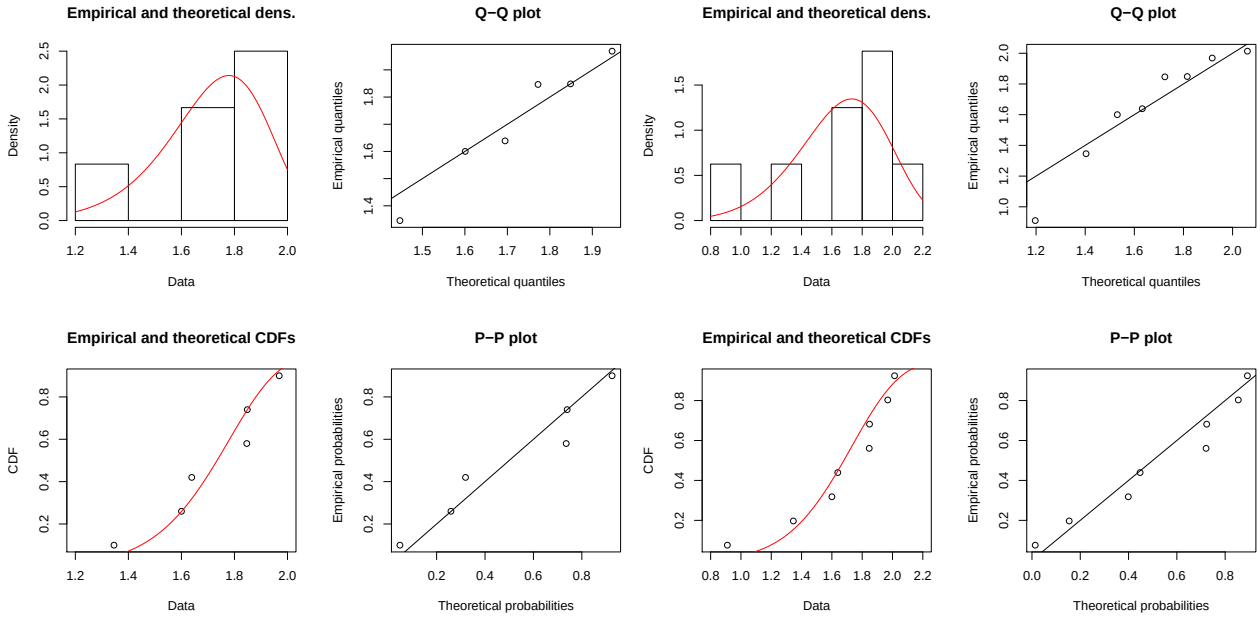
In parallel to this issue, we remind that we decided in Chapter 5 that for the MPP model we develop, events and marks are independent (*i.e.* intensity function and mark distribution are independent). Hence, while we only keep the N_{offer} eligible offers to compute the process rate, we can decide to fit the quality distribution on the whole history before thinning it. This history before thinning has $N_{\text{total}} \geq N_{\text{offer}}$ offers on it (refer to history building procedure in Section 4.1.1 and Figure 4.1 as a reminder). The rationale that justifies the possibility of this adjustment is based on the fact that the eligibility of an offer is not supposed to influence the mark distribution in our model.

We test this method modification later in this chapter, in Section 6.4.3.



(a) Patient $n°3$

(b) Patient $n°4$



(c) Patient $n°7$

(d) Patient $n°9$

Figure 6.5 Challenging quality fit for patient with low N_{offer}

6.2 Verification Methodology

We present the verification methodology we applied to check the validity of the MPP method we developed. It includes the use of specific verification tools, such as statistical indicators, and the comparison of performances with other methods.

6.2.1 Notations

In this chapter, we will manipulate several different notions that correspond to waiting times. Also, we will have to consider several data at the same time, to compute loss functions for instance, so that variables or parameters (T, q_0, etc.) do not necessarily refer to the same initial offer anymore. In order to make it perfectly clear, we introduce the following notations:

- For one given initial offer, we fit a MPP. As a result, we have an estimated distribution for the time before next offer.
  - When necessary, we will refer to the corresponding random variable with the notation $T^{\approx}$.
  - When referring to the observed time, the target we have, we will use the notation T^{*}.
- Now, when we work with several initial offers at the same time, we differentiate them by the subscript notation i for the variables corresponding to the i^{th} initial offer $(x_i, y_i, t_{0,i}, q_{0,i})$. Combining this and the previous notations, it gives
  - The random variable $T^{\approx}_{(i)}$ is the time to next offer we fitted for the i^{th} initial offer, and $\mathbb{E}\left[T^{\approx}_{(i)}\right]$ is the corresponding expected waiting time.
  - The real number $T^{*}_{(i)}$ is the observed waiting time for the i^{th} initial offer.
  - The random variable $T_{(i)}$ without exponent is the actual time to next offer, that is to say the real random variable we want to approximate with $T^{\approx}_{(i)}$.

6.2.2 Data particularity

Validation is difficult in our problem. In a classic statistical problem, we usually have several targets (or outputs) drawn from the same distribution, which makes it possible to compute statistics from the sample of targets, and then compare them to the predictions we made using our model. The larger the sample is, the more it reduces randomness of the statistic (its variance), and the closer the statistic is to the true value we are searching for. The

convergence of many usual estimators comes from various theorems like the central limit theorem (See [37]).

To illustrate a classical situation, let us take an example. Assume that we have a large number n_{sample} of machines of the exact same type, and the goal is to estimate their time before failure. We record for each machine, its observed waiting time before failure. For $i \in \{1, \ldots, n_{sample}\}$, it is denoted as $T^{*,ex}_{(i)}$.
We get a sample of n_{sample} observations from which we can get information about the distribution of the time before failure T^{ex}. To illustrate and analyze the sample distribution, one of the first things to do is to draw an histogram, as provided in Figure 6.6. We could also estimate the expected waiting time using a statistic: the sample mean

$$\bar{T}^{ex} = \frac{1}{n_{sample}} \sum_{i=1}^{n_{sample}} T^{*,ex}_{(i)}.$$

The statistic $\bar{T}^{ex}$ is represented as the point estimation in Figure 6.6. As a consequence, we could compare $\bar{T}$ to the waiting time $\mathbb{E}[T^{\approx}]$ we predicted (without using the targets but only the input variables), in order to validate the method. If the two values are sufficiently close, it means the method is valid *a priori*.

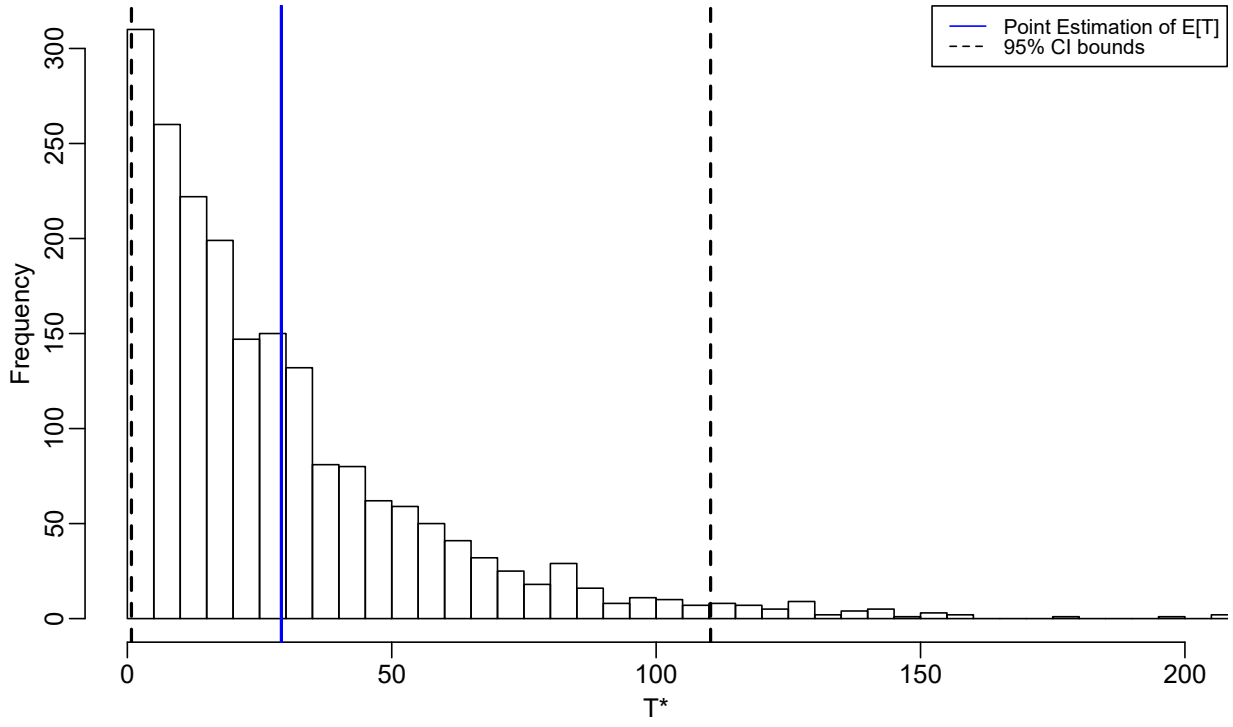


Figure 6.6 Histogram of observed sample

In our case, it is different since we only have one target. Indeed, for each initial offer, we fit a complete MPP, and consequently a distribution for $T^{\approx}$ at the same time. But to validate this distribution, we only have access to one observed time T^* which is the time before next offer for this patient and initial offer.

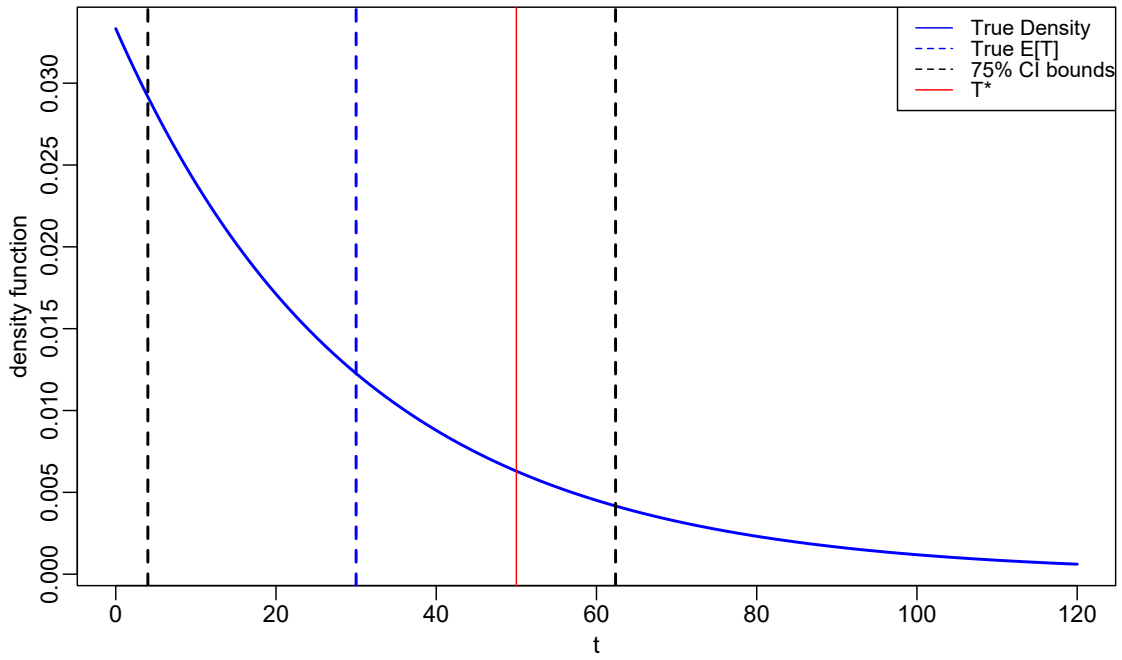


Figure 6.7 True PDF of the waiting time T

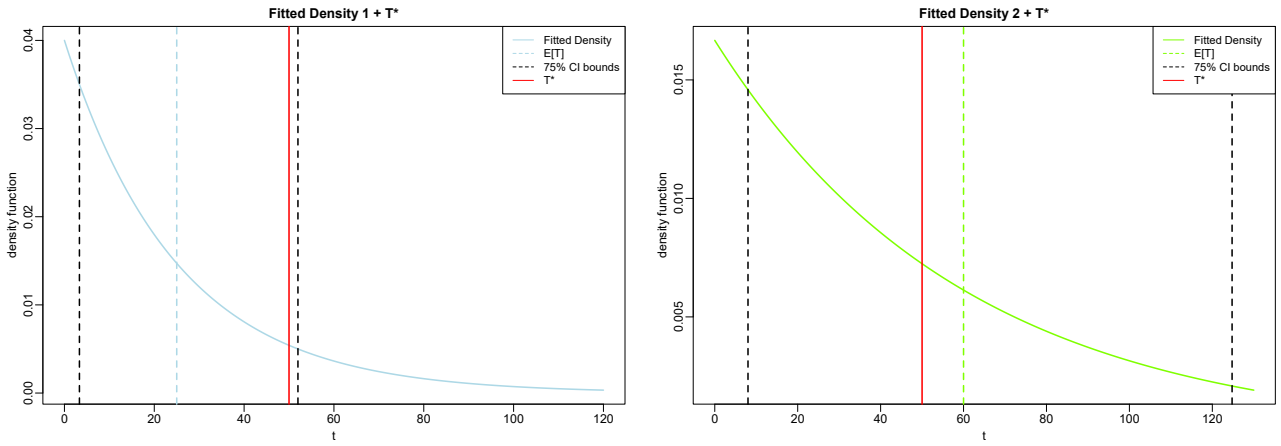
To illustrate the difficulty it represents, we take the previous example and adapt it to our current problem. Consider the following model is true: for the process corresponding to the initial offer, the waiting time T is exponentially distributed and its density is represented in Figure 6.7 with the blue line. It is the real exact distribution we want to predict to answer TNO issue, and that we could have guessed from the histogram shape in Figure 6.6 (indeed, the histogram converges to the true distribution as the sample size becomes larger [41]). However, in our case, we do not have access to the sample of observed times, and consequently cannot plot any histogram or compute any statistics on it. We only have access to the one observation T^*, which was drawn from the true distribution at random. We represent it with the red line on Figure 6.7.

With this value alone, there is few information about the distribution we are looking for. It is not necessarily near to the expected value $\mathbb{E}[T]$ (the blue line in Figure 6.7). We already understand clearly how easier it would have been to validate the model if we had access to a sample of observations instead of this one value.

However, let us assume that we fitted the distribution of the waiting time T using the MPP methodology of Chapter 5. We represent two hypothetical fitted distributions in Figure 6.8, that we could have fitted following our method. We can see that the two distributions are quite different from each other, but the observed value is still contained into the 75% CI for both.

This means that even if we predict the exact true distribution (the one in Figure 6.7), it will not guarantee that the observed value will be closer to the expected value we predict. Consequently, it means that we cannot validate the fitted distribution, and then the global methodology, by only considering one data. We need to use all the initial offers to cancel the randomness of the draws of the observed times $\left\{T^*_{(i)}\right\}_i$.

Moreover, we can predict that the difference $T^*_{(i)} - \mathbb{E}\left[T^{\simeq}_{(i)}\right]$ between our prediction and the observed time could be large, even if the prediction is good. Then, validation indicators that use this difference over all data will probably be large too. As a result, we will focus on the variation of the indicator value from one method to another, instead of looking at its absolute value. For instance, a decrease in an error score would mean a global improvement of the prediction power.



(a) Fitted distribution 1

(b) Fitted distribution 2

Figure 6.8 Two examples of possible fitted distributions

Censored observations. Another point to consider in the verification methodology is the presence of censored observations in the dataset.

For an uncensored data point, we have access to the target we want to predict: it is the time before next offer.

By opposition, a censored data point does not have the value of this target. It is missing since

it could not have been observed in practice. In the kidney transplant problem, a censored data point means the corresponding patient never received another offer after the current one, because something occurred before it could happen (*e.g.* patient removed from the waiting list, deceased patient). This results in the absence of a waiting time before next offer, and thus the data point has no target to compare the predicted waiting time to.
The Figure 6.9 illustrates the difference between a censored observation and an uncensored one.

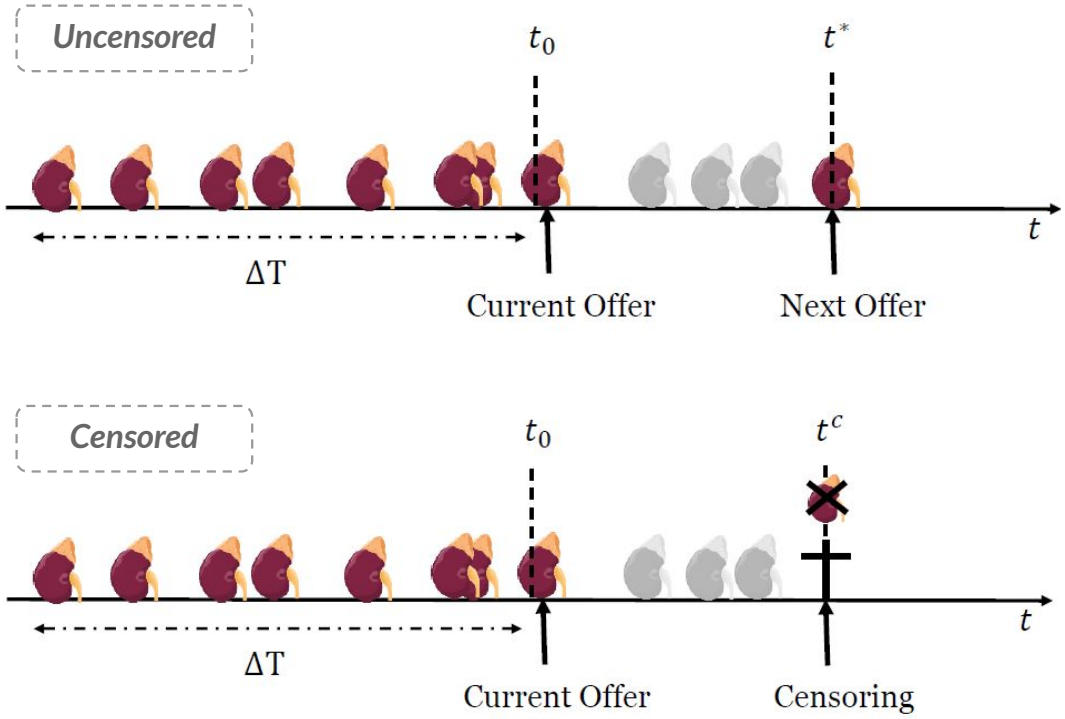


Figure 6.9 Uncensored and censored observations (figures from [1])

In the light of these difficulties, the objective is to try to make the expected value of our distribution and the observed value match, on average. However, due to the particularity of the data targets, the absolute error value is not that useful. Then, to measure improvement brought by the model we built, it is necessary to compare it to other methods, and consider relative performances. In addition, despite the fact censored data are generally not used since it is incomplete data, we propose an approach to include them anyway in the verification process.
But first, it is necessary to introduce appropriate statistical tools in order to measure performances.

