Estimating a distribution function for discrete data subject to random truncation with an application to structured finance *

Jackson P. Lautier[‡] Vladimir Pozdnyakov[†] Jun Yan[†]

August 10, 2021

Abstract

The literature for estimating a distribution function from truncated data is extensive, but it has given little attention to the case of discrete data over a finite number of possible values. We examine the Woodroofe-type estimator in this case and prove that the resulting vector of hazard rate estimators is asymptotically normal with independent components. Asymptotic normality of the survival function estimator is then established. Sister results for the truncation random variable are also proved. Further, a hypothesis test for the shape of the distribution function based on our results is presented. Such a test is useful to formally test the stationarity assumption in length-biased sampling. The finite sample performance of the estimators are investigated in a simulation study. We close with an application to an automotive lease securitization.

Keywords— Asymptotic normality, product-limit estimator, reverse hazard rate, stationarity, survival

1 Introduction

Consider two independent positive random variables X and Y with distribution functions F and G such that we only observe the pairs (X, Y) for which $Y \leq X$. The pairs (X, Y) are assumed to be independent and identically distributed (iid). This is the structure of the classical problem of estimating a distribution function from data subject to random truncation, and it has been thoroughly studied in the literature (e.g., Lynden-Bell, 1971; Woodroofe, 1985; Wang et al., 1986; Keiding and Gill, 1990; Stute, 1993; He and Yang, 1998a). We further assume, however, that X and Y are integer-valued random variables with a finite number of possible values.

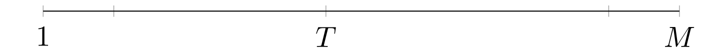
1.1 Motivation

The model of the opening paragraphs occurs frequently within finance. Expected payments are often due on a periodic basis, such as monthly, quarterly, or annually. A monthly frequency is common for insurance products and debt instruments, such as insurance premiums, credit card payments, mortgages, auto loans, and the like. Additionally, insurance contracts and debt typically have a fixed, finite term, such as a 36-month automobile lease contract. Even whole life insurance, which is technically written with payments due in perpetuity is, in actuality, a fixed-length contract of unknown duration (one may comfortably cap assumed lifetimes at 130 years, for example). Often,

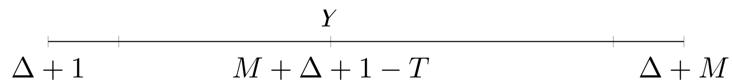
*This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DHE 1747453.

†Department of Statistics, University of Connecticut

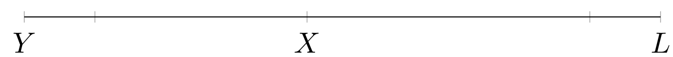
‡Corresponding to jackson.lautier@uconn.edu.



(a) Symbolically, T represents a random lease start time. Nonrandom time M is the origination time of the youngest lease in the trust as of the beginning of the trust observation window. Thus, $1 \leq T \leq M$.



(b) We call the time that the trust observation window begins $\Delta + M$, and so nonrandom Δ denotes the minimum age of a lease in the trust as of time $\Delta + M$. Defining $Y = M + \Delta + 1 - T$ implies $\Delta + 1 \leq Y \leq \Delta + M$.



(c) We only observe the random lease termination time, X, if $X \geq Y$. Nonrandom L represents the termination time of the lease with the longest active ongoing payments, and it coincides with the close of the trust observation window. Thus, $\Delta + 1 \leq X \leq L$.

Figure 1: The connected random variables T, Y, X and the associated timelines for truncated data from an auto lease securitization.

these financial products are structured or securitized into a trust. Therefore, investors or risk professionals will only observe those contracts that survive long enough to be included into the financial trust.

We elaborate with a specific example from structured finance. Consider an automotive lease securitization, such as Mercedes-Benz (2017), in which consumer automotive lease contracts are pooled together into a trust. Standard automotive lease contracts have a fixed and known duration, such as 36 months, with required monthly payments. Further, the payment performance of the lessee will be reported monthly, so the observed survival times of the lease contracts will be discrete and equally spaced. Truncation occurs because only those leases that survive long enough to be collected into the trust will be observable by the investor; that is, a form of bias under the general umbrella of delayed entry or length-biased sampling.

