Electronic Journal of Statistics
Vol. 18 (2024) 167–190
ISSN: 1935-7524
https://doi.org/10.1214/23-EJS2205

Random permutations generated by delay models and estimation of delay distributions

Ludwig Baringhaus and Rudolf Grübel

Institute of Actuarial and Financial Mathematics
Leibniz Universität Hannover
Welfengarten 1
D-30167 Hannover
Germany
e-mail: lbaring@stochastik.uni-hannover.de; rgrubel@stochastik.uni-hannover.de

Abstract: Objects arrive at a system at times $U_1, U_2, \ldots$ and leave at times $U_1 + X_1, U_2 + X_2, \ldots$, where we assume that the arrivals are independent and uniformly distributed on the unit interval, that the delay times are independent with distribution function G, and that arrival and delay times are independent. Let Π_n be the random permutation that connects the ranks of the first n arrivals and departures. We investigate the use of Π_n for estimating G. We consider empirical copulas in the nonparametric and pattern frequencies in the parametric situation.

MSC2020 subject classifications: Primary 62G20; secondary 05A05.
Keywords and phrases: Service time distribution, random permutation, rank plot, delay copula, nonparametric estimation, monotone minorant estimator, parametric estimation, pattern.

Received March 2023.

1. Introduction

The following general *delay model* appears in several applied situations: Customers or particles etc. enter a system at times U_i and depart at times $Y_i := U_i + X_i$, $i \in \mathbb{N}$, where the random variables $U_1, U_2, \ldots, X_1, X_2, \ldots$ are independent; further, for all $i \in \mathbb{N}$, U_i is uniformly distributed on the unit interval, and $X_i \geq 0$ has distribution function G, so that $G(0-) = 0$. Examples include queuing models, notably the $M/G/\infty$ queue, but also models from statistical physics and mathematical biology; see the discussions in [3] and [5]. In particular, the uniform distribution of arrival times appears if the system is fed by a Poisson process with constant intensity.

We are mainly interested in the delay distribution. If the arrival and departure time for each customer are known then so are the individual service times, and the problem is simply another instance of the classical situation, with a sample from an unknown distribution. A typical variant involving data loss appears if we only know the set of arrival and departure times, or the number of customers in the system as a function of time over a fixed time interval, or the duration

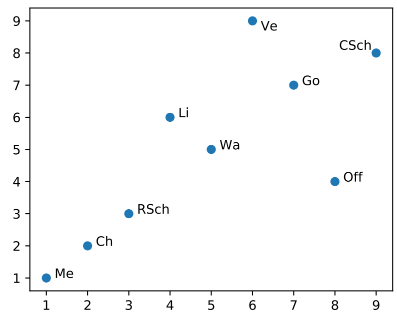

Fig 1. *The rank plot for the lifetime data.*

of busy periods, which further reduces the number of customers to a binary variable; see e.g. [6, 22, 4, 15, 14, 5, 10]. If only the sets of arrival and departure times are known then the bijection between arrival and departure indices may be regarded as the missing part of the complete data, i.e. the individual service times. Maximum a posteriori and minimum distance prediction of this bijection are discussed in [14]. An interesting approach is given in [6], where on the basis of a segment of a path of the $M/G/\infty$ queue departure times are matched to the last previous arrival times, and a functional relationship between the distribution of the differences and the service time is used to obtain a consistent estimator of the latter; see also the end of Sect. 3.

In the present paper we consider the complementary problem where the bijection is known but the arrival and departure times are not. For example, arriving customers might receive consecutively numbered tickets which they return on departure, and the tickets are put on a stack. Of course, the point is that the order of arrivals will generally not be the same as the order of departures.

In [14] the dates of birth and death for some famous composers were used as a running example. In the same vein, Chopin (Ch), Gounod (Go), Liszt (Li), Mendelssohn (Me), Offenbach (Off), Clara Schumann (CSch), Robert Schumann (RSch), Verdi (Ve) and Wagner (Wa), all born between 1809 and 1819, give rise to the rank plot in Fig. 1. For example, Offenbach was born later than Verdi but died earlier.

Our main concern in this paper is the following question: What information about the delay distribution can be extracted from these random permutations? As usual, the answers that we obtain in a statistical context, such as estimation of the delay distribution, refer to asymptotics, where the sample size n tends to infinity. A methodological aspect worth mentioning is the different view of permutations that we use in the nonparametric and the parametric case respectively. In the nonparametric situation we regard the data as representing an empirical copula and we obtain consistency of a specific estimator. In the parametric situation we base the asymptotic analysis on a view proposed in [16] where limits of permutations refer to pattern frequencies and lead to *permutons* (two-dimensional copulas) as limit objects. Copulas are often used in dependence modeling; see [20] for an introduction. Copulas predate permutons, but were in turn used avant la lettre in nonparametric statistics, for example in [23] as *rank order statistics*, or in [8] as *dependence functions*; see also [13] for a recent review.

In Sect. 2 we introduce some basic notation, we relate rank plots such as in Fig. 1 to empirical variants of a specific set of copulas, and we discuss model identifiability (Theorem 4). In Sect. 3 we propose and analyze a nonparametric estimator for the delay distribution (Theorem 6) and generally remark on the difficulty of this problem. In Sect. 4 we apply the pattern view in the context of parametric estimation. Here we can go beyond consistency and obtain asymptotic normality of the estimators. We work out the case of exponential delay distributions (Theorem 10) and determine the asymptotic efficiency of the delay model estimator as a function of the parameter $\lambda = 1/EX_1$.

There is a rough analogy between using pattern frequencies in the present context and the classical moment method, where sample moments are equated to moments arising in the parametric model in order to arrive at parameter estimates. For finite-dimensional parameters it is enough to consider a finite set of patterns. In a separate paper [2] we deal with nonparametric goodness-of-fit and two-sample tests for general permutation data. There, a 'functional view' turns out to be useful: All patterns are considered simultaneously, which leads to infinite-dimensional spaces and functional central limit theorems.

2. Delay copulas

We first formalize rank plots and relate them to permutations. Let $\mathbb{S}_n$ be the set of permutations of $[n] := \{1, \ldots, n\}$. An element π of $\mathbb{S}_n$ may be described by the list $(\pi(1), \pi(2), \ldots, \pi(n))$ of its values. For $n < 10$ we often use a condensed form, such as 21 instead of $(2, 1)$.

In our model arrival and departure distributions are continuous. Hence, ignoring a set of probability 0 we may assume that $U_1, U_2, \ldots$ and $Y_1, Y_2, \ldots$ are pairwise different. Given the first n variables $Z_i = (U_i, Y_i)$, $i = 1, \ldots, n$, let $Q_{ni} = \sum_{j=1}^n 1(U_j \le U_i)$ and $R_{ni} = \sum_{j=1}^n 1(Y_j \le Y_i)$ be the ranks of the respective arrival and departure times in the sample. The rank statistics $Q_n = (Q_{n1}, \ldots, Q_{nn})$ and $R_n = (R_{n1}, \ldots, R_{nn})$ then give rise to a unique random permutation Π_n with the property that

$$\Pi_n(Q_{ni}) = R_{ni} \quad \text{for all } i = 1, \ldots, n. \tag{1}$$

Thus the permutation plot of Π_n is same as the rank plot for the first n pairs of arrival and departure times. Note that Π_n is invariant under permutations of $Z_1, \ldots, Z_n$.

We regard the values of these permutations as our data. Later, asymptotic considerations with $n \to \infty$ require a notion of convergence for the permutations Π_n, together with a description of the possible limits. We recall that a (two-dimensional) copula is a distribution function $C : [0, 1] \times [0, 1] \to [0, 1]$ with uniform marginals, i.e. $C(u, 1) = C(1, u) = u$ for all u, $0 \le u \le 1$. Generally, if X

and Y are real random variables with continuous distribution functions F_X, F_Y defined on the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ and joint distribution function $F_{X,Y} : \overline{\mathbb{R}} \times \overline{\mathbb{R}} \to \mathbb{R}$, then the transformed random variables $U = F_X(X)$ and $V = F_Y(Y)$ are uniformly distributed on the unit interval, and the associated copula is given by

$$C_{X,Y}(u,v) = P(U \le u, V \le v) = F_{X,Y}\big(F_X^{-1}(u), F_Y^{-1}(v)\big), \quad 0 \le u, v \le 1, \quad (2)$$

where $F^{-1} : [0,1] \to \overline{\mathbb{R}}$, with $F^{-1}(w) = \inf\{z \in \overline{\mathbb{R}} : F(z) \ge w\}$ for $0 \le w \le 1$, denotes the quantile function associated with a distribution function F. As F_X, F_Y are continuous, $C_{X,Y}$ is the unique copula C that satisfies

$$F_{X,Y}(x,y) = C\big(F_X(x), F_Y(y)\big) \quad \text{for all } (x,y) \in \overline{\mathbb{R}} \times \overline{\mathbb{R}}.$$

For our purposes, a specific family of copulas will be of central importance.

Definition 1. Let G be a distribution function with $G(0-) = 0$. Then the associated *delay copula* $C = C[G]$ is the copula given by the joint distribution μ of U and $Y := U + X$, where U and X are independent, U is uniformly distributed on the unit interval, and G is the distribution function of X.

Example 2. (a) Given G and $a > 0$ let G_a be defined by $G_a(x) = G(x - a)$, $x \in \mathbb{R}$. If X has distribution function G then G_a is the distribution function of $X + a$. Writing C and C_a for the respective delay copulas and using $F_{a+Y}(z) = F_Y(z - a)$, $F_{a+Y}^{-1}(v) = F_Y^{-1}(v) + a$, we get

$$\begin{aligned} C_a(u,v) &= F_{U,a+Y}\big(u, F_{Y+a}^{-1}(v)\big) = F_{U,a+Y}\big(u, F_Y^{-1}(v) + a\big) \\ &= F_{U,Y}\big(u, F_Y^{-1}(v)\big) = C(u,v), \end{aligned}$$

hence a positive shift of the delay distribution does not change the delay copula.