6.2.3 Validation tools

In order to measure performances of methods, we use two main indicators. The first one is the Mean Absolute Percentage Error (MAPE).

Definition 6.1 (Mean Absolute Percentage Error)**.** For $i \in 1, \ldots, n$, let X_i be an independent random variable of predicted value $\hat{X}_i$. The Mean Absolute Percentage Error is

$$\frac{100\%}{n} \sum_{i=1}^{n} \frac{\left| X_i - \hat{X}_i \right|}{X_i}. \tag{6.1}$$

In our problem, it is written as

$$\frac{100\%}{n} \sum_{i=1}^{n} \frac{\left| T^*_{(i)} - \mathbb{E}\left[T^{\approx}_{(i)} \right] \right|}{T^*_{(i)}}. \tag{6.2}$$

The MAPE is an error score measuring the quality of a predictor. It evaluates the difference between the predictions and the observed values. Also,

- It does not over-penalize errors for large observed times since we normalize the error by the real observed value $T^*_{(i)}$.
- It could lead to exaggerated large penalties for small observed times, since we will divide by a small quantity. However, it will not be a problem here since we consider a 1-day granularity.

The second validation tool evaluates the ability of the method to order patients in terms of their waiting times. It is the Concordance Index (C-index).

Definition 6.2 (C-index in the continuous case)**.** For $i \in 1, \ldots, n$, let t_i be the observed time for a patient i, $\hat{t}_i$ the predicted time and $C_i = 0$ if the observation is censored and 1 otherwise. We assume t_i and $\hat{t}_i$ are continuous. The C-index is

$$C^{index} = \frac{\sum_{i:C_i=1} \sum_{j \neq i} 1_{t_i < t_j} 1_{\hat{t}_i < \hat{t}_j}}{\sum_{i:C_i=1} \sum_{j \neq i} 1_{t_i < t_j}}. \tag{6.3}$$

The C-index (also known as C-Statistic) is a well known indicator used to evaluate the predictive power of a model in survival analysis. The index utility allows to measure the ability of the survival model to predict the events in the right order. It means that if an

event happens in reality before another one, then the predicted times should be in the same order. One main characteristic of this indicator is that it can take into account censored data as well as uncensored ones.

- If the times are not assumed to be continuous, there are some special ways to handle the tie cases (equality between two times).
- In our problem, the predicted times $\hat{t}_i$ will be the times $\mathbb{E}\left[T^{\simeq}_{(i)}\right]$.

6.2.4 Baseline Policy

In order to evaluate the performance of the method we develop, we need to compare it to some baseline methods. We need to find a basic method that will give first approximations of the quantities we want to predict, such as waiting times.

Definition 6.3 (Baseline Policy: mean waiting time)**.** We chose a baseline method, to approximate the waiting time of a patient by the mean waiting time for all patients over the dataset we have. We will denote it as the mean waiting time

$$\bar{T}_{Global} = \frac{1}{n}\sum_{i=1}^{n} T^*_{(i)}. \tag{6.4}$$

This method uses the data targets to estimate a target. Theoretically, when it comes to computing the performances of the method on a dataset (validation set for instance), we cannot use the target of the tested data (the one for which we want to predict the real waiting time) to compute $\bar{T}_{Global}$. For example, if we want to compute the error of the k^{th} data $\epsilon_k = T^*_{(k)} - \bar{T}_{Global}$, we cannot use $T^*_{(k)}$ to compute $\bar{T}_{Global}$. It is inherently incorrect. Hence, to evaluate the performance of the method on a dataset, we will use the leave-one-out cross validation method. This method provides results that are really close to real test error in practice [42]. The leave-one-out error for the k^{th} data point is

$$\epsilon^{OOB}_k = T^*_{(k)} - \bar{T}_{Global,k}\ , \quad \text{where } \bar{T}_{Global,k} = \frac{1}{n-1}\sum_{i=1:i\neq k}^{n} T^*_{(i)}.$$

Another aspect of the Baseline Policy is that, as the method gives the same waiting time for each patient, it does not have any discriminative power. Indeed, all the predicted times are equal so the patients cannot be ordered. The resulting C-index is then 0.

Time before Next Offer: For TNO problem, the mean waiting time will be computed over the whole set, as for each offer we got the time before next offer. The censored data will not

be used.

Time before Next Better Offer: For the TNBO problem, the mean waiting time will only be computed over the offer with a next better offer. Censored data will not be used.

6.2.5 Data partition

We use the validation set/test set methodology to validate the model.

During all the validation methodology (Sections 6.3, 6.4 and 6.5), we will run experiments only on a subset of the data; the validation set. We use it to define the model, like choosing quality distribution, and also to make adjustments on the model, like with the estimator of the process rate.

The rest of the data, the test set, is kept unused until the end. Once the model is totally defined and adjusted, we use the test set to run a last verification procedure and check the performances of the MPP method we built (in Section 6.6). The performances we get at this point are supposed to reflect true performances of the model on any new data, since the test set is not used at any time in the model building process and is supposed to be representative of real data [42].

The dataset partition is presented in Table 6.5. In order to have similar datasets, we split data randomly so that both datasets have the same ratio between uncensored and censored data.

Table 6.5 Dataset partition

Data type	Validation set	Test set
Uncensored	627	626
Censored	153	152
Total	780	778

6.3 Define and estimate the GIF: the rate λ

In our MPP model, the GIF λ defines the rate of arrival of eligible kidneys. We need to infer it with an estimator that uses the history of the process. In the model we built, rate and quality are uncorrelated, so we can calibrate the rate alone first. This corresponds to answering the TNO problem, since no quality is used for the moment.

6.3.1 Estimating the process rate with $\widehat{\lambda}_1$

We first try the estimator $\widehat{\lambda}_1 = \frac{N_{\text{offer}}}{\Delta T}$ and evaluate it on the validation set. We will refer to this estimation method as Method 1. We check the validity of the model using an appropriate statistical tool, the normalized process and the corresponding Quantile-Quantile plot (Q-Q plot) diagnosis figure. Additionally, we compute results and compare performances with the Baseline Policy. In the end, we conclude that the rate estimator $\widehat{\lambda}_1$ has to be modified in order to improve predictions. Therefore, we will use the other rate estimator $\widehat{\lambda}_2$ in the upcoming Section 6.3.2.

6.3.1.1 Model checking: Q-Q plot of normalized process

One big issue with evaluating the model in our situation is the nature of our data. For each process we fit, we only have one target, the real observed time before the next offer. It is not convenient to use usual tools that require a larger sample to validate the quality of the process fit. In order to overcome this difficulty, we propose a method that enables us to use all the data at once. The idea is to normalize each process to gain homogeneous data that can be considered to come from the same unit rate process. Then a Q-Q plot can be generated using all the normalized data, making empirical quantiles appear from the normalized observed time. The Q-Q plot informs us about the validity and goodness of the model.

Mathematically, we need to normalize each process we fitted (for each initial offer), so each one is converted into a unit rate process. To do so, we have to normalize each process by using the value of its own rate.

Consider one of the processes we fitted with process rate λ. According to Proposition 3.3, in order to normalize the corresponding observed time T^*, we have to multiply it by the rate of the process. Since it is not accessible, we use instead the estimator of the rate we computed: $\widehat{\lambda}_1$. Then, we get the normalized observed time $Z^* = \widehat{\lambda}_1 \times T^*$. The process rates λ being always smaller than 1 in the kidney offer problem (1 would be equivalent to 1 offer per day on average), the normalization has the effect of reducing the observed time T^*.

By doing this for each process, *i.e.* each initial offer, we get the sample of normalized observed waiting times we wanted in the end. It can be considered to come from a unit rate Poisson process, which means we can draw a Q-Q plot with theoretical quantiles drawn from $\mathcal{E}xp(1)$.

Remark 6.1. We can see how we get here the same result we saw before in Section 5.2.2 when we modified the rate of the process. Indeed, when we normalize the process, we obtain a new rate λ_Z by multiplying the original rate by the factor $\frac{1}{\lambda}$: $\lambda_Z = \frac{1}{\lambda} \times \lambda$. And as we computed previously in Equation (5.19), this results in dividing the expected value by this same factor $1/\lambda$. In fact, we do have for the new interevent waiting time variable Z, the expectation $\mathbb{E}[Z] = \frac{\mathbb{E}[T]}{1/\lambda} = \frac{1/\lambda}{1/\lambda} = 1$. □

Remark 6.2. The study of the normalized waiting times of a process is a common verification procedure discussed in [2] and [43]. After undergoing a transformation involving the iGIF, the transformed times constitute the *residual process*. □

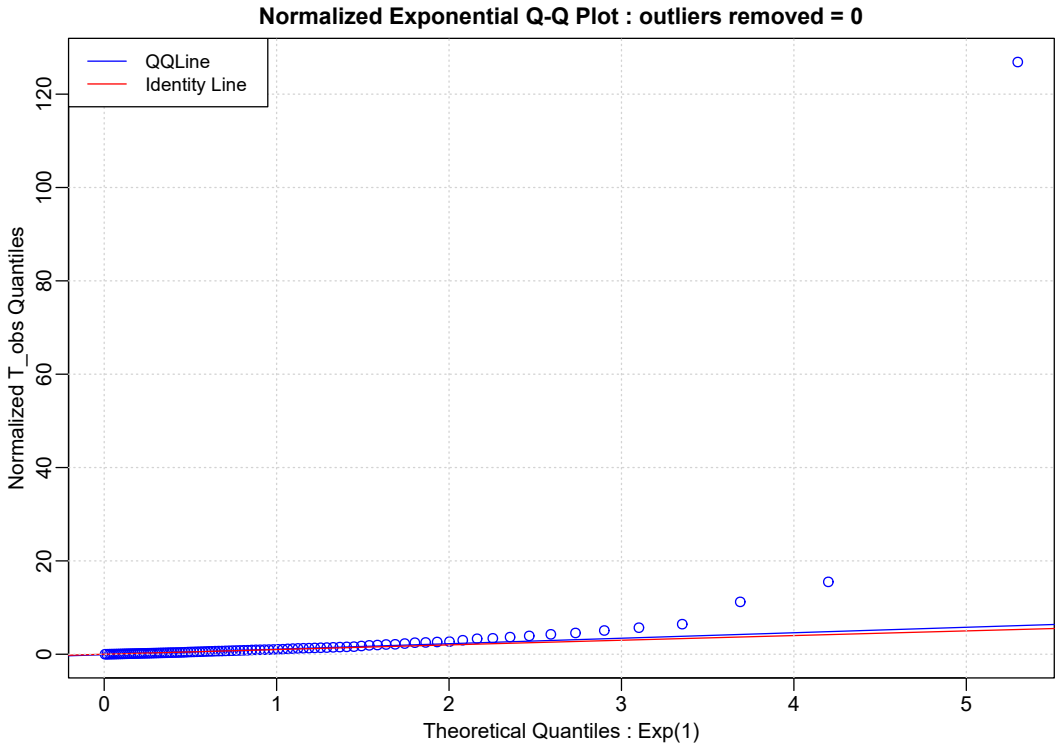


Figure 6.10 Normalized Q-Q plot: no outliers removed

Once we normalize all the observed times, we draw the Q-Q plot of the normalized sample in Figure 6.10. We plot 100 points, each one corresponds to an empirical quantile value computed on the sample (on the y-axis) and the corresponding theoretical quantile value (on the x-axis). For a perfect fit, that is to say for an empirical distribution that is the same as

the theoretical distribution we are comparing the sample to, the points of the Q-Q plot are all on the identity line. In this case, empirical and theoretical quantiles are all equal.

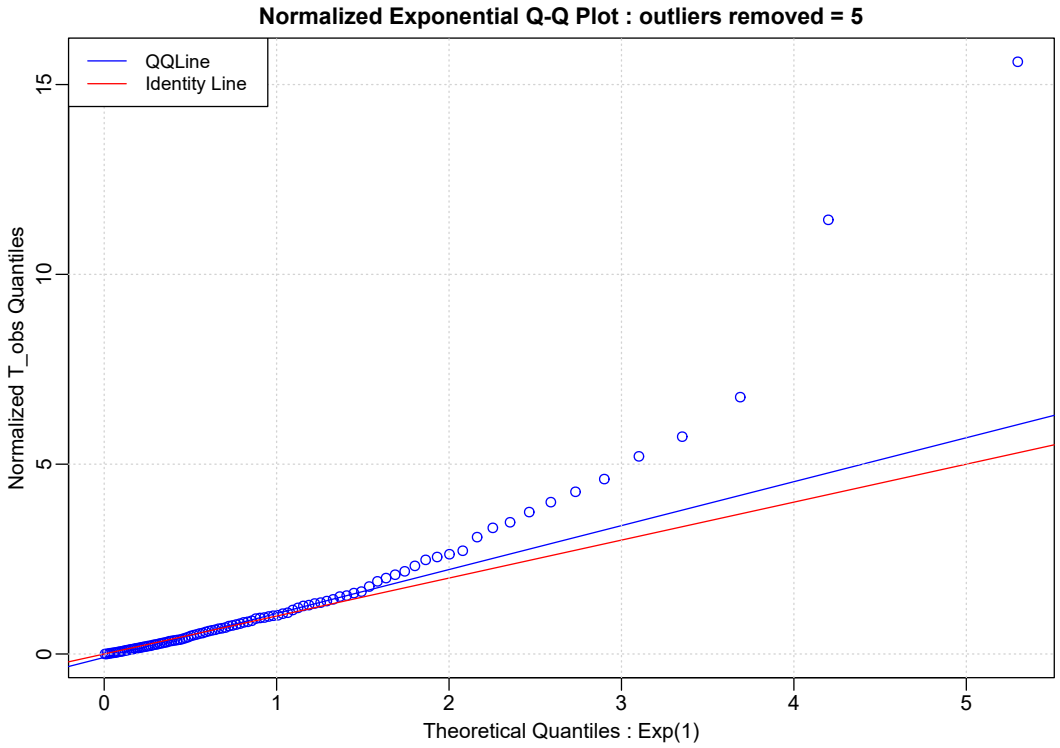


Figure 6.11 Normalized Q-Q plot: 5 outliers removed

As Figure 6.10 shows, there are abnormally large outlier values that are flattening the plot. Indeed, the resulting scale of the y-axis makes it hard to discern the rest of the points and analyze the plot. We remove the 5 largest outlier values and form the Q-Q plot in Figure 6.11. The analysis of this second Q-Q plots informs us that:

- The removal of only 5 values (out of a total of 626) makes the y-scale go from a 120 maximum to only 15. It means that the outlier values we removed were really far from the other values, and influenced a lot the graph.
- The 75 first empirical quantiles are the same as the theoretical quantiles. They are straight on the red identity line $y = x$. For the early quantiles, empirical and theoretical distributions are matching.
- Then, the quantiles start to shift appart from the line. The 10 following quantiles are near theoretical quantiles and have a quite constant shift from the reference line.
- The more we go towards the last quantiles, the more they are shifting from the reference line. The 2 last quantiles are far from it: we have a ratio empirical/theoretical quantile

that is around 3 (ratios of 12/4 and 15/5).

- The shift is always on the same side of the line, meaning that the empirical quantile values are always higher than the theoretical ones. This bias tends to increase when the values increase. A larger empirical quantile compared to the theoretical one means the empirical distribution of the sample spreads more towards large values than the theoretical exponential distribution.

The model has some difficulties with large values, for which it seems the sample of normalized times has larger values than a sample drawn from a $\mathcal{E}xp(1)$. Before digging deeper into this situation (like checking for causes), we check the performances of Method 1 in terms of predictions.

6.3.1.2 Performances

We test performances of the method by computing the indicators we detailed earlier during the validation methodology. We compare the method with the Baseline Policy. The results are reported in Table 6.6.

Table 6.6 Methods performances comparison

Method	MAPE	C-index
TNO - Baseline Policy	700	0
TNO - MPP: Method 1	677	0.66

For the C-index, the Method 1 gives a 66% rate of ordering the patient well (based on the observed time). Concerning the Baseline Policy, it does not offer the possibility to order patients. Since everyone is given the same waiting time expectation, then the C-index is null. As a consequence, Method 1 is a good approach for ordering the patients, since otherwise we will not be able to do it at all.

Compared to Method 1, the Baseline Policy increases the MAPE error by $100 \times \frac{700-677}{677} = 3.3\%$. Thus, Method 1 is only slightly better than the Baseline Policy in terms of waiting time prediction. There is a problem in the predictive power of this estimator.

As a conclusion, the first method $\widehat{\lambda}_1$ built to estimate the process rate presents limitations. In addition to the issues the Q-Q plot revealed, the predictive power appears to be weak. In order to address these limitations, we test in the following section another method to estimate the rate of the process, that is to say the estimator $\widehat{\lambda}_2$ that we presented in Section 5.1.5.3.

6.3.2 Addressing limitations of $\widehat{\lambda}_1$ with $\widehat{\lambda}_2$

First, we explain the coherence and the link that exists between both estimation methods for λ that we presented.
The first estimation method is $\widehat{\lambda}_1 = \frac{N_{\text{offer}}}{\Delta T}$. If we look at Figure 5.4, we can easily observe that

$$\begin{aligned} \Delta T &= (t_1 - 0) + (t_2 - t_1) + ... + (t_N - t_{N-1}) + (t_{N+1} - t_N) \\ &= \tau_0 + \tau_1 + \tau_2 + ... + \tau_N \\ &= \tau_0 + \sum_{j=1}^{N} \tau_j. \end{aligned}$$

Also, we know that

$$N = N_{\text{offer}},$$

so that we can rewrite

$$\widehat{\lambda}_1 = \frac{N_{\text{offer}}}{\Delta T} = \frac{N}{\tau_0 + \sum_{j=1}^{N} \tau_j}.$$

Now, we compare $\widehat{\lambda}_1$ with the expression of the Method 2 estimator

$$\widehat{\lambda}_2 = \frac{1}{\frac{1}{N}\sum_{j=1}^{N} \tau_j} = \frac{N}{\sum_{j=1}^{N} \tau_j}.$$

We can see that they have the same expression without the first term of the denominator. They are equivalent if we ignore a certain amount of time τ_0, which is the time elapsed between the beginning of the history and the first event.

As a result, the estimated rate $\widehat{\lambda}_2$ is greater than $\widehat{\lambda}_1$. One of the main objectives of developing this new estimator is particularly not to underestimate process rates for a not popular patient who would not have a lot of eligible offers in his history. For such a patient, the estimation of the process rate can be doubled easily by using Method 2 instead of Method 1.

Two examples are given in Figure 6.12. In the history above $N = 1$, there is only one offer that happens at time $t_1 \simeq \Delta T/2$. Then, Method 2 will only divide N by $\tau_1 \simeq \Delta T/2$, while Method 1 divides it by the entire time horizon ΔT anyway. By doing this, we can double the rate for a very unpopular patient. The same thing is demonstrated in the second example below with $N = 2$.