To formalize, let X denote the random lease lifetime and let T denote the random time of lease originations. If L represents the age of the last lease termination in a sample, then $X \leq L$. Since issuers of structured debt typically have a legal obligation to the trust to select lease contracts with a minimum history of on-time payments, the youngest lease in the trust will have a minimum age of Δ as of the onset of the trust, where Δ is a positive integer. Hence, each lease will have a minimum survival time of $\Delta + 1$, and so $\Delta + 1 \leq X \leq L$. If M is the origination time of the youngest lease in the trust, then $0 \leq T \leq M$ and the trust starting time is $M + \Delta$. For all practical purposes, $M + \Delta \leq L$. We emphasize here that the variables Δ, M, and L are non-random and known as of the onset of the problem. Notably, if we define $Y = M + \Delta + 1 - T$, then Y denotes a truncation random variable representing the minimum amount of time a lease must remain active to be observed in the trust. In other words, an investor will only observe those leases such that $X \geq Y$, and this is exactly the situation of the opening paragraph. For completeness, notice $\Delta + 1 \leq Y \leq \Delta + M$. We present a visualization of the connected random variables and timelines in Figure 1.

1.2 Literature Review

In reviewing the substantial body of literature for estimating a distribution function in the presence of random truncation, we were surprised to see the case of discrete F and G receive little attention. Two examples of seminal works in this field are of course Woodroofe (1985) and Wang et al. (1986). Woodroofe (1985) proves consistency results for the Lynden-Bell (1971) estimator and shows weak convergence to a Gaussian process but left the exact form of the covariance structure of the limiting process undefined. Throughout his work, Woodroofe (1985) assumes continuous distribution functions F and G. Wang et al. (1986) extends the results of Woodroofe (1985) with a precise description of the asymptotic covariance structure. It is noteworthy that this structure is the analogue of the covariance structure of the Kaplan–Meier estimator. Wang et al. (1986) alludes to the idea that F and G need not be continuous in establishing strong consistency for the product limit estimator of F, but they assume continuity of F and G in working to define the covariance structure.

Since Woodroofe (1985) and Wang et al. (1986), there has been many notable and significant contributions. Chronologically, Chao and Lo (1988) further study the estimator of F by expressing a hazard process as iid means of random variables and imposing the same conditions as Woodroofe (1985). The result is the ability to represent the difference of F and its estimator as iid means of random variables to obtain weak convergence, including the associated covariance structures. Keiding and Gill (1990) reparametrize the left truncation model as a three-state Markov process to invoke the statistical theory of counting processes by Aalen and Johansen (1978) to establish the nonparametric maximum likelihood estimator (NPMLE), consistency, asymptotic normality, and efficiency. Both papers derive results assuming continuity of F, however. Lai and Ying (1991) relax the continuity assumption of F in using martingale integral representations and empirical process theory to prove uniform strong consistency and weak convergence results, though they modify the product-limit estimator in doing so.

Somewhat more recently, Gürler and Wang (1993) examine hazard functions and their derivatives for nonparametric kernel estimators. Similarly, they again assume continuity of G in proving asymptotic normality. Stute (1993) derives an almost sure representation of the estimator for F with weaker distributional assumptions than Woodroofe (1985) and improved error bounds. Chen et al. (1995) prove the Lynden-Bell (1971) estimator is uniformly strong consistent over the whole half line, a problem left open by Woodroofe (1985). Both papers assume continuity of F and G throughout. In part one of a two-part sequence, He and Yang (1998a) find a simpler representation for the estimator of the truncation probability to show strong consistency and asymptotic normality via an iid representation. While, these results are true for arbitrary F and G, they do not consider the estimators for the distribution functions for F and G. In part two, He and Yang (1998b) prove that the estimator for F obeys the strong law of large numbers when estimating F_0 for arbitrary and not necessarily continuous F (we clarify the important distinction between F and F_0 momentarily). This relaxes the assumption of continuity but does not address asymptotic normality.