(b) Given G and $a > 0$ let G_a be defined by $G_a(x) = G(x/a)$, $x \in \mathbb{R}$. If X has distribution function G then G_a is the distribution function of aX. As in (a) let C and C_a be the delay copulas for G and G_a. We claim that C_a converges weakly if $a \to 0$ or $a \to \infty$ to the copulas C_0 and C_∞ given by $C_0(u,v) = u \wedge v$, $C_\infty(u,v) = uv$ for all $u, v \in (0,1)$ respectively, which are associated to the extreme cases $U = Y$ and U, Y independent. In particular, rescaling of the service time distribution changes the associated delay copula. Indeed, for $a \to 0$ continuity implies that $C_{U,U+aX} \to C_{U,U}$, and for $a \to \infty$ we use the invariance of copulas under continuous and strictly increasing transformations to obtain $C_{U,U+aX} = C_{U,a^{-1}U+X}$, and continuity now leads to the limit $C_{U,X}$ as $a \to \infty$.

(c) Next we consider delay distributions concentrated at two points. Specifically, let $G = G_{\alpha,c}$ be the distribution function for the mixture $\alpha\delta_0 + (1-\alpha)\delta_c$ of the Dirac distributions at the points 0 and c respectively, with parameters $\alpha \in (0,1)$ and $c > 0$. This models the situation where incoming particles or customers either leave the system immediately or stay for a fixed time c, with respective probabilities α and $1 - \alpha$. If $c \ge 1$ and $X \sim G_{\alpha,c}$ then $F_{U,U+X} = \alpha\, F_{U,U} + (1-\alpha)\, F_{U,U+c}$. With $C_{0,\alpha}$ and $C_{1,\alpha}$ denoting the restrictions of the distribution functions $F_{U,\alpha U}$ and $F_{U,\alpha+(1-\alpha)U}$ to the unit square, it is easily verified that $C_\alpha = \alpha\, C_{0,\alpha} + (1-\alpha)\, C_{1,\alpha}$ is the associated delay copula; remarkably, it does not depend on the shift $c \ge 1$. Let Π_n be the random permutation generated by the first n arrivals. Assuming that there are k customers that leave immediately, and ℓ customers that require service time $c \ge 1$, the permutation Π_n is obtained as follows. With $U_1 < \cdots < U_n$ denoting the ordered arrival times, there are disjoint subsets $\{r_1, \ldots, r_k\}$ and $\{s_1, \ldots, s_\ell\}$ of $\{1, \ldots, n\}$, with $1 \le r_1 < \cdots < r_k \le n$, $1 \le s_1 < \cdots < s_\ell \le n$, and $k + \ell = n$ such that $Y_{r_m} = U_{r_m}$, $m = 1, \ldots, k$, are the departure times of the customers that leave immediately, and $Y_{s_m} = U_{s_m} + c$, $m = 1, \ldots, \ell$, are the departure times of the customers that require service time $c \ge 1$. Then $\Pi_n(r_m) = m$ for $m = 1, \ldots, k$ and $\Pi_n(s_m) = m + k$ for $m = 1, \ldots, \ell$. This also shows that the length of the delay disappears in the step from the time data to the permutation, as all $c \ge 1$ lead to the same sequence.

(d) In view of its later importance we consider the case of exponential distributions $\mathrm{Exp}(\lambda)$, where we have $G(x) = 1 - \exp(-\lambda x)$, $x \ge 0$, for some $\lambda > 0$. The joint distribution function $F_{U,Y}$ of U and $Y = U + X$ with U uniformly distributed on the unit interval, $X \sim \mathrm{Exp}(\lambda)$, and U and X independent, is readily calculated as

$$F_{U,Y}(u,y) = \begin{cases} y - \bigl(1 - e^{-\lambda y}\bigr)/\lambda, & \text{if } y < u, \\ u - \bigl(e^{-\lambda(y-u)} - e^{-\lambda y}\bigr)/\lambda, & \text{if } y \ge u, \end{cases} \tag{3}$$

and F_Y can be obtained from this using $F_Y(y) = F_{U,Y}(1,y)$. As $F_Y(y) = y - (1 - e^{-\lambda y})/\lambda$ for $y < 1$ we have that $F_Y^{-1}(v)$, with $v \in (0, 1 - (1 - e^{-\lambda})/\lambda)$, is the unique solution $y \in (0,1)$ of the equation $y - (1 - e^{-\lambda y})/\lambda = v$. This solution can suitably be expressed with the help of the Lambert W-function $t \mapsto W(t)$, with $W(t)$ for $t \in [-e^{-1}, \infty)$ as the unique solution $W(t) = y \in [-1, \infty)$ of the equation $ye^y = t$. On $[-e^{-1}, e^{-1}]$ the function W has the absolutely convergent series expansion

$$W(t) = \sum_{k=1}^{\infty} \frac{(-k)^{k-1}}{k!}\, t^k, \quad |t| \le e^{-1};$$

see, e.g. [7]. We get by an easy calculation that

$$F_Y^{-1}(v) = v + \frac{1}{\lambda} + \frac{1}{\lambda}\, W\bigl(-e^{-(\lambda v + 1)}\bigr) = \frac{1}{\lambda}\left(\lambda v + 1 - \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!}\, e^{-k(\lambda v + 1)}\right)$$

for $v \in (0, 1 - (1 - e^{-\lambda})/\lambda)$. This representation can also be obtained directly by using Lagranges's series expansion; see [27, p. 133].

Obviously,

$$F_Y^{-1}(v) = \frac{1}{\lambda} \log \frac{e^\lambda - 1}{\lambda(1-v)}, \quad \text{for } v \in [1 - \bigl(1 - e^{-\lambda}\bigr)/\lambda, 1).$$

Putting for abbreviation $e_\lambda(u) := u - (1 - e^{-\lambda u})/\lambda$, $u \in [0,1]$, we get using (3) that $C[G](u,v) = F_{U,Y}(u, F_Y^{-1}(v))$ can be written as

$$C[G](u,v) = \begin{cases} v, & v < e_\lambda(u), \\ u + W\bigl(-e^{-(\lambda v+1)}\bigr)\bigl(e^{\lambda u}-1\bigr)/\lambda, & e_\lambda(u) \le v < e_\lambda(1), \\ u - \frac{1-v}{e^\lambda - 1}\bigl(e^{\lambda u}-1\bigr), & v \ge e_\lambda(1). \end{cases}$$

Note that part (b) above applies in the present situation. ◁

A central question concerns the amount of information about G contained in the associated delay copula $C[G]$. In particular, is G specified by $C[G]$? In the literature this is known as the *identifiability* of the model. In the above example we have seen that this does not hold in general, and that we have to assume, at least, that G satisfies the conditions $G(x) > 0$ and $G(x+1) > G(x)$ for all $x > 0$. In words, $x = 0$ is a point of (right) increase of G, and there is a point of increase in every interval of length 1. As we will see below, a slight amplification leads to the sufficient condition

$$G(0) = 0, \quad G(n+\epsilon) > G(n) \text{ for all } \epsilon > 0 \text{ and } n \in \mathbb{N}_0 \text{ with } G(n) < 1. \tag{4}$$

This is obviously satisfied if G has a positive density on $(0,\infty)$, for example.

As a preparation for the proof we note that the distribution function G of a random variable X with $G(0) = 0$ is determined by the function

$$\tilde{G}(x) = \int_0^1 G(x-u)\,du = \begin{cases} \int_0^x G(u)\,du, & \text{if } x \le 1, \\ \int_{x-1}^x G(u)\,du, & \text{if } x > 1, \end{cases} \tag{5}$$

which is the distribution function F_Y of $Y = U + X$ if U is uniformly distributed on the unit interval and independent of X. In fact, even more is true.

Lemma 3. *Let G and H be distribution functions with $G(0) = H(0) = 0$. Then, for all $x > 0$, $\tilde{G}(y) = \tilde{H}(y)$ for $0 \le y \le x$ implies that $G(y) = H(y)$ for $0 \le y \le x$.*

Proof. It follows from (5) that $\tilde{G}$ has the right continuous density

$$G(x)\,1_{[0,1]}(x) + \bigl(G(x) - G(x-1)\bigr)1_{(1,\infty)}(x), \quad x \in \mathbb{R}. \tag{6}$$

If it exists, the right continuous density of a distribution is unique. □

For general copulas, i.e. with no restrictions on the structure of the constituting random vector (X,Y), the passage from the joint distribution to $C_{X,Y}$ as in (2) is of course not invertible; indeed, for any two strictly increasing $\Psi, \Phi : \mathbb{R} \to \mathbb{R}$, the random vector $(\Psi(X), \Phi(Y))$ would lead to the same copula. In contrast, for the smaller class of delay copulas we have the following invertibility result.

Theorem 4. *If G and H both satisfy (4) then $C[G] = C[H]$ implies $G = H$.*

Proof. In view of $F_{U,U+X}(u,y) = \int_0^u P(X \le y-w)\,dw$ the delay copula $C[G]$ may be written as

$$C[G](u,v) = \int_0^u G\big(\tilde{G}^{-1}(v) - w\big)\,dw, \quad u \in [0,1],\ v \in (0,1). \tag{7}$$

Here $\tilde{G}^{-1}(v) = \inf\{y : \tilde{G}(y) \ge v\}$ with $\tilde{G}$ as in (5). Due to (4) the density (6) of $\tilde{G}$ is positive on the interval $(0,x_0)$, where $x_0 := \sup\{x \in \mathbb{R} : \tilde{G}(x) < 1\}$. We deduce from this that the continuous distribution function $\tilde{G}$ is strictly increasing on the interval $(0,x_0)$, $\tilde{G}^{-1}$ is continuous and strictly increasing on $(0,1)$, and $\tilde{G}(\tilde{G}^{-1}(v)) = v$ for all $v \in (0,1)$.

Suppose now that $C[G] = C[H]$. Then taking the left derivative in (7) with respect to u leads to

$$G\big(\tilde{G}^{-1}(v) - u\big) = H\big(\tilde{H}^{-1}(v) - u\big) \quad \text{for all } u \in (0,1],\ v \in (0,1). \tag{8}$$

In particular, whenever $v \in (0,1)$ is such that $y := \tilde{G}^{-1}(v) \ne \tilde{H}^{-1}(v) =: z$ then $G(y-u) = H(z-u)$ for all $u \in (0,1]$.