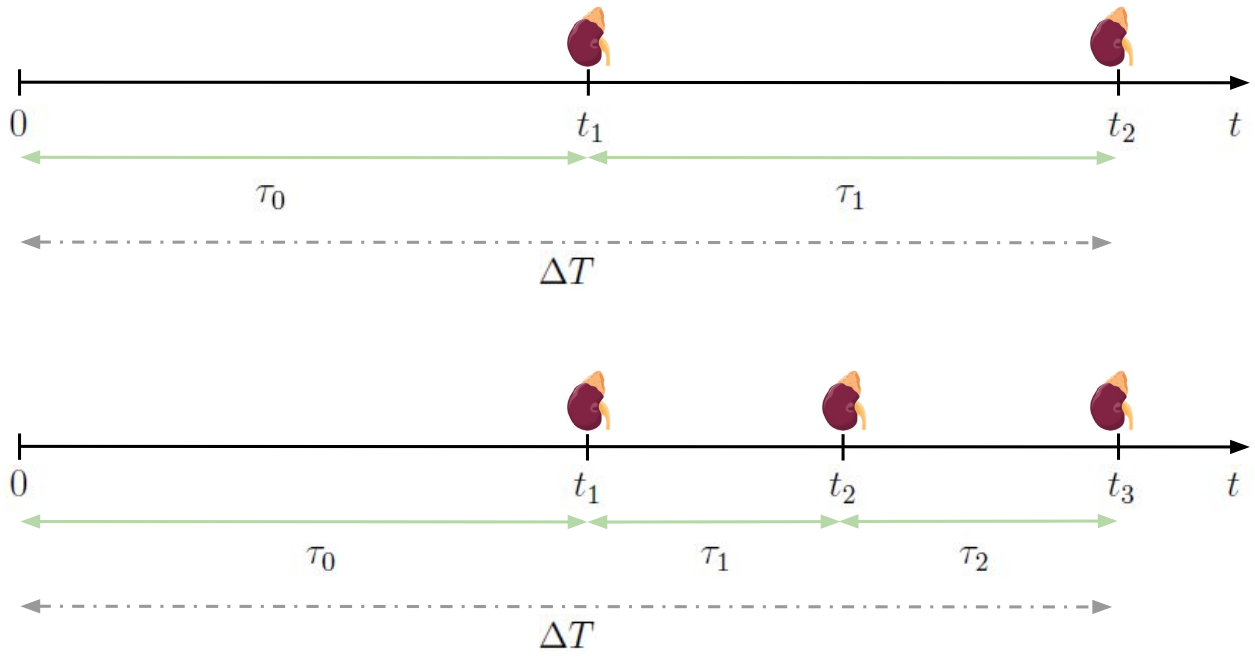


Figure 6.12 Example 1: Low number of eligible offers

We chose to develop the second method since we saw on diagnosis plots (for example on error diagrams) that too low rates are not representative of the true corresponding rates. For a patient with one offer, the rate with Method 1 is $\widehat{\lambda}_1 = N_{\text{offer}}/\Delta T = 1/1000$, which is very low, taking into account that the patient is not inactive since he already received a first offer at t_0 (the systematic configuration in our problem).

While Method 2 can fix low rates by increasing them, it has no major impact on the other higher rates. Indeed, the estimator tends to the true value (by consistency property of the estimator), and this intentional difference is caused by the bias of the estimator $\widehat{\lambda}_2$, $Bias\left[\widehat{\lambda}_2\right] = \frac{1}{N-1}\lambda$, which decreases inversely proportionally with the number N of offers in the history. With $N = 3$ offers, the bias is half the rate (the rate is increased by half on average), but with $N \geq 11$, we have a bias less than 10% of the value of the rate (only a small increase above the real rate value λ on average).

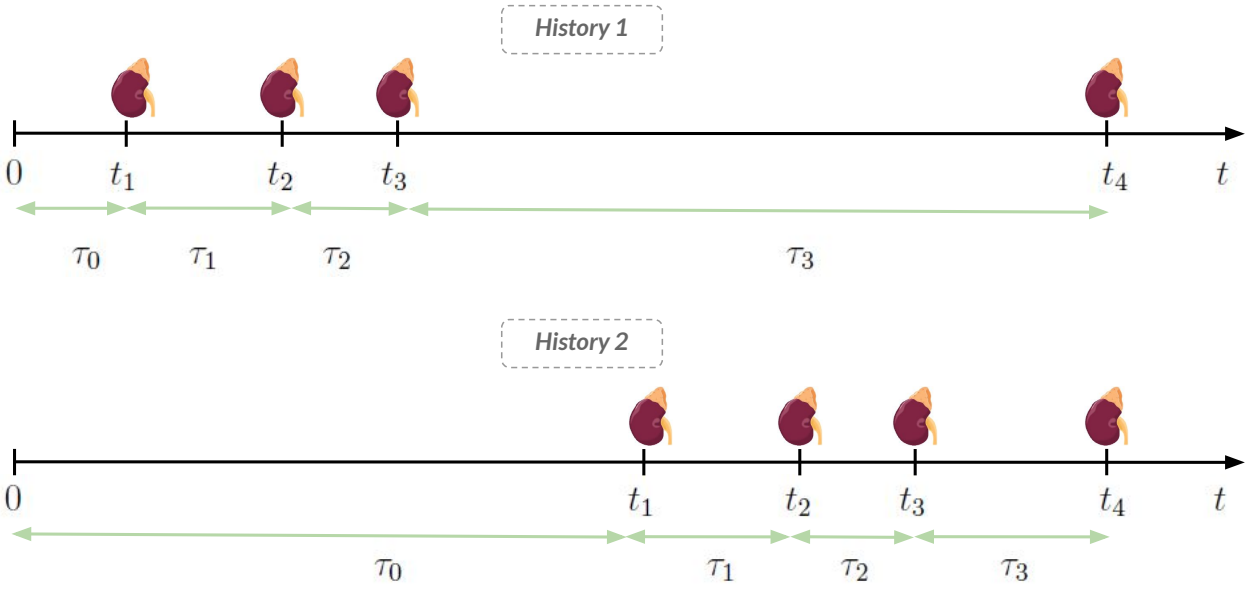


Figure 6.13 Example 2: Recent activity

Furthermore, Method 2 is more adaptive to the recent activity of the patient than Method 1. Let us consider a patient who does not have any eligible offer in the beginning of the history, but who starts to become "popular" and receives several offers in the end of the history. The process rate will be low on average in the whole history period with Method 1, while with Method 2, we only start to count time from the first eligible offer. The process rate will be higher, and more representative of the recent actual situation for the patient. We illustrate this point in Figure 6.13. History 1 and 2 give the same rate with Method 1, while the History 2 with a recent activity of the patient gives a higher rate with Method 2.

Additionally, Method 2 allows to have more continuous values for expected time predictions. For low values of N_{offer}, the corresponding expectation is really discrete in the sense that possible values are far away from each other. Namely,

$$N_{\text{offer}} = 1 \Rightarrow \mathbb{E}[T] = \frac{1}{\widehat{\lambda}_1} = \frac{\Delta T}{N_{\text{offer}}} = \frac{1000}{1} = 1000,$$

$$N_{\text{offer}} = 2 \Rightarrow \mathbb{E}[T] = 500,$$

$$N_{\text{offer}} = 3 \Rightarrow \mathbb{E}[T] = 333, \textit{etc.}$$

Patients with the same N_{offer} will get the exact same predicted waiting time, even if the pattern of their history is not the same at all. This categorization of patients results in

the high discretization of the expected times, and results in giving the same prediction for patients who do not necessarily have the same profile.

We demonstrate both these limitations in Figure 6.14 and the corresponding results in Table 6.7, and show how the pinpointed issues are corrected by using $\widehat{\lambda}_2$ instead of $\widehat{\lambda}_1$.

Table 6.7 Solving clustering and discretization with Method 2: estimators and predictions

		Method 1		Method 2	
Patient n°	N_{offer}	$\widehat{\lambda}_1$	$\mathbb{E}[T]$	$\widehat{\lambda}_2$	$\mathbb{E}[T]$
1	1	0.001	1000	0.0011	897
2	1	0.001	1000	0.0014	724
3	1	0.001	1000	0.0021	479
4	2	0.002	500	0.0028	357
5	2	0.002	500	0.0042	240
6	3	0.003	333	0.0046	216
7	3	0.003	333	0.0071	141

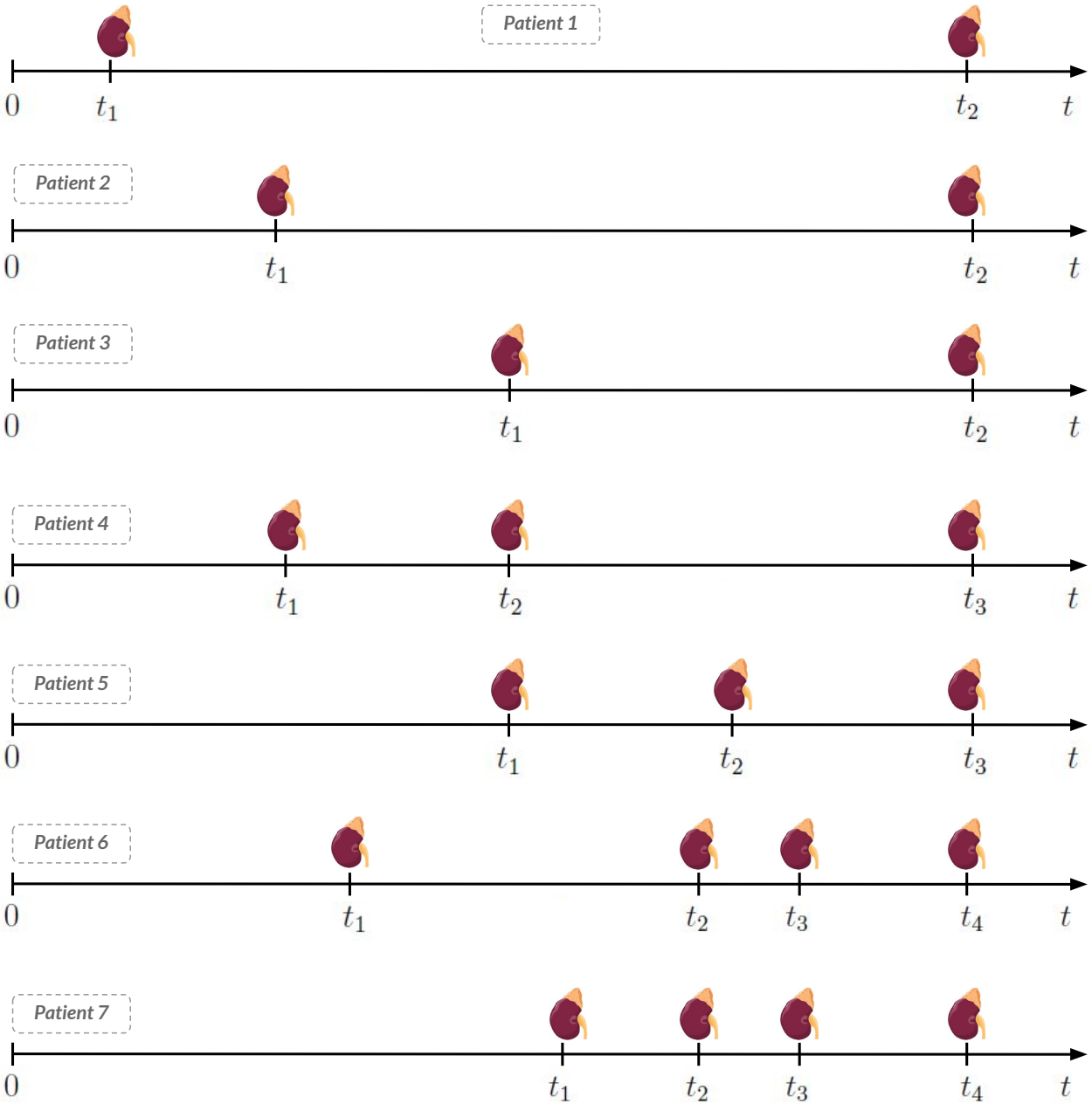


Figure 6.14 Solving clustering and discretization with Method 2: Examples of histories

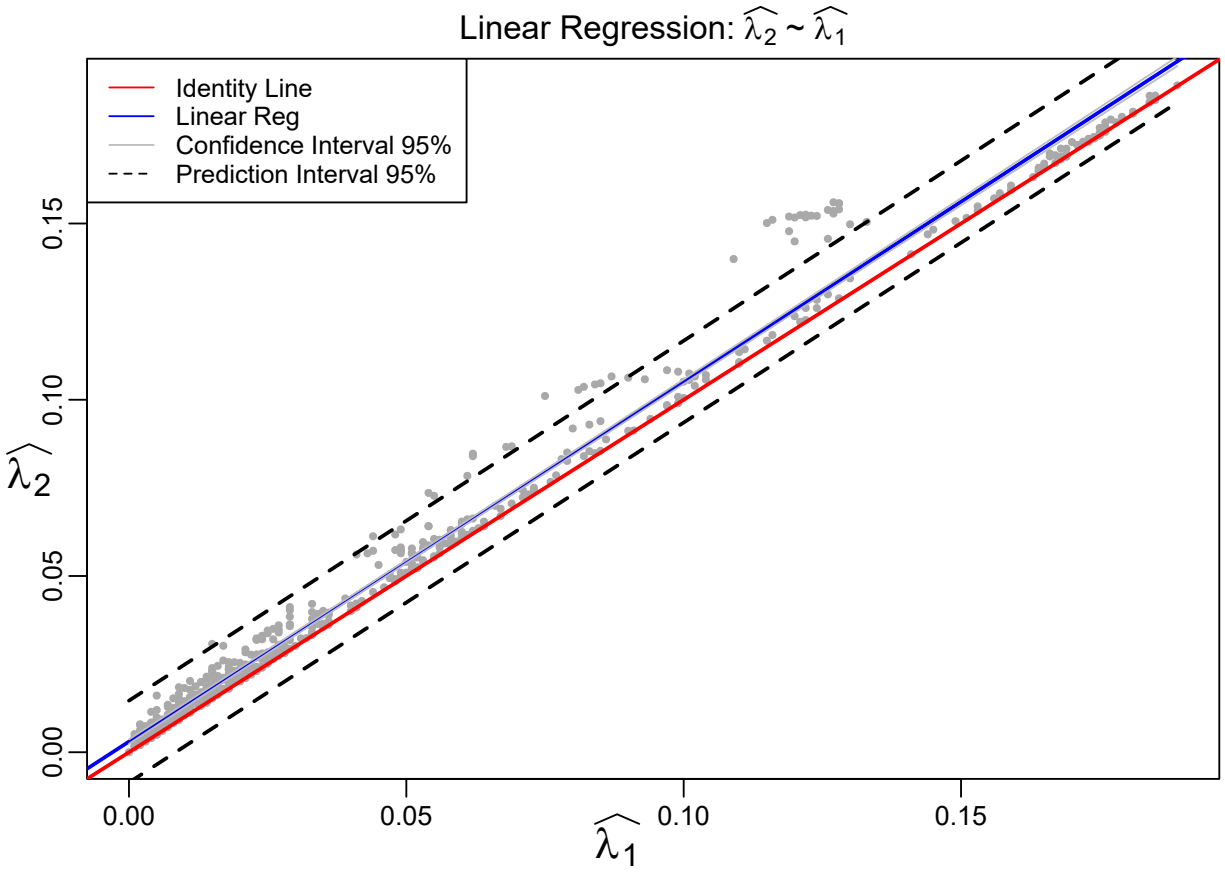


Figure 6.15 Correlation between $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$

Linear Regression We compare the rates λ computed with both methods, by plotting a scatter plot and fitting a linear regression in Figure 6.15. We list important points to consider in this figure.

- First, it is clear that we have a real coherence between both methods. It is confirmed statistically: the linear regression model has a p-value for ANOVA test such that p-value $< 2.2 \times 10^{-16}$, which is really significant.
- As we mentioned above, the value of $\widehat{\lambda}_2$ is always larger than the value of $\widehat{\lambda}_1$ (all the points are above the identity line).
- We see with the regression that we have a global, non negligible, translation to higher values. The slope is close to 1 (exact value being 1.02) and the intercept is 0.003, which is equivalent (on average) to adding $0.003 \times \Delta T = 0.003 \times 1000 = 3$ eligible offers to the history of each patient.

- Method 2 address the low rate limitations we discussed earlier. The rates are increased, although they still stay reasonably low, but the increase is significant proportionally to their low initial value with Method 1.
- The other theoretical major improvement of the better representation of the recent situation for a patient seems to work too. For higher rates of the process, we could expect (as said and illustrated previously) that values would not differ so much from one method to the other. This is the case if offers are distributed regularly in the history, but we see that some groups of points are far away from the identity line (*e.g.* detached group at $\widehat{\lambda}_1 = 0.12$). In this case, the big difference comes from the recent popularity of the patient that Method 2 managed to get to adapt the rate in consequence.

Now that we have illustrated the limitations addressed by the estimator of Method 2, we apply the solution to answer TNO issue and analyze its performances in the next section.

6.3.3 Estimating the process rate with $\widehat{\lambda}_2$

We check Method 2 based on the same verification methodology we carried out for Method 1. We first use the normalized processes sample technique and draw corresponding diagnosis plots from it. We analyze the Q-Q plot, as well as another diagnosis plot: the Probability-Probability plot (P-P plot). After this, we check on performances and compare them with Method 1 as well as with Baseline Policy.

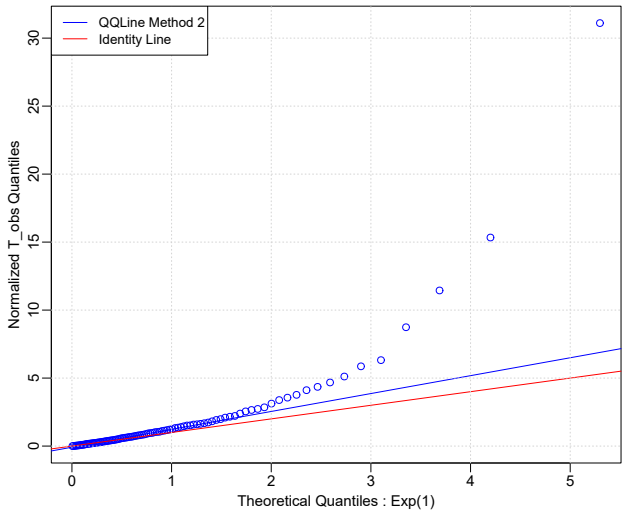
6.3.3.1 Model checking: Q-Q plot of normalized process

As in Section 6.3.1, we proceed to the Q-Q plot verification. We plot three different Q-Q plots in Figure 6.16.