The classical problem of estimating F from truncated data has by now become commonplace in textbooks (e.g., Karr, 1991; de la Peña and Giné, 1999; Owen, 2001; Hu, 2013), but any extended treatment assumes continuity of F (e.g., de la Peña and Giné, 1999, §5.5.3).

We expanded our review to consider the random truncation model along with right censoring. A seminal work in this field is Tsai et al. (1987), which gives asymptotic results when left truncated data are also subject to right censoring. Nonetheless, the authors also assume continuous F. The continuity of F and G is assumed in related works (Uzoḡullari and Wang, 1992; Gijbels and Wang, 1993; Gürler, 1996; Zhou, 1996; Zhou and Yip, 1999; Asgharian and Wolfson, 2005; Huang and Qin, 2011). Given its unexpected absence, therefore, the literature gap we endeavor to fill is to prove asymptotic normality of the Woodroofe-type estimator in the case of discrete random variables X and Y with a finite number of possible values.

1.3 Outline

The paper proceeds as follows. In Section 2, we review some of the major results of Woodroofe (1985) to demonstrate that there is enough information from the sample of truncated data to draw inference about F and G. Next, in Section 3, we prove our major results: the asymptotic normality and independence of the estimation vector of the hazard rates for F and its analog for G; and, asymptotic normality of the estimator for the survival function of X and the estimator of the distribution function of Y. We also present a convenient hypothesis test for the shape of the distribution function of Y. In Section 4, we present simulation studies for a closer examination of the results in Section 3. In Section 5, we apply our results to auto lease securitization data. The paper closes with a brief discussion.

2 Estimation

We begin by reviewing notation and key results from Woodroofe (1985). Next, we emphasize that it is possible to express the probability mass function (pmf) of F (or G) in terms of a useful joint conditional distribution representation. We then suggest that, because of the discrete nature of X and Y, it is preferable to work in terms of the hazard rate of X and the *reverse hazard rate* of G. Estimators for both the hazard and reverse hazard rates are formally defined. Finally, we state the maximum likelihood estimators of F, G, the hazard rate, and the reverse hazard rate.

2.1 Preliminaries

Working from the notation of Woodroofe (1985), let F and G be the distribution functions of non-negative independent random variables X and Y, respectively. Let H_* denote the joint distribution function of X and Y given $Y \leq X$, and let F_* and G_* denote the marginal distributions functions given $Y \leq X$ of X and Y, respectively. That is,

$$H_*(F, G, x, y) = \Pr(X \leq x, Y \leq y \mid X \geq Y)$$

is the joint conditional distribution function with conditional marginal distributions F_* and G_*. We include F and G within the definition of H_* to stress which F and G are employed to construct distribution H. For convenience, we may drop the x and y from the notation for H_* when the meaning is clear or if the clarification is nonessential; that is, we may simply write: $H_*(F, G)$.

We now review key observations made by Woodroofe (1985). Define

$$a_F = \inf\{z > 0 : F(z) > 0\} \geq 0,$$

and

$$b_F = \sup\{z > 0 : F(z) < 1\} \leq \infty.$$

That is, (a_F, b_F) is the interior of the convex support of F and similarly (a_G, b_G) for G. Naturally, to avoid complete truncation and full data loss, we must have $a_G < b_F$.

Next, we need to introduce two classes of distribution pairs (F, G). The first class includes all pairs of F and G that allow the construction of the two-dimensional distribution H_*:

$$\mathcal{K} = \{(F, G) : F(0) = 0 = G(0), \quad \Pr(Y \leq X) > 0\}.$$

The second class includes those pairs (F, G) that can be recovered from H:

$$\mathcal{K}_0 = \{(F, G) \in \mathcal{K} : a_G \le a_F, \quad b_G \le b_F\}.$$

Further, we note that Woodroofe (1985) demonstrated in his Lemma 1 that if we take any $(F, G) \in \mathcal{K}$ and let $F_0 = \Pr(X \le x \mid X \ge a_G)$ and $G_0 = \Pr(Y \le y \mid Y \le b_F)$, then $(F_0, G_0) \in \mathcal{K}_0$ and $H_*(F_0, G_0) = H_*(F, G)$. This subtle but important result implies that, if given H_*, we may not be able to recover the pair (F, G). This is because there is another pair, (F_0, G_0), that gives us exactly the same H_*. It is not surprising. For example, in the context of our motivating problem, we only observe X when it is equal or greater than $\Delta + 1$. Hence, it is impossible to get any information on the distribution of X over values less than $\Delta + 1$.