If $\tilde{G} = \tilde{H}$ on $[0,1]$, then $G = H$ on $[0,1]$ by Lemma 3, so for G and H not to be the same on the unit interval we would need some $y \in (0,1]$ with $v := \tilde{G}(y) \ne \tilde{H}(y) =: w$; due to (4) we would have $v, w \in (0,1)$. If $v > w$,

$$y = \tilde{G}^{-1}(v) = \tilde{H}^{-1}(w) < \tilde{H}^{-1}(v) =: z.$$

With $u = y$ in (8) this would lead to $0 = G(0) = H(z-y)$, in contradiction to $H(z-y) > H(0) = 0$. Similarly, if $w > v$, then G would not be strictly increasing in 0.

Suppose now that G and H are identical on the interval $[0,n]$ for some $n \in \mathbb{N}$. If $G(n) = H(n) = 1$ then clearly $G = H$. Otherwise, for G and H to differ on $[0,n+1]$ we would need a $y \in (n,n+1]$ with $v := \tilde{G}(y) \ne \tilde{H}(y) =: w$; again we would have $v, w \in (0,1)$. If $v > w$ then, with $z := \tilde{H}^{-1}(v)$ would lead to the contradiction $H(n) < H(z-(y-n)) = G(n) = H(n)$. Similarly, if $v < w$ then, with $z := \tilde{G}^{-1}(w)$ we would get the contradiction $G(n) < G(z-(y-n)) = H(n) = G(n)$. □

In the delay models, U_i and $Y_i = U_i + X_i$ are the arrival and departure times respectively of the ith customer, $i \in \mathbb{N}$. Let $V_i := F_Y(Y_i)$. Then (U_i, V_i), $i \in [n]$ is a sample from the joint distribution μ of U_1 and V_1. Somewhat analogous to the step from μ to the empirical distribution

$$\mu_n := \frac{1}{n}\sum_{k=1}^{n} \delta_{(U_i,V_i)} \tag{9}$$

we pass from the copula C to the *empirical copula* C_n, which we define as the distribution function associated with

$$\nu_n := \frac{1}{n}\sum_{k=1}^{n} \delta_{(Q_{ni}/n,R_{ni}/n)} = \frac{1}{n}\sum_{k=1}^{n} \delta_{(i/n,\Pi_n(i)/n)}. \tag{10}$$

Apart from a scaling factor $1/n$ this is the discrete uniform distribution on the points which constitute the graph of the rank plot or, equivalently, the graph of Π_n. Obviously the marginals of ν_n are uniformly distributed on the set of values i/n, $i = 1, \ldots, n$, which justifies the interpretation of C_n as a discrete copula.

Asymptotics for C_n can be translated into asymptotics for Π_n. The former have been studied by various authors, see e.g. [9]. For example, as $n \to \infty$, $C_n \to C$ almost surely (a.s.) with respect to weak convergence of probability measures on the unit square, and this implies that the delay copula is asymptotically determined, with probability 1, by the sequence $(\Pi_n)_{n\in\mathbb{N}}$ of random permutations.

3. Nonparametric estimation of delay distributions

We consider the delay model with delay distribution function G, where we assume that $G(0-) = 0$ and that $G(\epsilon) > 0$ for all $\epsilon > 0$. As $F_Y(y) = \int_0^y G(w)\,dw$ for all $y \in [0,1]$, it then follows that the distribution function F_Y of the departure times is continuous and strictly increasing on the interval $[0,1]$, and that its quantile function F_Y^{-1} is continuous and strictly increasing on the interval $(0, F_Y(1)]$. Further, see also (7), the delay copula $C = C[G]$ is given by

$$C(u,v) = \int_0^u G\big(F_Y^{-1}(v) - w\big)dw \quad \text{for } (u,v) \in [0,1]^2. \tag{11}$$

As in Sect. 2, let μ be the distribution of (U,V), where $V := F_Y(Y) = F_Y(U + X)$, so that C is the distribution function of μ (we freely switch between viewing μ as a probability measure on the Borel subsets of $\mathbb{R}^2$ or $[0,1]^2$). As $F_Y(Y) = F_Y(U + X) \geq F_Y(U)$ it follows that the support H of μ is a subset of

$$H_+ := \big\{(u,v) \in [0,1]^2 : v \geq F_Y(u)\big\}.$$

Define $H_=$ and H_- similarly, with $v = F_Y(u)$ and $v < F_Y(u)$ respectively. Then each point of the graph $H_=$ of F_Y on the unit interval $[0,1]$ is an element of H. To see this we first consider $u < 1$; let $v = F_Y(u)$. Then for each $\epsilon > 0$ with $v + \epsilon < F_Y(1)$ there is a unique $0 < \delta_\epsilon < 1 - u$ such that $F_Y(u + \delta_\epsilon) = F_Y(u) + \epsilon$. With

$$\begin{aligned} D_{u,\epsilon} &:= \big([u, u+\delta_\epsilon] \times [v, v+\epsilon]\big) \cap H_+ \\ &= \big\{\big(u', v'\big) \in [0,1]^2 : u \leq u' \leq u + \delta_\epsilon, F_Y\big(u'\big) \leq v' \leq F_Y(u) + \epsilon\big\} \end{aligned} \tag{12}$$

and $\mu(H_+^c) = 0$ we obtain

$$\mu(D_{u,\epsilon}) = \mu\big([u, u+\delta_\epsilon] \times [v, v+\epsilon]\big) = \int_u^{u+\delta_\epsilon} G(u + \delta_\epsilon - w)\,dw > 0. \tag{13}$$

This shows that the pair (u,v) is an element of the closed set H, and a similar argument works for $u = 1$. Hence $H_- \subset H^c$ and $H_= \subset H$.

Taken together this implies that the subdistribution function $u \mapsto F_Y(u)$, $0 \le u \le 1$, is the supremum of all increasing functions that lie below the support of μ. Replacing the delay copula by its empirical counterpart we obtain the *monotone minorant estimator* $\hat{F}_{n,Y}$ as the supremum of all increasing functions that lie below the support of ν_n as defined in (10). This leads to

$$\hat{F}_{n,Y}(u) := \frac{1}{n} \min\bigl\{\Pi_n(i) : \lceil un \rceil \le i \le n\bigr\}, \quad u \in (0,1],$$

which we augment by $\hat{F}_{n,Y}(u) = 0$ for all $u \le 0$. The monotone minorant estimator is a nonnegative and left continuous step function with a jump at 0 and jumps at a subset of the first coordinates of the support points of the empirical copula. As such it can be regarded as the (left continuous) distribution function associated with a subprobability measure on the unit interval. In our running example, see Fig. 1, apart from the jump at 0 there are jumps at the first coordinates associated with Me, Ch, RSch, and Off, and the subprobability measure assigns the values $1/9, 1/9, 1/9, 1/9, 4/9$ to the atoms at $0, 1/9, 2/9, 3/9, 8/9$.

In the proof of our next theorem it will be helpful to regard the unit square as a compact metric space, with metric $d_1(a,b) = |b_1 - a_1| + |b_2 - a_2|$, $a = (a_1, a_2), b = (b_1, b_2) \in [0,1]^2$. For $x \in [0,1]^2$, $\emptyset \ne K \subset [0,1]^2$, $\rho > 0$, let

$$d_1(x,K) := \inf\bigl\{d_1(x,c) : \; c \in K\bigr\} \quad \text{and} \quad K_\rho := \bigl\{x \in [0,1]^2 : \; d_1(x,K) \le \rho\bigr\}.$$

We use the corresponding Hausdorff distance d_0 of nonempty subsets A, B of the unit square,

$$d_0(A,B) := \max\Bigl\{\sup\bigl\{d_1(a,B) : a \in A\bigr\}, \sup\bigl\{d_1(b,A) : b \in B\bigr\}\Bigr\}, \tag{14}$$

which may also be written as

$$d_0(A,B) := \inf\{\rho > 0 : \; A \subset B_\rho \text{ and } B \subset A_\rho\}. \tag{15}$$

The following theorem implies that $\hat{F}_{n,Y}$ is a strongly consistent estimator for F_Y with respect to the supremum norm distance on each interval $[0, 1-\delta]$, $0 < \delta < 1$.

Theorem 5. *With $\hat{F}_{n,Y}$ and F_Y as above it holds that*

$$\limsup_{n\to\infty} \sup_{0 \le u \le 1} \bigl(F_Y(u) - \hat{F}_{n,Y}(u)\bigr) \le 0 \quad a.s., \tag{16}$$

and that, for all $0 < \delta < 1$,

$$\limsup_{n\to\infty} \sup_{0 \le u \le 1-\delta} \bigl(\hat{F}_{n,Y}(u) - F_Y(u)\bigr) \le 0 \quad a.s. \tag{17}$$

Proof. As above, let U_i, $Y_i = U_i + X_i$, $V_i = F_Y(Y_i)$ for all $i \in \mathbb{N}$. Then μ_n and ν_n are the discrete uniform distributions on the random sets

$$A_n := \bigl\{(U_i, V_i) : \; i \in [n]\bigr\} \tag{18}$$

and

$$B_n := \{(Q_{ni}/n, R_{ni}/i) : i \in [n]\} = \{(i/n, \Pi_n(i)/n) : i \in [n]\} \tag{19}$$

respectively. Our proof relies on a comparison between μ_n and ν_n and the known asymptotic behavior of the former.