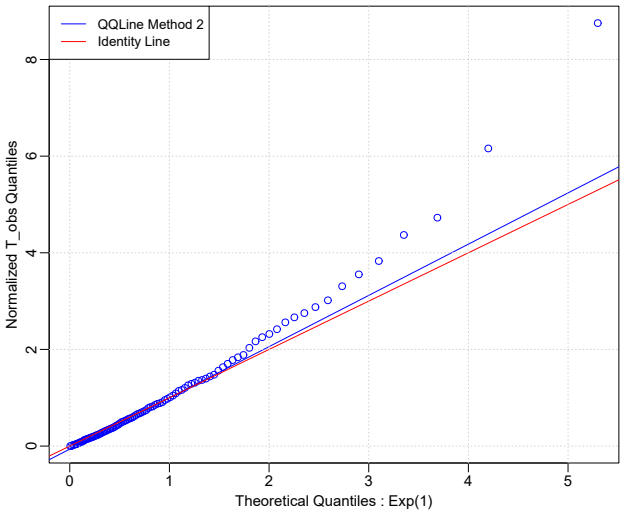
The first plot in Figure 6.16a is drawn without removing any of the outliers values, like in Figure 6.10. The two plots genuinely look alike. We can also observe that the last empirical quantiles are far from the theoretical ones, so the points drift apart from the red line. Also, the global shape of the graph is the same as in Figure 6.10. The value of y-scale is significantly reduced, when compared to the one in Figure 6.10. It means we successfully reduced large outlier values by using Method 2 instead of Method 1.

However, outliers values still exist and provoke the shift of last quantiles as well as the impossibility to read the graph. Thus, we will remove some of them to see if there is an improvement in the Q-Q plot, and judge if the method works well on the remaining values. In the same way we did in Figure 6.11, we remove the 15 largest outliers values and show the result in Figure 6.16b. There is a significant improvement in the Q-Q plot, as the points

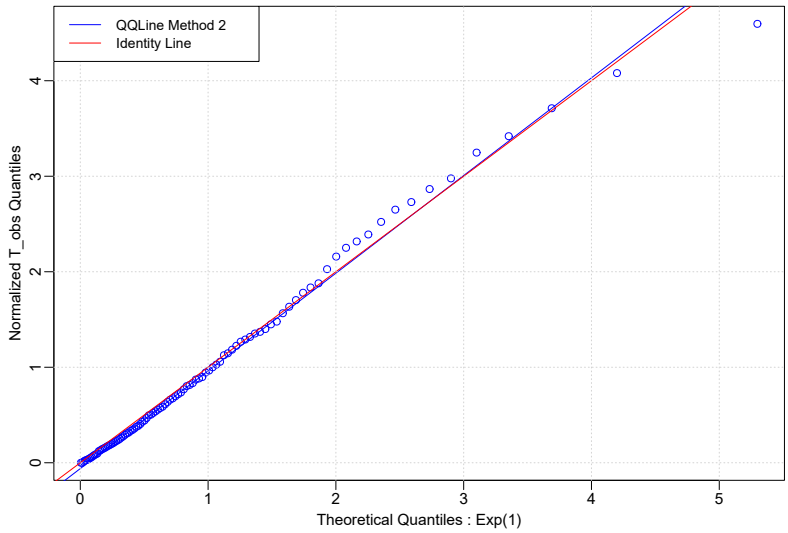
are getting closer to the identity line. It indicates that the empirical quantiles match their corresponding theoretical values better. Moreover, the points are now more aligned than before, which means that the exponential model is a fairly valid way of modeling the problem.



(a) Removed = 0



(b) Removed = 15



(c) Removed = 30

Figure 6.16 Normalized Q-Q plot for Method 2

Remark 6.3 (Q-Q plot). In a Q-Q plot, a straight line for the quantiles means the theoretical distribution we should have (exponential distribution) is the right type of distribution the sample is drawn from. The slope of the line formed by the quantiles gives information about the rate of the process (the rate of the exponential distribution). If the slope is 1, then we have the theoretical rate 1, and the theoretical model matches perfectly the observed data.

If we have a slope slightly greater than 1, then the rate of the process is slightly smaller than 1 (the events happen less frequently, thus the quantiles are bigger than theoretical ones, so the points in the Q-Q plot are above the red line). □

To match perfectly the red line, we remove 30 outliers in Figure 6.16c. In this case, the theoretical model matches perfectly the normalized sample.

It appears that the model we built is really coherent and valid if we remove a little proportion of the last largest values of the normalized sample. In the Q-Q plot verification, the data we use are composed of the uncensored values, and non null λ (because otherwise it would give an infinite expected waiting time), making a total of 626 values. Removing 30 of them represents a 4.8 % decrease of the total. In Table 6.8, we indicate the proportions in which some given numbers of removed outliers account for in comparison with the total size of the validation set.

Table 6.8 Removed outliers proportions - Validation set

Number of outliers removed	5	10	15	20	25	30	35	40	45
Proportion of dataset (%)	0.8	1.6	2.4	3.19	3.99	4.79	5.59	6.39	7.19

Outliers removed. The number of outliers we removed to reach a perfect match between the model and the data does not account for a significant proportion of the data, less than 5%. It is quite impossible that all the values fit perfectly a model we build, since there is always a difference between reality and the model. In our case, only the very tail of the empirical distribution does not fit the model, which makes the other 95% remaining data validate the model. We will not focus on the analysis of this drift in quantiles, but we present some considerations about the outliers values we removed.

- The strategy we adopted to correct the issue is a very simple one. We only removed the highest values of the sample without any selection process. Some of these values seemed fairly aberrant given how much they stood out from the rest of the sample, as Figure 6.16a shows. We remind we are referring to values after normalization.
- In order to draw the normalized Q-Q plot, we normalized the observed times T^* of the data using the rate $\widehat{\lambda}_2$ we computed : $Z^* = \widehat{\lambda}_2 \times T^*$. So in the end we have a new normalized waiting time Z^*, which is supposed to come from a unit rate Poisson process, that is to say $Z^* \sim \mathcal{E}xp(1)$. A "large" value means a large value compared to this specific distribution.

- Given that, we see that a large value can occur for different reasons. It may come from a large original T^*, an estimation $\widehat{\lambda}_2$ of the real rate that is not small enough, or both at the same time. For example, if the $\widehat{\lambda}_2$ we compute is not small enough, the observed time T^* (that is large because it comes from a low rate Poisson process) will not be shrinked enough when normalized by the estimated rate $\widehat{\lambda}_2$.
  Incoherent values can be hard to identify then, when nothing suggests that the patient will wait a long time (there are lots of offers in the history), but still has a long observed waiting time. The kidney transplant problem involves a lot of hidden variables that bring a lot of variability and cannot always be taken into account correctly in the model we build.

6.3.3.2 Model checking: P-P plot of normalized process

We pursue the verification methodology with the analysis of the normalized sample through a new type of diagnosis plot, the P-P plot.

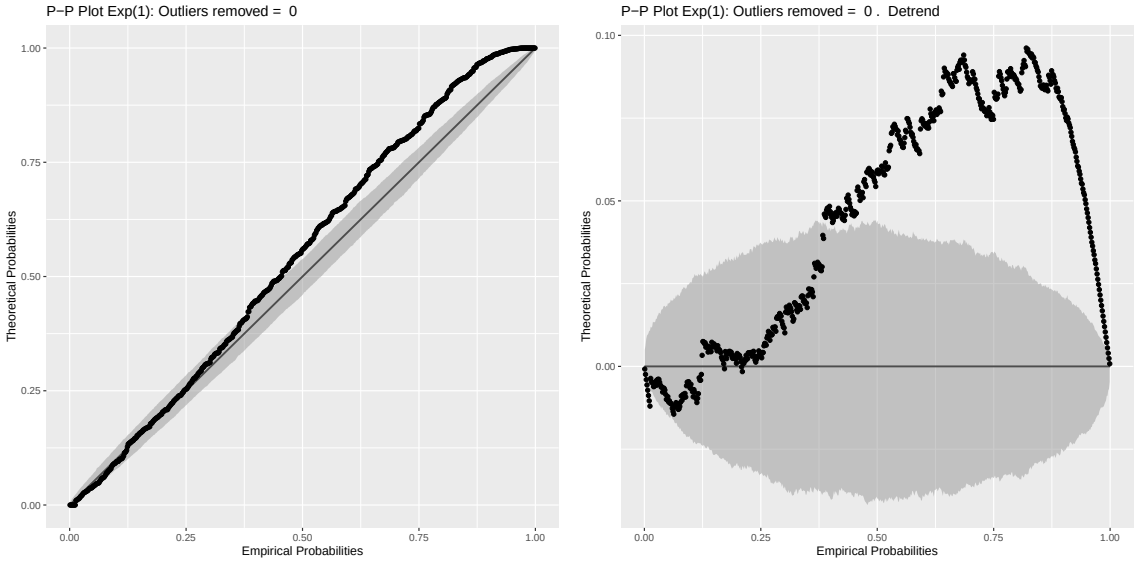
While the Q-Q plot represents quantiles, the P-P plot allows to represent theoretical and empirical probabilities. A P-P plot is a probability plot used to assess how closely two datasets agree, and that plots the two CoDFs against each other. P-P plots are vastly used to evaluate the skewness of a distribution.

As a Q-Q plot, the P-P plot allows to assess the goodness of fit for the normalized sample as well as the validity of the GIF of the MPP model we build. The plot uses the normalized sample, so the proof that supports it is the same one as for Q-Q plot, that is to say Proposition 3.3.

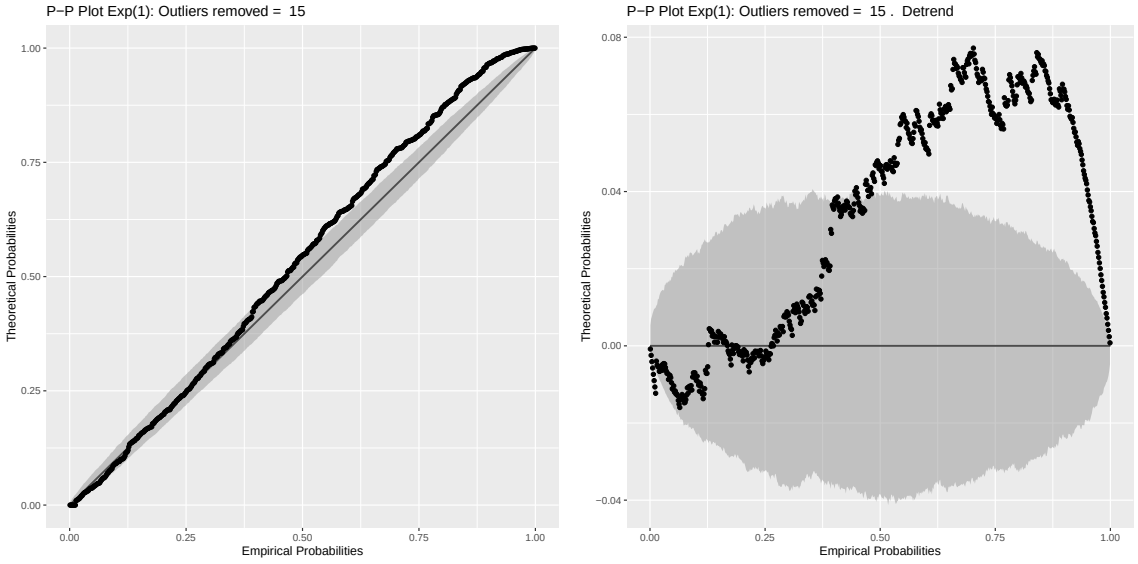
We display the P-P plots in Figure 6.17 and Figure 6.18 for different numbers of outliers we removed (as we did previously with Q-Q plots). We add confidence bands (in grey) to detect natural random little deviations from larger and more abnormal ones. For each subfigure, on the left we have the classic P-P plot, and on the right we have the corresponding P-P plot that is detrended in order to reduce visual bias caused by the orthogonal distances from points to the reference lines [44]. This bias may cause wrong conclusions to be drawn via visual inference of the plot. With the detrended plot, we can see where and how much the points deviate from the line or the confidence bands more easily and precisely.

We observe the shift in probabilities in the ending tail of the distribution, like we did before with Q-Q plots. As before, we remove outliers to see if it reduces the issue. In Figure 6.18b, the deviation is totally eliminated and the remaining deviation can be considered to be the result of inherent randomness of the model (because the dots are included in the grey

confidence bands).



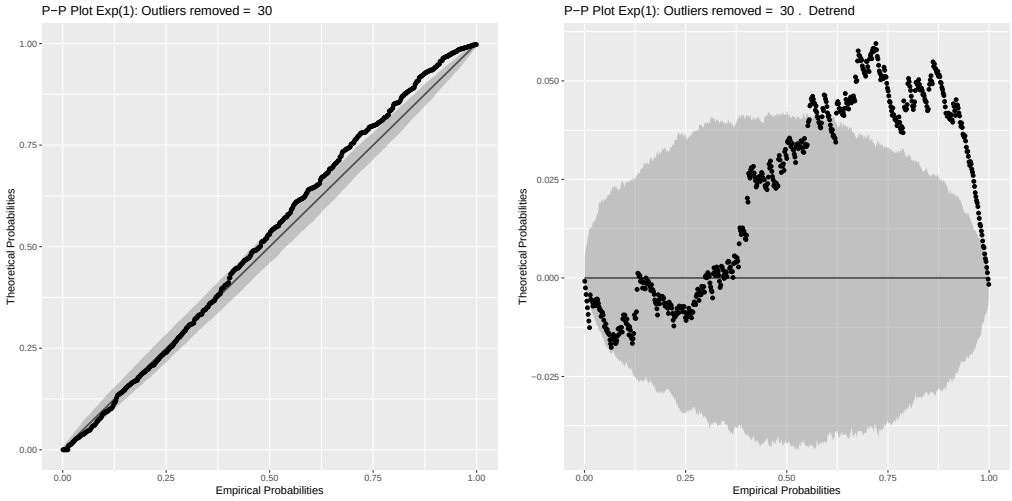
(a) Removed = 0



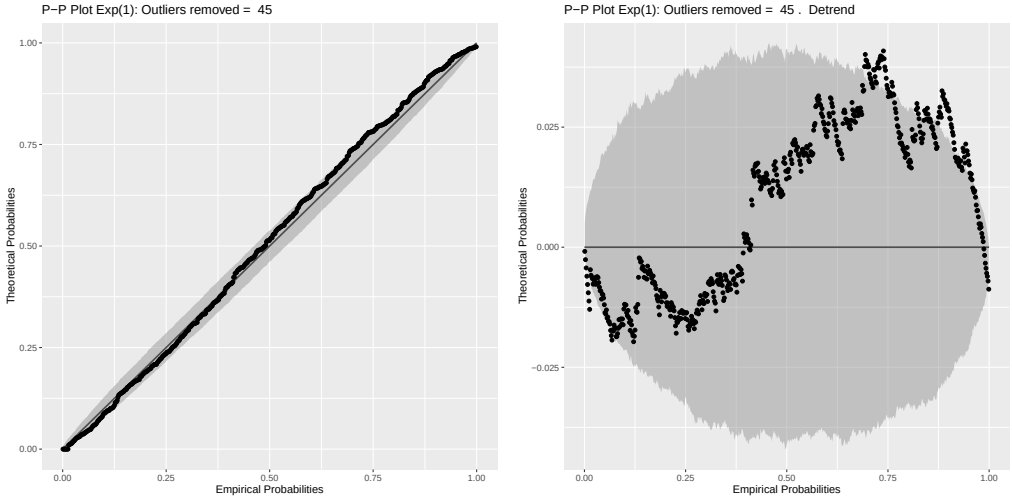
(b) Removed = 15

Figure 6.17 P-P Plot for different number of removed outliers

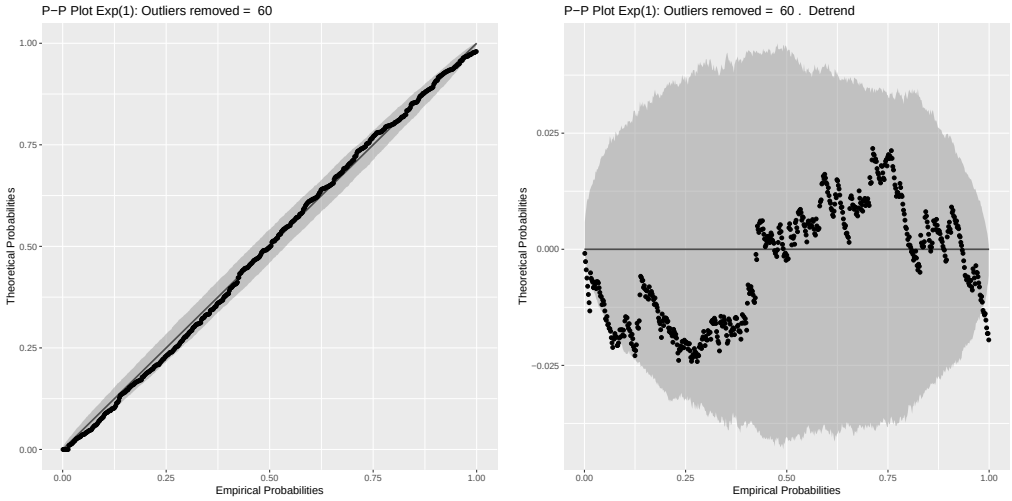
Also, if we start to remove too many of the highest values, as in Figure 6.18c, the probability points begin to shift down too much, and to go beyond the confidence bands. This is a good sign which means we also need the highest values of the sample to have a good fit. It confirms that only the very ending tail of the sample does not fit well the MPP model we develop.



(a) Removed = 30



(b) Removed = 45



(c) Removed = 60

Figure 6.18 P-P Plot for different number of removed outliers

6.3.3.3 Model checking: fit a distribution on normalized data

We can also think of another way to validate our model. Until now, we compared the normalized sample to a unit rate Poisson process and checked how good the match was. However, we can try another approach in which we simply fit a Poisson process on the normalized sample, without fixing any particular rate, and evaluate how probable it is that the sample comes from a Poisson process. We present the diagnosis plots of the exponential fit in Figure 6.19.