2.2 Recovery

Woodroofe (1985) shows in his Theorem 1 that if we restrict our construction of H_* to class $\mathcal{K}_0$, then this operation is "invertible". More specifically, for every H based on some $(F, G) \in \mathcal{K}$ there is only one pair $(F_0, G_0) \in \mathcal{K}_0$ such that $H_*(F_0, G_0) = H_*(F, G)$ and this pair is given by F_0 and G_0. Moreover, this theorem gives specific instructions on how to recover the cumulative hazard functions of F_0 and G_0 (and, therefore, F_0 and G_0 as well).

Once again, in the context of our example, we have $F_0(x) = \Pr(\Delta+1 \le X \le x)/\Pr(X \ge \Delta+1)$ (that is, the range of F_0 is $[\Delta + 1, L]$) and $G_0(y) = \Pr(Y \le y) = G(y)$ because $\Delta + M \le L$ by assumption. The range of G_0 is $[\Delta + 1, \Delta + M]$. Thus, from H_* based on the original F and G, we can recover G and only the F_0 portion of F.

We have discussed X and Y at length thus far, but we now do so with some additional precision. Specifically, let X and Y be independent integer-valued random variables with ranges $[\Delta + 1, L]$ and $[\Delta + 1, \Delta + M]$, respectively. We will assume that $\Pr(X = \Delta + 1)$, $\Pr(Y = \Delta + 1)$, $\Pr(X = L)$, and $\Pr(Y = \Delta + M)$ are strictly positive, and $\Delta + M \le L$. Let A be a set of points on the plane with integer-valued coordinates (u, v) such that $u \in [\Delta + 1, L]$, $v \in [\Delta + 1, \Delta + M]$, and $v \le u$. See Figure 2.

Let

$$f(u) = \Pr(X = u), \quad g(v) = \Pr(Y = v), \quad \text{and} \quad \alpha = \Pr(Y \le X).$$

The bivariate distribution function H_* over the trapezoid A has pmf

$$\begin{aligned} h_*(u, v) &= \Pr(X = u, Y = v \mid Y \le X) \\ &= \frac{\Pr(X = u, Y = v, Y \le X)}{\Pr(Y \le X)} \\ &= \frac{\Pr(X = u, Y = v)}{\Pr(Y \le X)} \\ &= \frac{\Pr(X = u)\Pr(Y = v)}{\Pr(Y \le X)} \\ &= \frac{f(u)g(v)}{\alpha}. \end{aligned} \tag{1}$$

This simple observation tells us that not every distribution over A can be a result of our truncation procedure. The marginal distributions of H_* are given by