By construction, μ has uniform marginals, and the Dvoretsky–Kiefer–Wolfowitz inequality, see [19], gives

$$P\Big(\sqrt{n} \sup_{0\le u\le 1} \big|\hat{F}_{n,U}(u) - u\big| \ge \lambda\Big) \le 2e^{-2\lambda^2} \quad \text{for all } n \in \mathbb{N},\ \lambda > 0. \tag{20}$$

Here $\hat{F}_{n,U}$ denotes the the empirical distribution function associated with the variables $U_1, \dots, U_n$. Of course, the analogue statement for the V-components also holds. By (15) the Hausdorff distance between the sets A_n and B_n defined in (18) and (19) is bounded from above by the maximum of the distances between the individual points in any pairing of their elements. Hence, using $\hat{F}_{n,U}(U_i) = Q_{ni}/n$ and $\hat{F}_{n,V}(V_i) = R_{ni}/n$, $1 \le i \le n$, we get

$$\begin{aligned} d_0(A_n, B_n) &\le \max\{d_1\big((Q_{ni}/n, R_{ni}/n), (U_i, V_i)\big) : i \in [n]\} \\ &\le \sup_{0\le u\le 1} \big|F_{n,U}(u) - u\big| + \sup_{0\le v\le 1} \big|F_{n,V}(v) - v\big|. \end{aligned}$$

Taking $\lambda = \lambda_n = n^{-1/4}$ in (20) we obtain a summable upper bound, and the Borel-Cantelli lemma leads to

$$\lim_{n\to\infty} d_0(A_n, B_n) = 0 \quad \text{a.s. as } n \to \infty. \tag{21}$$

For a given $\epsilon > 0$ we now choose $\rho > 0$ such that

$$H_\rho \subset \{(u, v) \in [0, 1]^2 : v \ge F_Y(u) - \epsilon\}.$$

We have $A_n \subset H$ for all $n \in \mathbb{N}$ on a probability 1 set E, and then $\mu_n(H) = 1$. Hence it follows from (21) that, on E and for n large enough, $\nu_n(H_\rho) = 1$, which implies that $\hat{F}_{n,Y}(u) \ge F_Y(u) - \epsilon$. Since $\epsilon > 0$ was arbitrary, this proves (16).

Now let $0 < \delta < 1$, $u \in [0, 1-\delta]$, and $0 < \epsilon < F_Y(1) - F_Y(u)$ be given. As $u \mapsto F_Y(u)$ is a convex function on the unit interval $[0, 1]$, the compact subset $D_{u,\epsilon}$ of the unit square $[0, 1]^2$ defined in (12) is also convex. Recall (13), according to which $\mu(D_{u,\epsilon}) > 0$. Obviously, there exists a compact and convex subset D° of $D_{u,\epsilon}$ with $\mu(D^\circ) > 0$ and $(D^\circ)_\rho \subset D_{u,\epsilon}$ for some $\rho > 0$. By construction, and noting that $\hat{F}_{n,Y}$ is increasing, $\hat{F}_{n,Y}(u) > F_Y(u) + \epsilon$ implies that $\nu_n(D_{u,\epsilon}) = 0$. From the above comparison we obtain that, on a set of probability 1, $\nu_n(D_{u,\epsilon}) = 0$ implies that $\mu_n(D^\circ) = 0$ for all sufficiently large n, contradicting

$$\lim_{n\to\infty} \mu_n\big(D^\circ\big) = \mu\big(D^\circ\big) > 0 \quad \text{a.s.}$$

Thus we have

$$\limsup_{n\to\infty} \hat{F}_{n,Y}(u) \le F_Y(u) + \epsilon \quad \text{a.s.}$$

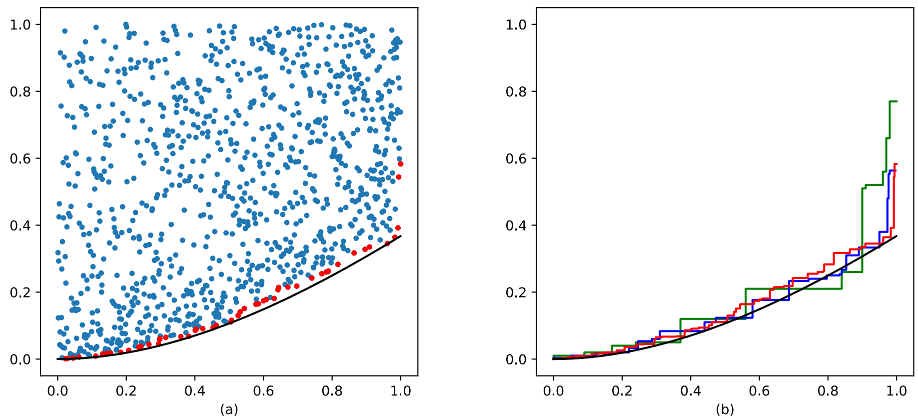

FIG 2. *Empirical copula and monotone minorant estimator (see text).*

As $\epsilon > 0$ was arbitrary, this gives (17), albeit at a single value u. Finally, for the step from the pointwise to the uniform statement we may proceed as in the proof of the Glivenko-Cantelli lemma, using monotonicity of F_Y and $\hat{F}_{n,Y}$ together with pointwise convergence in a finite set of suitably chosen quantiles of F_Y. □

The theorem implies that, for all $\delta \in (0,1)$,

$$\lim_{n\to\infty} \sup_{0\le u\le 1-\delta} \bigl|\hat{F}_{n,Y}(u) - F_Y(u)\bigr| = 0 \quad \text{a.s.} \tag{22}$$

Figure 2 illustrates this with an artificial data set of size $n = 1000$ and with $G(x) = 1-e^{-x}$, $x \ge 0$. Part (a) shows the empirical copula, with the jump points of the monotone minorant in red. Part (b) shows the corresponding monotone minorant estimators for the first 100 (green) and 300 (blue) data values and for the full data set (red). In both plots, the black line is the graph of the function $u \mapsto u+e^{-u}-1$, which is the restriction of F_Y to $0 \le u \le 1$; see Example 2 (d). It is apparent that the variability of the estimates increases as u increases towards the right endpoint of the interval.

We next deal with the step from F_Y to the delay distribution function G. The departure time distribution is the convolution of the delay distribution and the uniform distribution on the unit interval, which displays this step as a deconvolution problem. In particular, the restriction of F_Y to the unit interval is convex, and the restriction of G to the unit interval is the associated density. We thus obtain an estimator $\hat{G}_n$ for G on the interval $[0,1)$ as the right derivative of the *convex minorant* $\hat{F}^{\text{cm}}_{n,Y}$ of the points of the rescaled rank plot or, equivalently, the rescaled permutation plot for Π_n. The connection to the Grenander estimator [11] for a decreasing density, which is the derivative of the concave majorant of the empirical distribution function, appears if we note that $u \mapsto F_Y(1) - F_Y(1-u)$ has density $u \mapsto G(1-u)$. Here, however, we only have ν_n rather than μ_n.

Informally, the convex minorant arises by 'tightening a rubber band' below the respective colored step function in Fig. 2 (b). Formally, we obtain $\hat{F}^{\text{cm}}_{n,Y}$ and the associated estimator $\hat{G}_n$ of G, both on the unit interval, as follows. First, given $n \in \mathbb{N}$ and Π_n we put $k_0 = 0$ and

$$f_{i,j} := \frac{\Pi_n(j) - \Pi_n(i)}{j - i}, \quad 0 \le i < j \le n,$$

with $\Pi_n(0) := 0$. We then determine m and $1 \le k_1 < \cdots < k_m \le n$ inductively: If k_l is already defined for some $l \ge 0$ and $k_l = n$, then let $m = l$ and stop. Otherwise, if $k_l < n$, let

$$h_l := \min\{f_{k_l,j} : k_l < j \le n\} \quad \text{and} \quad k_{l+1} := \max\{j : k_l < j \le n,\ f_{k_l,j} = h_l\}.$$

Then, $\hat{F}^{\text{cm}}_{n,Y}(x) = \int_0^x \hat{G}_n(u)\,du$, where $\hat{G}_n$ is the increasing, right continuous step function given by

$$\hat{G}_n := \sum_{l=0}^{m-1} f_{k_l,k_{l+1}} 1_{[k_l/n, k_{l+1}/n)}. \tag{23}$$

Theorem 6. *Suppose that the delay distribution function G is continuous on the interval $[0,1)$, and let $\hat{G}_n$ be as defined in (23). Then, for each $0 < \delta < 1$,*

$$\lim_{n\to\infty} \sup_{0\le u\le 1-\delta} \bigl|\hat{G}_n(u) - G(u)\bigr| = 0 \quad \textit{a.s.} \tag{24}$$

Proof. We first relate the convex minorant to the monotone minorant, using an argument that is also known as Marshall's lemma [18] in order restricted statistical inference.

As the restrictions of F_Y and $F_Y - \epsilon$ to the interval $[0,1]$ are convex, it follows from Theorem 5 that, on a set of probability 1, for all $\epsilon > 0$ there exists an n_0 such that for all $n \ge n_0$,

$$\hat{F}_{n,Y}(u) \ge \hat{F}^{\text{cm}}_{n,Y}(u) \ge F_Y(u) - \epsilon \quad \text{for all } u \in [0, 1-\delta]$$

for all $\delta \in (0,1)$. As $\epsilon > 0$ was arbitrary, (22) now implies that, for all $\delta \in (0,1)$,

$$\lim_{n\to\infty} \sup_{0\le u\le 1-\delta} \bigl|\hat{F}^{\text{cm}}_{n,Y}(u) - F_Y(u)\bigr| = 0 \quad \text{a.s.} \tag{25}$$

Now let $0 < \delta < 1$ and $0 < h < \delta/2$ be given. Using

$$\begin{aligned}\int_u^{u+h} \bigl(G(t) - G(u)\bigr)\,dt &= \int_u^{u+h} \bigl(\hat{G}_n(t) - G(u)\bigr)\,dt \\ &\quad - \bigl(\hat{F}^{\text{cm}}_{n,Y}(u+h) - F_Y(u+h)\bigr) + \bigl(\hat{F}^{\text{cm}}_{n,Y}(u) - F_Y(u)\bigr) \\ &\ge h\bigl(\hat{G}_n(u) - G(u)\bigr) \\ &\quad - \bigl(\hat{F}^{\text{cm}}_{n,Y}(u+h) - F_Y(u+h)\bigr) + \bigl(\hat{F}^{\text{cm}}_{n,Y}(u) - F_Y(u)\bigr)\end{aligned}$$

for all $u \in [0, 1-\delta]$ we get

$$\begin{aligned} h \sup_{0\le u\le 1-\delta} &\bigl|G(u+h) - G(u)\bigr| \\ &\ge h \sup_{0\le u\le 1-\delta} \bigl(\hat{G}_n(u) - G(u)\bigr) - 2 \sup_{0\le u\le 1-\delta/2} \bigl|\hat{F}^{\,\mathrm{cm}}_{n,Y}(u) - F_Y(u)\bigr|. \end{aligned}$$