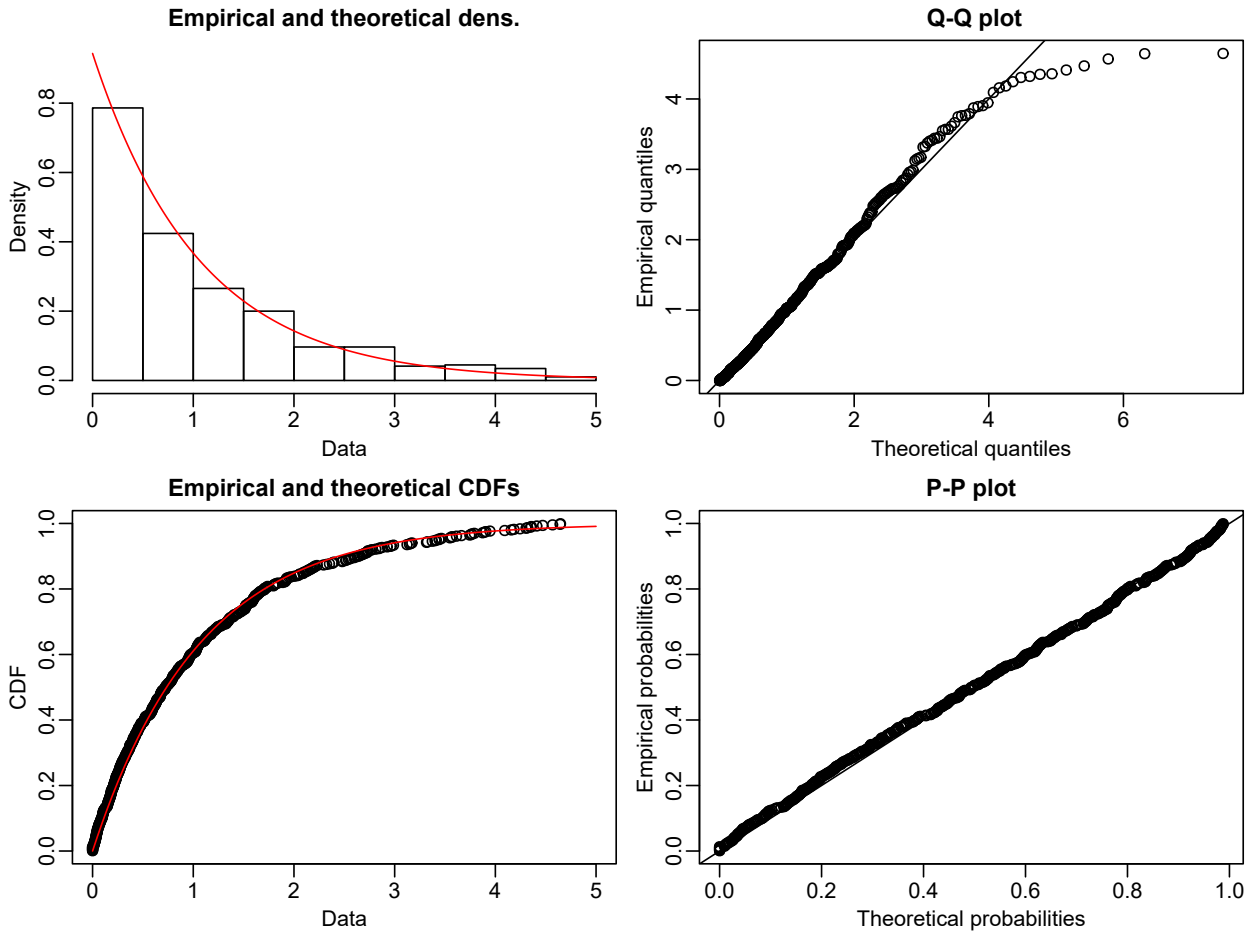


Figure 6.19 Exponential Fit on Normalized data

The diagnoses plots show a really good fit. The histogram matches definitely well the exponential distribution, and the CoDF and P-P plots show a genuine match between theory and observation. Concerning the Q-Q plot, it is correct except for the tail of the distribution that shows again the recurrent issue we had before. The highest values seem to have a particular behaviour, different from the rest of the sample.

Concerning the estimated rate of the process, the distribution of the sample is theoretically supposed to be the same as a unit rate Poisson process, *i.e.* a $\mathcal{E}xp(1)$. Meanwhile, the fit gives a rate $\lambda = 0.86 \pm 0.04$ which is really close to the theoretical value 1 that represents a situation where everything in practice is the same as in theory (model assumptions, estimated rates, not polluted observed values, ...). Also, when fitting the exponential distribution, we did not remove any outlier values, which has the effect of reducing the estimated rate of the process.

The fact that the exponential distribution fits so well the normalized sample is a sign of a genuine coherent model.

6.3.3.4 Performances of Method 2

We test performances of Method 2 by computing indicators we detailed earlier on the verification methodology. We compare it with Method 1 and the Baseline Policy.

Table 6.9 Methods performances comparison

Method	MAPE	C-index
TNO - Baseline Policy	700	0
TNO - Method 1: $\widehat{\lambda}_1$	677	0.66
TNO - Method 2: $\widehat{\lambda}_2$	480	0.66
TNO - Baseline Policy $\rightarrow$ Method 2	- 31.4%	$+\infty$
TNO - Method 1 $\rightarrow$ Method 2	- 29.1%	=

We give the results in Table 6.9, where we can observe that:

- Choosing the Baseline Policy over the Method 2 increases the MAPE error by $100 \times \frac{700-480}{480} = 45.8\%$. The Method 2 is way better than the Baseline Policy, since it makes the MAPE drop by 31.4%. Method 2 outperforms Method 1, with a 29.1% MAPE decrease.
- For the C-index, Method 1 and Method 2 give the same performance with a 66% rate of ordering well the patients (based on the observed time).

As a conclusion, Method 2 gives better results than Method 1 on several and complementary aspects. We decide then to choose the statistic $\widehat{\lambda}_2$ to estimate the GIF λ of the MPP model we develop.

6.4 Define and estimate the mark distribution: the quality Q

The second step to completely define the MPP is to choose the mark CoDF of the process, which stands for the quality of kidneys. We have to select a probability law that fits well the quality marks on patients' histories. We compare the different candidate distributions we introduced in Section 5.1.4.1, using indicators like Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC). Once we define the mark CoDF, the MPP is totally defined. The marked aspect of the MPP makes it possible to answer the TNBO problem.

6.4.1 Selection procedure of the quality distribution

We already employed Gamma and Weibull distributions during experiments in Section 6.1, as they are good candidates to fit well the distribution of kidney quality. As we saw, it seems they both gave similar results, and both fit quite well. However, the systematic selection between the two has not been discussed. We need to select only one of the distribution to define the mark CoDF of the MPP. We will use statistical criteria to evaluate the quality of fit and obtain the better one. For that, we introduce the BIC,

Definition 6.4 (Bayesian Information Criterion)**.** For a k-parameters model $\mathcal{M}$ fitted on a n-sample X,

$$BIC = k\ln(n) - 2\ln(\hat{L}), \tag{6.5}$$

where $\hat{L}$ is the maximized value of the likelihood function of the model $\mathcal{M}$, *i.e.*

$$\hat{L} = \max_{\theta}\left[p(X \mid \theta, M)\right] = p(X \mid \widehat{\theta}^{MLE}, M). \tag{6.6}$$

And we also introduce the AIC,

Definition 6.5 (Akaike Information Criterion)**.** For a k-parameters model $\mathcal{M}$ fitted on a n-sample X,

$$AIC = 2k - 2\ln(\hat{L}), \tag{6.7}$$

where $\hat{L}$ is defined in the same way as in Equation (6.6) in Definition 6.4.

In our case, we need to compare Gamma and Weibull distribution, which both have $k = 2$ parameters. Then, for a given n-sample of qualities, the term $k\ln(n)$ of the BIC expression in Equation (6.5) is fixed, whatever the distribution is. In the same way, for a fixed sample, the term $2k$ of the AIC expression in Equation (6.7) is fixed from one distribution to the other. Hence, the difference in goodness of fit from one model to the other will only come

from the maximization of the likelihood $\hat{L}$.
Therefore, the 2 criteria will order the two tested distributions in the same positions, resulting in the choice of the same distribution. Consequently, we can rely on only one of the criteria for our experiments. Here, we select the BIC.

Remark 6.4. the AIC and BIC criteria are proved to be able to estimate how much more (or less) information is lost by choosing one model or another, which makes them convenient tools to decide between possible models. However, these estimates are only valid asymptotically, when the number of data points is large [45, 46].
This could cause problems with some patients, whose histories only have a few data points (low number of offers $n = N_{\text{offer}} = 10$ for instance). In this case, and even in the case $n/k < 40$ according to [47], it is recommended to use rectified criteria, that concentrate on the difference in the number of parameters between the models we are testing (like the rectified second-order criterion $AICc$).
Nonetheless, in our situation, we simply need to compare two distributions that both have the same number of parameters ($k = 2$), and each time we compare them, it is on the same sample with n fixed (a given history). Therefore, we can stick to non-rectified criteria, such as AIC and BIC, to compare the candidate distributions. □

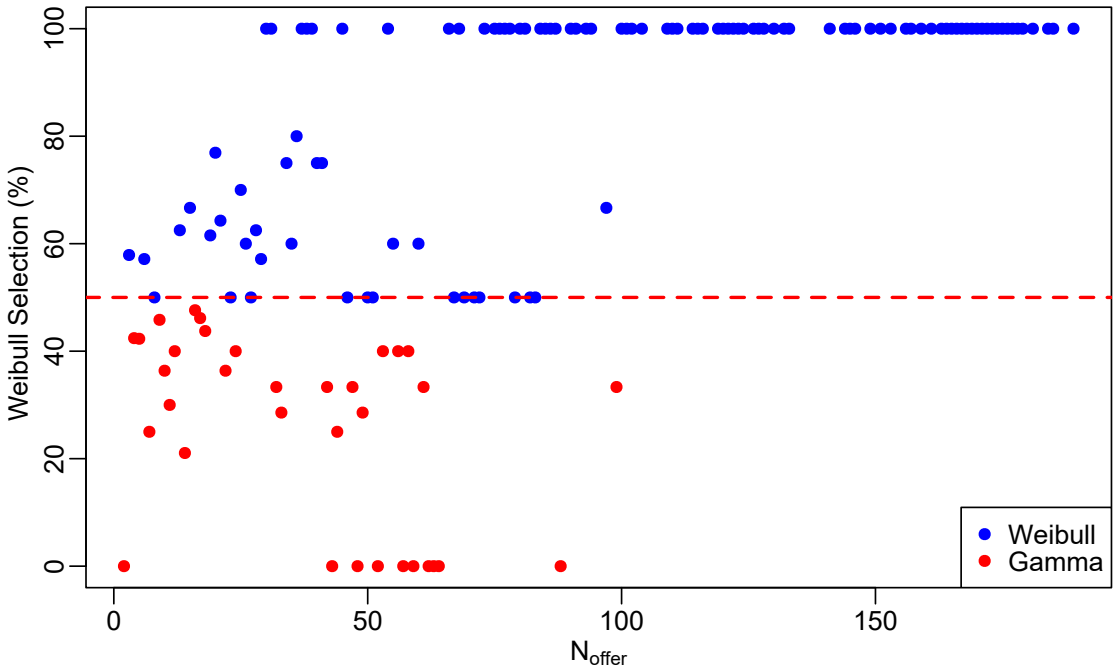


Figure 6.20 Weibull vs Gamma selection over BIC criteria, in terms of N_{offer}

Test over the validation set. We test if $BIC_{Weibull} < BIC_{Gamma}$ on the histories of the patients of the validation set. We find that $BIC_{Weibull} < BIC_{Gamma}$ 57.1% of the time. We plot the percentage in which Weibull law was chosen over the Gamma law in Figure 6.20, in terms of the number of offers N_{offer} on considered histories. The red-dotted line shows the 50-50 percentage where Weibull is selected as often as Gamma. If the quality of fit between the two distributions is rather the same for low number of offers, it is definitely better with the Weibull law for a high N_{offer}. Hence, the right choice would be to take Weibull distribution to model the quality distribution of the MPP.

6.4.2 Testing Lognormal distribution

Even if Weibull distribution seemed to have a good fit, it is still reasonable to confirm it by trying another possible distribution. Thus, we apply the Lognormal distribution over the validation set to see if it outperforms the Weibull distribution. The Lognormal distribution is denoted as Lognormal(θ_1, θ_2), with parameters $\theta_1 \in \mathbb{R}$ and $\theta_2 > 0$. It was also introduced in Section 5.1.4.1.

Selection over validation set among 3 laws. We did the selection using BIC on the validation set, as what we did previously with only Gamma and Weibull. For each data point, we choose the distribution that gives the lowest BIC, which is the one supposed to fit best the quality sample (the qualities on the history of the corresponding patient). For each data, the best distribution is said to be "selected".

If we count each selection individually, *i.e.* for each one of the data point, we get the followings figures: the proportions in which Weibull, Gamma and Lognormal are selected are respectively 51.4%, 0.5% and 46.4%. Figures are summarized in Table 6.10.

Table 6.10 Best density fit selected over BIC

	Weibull	Gamma	Lognormal	No fit
Number of selection	401	4	362	13
Proportion of selection (%)	51.4	0.5	46.4	1.7

For the patients with no more than 2 offers on their history ($N_{\text{offer}} < 2$), no distribution fit can be performed at all, as the column 'No fit' in Table 6.10 shows. This issue has already been risen in Section 6.1.

In Figure 6.21 we are now reasoning by clusters, grouping data by same N_{offer}. For a given N_{offer}, we fit all three laws on each data of the group. Then, we compute the proportion in

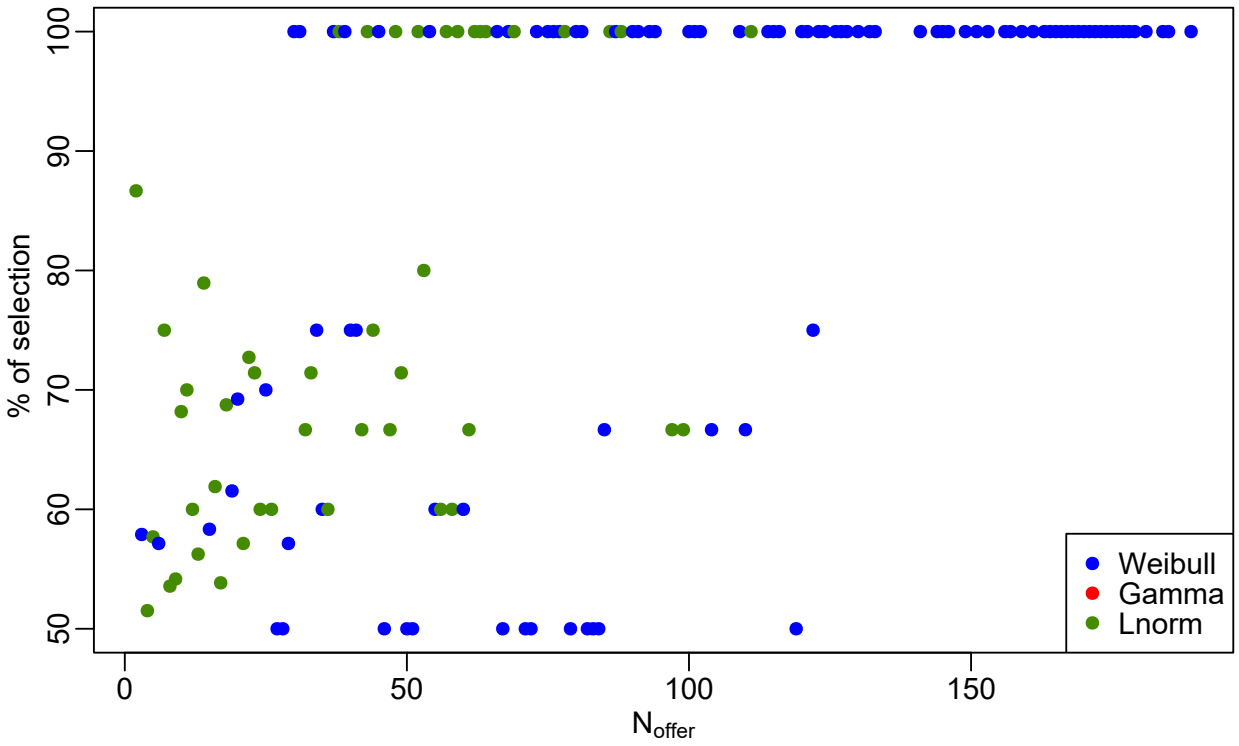


Figure 6.21 Weibull/Gamma/Lognormal: Best density fit selected over BIC

which each distribution has been selected, and plot the one which has been selected the most on the graph, with corresponding proportion.
We can see in Figure 6.21 that:

- Gamma is never the best fit. Weibull or Lognormal are always preferred.
- For low N_{offer}, Lognormal and Weibull are chosen half of the time approximately. They are tied. With increasing N_{offer}, Weibull tends to be chosen more often, to the point it is always the best of the three.
- One can argue that a larger number of offers in a history can lead to a better and more faithful representation of the quality. Hence with a large N_{offer}, randomness is reduced and the sample is closer to the "real" shape of the quality distribution. In this case, we should probably go with the systematic selection of Weibull and decide to consider Weibull as the most appropriate distribution for our problem.

In the light of its clear superiority during the experiments, we decide to stick with the Weibull distribution to model the mark CoDF of the MPP.

6.4.3 Method adjustment to tackle hard to fit patients challenge

We apply the method adjustment we discussed in Section 6.1.3 to fix the challenge of fitting the mark distribution for patients with a low N_{offer}. Consequently, we adapt the algorithm and then proceed to the same methodology we did before in Section 6.4.2 to select one of the three candidate distributions.

If we count each selection individually, *i.e.* for each one of the data point, we get the followings figures: the proportions in which Weibull, Gamma and Lognormal are selected are respectively 97.2%, 0% and 2.8%. Figures are summarized in Table 6.11.

Table 6.11 Best density fit selected over BIC - Adjustment

	Weibull	Gamma	Lognormal	No fit
Number of selection	758	0	22	0
Proportion of selection (%)	97.2	0	2.8	0

It is worth noting that the column 'No fit' in Table 6.11 is empty. It means a quality distribution has been fitted to each one of the patients' history. The previously pinpointed problem about patients with a too low number of offers ($N_{\text{offer}} < 2$) has been solved. The number of offers N_{total} before thinning is always greater than 2.

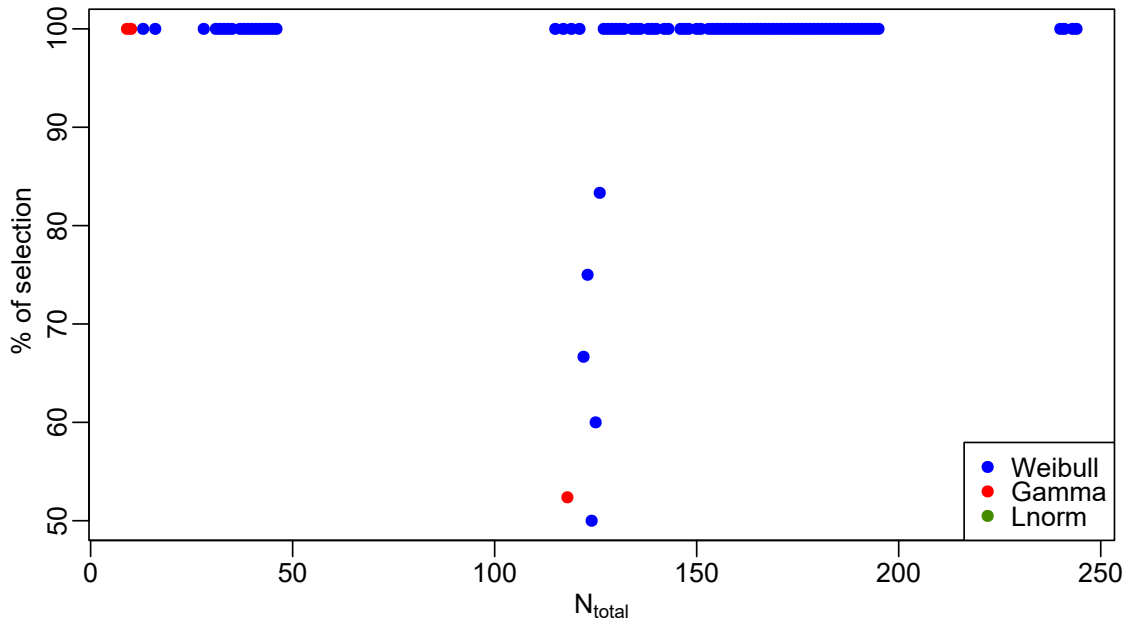


Figure 6.22 Weibull/Gamma/Lognormal: Best density fit selected over BIC - Adjustment

Also, we produce in Figure 6.22 the same diagnostic graph as in Figure 6.21 to represent which distribution is the most selected one in terms of the total number N_{total} of offers in histories before thinning. The conclusion is clearer than before: the Weibull distribution is the best option to model the quality distribution.