$$f_*(u) = \Pr(X = u \mid Y \le X) = \sum_v h_*(u, v),$$

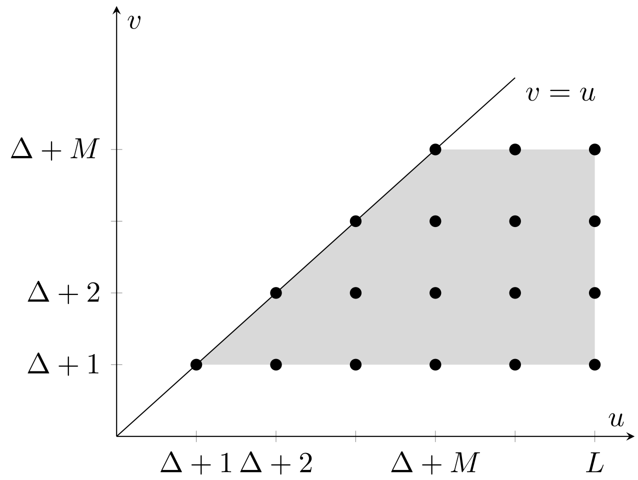


Figure 2: The set of points on the plane with integer-valued coordinates (u, v) such that $u \in [\Delta + 1, L]$, $v \in [\Delta + 1, \Delta + M]$, and $v \leq u$. The shaded region is the sample space of H_* and is denoted by trapezoid A. Since X and Y are discrete, all of the probability is contained in masses on the discrete points within the shaded region. Also note it is assumed that $\Pr(X = \Delta + 1)$, $\Pr(Y = \Delta + 1)$, $\Pr(X = L)$, and $\Pr(Y = \Delta + M)$ are strictly positive.

and

$$g_*(v) = \Pr(Y = v \mid Y \leq X) = \sum_u h_*(u, v).$$

For our forthcoming results to be meaningful, it must be possible to express the pmf f (or g) in terms of the pmf h_*. Notably, Woodroofe (1985) shows us by his Theorem 1 that we can indeed do so by expressing the cumulative hazard rate function in terms of joint cdf H_*. However, since we deal only with discrete random variables, it is more convenient to work with the hazard rate for X:

$$\lambda(x) = \frac{\Pr(X = x)}{\Pr(X \geq x)},$$

where $x \in \{\Delta + 1, \Delta + 2, \ldots, L\}$. One can show that

$$\lambda(x) = \frac{f_*(x)}{C(x)}, \tag{2}$$

where

$$C(x) = \Pr(Y \leq x \leq X \mid Y \leq X) = \sum_{v \leq x \leq u} h_*(u, v). \tag{3}$$

Indeed, first observe that

$$\begin{aligned} C(x) =& \Pr(Y \leq x \leq X \mid Y \leq X) = \frac{\Pr(Y \leq x \leq X, Y \leq X)}{\Pr(Y \leq X)} \\ =& \frac{1}{\alpha} \Pr(Y \leq x \leq X) = \frac{1}{\alpha}(\Pr(Y \leq x) - \Pr(X < x, Y \leq x)) \\ =& \frac{1}{\alpha}(\Pr(Y \leq x) - \Pr(X < x) \Pr(Y \leq x)) = \frac{1}{\alpha} \Pr(Y \leq x) \Pr(X \geq x). \end{aligned}$$

Hence,

$$\lambda(x) = \frac{\Pr(X = x)}{\Pr(X \geq x)} = \frac{\Pr(X = x) \Pr(Y \leq x)}{\Pr(X \geq x) \Pr(Y \leq x)}$$

$$
\begin{aligned}
=&\frac{\Pr(X=x,Y\le x)}{\Pr(X\ge x)\Pr(Y\le x)}=\frac{\Pr(X=x,Y\le X)}{\Pr(X\ge x)\Pr(Y\le x)}\\
=&\frac{\Pr(X=x,Y\le X)}{\Pr(Y\le X)}\frac{\Pr(Y\le X)}{\Pr(X\ge x)\Pr(Y\le x)}\\
=&f_*(x)\frac{\alpha}{\Pr(X\ge x)\Pr(Y\le x)}=\frac{f_*(x)}{C(x)}.
\end{aligned}
$$

Note that having $C(x)$ in the denominator is not a concern, because for any x

$$
C(x)\ge h_*(L,\Delta+1)=\frac{f(L)g(\Delta+1)}{\alpha}>0.
$$

The re-construction of the cdf F from the hazard rate λ is based on the following standard result of survival analysis. For any integer x such that $\Delta+1\le x\le L$, then

$$
\begin{aligned}
&\prod_{\Delta+1\le k<x}[1-\lambda(k)]\\
&=\left[1-\frac{\Pr(X=\Delta+1)}{\Pr(X\ge\Delta+1)}\right]\left[1-\frac{\Pr(X=\Delta+2)}{\Pr(X\ge\Delta+2)}\right]\cdots\left[1-\frac{\Pr(X=x-1)}{\Pr(X\ge x-1)}\right]\\
&=\left[\frac{\Pr(X\ge\Delta+2)}{\Pr(X\ge\Delta+1)}\right]\left[\frac{\Pr(X\ge\Delta+3)}{\Pr(X\ge\Delta+2)}\right]\cdots\left[\frac{\Pr(X\ge x)}{\Pr(X\ge x-1)}\right]\\
&=\Pr(X\ge x). \qquad (4)
\end{aligned}
$$

Since X is discrete, it is enough to know F at the jump points.