Thus, by (25),

$$\sup_{0\le u\le 1-\delta} \bigl|G(u+h) - G(u)\bigr| \ge \limsup_{n\to\infty} \sup_{0\le u\le 1-\delta} \bigl(\hat{G}_n(u) - G(u)\bigr)$$

for all $0 < h < \delta/2$. Let $h \downarrow 0$ to deduce from this and the uniform continuity of G on $[0, 1-\delta/2]$ that

$$0 \ge \limsup_{n\to\infty} \sup_{0\le u\le 1-\delta} \bigl(\hat{G}_n(u) - G(u)\bigr). \tag{26}$$

Similarly, using $G(u) = F_Y(u) = 0$ for $u < 0$, and defining $\hat{G}_n(u) := 0$, $F^{\,\mathrm{cm}}_{n,Y}(u) := 0$ for $u < 0$, we get

$$\begin{aligned} \int_{u-h}^{u} \bigl(G(u) - G(t)\bigr)\, dt = &\int_{u-h}^{u} \bigl(G(u) - \hat{G}_n(t)\bigr)\, dt \\ &+ \bigl(\hat{F}^{\,\mathrm{cm}}_{n,Y}(u) - F_Y(u)\bigr) - \bigl(\hat{F}^{\,\mathrm{cm}}_{n,Y}(u-h) - F_Y(u-h)\bigr) \\ \ge\; & h\bigl(G(u) - \hat{G}_n(u)\bigr) \\ &+ \bigl(\hat{F}^{\,\mathrm{cm}}_{n,Y}(u) - F_Y(u)\bigr) - \bigl(\hat{F}^{\,\mathrm{cm}}_{n,Y}(u-h) - F_Y(u-h)\bigr) \end{aligned}$$

for all $u \in [0, 1-\delta]$. Arguing as above we deduce from this that

$$0 \ge \limsup_{n\to\infty} \sup_{0\le u\le 1-\delta} \bigl(G(u) - \hat{G}_n(u)\bigr). \tag{27}$$

Combining (26) and (27) we obtain the assertion of the theorem. □

Figure 3 illustrates the estimator for delays that are uniformly distributed on the interval $[0, 1/2]$. The left part shows the passage from monotone minorant to convex minorant for two simulated data sets of size $n = 10$ (red) and $n = 100$ (blue) respectively. The right part shows G (black) and two simulation results for $n = 1000$ (red) and $n = 10000$ (blue). Our impression derived from several such experiments indicates that a rather large value of n is needed for $\hat{G}_n$ to become a useful estimator for G.

We next discuss the problem of extending the above estimator for G to the range beyond the unit interval. If we assume for simplicity that the delay distribution has a positive density then Theorem 4 implies that there is only one G for a given delay copula $C = C[G]$, and we know that C can be estimated consistently. Thus, the (informal) question is if this inverse problem is statistically well-posed.

From (11) it follows that, for any fixed $v \in (0,1)$, the function $u \mapsto C(u,v)$ has derivative $u \mapsto G(F_Y^{-1}(v) - u)$, which is decreasing and may be regarded

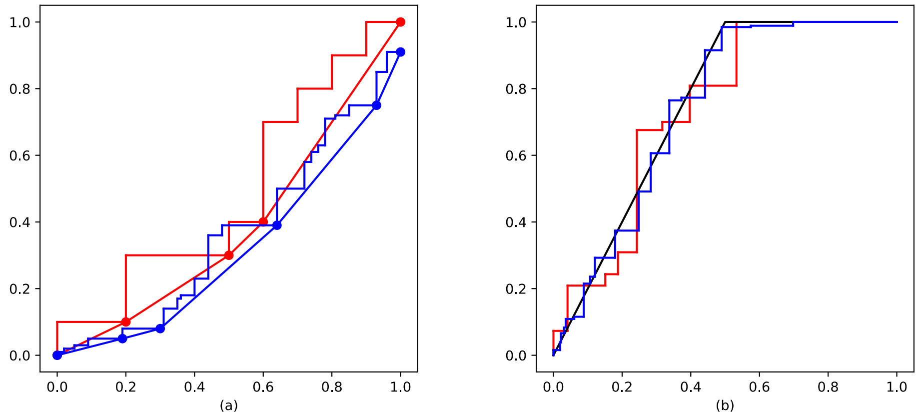

FIG 3. *Monotone and convex minorant, and estimates of G (see text).*

as a density of the subprobability with subdistribution function $u \mapsto C(u,v)$, $0 \le u \le 1$. As before, let C_n be the empirical copula associated with a sample of size n from the delay copula $C[G]$. As C_n converges to C in supremum norm, see e.g. [9, p. 51], it holds that, again a.s.,

$$\lim_{n\to\infty} \sup_{0\le u\le 1} \bigl|C_n(u,v) - C(u,v)\bigr| = 0.$$

The arguments used in the proof of Theorem 6 now lead to the following result.

Proposition 7. *Suppose that $v \in (0,1)$ is such that*

$$a(v) := F_Y^{-1}(v) - 1 > 0. \tag{28}$$

Let H_n be the right derivative of the concave majorant of the function $u \mapsto C_n(u,v)$ on the interval $[0,1)$. Then, for each $0 < \delta < 1$,

$$\lim_{n\to\infty} \sup_{0\le u\le 1-\delta} \bigl|H_n(u) - G\bigl(a(v)+1-u\bigr)\bigr| = 0 \quad a.s. \tag{29}$$

As it stands, this cannot be used to estimate the delay distribution function on the interval $(a(v), a(v)+1]$ by $\hat{G}_n(x) = H_n(a(v)+1-x)$, $x \in (a(v), a(v)+1]$, as $a(v)$ is in general not known. However, given a consistent estimate $\hat{G}_n$ of G on some interval $[0,c]$ and choosing v small enough for $a = a(v) < c$ and large enough for $a+1 > c$ it is possible to glue together the two estimates, leading to an extension from $[0,c]$ to $[0,a+1]$. Of course, we then need assumptions on G to exclude 'flat pieces'. For example, we may suppose that the delay distribution has support $[0,\infty)$ or $[0,a]$ for some $a > 0$, and that it has a strictly positive density on its support. The whole procedure could then, at least in principle, be iterated. Note the rough parallel to the stepwise argument in the proof of Theorem 4.

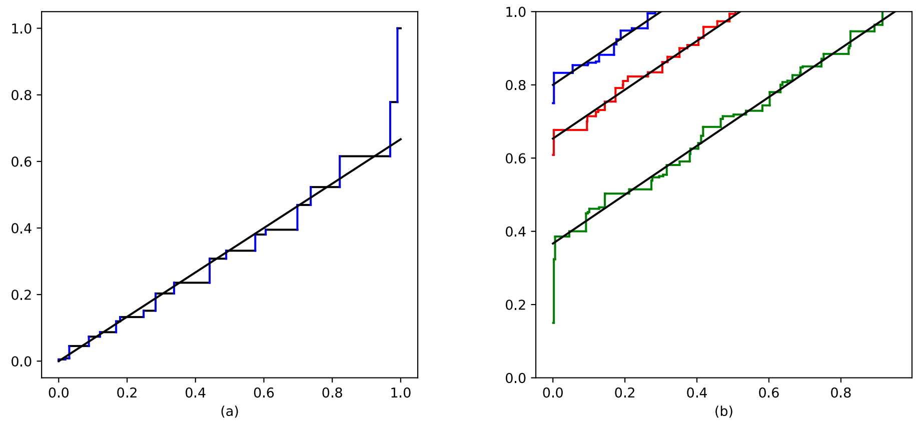

FIG 4. *Simulation results for* unif$(0, 3/2)$ *(see text).*

We illustrate this in Fig. 4 with another simulation example, with $n = 10000$ and unif$(0, 3/2)$ as delay distribution. Part (a) shows the estimator for G on the unit interval, with the restriction of G to the unit interval given in black. Part (b) displays the extensions obtained with $v = 0.97$ (blue), $v = 0.91$ (red), and $v = 0.7$ (green). These correspond to the respective 'true' intervals $(1.2, 2.2]$, $(0.98, 1.98]$, and $(0.55, 1.55]$, where the left endpoints of the intervals were obtained from F_Y, which is known in the simulation example. Again, in each case the black line is the respective piece of the graph of G, where $x = 0$ corresponds to the respective left endpoint 1.2, 0.98 and 0.55. It seems that a judicious choice of v, combined with a formal or informal matching of the two functions, may result in a practicable procedure.

In the classical case, with a sample from a distribution on $[0, \infty)$ with decreasing density, the Grenander estimator has been investigated in considerable detail. It is, for example, well known that the estimator is not consistent at $x = 0$; see e.g. [1] for a discussion. In the present setup this is responsible for the restriction $0 \le u \le 1 - \delta$ in (29); see Fig. 4 (a). Similarly, for the extensions beyond the unit interval, these difficulties appear at the left end; see Fig. 4 (b).

As pointed out in the introduction the present problem may be regarded as complementary to the problem where only the two sets of arrival times and departure times are known. Brown [6] gave an interesting estimator for this later case, together with a consistency result: From a path of the stationary $M/G/\infty$ queue successive departure times $0 < Y_1 < \cdots < Y_n$ and the associated last previous arrival times $X_1 \le X_2 \le \cdots \le X_n$ are extracted. The differences $Z_i = Y_i - X_i$, $1 \le i \le n$, while not independent, are identically distributed, and their distribution function H is related to the service time distribution function G as follows,

$$H(z) = 1 - e^{-\eta z}\bigl(1 - G(z)\bigr), \quad z \ge 0, \tag{30}$$

where η denotes the rate of the Poisson process of arrivals. Using (30) an es-

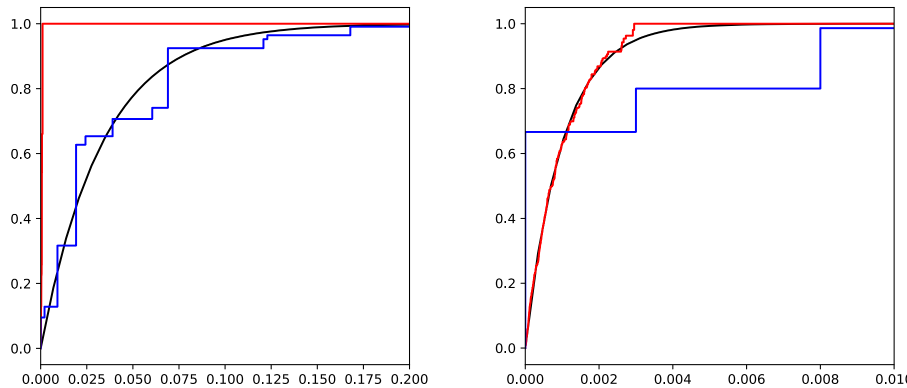

Fig 5. *Comparison of Brown's estimator (red) with the permutation based estimator (blue) and the true distribution function (black) (see text).*

timator $\hat{G}_n$ of G can be obtained that is based on the empirical distribution function $\hat{H}_n$ of the differences $Z_1, \ldots, Z_n$ and an estimator $\hat{\eta}_n$ of η,

$$\hat{G}_n(x) = \sup_{0 \le z \le x} \big(1 - e^{\hat{\eta}_n z}\big(1 - \hat{H}_n(z)\big)\big), \quad x \ge 0.$$