In addition to tackle hard to fit patients challenge, this adjustment in the method also emphasizes the relevance of the use of Weibull distribution to model quality.

6.5 Combining rate and quality: the answer to Time before Next Better Offer

In this section, we apply the solution presented in Section 5.2.2. We proceed to the MPP thinning by using the quality distribution, so we can predict waiting times before a better kidney proposal, and thus answer the TNBO problem. We refer to this method as Method 2 MPP, by opposition to Method 1 of previous work [1], and in order to indicate the difference with Method 2 (that does not use the marked aspect of the process). To validate the solution, we use the same verification methodology as in Section 6.3 with the TNO solution.

6.5.1 Result on original dataset

We first apply the TNBO solution we developed on the original dataset, the one we have been using from the beginning. The experiments are still carried out on the validation set we defined in Section 6.2.

Data proportions On the whole validation set, we can have different types of data. There are censored and uncensored data, but there are also data that correspond on the one hand to a TNO, and on the other hand to a TNBO. For the verification process of the TNBO solution, we will run the solution only on TNBO data points. Then, it is fair to check if there are enough TNBO data points available to validate our solution. We present the proportions of all different data types in Figure 6.23.

Figure 6.23 Data proportions in terms of their nature

Given the reasonable proportion of TNBO data points, it seems that we have the required amount of resources to test the MPP method we developed to tackle TNBO issue.

Performances We run the TNBO solution to predict waiting TNBO. Then, we evaluate performances and compare them to the Baseline Policy. We give the results in Table 6.12. From the results and complementary analyses (*e.g.* residuals analysis), we conclude that:

- We obtain abnormally bad results with the original dataset, in fact even worse results with our MPP method than with the Baseline Policy. The MAPE score was 703 for our method against 660 for the Baseline Policy, which accounts for a 6.5% increase.
- The process thinning we do to answer TNBO is too strong and leads to overestimating the real observed TNBO (the target we want to predict).

Table 6.12 Methods performances comparison - on original dataset

Method	MAPE	C-index
TNBO - Baseline Policy	659.5	0
TNBO - Method 2 MPP	702.6	0.627

These unexpected results led us to search for an explanation by checking meticulously the nature of the dataset we use. Finally, we pinpointed one aspect about the way the dataset is built, which explains the issue we just encountered.

6.5.2 Rebuilding dataset

As we explain in Section 4.2.1 and Section 6.1.1, the building process of the dataset is retrieved from the former work [1]. During this building process, only pairs of immediate consecutive offers are retrieved from raw data to build the interevent waiting time targets. We needed to change the way the dataset is built to retrieve true next better offers. We refer to the Section 4.2.1 where this issue is already discussed. We illustrate the point in Figure 6.24 to summarize it.

CAN_ID	DON_ID	CENSORED	UPDATE_TM	TIME_NEXT_OFFER	KDRI	KDRI_NEXT
694	1204	1	2015-04-20	4	1.5151059	2.3535731
694	1209	1	2015-04-24	2	2.3535731	1.4091363
694	1211	1	2015-04-26	13	1.4091363	1.1232409

Real Next_Better_Offer

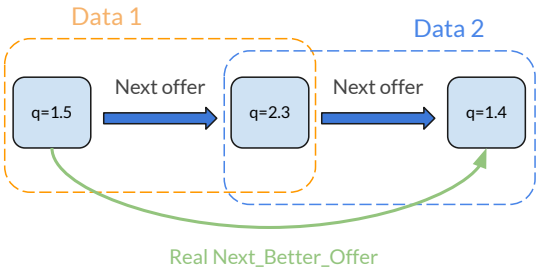


Figure 6.24 Build the new targets

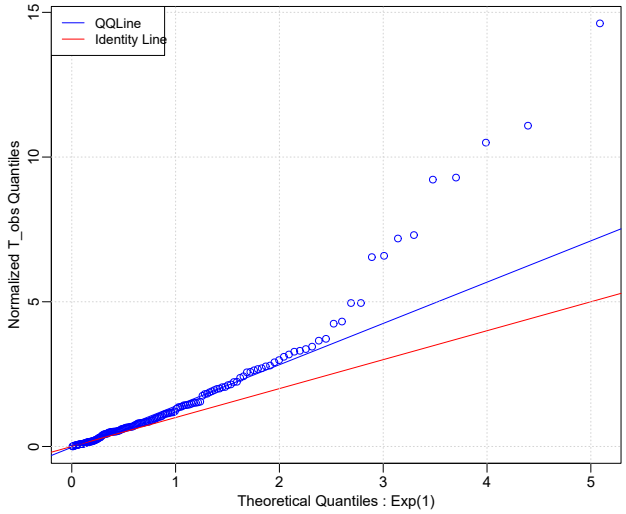
Once we tackled this issue, we are able to run again the verification methodology on the new dataset with valid TNBO targets. As what we did with the original dataset in Section 6.2.5, we partition the new dataset in a validation set and a test set, at a rate of 50% of total data each. In this section, experiments are run on the validation set.

6.5.3 Q-Q plots an P-P plots of normalized process

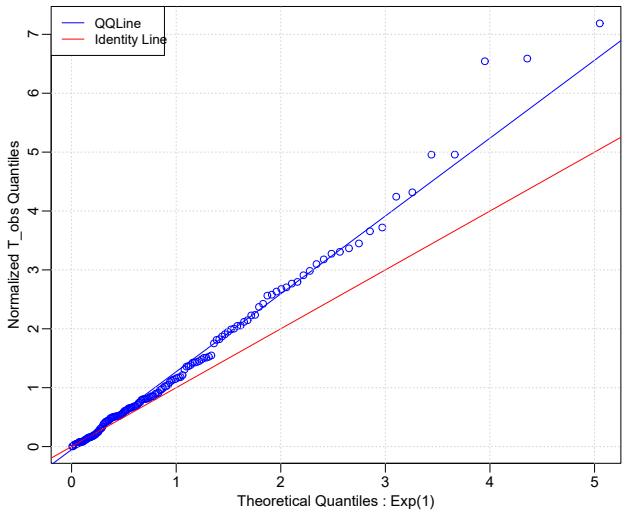
As a first step, we proceed to the Q-Q plot verification. We draw Q-Q plot of the normalized sample in Figure 6.25. For each subfigure, we removed a given number of outliers, in order to evaluate if we get a better fit.

In Figure 6.25c we remove 12 outliers, accouting for $12/164 = 7\%$ of the whole dataset. The quantiles were all brought closer to the red line. In addition, they form quite a straight line meaning that the theoretical model (exponential distribution) is genuinely correct for the normalized sample we are testing.

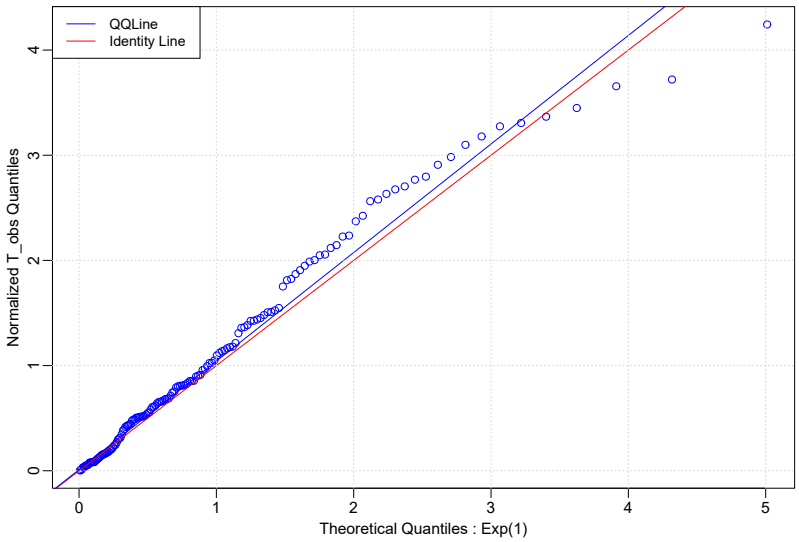
We can see that there is a lot more noise in the graphs in Figure 6.25 than in the ones in Figure 6.16. This is partly because of the significant difference in the amount of data. Indeed, less data generally means more noise.



(a) Removed = 0



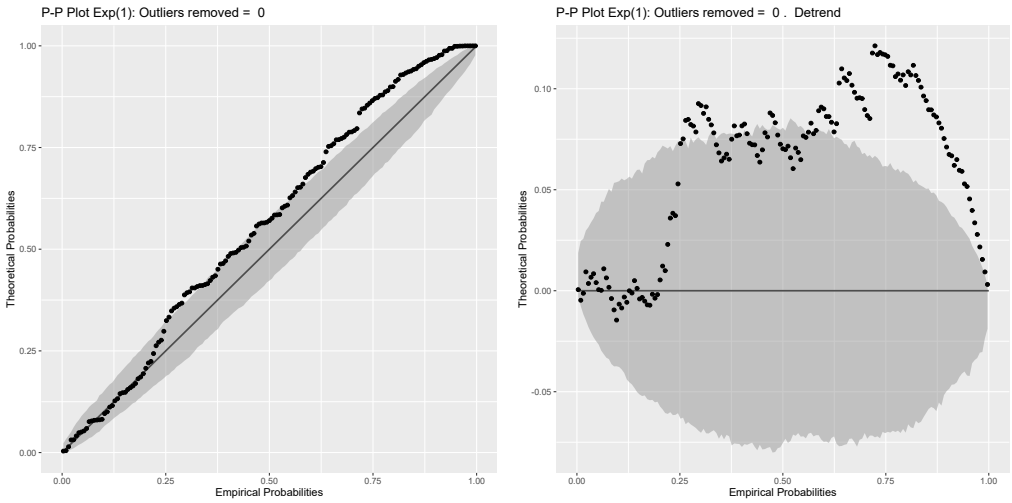
(b) Removed = 6



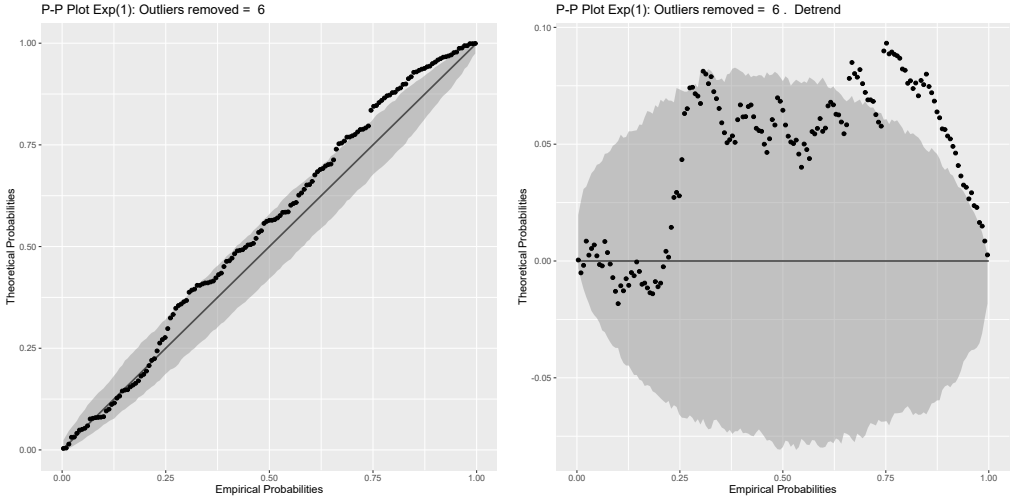
(c) Removed = 12

Figure 6.25 Normalized Q-Q plot for TNBO

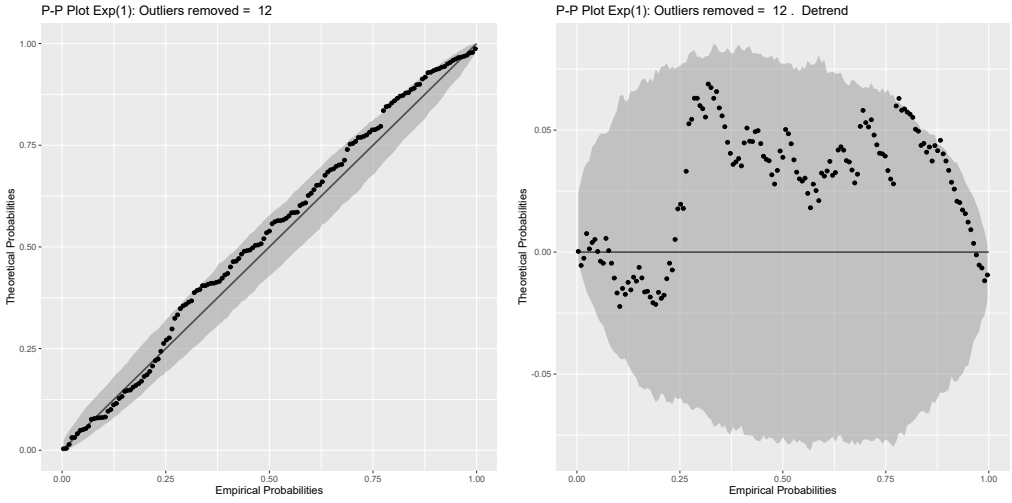
As a second step, we draw the P-P plots of the normalized sample in Figure 6.26, as in Figure 6.17 before. Like in the Q-Q plots, we observe more variability than in Figure 6.17, but we can see the pattern is quite the same. The beginning of the points distribution is close to the theoretical line, and then we observe a deviation of the empirical points from the theory, which results in points outside the gray confidence bands. However, removing some outlier values here also solves the shift by translating the points into the confidence bands. Which means the model is well adapted to the remaining 93% of the data.



(a) Removed = 0



(b) Removed = 6



(c) Removed = 12

Figure 6.26 P-P Plot for different number of removed outliers

6.5.4 Performances of MPP Method - Comparison

We now test performances of the method by computing indicators, in the case of the TNBO problem. We compare the Method 2 MPP with both the Baseline Policy and the former Method 1. We give results in Table 6.13, where it can be seen that:

- The Baseline Policy is outperformed by far by both Method 1 and Method 2 MPP. The Method 1 reduces the MAPE error by $100 \times \frac{649.2-335.5}{649.2} = 48.3\%$, while the Method 2 MPP reduces it by nearly 60%. This means both Method 1 and Method 2 MPP predictions of waiting times are closer to real observed values.
- However, Method 2 MPP gives a significant better score than Method 1, with a 21.8% reduction of the MAPE error score.
- Concerning C-index, Method 1 and Method 2 MPP both give almost the same percentage of ordering well the patient (based on the observed time). We can note a slightly better performance with Method 2 MPP, a 2.2% increase (proportionally). Concerning the Baseline Policy, it still does not offer the possibility to order patients so the C-index is null.

Table 6.13 Methods performances comparison for TNBO

Method	MAPE	C-index
TNBO - Baseline Policy	649.2	0
TNBO - Method 1	335.5	65.0
TNBO - Method 2 MPP	262.2	66.4
TNBO - Baseline Policy $\rightarrow$ Method 2 MPP	- 59.6%	∞
TNBO - Method 1 $\rightarrow$ Method 2 MPP	- 21.8%	+2.2%

In conclusion, the verification procedure we conducted on Method 2 MPP shows it is a viable way of modeling the TNBO problem. Moreover, in terms of performances, Method 2 MPP has a better predictive power than Method 1, and is much better than Baseline Policy. The predictions of the waiting TNBO made by the new MPP model are closer to reality.

6.6 Results on the test set

During all previous sections in current Chapter 6, the validation set is used to define the model (selection of quality distribution), assess the current model, and bring modifications to it if the tests reveal limitations. In the end, we have a fixed definitive MPP model. In this section, we evaluate the performances of this final MPP model. For this purpose, we use the test set, which we kept unused so far. In order to confirm the validity of the model, we run the same verification methodology we carried out on the validation set in the Section 6.3.3 for TNO problem and in the Section 6.5 for TNBO problem.

6.6.1 Results for TNO

We first test the model for the TNO problem.

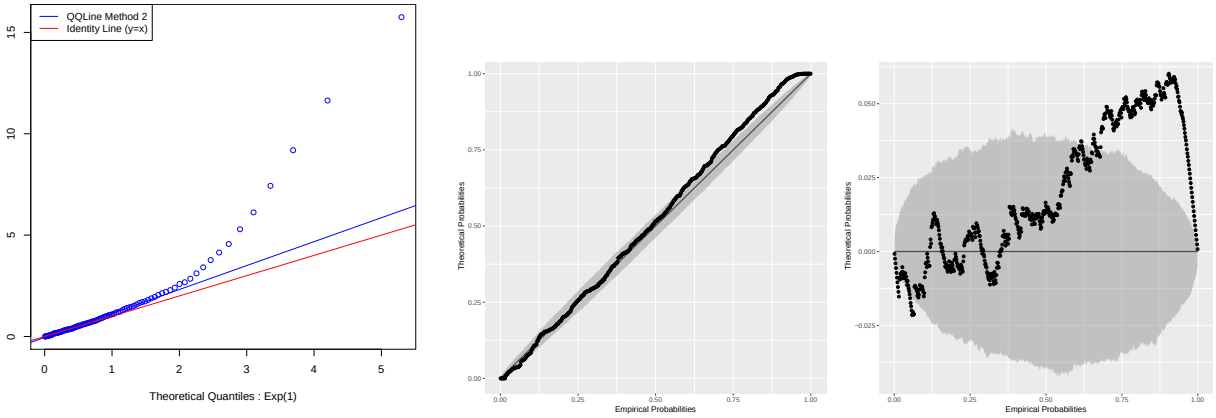
Performances. We indicate the performances of Method 2 as well as Method 1 and Baseline Policy in Table 6.14.

The indicators values (MAPE and C-index) are close to the ones we obtained with the validation set (see Table 6.9). The improvement percentages from one method to another are also similar. This is a good indicator that the method is consistent, and that it does not fluctuate from one dataset to another. Also, Method 2 still performs better than Method 1 and the Baseline Policy. It improves significantly the predictive power of the model.