In a similar fashion, one can derive an analog of formula (2) for what is known as the *reverse hazard rate function* (for a nice introduction, see Block et al. (1998)). The reverse hazard rate is effectively analogous to the hazard rate in (2) but backwards-looking. That is, the reverse hazard rate is the probability of the event of interest occurring in the current interval, given we know the event of interest occurred prior to the current interval. Formally, the reverse hazard rate is defined as:

$$
\beta(y)=\frac{\Pr(Y=y)}{\Pr(Y\le y)}=\frac{g_*(y)}{C(y)}, \qquad (5)
$$

where $y\in\{\Delta+1,\Delta+2,\dots,\Delta+M\}$. As a consequence, we get the following formula for the cdf G:

$$
\Pr(Y\le y)=\prod_{\Delta+M\ge k>y}[1-\beta(k)],
$$

where $\Delta+1\le y\le\Delta+M$.

2.3 Estimators

There are different ways to think about truncation. For example, Woodroofe (1985) assumes that there is a population of Xs and Ys, from which we take a sample of size N. Then we apply truncation to the sample, and this gives a sample of truncated pairs of *random* sample size n. Our thinking, however, is different. We assume that there is the original population of Xs and Ys. We apply truncation to the entire population to get a population of truncated pairs. Then we extract a sample of *deterministic* size n from the truncated population. That is, our observations are directly from the distribution H_*.

Let $\{(X_i, Y_i)\}_{1\le i\le n}$ be independent identically distributed pairs of random variables with distribution H_*. That is, a sample from distribution H_*. Our main task is to provide interval estimates for the hazard rates of F_0 and the reverse hazard rates of G, and the cdfs F_0 and G.

Examination of (2) tells us that the hazard rate $\lambda(x)$ is a ratio of two probabilities of some events related to random variables (X_i, Y_i). We have natural estimates in terms of observed frequencies for these probabilities. This suggests the following estimate for the hazard rate:

$$\hat{\lambda}_n(x) = \frac{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i=x}}{C_n(x)}, \tag{6}$$

where

$$C_n(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{Y_j\le x\le X_j}, \tag{7}$$

and

$$\mathbf{1}_\xi = \begin{cases} 1, & \text{if the condition } \xi \text{ is true;} \\ 0, & \text{otherwise.} \end{cases}$$

By employing the same method of (4), we immediately get an estimate for the cdf F_0:

$$\hat{F}_n(x) = 1 - \prod_{\Delta+1\le k\le x} \left[1 - \hat{\lambda}_n(k)\right]. \tag{8}$$

It is interesting that one can demonstrate (e.g., Woodroofe, 1985; Keiding and Gill, 1990) that $\hat{F}_n$ in (8) is the NPMLE of F_0.

In similar fashion we can produce the maximum likelihood estimate (MLE) of the reverse hazard rate $\beta(y)$ of Y,

$$\hat{\beta}_n(y) = \frac{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{Y_i=y}}{C_n(y)}, \tag{9}$$

and the MLE of cdf G,

$$\hat{G}_n(y) = \prod_{\Delta+M\ge k>y} \left[1 - \hat{\beta}_n(k)\right]. \tag{10}$$

3 Asymptotic Results

We now establish asymptotic normality of the hazard rate and reverse hazard rate estimators, along with the unanticipated result of independence. We also prove asymptotic normality of the survival function estimator for X and distribution function estimator for Y. Finally, this section closes with a hypothesis test for the shape of G. Prior to the proofs, however, it is useful to carefully set the stage.