This and other estimators in the $M/G/\infty$ context are discussed and compared in [10] where it was noted that the quality of Brown's estimator depends on η. Specifically, for large arrival rates its performance can be poor, as is intuitively plausible from the fact that the support of the estimator is contained in the set of Z-values; see also [5]. While the delay model is not the same as the situation where observations are from a single path of increasing length a comparison can be made if we take X_1 to be 0 and restrict the estimator to the unit interval. Rescaling the length of the observation interval in the $M/G/\infty$ situation amounts to a rescaling of the service times, and the limit $n \to \infty$ in the delay model then corresponds to the heavy traffic limit $\eta \to \infty$. Figure 5 illustrates the different performances with two simulation results, where $n = 10000$ and $\lambda = 30$ in its left part and $n = \lambda = 1000$ on its right.

The above concerns the consistency of the nonparametric estimators. It is natural to ask for an associated rate, or even a second order result. In the context of Grenander's estimator it is known that the step from a concave majorant to its derivative, which parallels the transition displayed in Fig. 3, leads to a decrease in rate from $n^{-1/2}$ to $n^{-1/3}$; see e.g. [21, 12]. In the present setup this would explain that n has to be large in order for 'the asymptotics to set in', as suggested by simulation experiments. In the complementary problem, where only the separate sets of arrival and departure times are known, such a loss of rate might not occur; see e.g. [4, Theorem 6.3]. Also, it is expected that Brown's estimator has rate $n^{-1/2}$. For a related discrete time queuing model this has indeed been proved in [25].

4. Patterns, permutons and parametric models

The nonparametric results in the previous section were based on regarding permutations as two-dimensional distribution functions, which provides a connection to empirical process theory. For samples, a classical approach in a parametric context is the moment method, where estimating equations for the parameters are obtained by equating theoretical moments with their empirical versions, the sample moments. In the present situation the role of moments may be taken over by *pattern probabilities* and their data analogues, *pattern frequencies*. In [16], a topology based on pattern counting is introduced; the resulting limit objects are permutons, in fact two-dimensional copulas. We discuss this in the general case and then apply the results to delay copulas.

We need a suitable notion of restriction for permutations: For $\pi \in \mathbb{S}_n$, $A = \{i_1, \dots, i_m\} \subset [n]$ with $1 \le i_1 < \cdots < i_m \le n$, let $\pi_A \in \mathbb{S}_m$ be defined by the requirement that, for all $j, k \in [m]$ with $j \neq k$,

$$\pi_A(j) < \pi_A(k) \quad \Longleftrightarrow \quad \pi(i_j) < \pi(i_k). \tag{31}$$

For $\sigma \in \mathbb{S}_m$ and $\pi \in \mathbb{S}_n$ we say that σ occurs in π at the position vector given by $A \subset [n]$ if $\pi_A = \sigma$, and the *density* of σ in π is defined as the corresponding proportion,

$$t(\pi, \sigma) = \frac{\#\{A \subset [n] : \#A = m,\, \pi_A = \sigma\}}{\binom{n}{m}}. \tag{32}$$

This is also the probability of observing the pattern $\sigma \in \mathbb{S}_m$ in the permutation $\pi \in \mathbb{S}_n$ if $A \subset [n]$ with $\#A = m$ is chosen uniformly at random. We augment this by putting $t(\pi, \sigma) := 0$ if $n < m$. For example, the data in Fig. 1 lead to $\pi = 123659748 \in \mathbb{S}_9$, and with $A = \{3, 6, 7, 8\}$ we obtain $\pi_A = 1432 \in \mathbb{S}_4$. Further, each occurrence of the transposition $\tau = 21 \in \mathbb{S}_2$ corresponds to a pair where the earlier-born outlives the other person. There are seven such pairs, hence $t(\pi, \tau) = 7/\binom{9}{2} = 7/36$. Similarly, the permutations Π_n discussed at the end of Example 2 (c) contain the identity permutation $\sigma = (1, 2, \dots, k) \in \mathbb{S}_k$ as a pattern.

Considering all pattern densities simultaneously we may regard any permutation π as a real-valued function on $\mathbb{S} := \bigcup_{n=1}^{\infty} \mathbb{S}_n$ via

$$\mathbb{S} \ni \pi \mapsto \big(t(\pi, \sigma)\big)_{\sigma \in \mathbb{S}} \in \mathbb{R}^{\mathbb{S}}.$$

Given a random permutation Π we then obtain a random function $T : \mathbb{S} \to \mathbb{N}_0$ by putting $T(\sigma) = t(\Pi, \sigma)$, $\sigma \in \mathbb{S}$. This may be interpreted as an analogue of the sequence of sample moments. To obtain the theoretical counterpart for a given (general) copula C, we construct the function $C^{\mathbb{S}} : \mathbb{S} \to [0, 1]$ as follows: Let $Z_i = (U_i, V_i)$, $i \in \mathbb{N}$, be a sequence of independent bivariate random vectors with distribution function C. For each $k \in \mathbb{N}$ let Π_k be the random element of $\mathbb{S}_k$ that connects the ranks of $U_1, \dots, U_k$ and $V_1, \dots, V_k$ as in (1). Then we define the function $C^{\mathbb{S}} : \mathbb{S} \to [0, 1]$ by

$$C^{\mathbb{S}}(\sigma) = \mathbb{P}(\Pi_k = \sigma) \quad \text{for all } \sigma \in \mathbb{S}_k,\ k \in \mathbb{N}. \tag{33}$$

Clearly, C determines the distribution of Π_k for all $k \in \mathbb{N}$, and thus $C^{\mathbb{S}}$ depends on C only. Remarkably, as pointed out in [16, Lemma 5.1], the copula C is in turn determined by $C^{\mathbb{S}}$. A somewhat streamlined version of their argument is the following: As in Sect. 2, let ν_n and μ be the distributions on the unit square with distribution functions C_n and C. The empirical copulas C_n converge almost surely to C as $n \to \infty$, which implies that the laws $\mathcal{L}(\nu_n)$ converge in the space of probability measures on the unit square to the distribution δ_μ concentrated at μ with respect to the weak topology. Now we note that $\mathcal{L}(\nu_n)$ is specified by the function $\mathbb{S}_n \ni \sigma \to C^{\mathbb{S}}(\sigma)$. In particular, this settles the analogue of the classical moment problem.

Results on the convergence of Π_n in the pattern counting topology are most easily derived by using U-statistics. Recall that, for an i.i.d. sequence $(Z_i)_{i\in\mathbb{N}}$ with values in the measurable space $(E, \mathcal{E})$, and a symmetric function $h : E^k \to \mathbb{R}$ that is measurable with respect to the product σ-field $\mathcal{E}^{\otimes k}$, the associated sequence $T_n, n \geq k$, of U-statistics with kernel h is given by

$$T_n = \frac{1}{\binom{n}{k}} \sum_{1\leq i_1<\cdots<i_k\leq n} h(Z_{i_1}, \ldots, Z_{i_k}), \quad n \geq k.$$

The relative frequency $t(\Pi_n, \sigma)$ of a pattern $\sigma \in \mathbb{S}_k$ then arises as a U-statistic if $h = h_\sigma$ is the indicator function of σ, applied to the permutation associated with $z_i = (u_i, v_i)$, $i \in [k]$. Further, with this choice of h, it holds that $C^{\mathbb{S}}(\sigma) = ET_k$.

For computational purposes such as the calculation of moments of $T_n(\sigma) = t(\Pi_n, \sigma)$ it is helpful to express $T_n(\sigma)$ directly in terms of the sample variables $Z_i = (U_i, V_i)$, $i = 1, \ldots, n$. We regard $E := [0, 1] \times [0, 1]$ as the sample space for the random variables Z_i. Of course, for the single element $\sigma \in \mathbb{S}_1$ we have $h_\sigma \equiv 1$ and $T_n(\sigma) = 1 = C^{\mathbb{S}}(\sigma)$ for all $n \in \mathbb{N}$. For $\sigma \in \mathbb{S}_k$, $k \geq 2$, with inverse $\sigma^{-1} \in \mathbb{S}_k$, defining

$$\begin{aligned} &h_\sigma\big((u_1, v_1), \ldots, (u_k, v_k)\big) \\ &\quad = \sum_{\tau\in\mathbb{S}_k} 1(u_{\tau(1)} < \cdots < u_{\tau(k)}, v_{\tau\circ\sigma^{-1}(1)} < \cdots < v_{\tau\circ\sigma^{-1}(k)}) \end{aligned} \tag{34}$$

for $z_j = (u_j, v_j) \in E$, $j = 1, \ldots, k$, we get

$$T_n(\sigma) = \frac{1}{\binom{n}{k}} \sum_{1\leq j_1<\cdots<j_k\leq n} h_\sigma(Z_{j_1}, \ldots, Z_{j_k}), \quad n \geq k,$$

and

$$C^{\mathbb{S}}(\sigma) = \mathbb{P}(\Pi_k = \sigma) = k!\mathbb{P}(U_1 < \cdots < U_k, V_{\sigma^{-1}(1)} < \cdots < V_{\sigma^{-1}(k)}). \tag{35}$$

With this background we have almost sure convergence of the sequence $(\Pi_n)_{n\in\mathbb{N}}$ in the pattern counting topology, meaning that for all $\sigma \in \mathbb{S}$,

$$T_n(\sigma) = t(\Pi_n, \sigma) \to C^{\mathbb{S}}(\sigma) \quad \text{a.s. as } n \to \infty. \tag{36}$$

This is [16, Corollary 4.3]. The connection to U-statistics leads to a direct argument: As for each $k \in \mathbb{N}$ and each $\sigma \in \mathbb{S}_k$ the expectation $Eh_\sigma(Z_1, \dots, Z_k) = \mathbb{P}(\Pi_k = \sigma)$ is finite, the process $(T_n(\sigma), \mathcal{F}_n)_{n \ge k}$ with $\mathcal{F}_n := \sigma(\{Z_m : m \ge n\})$ is a backwards martingale, and the limit theorem for these structures implies that $T_n(\sigma)$ converges a.s. to $Eh_\sigma(Z_1, \dots, Z_k)$ as $n \to \infty$; see e.g. [26, Problems 12.13 and 12.15] or [17, Abschnitt 10.3].