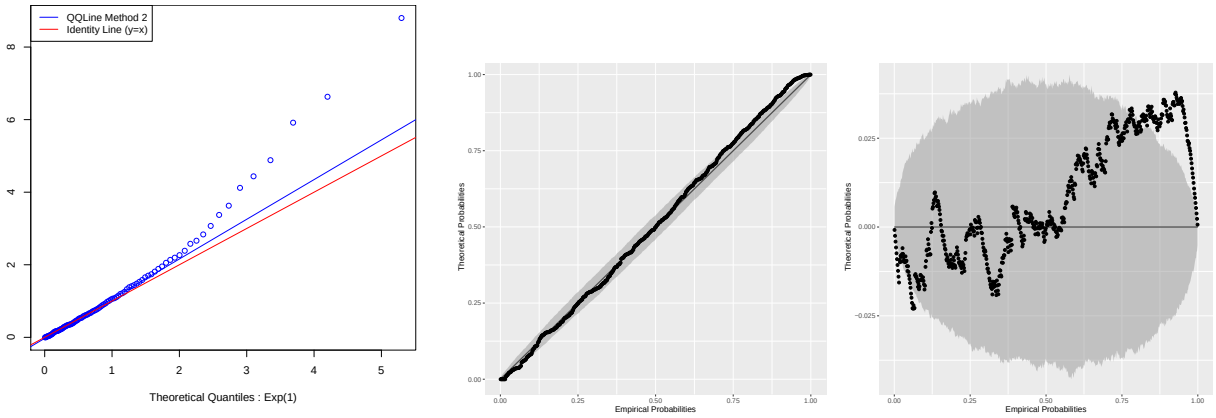
Table 6.14 TNO - Methods performances comparison

Method	MAPE	C-index
TNO - Baseline Policy	817.1	0
TNO - Method 1: $\widehat{\lambda}_1$	688.9	0.718
TNO - Method 2: $\widehat{\lambda}_2$	513.3	0.722
TNO - Baseline Policy $\rightarrow$ Method 2	- 37.2%	$+\infty$
TNO - Method 1 $\rightarrow$ Method 2	- 25.5%	+0.6%

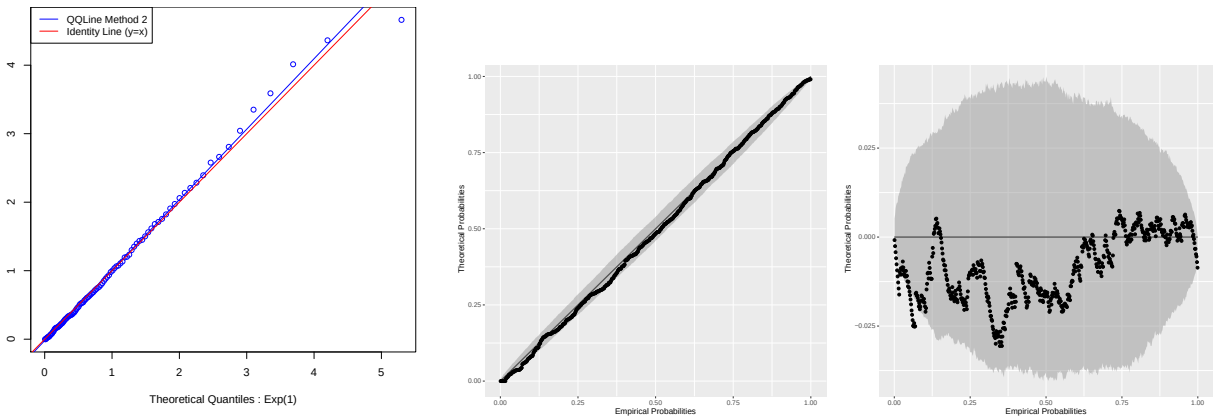
Q-Q and P-P plots. We give the Q-Q and P-P plots for different relevant numbers of outliers removed in Figure 6.27. With 35 outliers removed in Figure 6.27c, we reach a quite perfect straight line and a unit slope in the Q-Q plot. In the same way, the P-P plot shows a quite perfect line, inside the gray confidence bands. We get the same results as with the validation set. A removal of 35 values accounts for 5.6% of the total 624 data points we use in the normalization verification process.



(a) Removed = 0



(b) Removed = 15



(c) Removed = 35

Figure 6.27 Q-Q and P-P Plot for different number of removed outliers

Fit an exponential distribution on the normalized data. As in Section 6.3.3.3, we fit an exponential distribution on the normalized sample we get with the test set. We present the diagnosis plots in Figure 6.28. The plots look exactly the same as in Figure 6.19 for the validation set, so the same analysis can be conducted and the same conclusions can be drawn. We can conclude that the model matches the test set well too.

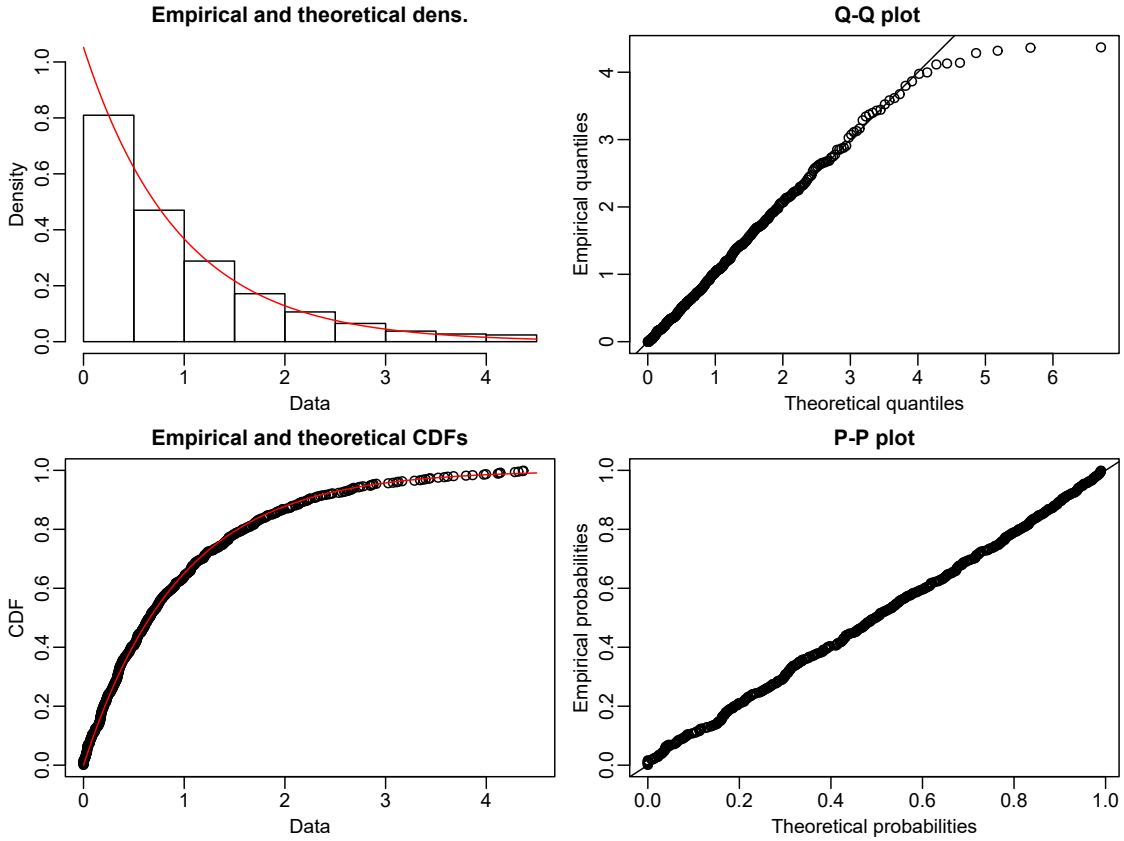


Figure 6.28 Exponential Fit on Normalized data - Test Set

In conclusion, we get exactly the same results on the test set as on the validation set. In both cases, the model we built improves significantly the predictions for TNO issue compared to Method 1. Moreover, this similarity in the results between validation set and test set also means the method we developed is consistent for the type of dataset we build it for.

6.6.2 Results for TNBO

In this section, we test the model for the TNBO problem.

Performances. We demonstrate the performances of Method 2 MPP as well as Method 1 and Baseline Policy in Table 6.15.

Method 2 MPP performs better than the two others methods. A solid MAPE decrease of 22.3% is observed when using Method 2 MPP instead of Method 1. The C-index is quite the same for both Method 2 MPP and Method 1. As with the TNO solution, we cannot notice a significant improvement in the discriminative power.

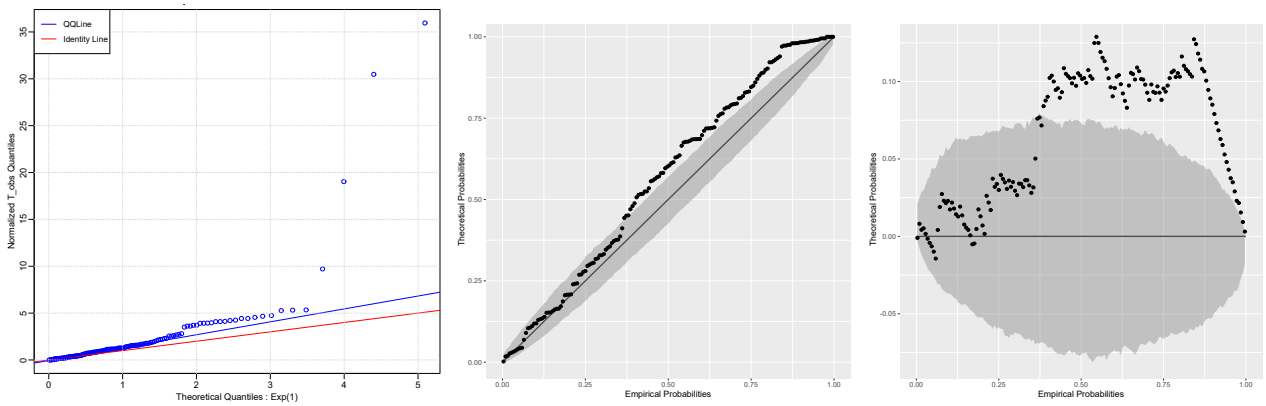
The indicators values (MAPE and C-index) are not the exact same ones as we obtained with the validation set (see Table 6.13), but they remain comparable. This difference was predictable since the validation and test sets are different. Moreover, the small size of the sets increases the variability and the difference between them (compared to bigger sets that would tend to be more similar).

However, the relative improvements in performances from one method to another are comparable to the ones in Table 6.13. This proves the consistency of the TNBO solution performances.

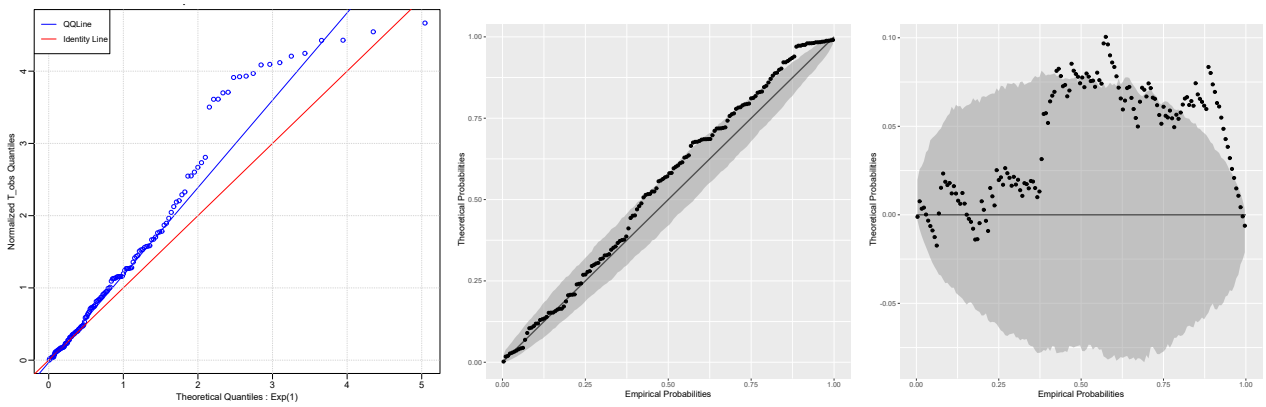
Table 6.15 Methods performances comparison for TNBO

Method	MAPE	C-index
TNBO - Baseline Policy	564.3	0
TNBO - Method 1	212.4	63.8
TNBO - Method 2 MPP	165.1	65.3
TNBO - Baseline Policy $\rightarrow$ Method 2 MPP	- 70.7%	∞
TNBO - Method 1 $\rightarrow$ Method 2 MPP	- 22.3%	+2.4%

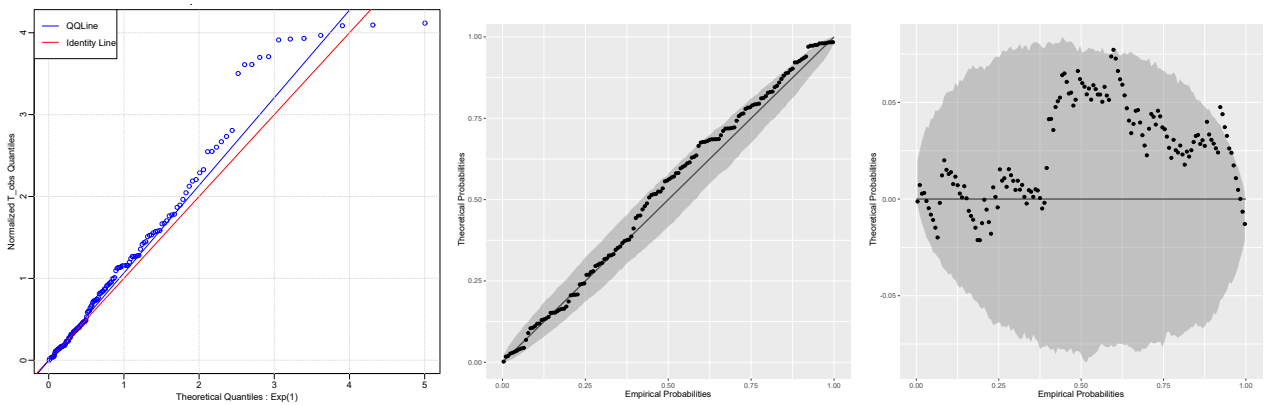
Q-Q and P-P plots. We present the Q-Q and P-P plots for different relevant numbers of outliers removed in Figure 6.29. In the same way as with previous Q-Q and P-P plots for the validation set in Section 6.5.3, removing outliers improves the match with the theoretical unit rate exponential distribution. For 14 outliers removed in Figure 6.29c (accounting for 8.6 % of the set), the P-P plots points are all in the gray confidence bands, and the quantiles of the Q-Q plot are quite aligned. More generally, as with validation set in Figure 6.25, high variability remains, particularly in the ending tail. Moreover, the Q-Q plot points pattern is similar to the one in Figure 6.25.



(a) Removed = 0



(b) Removed = 8



(c) Removed = 14

Figure 6.29 Q-Q and P-P Plot for different number of removed outliers

In conclusion, the Method 2 MPP developed to answer TNBO issue also proves its performances and its validity on the test set. The improvement in terms of predictions compared to Method 1 is comparable to the improvement we observed on the validation set (-22.3% MAPE on test set, and -21.8 % MAPE on validation set).

CHAPTER 7 CONCLUSION AND RECOMMENDATIONS

In this chapter, we first summarize the work achieved throughout the thesis and give concluding remarks. Then, we discuss limitations of the proposed method. Finally, we point out the potential future research that could be conducted, particularly in order to address current method weaknesses.

7.1 Summary of work

This thesis addresses the problem of informing an ESKD patient on his perspectives about a next kidney offer if he rejects the current one. The study of the problem is restricted to non-paediatric patients on a general scoring waiting list, without taking account special priorities. This kidney transplantation problem is a decision-making issue. Indeed, a patient and his nephrologist need to decide, during the SDM process, if they accept the kidney transplantation.

The current work intervenes in this SDM process, its objective is to help the patient decide if he should accept the current kidney transplant offer, or if he should rather wait for a next offer. The clinical tool developed enables him to access easily understandable information, *i.e.* waiting times. The method provides the expected time before a next offer, as well as the time to a next better offer. We think this kind of commonly understandable information can help patients in their decision, but also help them more generally in the disease fighting process as it helps them be proactive and empowered instead of passive. In addition, the access for the nephrologist to more complex information than waiting times could always be useful to better understand the situation himself, or to vulgarize it for the patient if he judges it relevant.

Many works in the literature addressed the organ transplantation issue but only a few took a one patient's side in the decision-aid tool they provided. Additionally, in the few examples we can find, information given by the tool still poses the question of interpretation for the patient, like in [22]. The work conducted in [1] palliated these limitations by answering the TNO issue, as well as laying a groundwork for the TNBO issue but without being conclusive about it.

The main objective of this thesis was to continue the work in [1] by designing a solution that consolidates the kidney quality aspect to answer the TNBO issue. To achieve this, the strategy adopted has been the MPP methodology that enables to model marked and

timestamped events. The methodology relies on a CIF that combines a GIF (intensity of the process) that describes the arrival of kidneys, and a mark CoDF that models the quality of kidneys.

The model we designed successfully gave personalized individual predictions of waiting times, taking into account patient characteristics through a particular history building as well as through the use of the KDRI quality index to carry out the MPP thinning. The predictions provide a solution to the *No question* issue from the patient aspect.

Moreover, the MPP method we developed provides other possibilities too, including giving customizable confidence intervals for waiting times, giving the possibility to change the desired quality threshold for the next kidney offer, and giving the expected quality of the next kidney. These options provide less easily understandable information and can be considered as *advanced* ones. If they should be offered or not to patients is a totally different question to answer. The aim of the tool is to inform the patient, not to confuse him. Besides, this personalized advanced information is understandable to the nephrologist, which can help him too during the SDM. Then, the solution answers the *No question* issue from the doctor aspect as well.

Performances obtained in both TNO and TNBO issues have proved that the new methodology completed the pinpointed missing aspects from previous work, and improved the predictive power of the clinical tool we are developing. Based on the MAPE error score on the test set, our proposed method outperformed an unmarked Poisson process by 25.5% for TNO problem and by 22.3% for TNBO problem. Moreover, we respectively observed a MAPE decrease of 37.2% and 70.7% when using our model instead of a baseline policy (a basic statistic that averages waiting times of all patients in the dataset). Additionally, the method has proved to be consistent between the validation set and the test set, in terms of absolute results, and in terms of relative improvements compared to competing methods.

7.2 Limitations

Obviously, the work we developed presents some limitations.

First, we demonstrated the validity of the model when removing a restricted number of outliers values during the verification procedure. Although we gave some potential causes, we could not identify the clear reason why these particular values were not shrunk correctly when normalized. It could potentially mean the process rates we estimated for these patients were not representative of their true process rates. It may be worth working on it to try to confirm potential causes, and potentially detect patients for which the method is not entirely

suitable. The access to better, more complete datasets and information about patients would help to achieve it.