In a statistical context this result can be used to obtain consistency of 'moment estimators'. A next important step would be the asymptotic normality of the scaled difference between Π_n and the limiting copula C, represented by the function $C^{\mathbb{S}}$. We require an extension of $C^{\mathbb{S}}$ to two arguments. Let $C^{\mathbb{S},2}(\sigma, \tau) = C^{\mathbb{S},2}(\tau, \sigma) = C^{\mathbb{S}}(\sigma)$ if $\tau = 1 \in \mathbb{S}_1$, and for $\sigma \in \mathbb{S}_k$, $\tau \in \mathbb{S}_j$ with $k, j > 1$ let

$$C^{\mathbb{S},2}(\sigma, \tau) = \mathbb{P}\bigl(\Pi_k(Z_1, Z_2, \dots, Z_k) = \sigma,\, \Pi_j\bigl(Z_1, Z_2', \dots, Z_j'\bigr) = \tau\bigr),$$

where $Z_1, Z_2, Z_2', \dots$ are independent copies of $Z = (U, V)$. For $\sigma \in \mathbb{S}$ let $|\sigma| := k$ if $\sigma \in \mathbb{S}_k$. For each $n \in \mathbb{N}$ and $\sigma \in \mathbb{S}$ let

$$W_n(\sigma) := \frac{\sqrt{n}}{|\sigma|}\bigl(T_n(\sigma) - C^{\mathbb{S}}(\sigma)\bigr) \quad \text{if } n \ge |\sigma|, \tag{37}$$

and $W_n(\sigma) := 0$ if $n < |\sigma|$.

Theorem 8. *Let $d \in \mathbb{N}$ and $\sigma_1, \dots, \sigma_d \in \mathbb{S}$. Then the d-dimensional random vector $(W_n(\sigma_1), \dots, W_n(\sigma_d))$ converges in distribution as $n \to \infty$ to the d-dimensional centered normal random vector $(\mathbb{W}_1, \dots, \mathbb{W}_d)$ with covariances $\mathrm{cov}(\mathbb{W}_i, \mathbb{W}_j) = \rho_C(\sigma_i, \sigma_j)$, $1 \le i, j \le d$, where*

$$\rho_C(\sigma, \tau) := C^{\mathbb{S},2}(\sigma, \tau) - C^{\mathbb{S}}(\sigma)\, C^{\mathbb{S}}(\tau), \quad \sigma, \tau \in \mathbb{S}. \tag{38}$$

Proof. For $\sigma \in \mathbb{S}$ with $k := |\sigma| > 1$ and $r = 1, \dots, k$ let $\hat{h}_{r,\sigma} : E^r \to \mathbb{R}$ be defined by

$$\hat{h}_{r,\sigma}(z) = Eh_\sigma(z_1, \dots, z_r, Z_{r+1}, \dots, Z_k) - C^{\mathbb{S}}(\sigma)$$

for all $z = (z_1, \dots, z_r) \in E^r$, and put $\hat{h}_{r,\sigma} \equiv 0$ if $|\sigma| = 1$. Then the random variables

$$\hat{W}_n(\sigma) := n^{-1/2} \sum_{j=1}^{n} \hat{h}_{1,\sigma}(Z_i), \quad n \in \mathbb{N},$$

are sufficiently close to the variables of interest in the sense that, for all $\sigma \in \mathbb{S}$,

$$W_n(\sigma) - \hat{W}_n(\sigma) \to 0 \quad \text{in probability as } n \to \infty, \tag{39}$$

see e.g. [26, Theorem 12.3]. Obviously, $\rho_C(\sigma, \tau) = \mathrm{cov}(\hat{h}_{1,\sigma}(Y_1), \hat{h}_{1,\tau}(Y_1))$ for $\sigma, \tau \in \mathbb{S}$. The multivariate central limit theorem and (39) together imply that

$$\bigl(W_n(\sigma_1), \dots, W_n(\sigma_d)\bigr) \to_{\mathrm{distr}} (\mathbb{W}_1, \dots, \mathbb{W}_d) \quad \text{as } n \to \infty, \tag{40}$$

where $(\mathbb{W}_1, \dots, \mathbb{W}_d)$ is a d-dimensional centered normal random vector with covariances $\rho_C(\sigma_i, \sigma_j)$, $1 \le i, j \le d$. □

Remark 9. In the context of parametric estimation for delay models it is enough to consider finite-dimensional random vectors, as in the above result. In a separate paper [2] we consider nonparametric tests for permutation data arising from general copulas. In order to obtain distributional limits for such procedures it is then important to regard all $\sigma \in \mathbb{S}$ simultaneously and to consider the stochastic processes $W_n = (W_n(\sigma))_{\sigma\in\mathbb{S}}$ as random elements of some suitable infinite-dimensional space. Theorem 8 then covers the convergence of the finite-dimensional distributions. ◁

Returning to the delay model we now apply the pattern frequency approach in a specific parametric context. We assume that the delays are exponentially distributed, so that $G(x) = 1 - \exp(-\lambda x)$ for all $x \geq 0$ with an unknown parameter $\lambda > 0$. The idea is to use the relative frequency T_n of inversions in Π_n as an estimator for the limiting probability $\phi(\lambda) = C[G]^{\mathbb{S}}(\tau)$ with $\tau = 21 \in \mathbb{S}_2$; see (35).

With U_1, U_2, X_1, X_2 independent, $U_1, U_2 \sim \text{unif}(0,1)$ and $X_1, X_2 \sim \text{Exp}(\lambda)$ we obtain by sequentially conditioning on U_1, U_2 and X_2,

$$\begin{aligned}\phi(\lambda) &= P(U_1 < U_2, U_1 + X_1 > U_2 + X_2) + P(U_1 > U_2, U_1 + X_1 < U_2 + X_2)\\ &= 2\int_0^1\int_u^1 P(X_1 > X_2 + v - u)\,dv\,du\\ &= 2\int_0^1\int_u^1\int_0^\infty P(X_1 > x + v - u)\,\lambda e^{-\lambda x}\,dx\,dv\,du\\ &= 2\int_0^1\int_u^1\int_0^\infty e^{-\lambda(x+v-u)}\,\lambda e^{-\lambda x}\,dx\,dv\,du = \frac{e^{-\lambda} - 1 + \lambda}{\lambda^2}.\end{aligned}$$

We note in passing that, in accordance with part (b) of Example 2, we have $\phi(\lambda) \to 1/2$ if $\lambda \to 0$, and $\phi(\lambda) \to 0$ if $\lambda \to \infty$. To see that the function ϕ is strictly decreasing on $(0,\infty)$, we argue that its derivative $\phi'(\lambda) = -\frac{(2+\lambda)e^{-\lambda}-(2-\lambda)}{\lambda^3}$ is negative for all $\lambda > 0$. In fact, this is obviously true for all $\lambda \geq 2$. For $0 < \lambda < 2$ we put $x = \lambda/2$, write

$$(2+\lambda)e^{-\lambda} - (2-\lambda) = 2(1-x)(1+x)e^{-x}\left(\frac{e^{-x}}{1-x} - \frac{e^{x}}{1+x}\right),$$

and note that

$$\frac{e^{-x}}{1-x} - \frac{e^{x}}{1+x} = 2\sum_{n=2}^{\infty}\left(\sum_{k=0}^{2n-1}\frac{1}{k!}(-1)^k\right)x^{2n-1},$$

where the sum $\sum_{k=0}^{2n-1}\frac{1}{k!}(-1)^k$ is known to be the (for $n \geq 2$) positive probability that a randomly chosen permutation in $\mathbb{S}_{2n-1}$ has no fixed point. If $T_n < 1/2$ we may thus define a unique estimator λ_n for λ via $\phi(\lambda_n) = T_n$. Almost sure convergence of the pattern frequencies, see (36), implies that the condition is satisfied with probability 1 from some finite $n \in \mathbb{N}$ onwards. Together with the continuity of the inverse function ϕ^{-1} at λ this also shows that λ_n is a consistent estimator for λ. We show next that this estimator is asymptotically normal.

Theorem 10. *With λ_n as defined above it holds that*

$$\sqrt{n}(\lambda_n - \lambda) \to_{\text{distr}} Z \quad \text{as } n \to \infty, \tag{41}$$

where Z has a centered normal distribution with variance

$$\operatorname{var}_\lambda(Z) = \frac{2\lambda^2(2\lambda^2 - 3\lambda - 6 + 2(3\lambda^2 + 2\lambda + 6)e^{-\lambda} - (\lambda+6)e^{-2\lambda})}{3((\lambda-2)^2 + 2(\lambda^2-4)e^{-\lambda} + (\lambda+2)^2 e^{-2\lambda})}. \tag{42}$$

Proof. We use Theorem 8 to obtain asymptotic normality for the relative number T_n of inversions in Π_n and then apply the delta method.

We can write T_n as

$$T_n = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \big(1(U_i < U_j)1(U_i + X_i > U_j + X_j) \\ + 1(U_i > U_j)1(U_i + X_i < U_j + X_j)\big),$$

where $U_1, U_2, \dots, X_1, X_2, \dots$ are independent, $U_i \sim \text{unif}(0,1)$, $X_i \sim \text{Exp}(\lambda)$ for all $i \in \mathbb{N}$.

We need some calculations. For $u, u' \in (0,1)$, $x, x' > 0$, let

$$h_1\big(u,x,u',x'\big) = \begin{cases} 1, & \text{if } u \le u' \text{ and } u + x \ge u' + x', \\ 0, & \text{otherwise,} \end{cases}$$

$$h_2\big(u,x,u',x'\big) = \begin{cases} 1, & \text{if } u > u' \text{ and } u + x < u' + x', \\ 0, & \text{otherwise.} \end{cases}$$

This displays T_n as a U-statistic with kernel $h := h_1 + h_2$ for the i.i.d. sequence (U_i, X_i), $i \in \mathbb{N}$, of two-dimensional random vectors. Let g, g_1, g_2 be defined as

$$g_1(u,x) := Eh_1(u,x,U_1,X_1), \quad g_2(u,x) := Eh_2(u,x,U_1,X_1)$$

and $g := g_1 + g_2$. Then $g(u,x) = Eh(u,x,U_1,X_1)$. For $u + x > 1$ we obtain

$$\begin{aligned} g_1(u,x) &= P(U_1 > u, U_1 + X_1 < u + x) \\ &= \int_u^1 P(X_1 < u + x - v)\, dv \\ &= \int_u^1 \big(1 - e^{-\lambda(u+x-v)}\big)\, dv = 1 - u - \frac{e^{-\lambda x}}{\lambda}\big(e^{\lambda(1-u)} - 1\big), \end{aligned}$$

and similarly for $u + x \le 1$,

$$g_1(u,x) = \int_u^{u+x} \big(1 - e^{-\lambda(u+x-v)}\big)\, dv = x - \frac{1}{\lambda}\big(1 - e^{-\lambda x}\big).$$

Further,

$$g_2(u,x) = P(U_1 < u, U_1 + X_1 > u + x)$$

$$= \int_0^u P(X_1 > u + x - v)\, dv$$
$$= \int_0^u e^{-\lambda(u+x-v)}\, dv = \frac{e^{-\lambda x}}{\lambda}\big(1 - e^{-\lambda u}\big).$$

Replacing u and x by independent random variables $U \sim \mathrm{unif}(0,1)$ and $X \sim \mathrm{Exp}(\lambda)$ and then taking expectations we obtain another proof of the above formula for $\phi(\lambda)$.

In order to apply Theorem 8 we need

$$\xi(\lambda) := 4\, E\big(h(U_1, X_1, U_2, X_2) h(U_1, X_1, U_3, X_3)\big),$$

where $(U_1, X_1), (U_2, X_2), (U_3, X_3)$ are i.i.d. with $U_1 \sim \mathrm{unif}(0,1)$ and $X_1 \sim \mathrm{Exp}(\lambda)$. Conditioning on U_1 and X_1 we get

$$\xi(\lambda) = 4 \int_0^1 \int_0^\infty g(u,x)^2\, \lambda e^{-\lambda x}\, dx\, du.$$

Inserting the expressions found above we obtain, after some calculations,

$$\xi(\lambda) = \frac{2}{3\lambda^3}\big(8\lambda - 15 + 2(3\lambda + 8)e^{-\lambda} - e^{-2\lambda}\big).$$

Together with $Eh(U_1, X_1, U_2, X_2) = \phi(\lambda)$ this yields

$$\rho(\lambda) := 4 \operatorname{cov}\big(h(U_1, X_1, U_2, X_2), h(U_1, X_1, U_3, X_3)\big)$$
$$= \frac{2}{3\lambda^4}\big(2\lambda^2 - 3\lambda - 6 + 2\big(3\lambda^2 + 2\lambda + 6\big)e^{-\lambda} - (\lambda + 6)e^{-2\lambda}\big),$$

which is the asymptotic variance of T_n.

In the second step we apply the delta method. As λ_n is given implicitly by $\phi(\lambda_n) = T_n$ we need the derivative of ϕ^{-1} at $\phi(\lambda)$, which is given by

$$\beta(\lambda) := \frac{1}{\phi'(\lambda)} = \frac{\lambda^3}{2 - \lambda - (2 + \lambda)e^{-\lambda}},$$

and $\rho(\lambda)\beta(\lambda)^2$ now leads to (42). □

If the sojourn times $X_1, \dots, X_n$ were all known then the maximum likelihood estimator $\hat\lambda_n = 1/\overline{X}_n$, with $\overline{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$, is an asymptotically efficient estimator of λ. The variance of its limit normal distribution is λ^2. From this we deduce that the asymptotic efficiency of λ_n with respect to $\hat\lambda_n$ is $\mathrm{eff}(\lambda) = \lambda^2/\mathrm{var}_\lambda(Z)$. From (42) we obtain

$$\mathrm{eff}(\lambda) = \frac{3}{4} - \frac{15}{8\lambda} + \frac{39}{16\lambda^2} + O\Big(\frac{1}{\lambda^3}\Big) \quad \text{as } \lambda \to \infty$$

and

$$\lim_{\lambda \to 0} \mathrm{eff}(\lambda) = 0.$$

The limits $\frac{3}{4}$ and 0 obtained as $\lambda \to \infty$ and $\lambda \to 0$ correspond to the extreme cases $U = Y$ and U, Y independent; see Example 2 (b).

Using Theorem 10 we can now augment the discussion at the end of Sect. 3: In the parametric model with exponential delay distributions we retain the standard rate $n^{-1/2}$ for the supremum norm distance between the corresponding estimate $\hat{G}_n$ of the distribution function G, where $\hat{G}_n(x) = 1 - \exp(-\lambda_n x)$ and $G(x) = 1 - \exp(-\lambda x)$, $x \geq 0$.

Finally, it should be clear that the above approach could be extended to parametric families of higher dimension. Again in analogy to the moment method, we would then count longer patterns. The necessary calculations may become cumbersome; indeed, those above were partially carried out with the help of the computer algebra program SageMath [24].

Acknowledgments

Constructive comments by an anonymous referee have led to a considerable improvement of the paper.

References

[1] Balabdaoui, F., Jankowski, H., Pavlides, M., Seregin, A., Wellner, J. (2011) On the Grenander estimator at zero. *Statist. Sinica* 21, 873–899. MR2829859

[2] Baringhaus, L., Grübel, R. (2023+) Pattern based tests for two-dimensional copulas. In preparation.

[3] Bingham, N.H., Dunham, B. (1997) Estimating diffusion coefficients from count data: Einstein-Smoluchowski theory revisited. *Ann. Inst. Statist. Math.* 49, 667–679. MR1621845

[4] Bingham, N.H., Pitts, S.M. (1999) Non-parametric estimation for the $M/G/\infty$ queue. *Ann. Inst. Statist. Math.* 51, 71–97. MR1704647

[5] Blanghaps, S., Nov, Y., Weiss, G. (2013) Sojourn time estimation in an $M/G/\infty$ queue with partial information. *J. Appl. Prob.* 50, 1044–1056. MR3161372

[6] Brown, M. (1970) An $M/G/\infty$ estimation problem. *Ann. Math. Statist.* 41, 651–654. MR0261722

[7] Corless, R.M., Gonnet, G.H., Hare, D.E.G.,Jeffrey, D.J.,Knuth, D.E. (1996) On the Lambert W function. *Adv. Comp. Math.* 5, 329–359. MR1414285

[8] Deheuvels, P. (1979) La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Bulletin de la Classe des sciences* 65, 274–292. MR0573609

[9] Gaenssler, P., Stute, W. (1987) *Seminar on Empirical Processes.* Springer, Basel. MR0902803

[10] Goldenshluger, A. (2018) The $M/G/\infty$ estimation problem revisited. *Bernoulli* 24, 2531–2568. MR3779694

[11] Grenander, U. (1956). On the theory of mortality measurement, Part II. *Skand. Aktuarietidskr.* 39, 125–153. MR0093415

[12] Groeneboom, P. (1985). Estimating a monotone density. In: *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Wadsworth, Belmont, 539–555. MR0822052

[13] Grübel, R. (2023) Ranks, copulas, and permutons. *Metrika*, to appear. doi: 10.1007/s00184-023-00908-2. MR4688341

[14] Grübel, R., Wegener, H. (2011) Matchmaking and testing for exponentiality in the $M/G/\infty$ queue. *J. Appl. Probab.* 48, 131–144. MR2809891

[15] Hall, P., Park, J. (2004) Nonparametric inference about the service time distribution from indirect measurements. *J. Roy. Statist. Soc. Ser. B* 66, 861–875. MR2102469

[16] Hoppen, C., Kohayakawa, Y., Moreira, C.G., Ráth, B., Sampaio, R.M. (2013) Limits of permutation sequences. *J. Combin. Theory Ser. B* 103, 93–113. MR2995721

[17] Luschgy, H. (2013) *Martingale in diskreter Zeit.* Springer, Berlin.

[18] Marshall, A. W. (1970) Discussion of Barlow and van Zwet's paper. In *Nonparametric Techniques in Statistical Inference.* Proceedings of the First International Symposium on Nonparametric Techniques held at Indiana University, June, 1969 (M.L. Puri, ed.), 174–176. Cambridge University Press, London. MR0273755

[19] Massart, P. (1990) The tight constant in the Dvoretsky-Kiefer-Wolfowitz inequality. *Ann. Probab.* 18, 1269–1283. MR1062069

[20] Nelson, R.B. (2006) *An Introduction to Copulas*, 2nd ed. Springer, New York. MR2197664

[21] Prakasa Rao, B. L. S. (1969) Estimation of a unimodal density. *Sankhyā, Ser. A* 31, 23–36. MR0267677

[22] Pickands, J., Stine, R.A. (1997) Estimation for an $M/G/\infty$ queue with incomplete information. *Biometrika* 84, 295–308. MR1467048

[23] Rüschendorf, L. (1976) Asymptotic distributions of multivariate rank order statistics. *Ann. Statist.* 4, 912–923. MR0420794

[24] SageMath. Available at www.sagemath.org.

[25] Schweer, S., Wichelhaus, C. (2015) Nonparametric estimation of the service time distribution in the discrete-time $GI/G/\infty$ queue with partial information. *Stoch. Proc. Appl.* 125, 233–253. MR3274698

[26] van der Vaart, A.W. (1998) *Asymptotic Statistics.* Cambridge University Press, Cambridge. MR1652247

[27] Whittaker, E.T., Watson, G.N. (1988) *A Course of Modern Analysis.* Reprint of the fourth (1927) edition. Cambridge University Press, Cambridge. MR1424469