Secondly, we used the KDRI indicator to evaluate the quality of a kidney, which is presented in [8]. The quality mark is a major characteristic in the MPP framework we used to answer the TNBO issue, thus the relevance of the quality indicator can influence directly performances of the method. For that reason, even if the KDRI indicator is a valid quality indicator, finding a better, more specific and adapted one could lead to a solution enhancement (*e.g.* the KDRI has been developed on American patients, while our work is about transplantation in Quebec). However, finding a better quality indicator for this kidney transplantation problem is an entire new research topic in its own which should be addressed at least partially by qualified nephrologists.

7.3 Future research directions

Beyond the improvements in terms of performances, our work aimed at setting a groundwork for future research in a longer-term vision. For instance, some of the limitations of this work could be addressed in future work.

In this research, all the different possibilities offered by the previous work [1] were not explored nor combined with current work. In this work, we limited the study to a given configuration of the previous solution from [1], and built the new methodology on top of it. Now that the MPP methodology that we developed proved to provide promising results for this one configuration, it could also be applied to other combinations of previous work options (another history building method for instance).

Also, we did not integrate the piecewise aspect of the previous work [1]. This piecewise aspect is supposed to update the rate of the Poisson process each year, accounting for the increase of patient's age and waiting time. With a full access to data and history building, this aspect could be integrated into the MPP methodology easily. It would involve the GIF and the mark CoDF of the process to be piecewise defined, and their parameters to be re-evaluated for each year elapsed since the time t_0 of the initial offer.

The MPP model we developed is not extremely complicated, so we have access to closed form results (expected waiting time, confidence intervals, etc.) as we demonstrated in Section 5.2. However, we also justified how the MPP framework provides powerful and convenient simulation possibilities in the case the MPP model becomes too complex to get exact analytical results. Simulation allows us to access empirical results that are good approximations of closed form results (refer to simulation Section 3.2.4.2 and the simulation application in

Section 5.2.3). As a consequence, even if the MPP theory introduced in Chapter 5 is not used to its full extent with the current model, a future research could explore more of its potential by designing a more advanced MPP model, as in [25]. One simple example of a possible improved model is described in Appendix A. Nonetheless, a model like this would require access to the real and complete history of the patients, when our current MPP model does not because of incompleteness of clinical data. CIHI annual report [3] claims that since 2011, Quebec is experiencing increased under-reporting due to administrative issues, and gives supporting statistics as a 40% completeness rate for Quebec dialysis data in 2013.

Finally, the possibility to develop these improvements, and more importantly the ability to test them, depend highly on the quality of the data we can access to. This is why a more systematic and complete collection and record of clinical data is a major issue for the research world to address more and more health problems efficiently. Especially with new methods like machine learning that require numerous quality data. It is worth noting that institutions like the data collect from the Canadian Institute for Health Information [48] are helping to pursue that objective as it is stated in the annual report [4] about organ replacement in Canada that: “it is only through the ongoing and systematic collection of data that sound information can be produced to assist with decision-making”.

REFERENCES

[1] J.-N. Weller, "Predicting next kidney offer for a kidney transplant candidate declining current one," Master's thesis, Dép. de Mathématiques et de Génie Industriel, École Polytechnique de Montréal, Montréal, QC, 2018. [Online]. Available: https://publications.polymtl.ca/3295/

[2] D. Harte, "Ptprocess: An r package for modelling marked point processes indexed by time," *Journal of Statistical Software, Articles*, vol. 35, no. 8, pp. 1–32, 2010. [Online]. Available: https://www.jstatsoft.org/v035/i08

[3] Canadian Institute for Health Information, "Canadian organ replacement register, annual report: Treatment of end-stage organ failure in Canada, 2004 to 2013," Canadian Institute for Health Information, Tech. Rep., 2015.

[4] ——, "Annual statistics on organ replacement in Canada: Dialysis, transplantation and donation, 2009 to 2018," Canadian Institute for Health Information, Tech. Rep., 2019.

[5] M. Carvalho *et al.*, "Robust models for the kidney exchange problem," *Informs Journal on Computing*, 2020. [Online]. Available: https://pdfs.semanticscholar.org/e1f1/b4bea7544a2ad34e9b8f0d1d78f62764aa8f.pdf

[6] M. Carvalho and A. Lodi, "Game theoretical analysis of kidney exchange programs," 2020. [Online]. Available: https://arxiv.org/abs/1911.09207

[7] Canadian Institute for Health Information, "Canadian organ replacement register, annual report: Treatment of end-stage organ failure in Canada, 2005 to 2014," Canadian Institute for Health Information, Tech. Rep., 2016.

[8] P. Rao *et al.*, "A comprehensive risk quantification score for deceased donor kidneys: The kidney donor risk index," *Transplantation*, vol. 88, pp. 231–6, 08 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19623019/

[9] S. Greenfield, S. Kaplan, and J. Ware, "Expanding patient involvement in care: Effects on patient outcomes," *Annals of internal medicine*, vol. 102, pp. 520–8, 05 1985.

[10] S. Greenfield *et al.*, "Patients' participation in medical care," *Journal of General Internal Medicine*, vol. 3, pp. 448–457, 09 1988.

[11] M. E. Brier, P. C. Ray, and J. B. Klein, "Prediction of delayed renal allograft function using an artificial neural network," *Nephrology Dialysis Transplantation*, vol. 18, no. 12, pp. 2655–2659, 12 2003. [Online]. Available: https://doi.org/10.1093/ndt/gfg439

[12] M. Luck *et al.*, "Deep learning for patient-specific kidney graft survival analysis," *arXiv preprint arXiv:1705.10245*, 2017. [Online]. Available: https://arxiv.org/abs/1705.10245

[13] O. Alagoz *et al.*, "Determining the acceptance of cadaveric livers using an implicit model of the waiting list," *Operations Research*, vol. 55, no. 1, pp. 24–36, 2007. [Online]. Available: https://doi.org/10.1287/opre.1060.0329

[14] ——, "Markov decision processes: a tool for sequential decision making under uncertainty," *Medical Decision Making*, vol. 30, no. 4, pp. 474–483, 2010.

[15] Z. Erkin *et al.*, "Eliciting patients' revealed preferences: An inverse markov decision process approach," *Decision Analysis*, vol. 7, no. 4, p. 358–365, Dec. 2010. [Online]. Available: https://doi.org/10.1287/deca.1100.0185

[16] B. Sandıkçı *et al.*, "Alleviating the patient's price of privacy through a partially observable waiting list," *Management Science*, vol. 59, no. 8, pp. 1836–1854, 2013. [Online]. Available: https://doi.org/10.1287/mnsc.1120.1671

[17] C. Harvey, "The kidney transplant process model," in *2015 Winter Simulation Conference (WSC)*, 2015, pp. 3176–3177.

[18] J. J. Abellan *et al.*, "Predicting the behaviour of the renal transplant waiting list in the pais valencia (spain) using simulation modeling," in *Proceedings of the 2004 Winter Simulation Conference, 2004.*, vol. 2, 2004, pp. 1969–1974 vol.2.

[19] Y. Chun and R. Sumichrast, "A rank-based approach to the sequential selection and assignment problem," *European Journal of Operational Research*, vol. 174, pp. 1338–1344, 02 2006.

[20] M. Bendersky and I. David, "The full-information best-choice problem with uniform or gamma horizons," *Optimization*, vol. 65, no. 4, pp. 765–778, 2016. [Online]. Available: https://doi.org/10.1080/02331934.2015.1080253

[21] ——, "Deciding kidney-offer admissibility dependent on patients' lifetime failure rate," *European Journal of Operational Research*, vol. 251, no. 2, pp. 686 – 693, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221715011182

[22] A. Wey *et al.*, "A kidney offer acceptance decision tool to inform the decision to accept an offer or wait for a better kidney," *American Journal of Transplantation*, vol. 18, no. 4, pp. 897–906, 2018.

[23] K. T. Islam *et al.*, "Marked point process for severity of illness assessment," ser. Proceedings of Machine Learning Research, F. Doshi-Velez *et al.*, Eds., vol. 68. Boston, Massachusetts: PMLR, 18–19 Aug 2017, pp. 255–270. [Online]. Available: http://proceedings.mlr.press/v68/islam17a.html

[24] C. C. T. Fok *et al.*, "A functional marked point process model for lupus data," *Canadian Journal of Statistics*, vol. 40, no. 3, pp. 517–529, 2012. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cjs.11136

[25] J. C. Weiss, "Clinical risk: wavelet reconstruction networks for marked point processes," 2019. [Online]. Available: https://openreview.net/forum?id=H1gNHs05FX

[26] W. Truccolo *et al.*, "A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects," *Journal of Neurophysiology*, vol. 93, no. 2, pp. 1074–1089, 2005, pMID: 15356183. [Online]. Available: https://doi.org/10.1152/jn.00697.2004

[27] M. Lefebvre, *Processus stochastiques appliqués*, 2nd ed. Presses inter Polytechnique, 2014. [Online]. Available: https://polymtl.on.worldcat.org/oclc/895812774

[28] J. Delmas and B. Jourdain, *Modèles aléatoires: Applications aux sciences de l'ingénieur et du vivant*, ser. Mathématiques et Applications. Springer Berlin Heidelberg, 2006. [Online]. Available: https://books.google.fr/books?id=hYLctAVONtsC

[29] R. Streit, *Poisson Point Processes: Imaging, Tracking, and Sensing*. Springer US, 2010. [Online]. Available: https://books.google.ca/books?id=KAWmFYUJ5zsC

[30] S. Meleard, *Modèles aléatoires en Ecologie et Evolution*, 01 2016, vol. 77.

[31] D. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*, ser. Probability and Its Applications. Springer, 2002. [Online]. Available: https://books.google.ca/books?id=LO2DQCS0TEQC

[32] ——, *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*, ser. Probability and Its Applications. Springer New York, 2007. [Online]. Available: https://books.google.ca/books?id=nPENXKw5kwcC

[33] J. G. Rasmussen, "Lecture notes: Temporal point processes and the conditional intensity function," 2018. [Online]. Available: https://arxiv.org/abs/1806.00221

[34] Y. Ogata, "On lewis' simulation method for point processes," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 23–31, January 1981.

[35] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008. [Online]. Available: https://www.R-project.org

[36] *Python Software Foundation*, 2020. [Online]. Available: https://www.python.org/

[37] Y. Caumel, *Probabilités et processus stochastiques*, ser. Statistique et probabilités appliquées. Springer Paris, 2011. [Online]. Available: https://books.google.ca/books?id=fvMs8DQsT90C

[38] J. Cook, "Inverse gamma distribution," 2008. [Online]. Available: https://www.johndcook.com/inverse_gamma.pdf

[39] J.-F. Delmas, *Introduction au calcul des probabilites et à la statistique.* les Presses de l'ENSTA, 2013. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00964982

[40] W. Hines *et al.*, *Probabilités et statistique pour ingénieurs.* Chenelière Education, 2011. [Online]. Available: https://books.google.ca/books?id=3F9MygAACAAJ

[41] R. Moché, "Estimation de la densité par les histogrammes et les estimateurs á noyau," 2016. [Online]. Available: http://gradus-ad-mathematicam.fr/documents/401_EstDensiteJAleHasard.pdf

[42] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*, ser. Springer Series in Statistics. Springer New York, 2009. [Online]. Available: https://books.google.ca/books?id=tVIjmNS3Ob8C

[43] Y. Ogata, "Statistical models for earthquake occurrences and residual analysis for point processes," *Journal of the American Statistical Association*, vol. 83, no. 401, pp. 9–27, 1988. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01621459.1988.10478560

[44] H. Thode, *Testing For Normality*, ser. Statistics, textbooks and monographs. CRC Press, 2002. [Online]. Available: https://books.google.ca/books?id=gbegXB4SdosC

[45] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[46] C. Giraud, *Introduction to High-Dimensional Statistics*, ser. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 2014. [Online]. Available: https://books.google.ca/books?id=qRuVoAEACAAJ

[47] K. P. Burnham and D. R. Anderson, "Multimodel inference: Understanding aic and bic in model selection," *Sociological Methods & Research*, vol. 33, no. 2, pp. 261–304, 2004. [Online]. Available: http://www.sortie-nd.org/lme/Statistical%20Papers/Burnham_and_Anderson_2004_Multimodel_Inference.pdf

[48] Canadian Institute for Health Information. (2020) CIHI's annual report 2019-2020. [Online]. Available: https://www.cihi.ca/sites/default/files/document/cihi-annual-report-2019-2020-en.pdf

[49] I. Kortchemski, "Notes de cours MAP 311: Rappels: Vecteurs de variables aléatoires," 2015. [Online]. Available: http://igor-kortchemski.perso.math.cnrs.fr/enseignement/2015-2016/PC3_rappels_vecteurs.pdf

[50] S. A. Zenios, G. M. Chertow, and L. M. Wein, "Dynamic allocation of kidneys to candidates on the transplant waiting list," *Operations Research*, vol. 48, no. 4, pp. 549–569, 2000. [Online]. Available: http://www.jstor.org/stable/222875

[51] D. Bertsimas, V. F. Farias, and N. Trichakis, "Fairness, efficiency, and flexibility in organ allocation for kidney transplantation," *Operations Research*, vol. 61, no. 1, pp. 73–87, 2013. [Online]. Available: https://doi.org/10.1287/opre.1120.1138

[52] C. Harvey and J. R. Thompson, "Exploring advantages in the waiting list for organ donations," in *2016 Winter Simulation Conference (WSC)*, 2016, pp. 2006–2017.

[53] A. Krasnosielska-Kobos, "Multiple-stopping problems with random horizon," *Optimization*, vol. 64, no. 7, pp. 1625–1645, 2015. [Online]. Available: https://doi.org/10.1080/02331934.2013.869808

[54] C. Bandi, N. Trichakis, and P. Vayanos, "Robust multiclass queuing theory for wait time estimation in resource allocation systems," *Management Science*, vol. 65, no. 1, pp. 152–187, 2019. [Online]. Available: https://doi.org/10.1287/mnsc.2017.2948

[55] D. Snyder and M. Miller, *Random Point Processes in Time and Space*, ser. Springer Texts in Electrical Engineering. Springer New York, 2012. [Online]. Available: https://books.google.ca/books?id=c_3UBwAAQBAJ

[56] C. Bélisle, "Notes de cours: Processus de poisson."

[57] H. Pham, *Springer Handbook of Engineering Statistics*, ser. Springer Handbook of Engineering Statistics. Springer, 2006. [Online]. Available: https://books.google.com.fj/books?id=XDfP63NCGU0C

[58] T. J. DiCiccio and B. Efron, "Bootstrap confidence intervals," *Statist. Sci.*, vol. 11, no. 3, pp. 189–228, 09 1996. [Online]. Available: https://doi.org/10.1214/ss/1032280214

[59] G. James *et al.*, *An Introduction to Statistical Learning: with Applications in R*, ser. Springer Texts in Statistics. Springer New York, 2014. [Online]. Available: https://books.google.ca/books?id=at1bmAEACAAJ

APPENDIX A MPP MODEL PROPOSAL

We give the outline of a possible MPP model that could be developed in a future work. Before that, we want to illustrate the point by providing an example of a MPP model that includes history dependencies.

Example A.1 (Marked Hawkes process)**.** We present a relevant example of MPP from [33] to show the relation between the CIF and the nature of the process. The Epidemic Type Aftershock Sequence (ETAS) model is a particular type of MPP for modeling earthquakes times and magnitudes. In this example, the mark $\kappa_j \in \mathbb{R}_+^* =]0, +\infty[$ stands for the magnitude of the earthquake occuring at time t_j. The ETAS model can be defined by its specific GIF

$$\lambda_G^*(t) = \mu + \alpha \sum_{t_j<t} e^{\beta\kappa_i} e^{-\gamma(t-t_j)},$$

where $\alpha, \beta, \gamma > 0$ are the GIF parameters. When looking closer, we can see how the past history influences the current rate of the process, through the sum over the whole history events $t_j < t$. A past mark κ_j increases the rate via the exponential contribution $e^{\beta\kappa_i}$, which itself is weighted by the attenuation exponential factor $e^{-\gamma(t-t_j)}$ that decreases with the age of the earthquake $t - t_j$. Concerning the magnitude, the mark CoDF is an exponential distribution

$$f^*(\kappa|t) = \delta e^{-\delta\kappa},$$

where $\delta > 0$ is the only parameter. In this model, marks do not depend on the past, and are exponentially distributed so that the higher the magnitude is, the lower the probability of such an earthquake to happen is. The exponential distribution parameter δ is essential to adjust the magnitude of the earthquakes. Finally, we get the corresponding CIF, including both marks and times by mutliplying the GIF and the mark CoDF together

$$\lambda^*(t, \kappa) = \lambda_G^*(t) f^*(\kappa|t) = \left(\mu + \alpha \sum_{t_j<t} e^{\beta\kappa_j} e^{-\gamma(t-t_j)}\right) \delta e^{-\delta\kappa}.$$

□

Back to our problem, we first believe to observe in some cases that a recent offer made to one patient could influences in reality the happening of a quick next offer (The model developed in the thesis is assumed to be a renewal process, which means the happening of an event does not influences the process rate, which is constant all along). The rationale is that when a patient is offered a kidney, it could mean that the attribution system situation at this

particular time (the position of the patient in the waiting list, and the situation of other patients) enables him to be offered more kidney proposals. In the current model, the rate computed before and after such an event does not change.

Secondly, in the model developed in this thesis the real history of the patient is not taken into account, and neither is the real current activity. The reason (as explained in the history building process in section 4.1.1) is the recurrent missing and/or incomplete data for a lot of patients that would not allow to develop a method that require real histories. The solution adopted in the current method is to create a pseudo history for each patient since this is a viable solution that could be applied for every one of them.

We could use the MPP mathematical framework to develop a model that includes the two aspects we just pinpointed. We propose a model with a more advanced GIF composed of two terms, as follows

$$\lambda_G^* = \lambda_{pseudo} + \lambda_{real},$$

where

- ▷ The term λ_{pseudo} is computed on the pseudo history $\mathcal{H}$ we used so far in the thesis. Then, we can take the estimator of the rate we used so far, namely: $\lambda_{pseudo} = \widehat{\lambda}_2$.
- ▷ The term λ_{real} depends on the real recent activity of the patient in its real history denoted as $\dot{\mathcal{H}}$. To take account of the first point we mentioned, the rate λ_{real} would include the self-exciting property. It could be defined similarly as in the example A.1, with an exponentially weighted sum:

  $$\lambda_{real} = \alpha \sum_j e^{-\gamma(t-\dot{t}_j)},$$

  where $\dot{t_j}$ is the real time of the j^{th} event in real history $\dot{\mathcal{H}}$, and α and γ are model parameters to infer. The parameter α has to be adjusted to account for the relative importance between the two rates λ_{pseudo} and λ_{real}. The parameter γ is the discount factor that values more recent events higher than older ones.