First, notice that throughout this section, as before, X and Y are positive discrete random variables with distribution functions F and G, respectively, and $\{(X_i, Y_i)\}_{1\le i\le n}$ are independent and identically distributed pairs of random variables with distribution H_*. More specifically, $\{(X_i, Y_i)\}_{1\le i\le n}$ are a random sample from a population with probability mass function h_* in Equation (1), spanning the finite set of points on the plane with integer-valued coordinates (u, v) such that $u \in [\Delta+1, L]$, $v \in [\Delta+1, \Delta+M]$, $\Delta+M \le L$, and $v \le u$ (trapezoid A). Additionally, we will continue to assume that $\Pr(X = \Delta + 1)$, $\Pr(Y = \Delta + 1)$, $\Pr(X = L)$, and $\Pr(Y = \Delta + M)$ are strictly positive. See Figure 2 as necessary.

Second, recall that we apply truncation to the entire population of X and Y, which yields a population of truncated pairs. From this truncated population, we draw a sample of deterministic size n. Therefore, in what follows we investigate the limiting behavior as $n \to \infty$.

Third and finally, to state our asymptotic results it is convenient to introduce the following notation. Where appropriate, we include summation representations to help build intuition about these probabilities. Let

$$\begin{aligned} c(u, v) &= \Pr(Y_i \le u \le X_i, Y_i \le v \le X_i) \\ &= \Pr(Y_i \le \min(u, v), X_i \ge \max(u, v)) \\ &= \sum_{y=\Delta+1}^{\min(u,v)} \sum_{x=\max(u,v)}^{L} h_*(x, y) \\ &= \Pr(Y \le \min(u, v), X \ge \max(u, v) \mid Y \le X) \\ &= \Pr(Y \le \min(u, v), X \ge \max(u, v), Y \le X) / \Pr(Y \le X) \\ &= \frac{1}{\alpha} \Pr(Y \le \min(u, v)) \Pr(X \ge \max(u, v)). \end{aligned} \tag{11}$$

Notice $c(z, z) = C(z)$ and $c(u, v) = c(v, u)$. Also, let

$$\begin{aligned} r(u, v) &= \Pr(X_i = \max(u, v), Y_i \le \min(u, v)) \\ &= \Pr(X = \max(u, v), Y \le \min(u, v) \mid Y \le X) \\ &= \sum_{y=\Delta+1}^{\min(u,v)} h(\max(u, v), y) \\ &= \Pr(X = \max(u, v), Y \le \min(u, v), Y \le X) / \Pr(Y \le X) \\ &= \frac{1}{\alpha} \Pr(X = \max(u, v)) \Pr(Y \le \min(u, v)). \end{aligned} \tag{12}$$

Notice $r(z, z) = f_*(z)$ and $r(u, v) = r(v, u)$.

3.1 Hazard and Reverse Hazard Rate

We first inspect the estimation of C_n in the denominator of the Woodroofe-type estimator (6) with the multivariate Central Limit Theorem (CLT).

Lemma 1 ($\hat{\mathbf{C}}_n$ Asymptotic Normality)**.** *Define* $\hat{\mathbf{C}}_n = (C_n(\Delta + 1), \dots, C_n(L))^\top$*. Then,*

$$\sqrt{n}(\hat{\mathbf{C}}_n - \mathbf{C}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{\Sigma}_c), \text{ as } n \to \infty,$$

where $\mathbf{C} = (C(\Delta + 1), \dots, C(L))^\top$ *and* $\mathbf{\Sigma}_c$ *is covariance matrix* $\|\sigma_{k',k}\|$ *such that*

$$\sigma_{k',k} = \begin{cases} C(k)[1 - C(k)], & k' = k \\ c(k', k) - C(k')C(k), & k' \ne k \end{cases},$$

for $k', k = \Delta + 1, \Delta + 2, \dots, L$.

Proof. Observe

$$\hat{\mathbf{C}}_n = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{Y_i \le \Delta+1 \le X_i} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{Y_i \le L \le X_i} \end{bmatrix} = \frac{1}{n}\sum_{i=1}^{n} \begin{bmatrix} Y_{\Delta+1(i)} \\ \vdots \\ Y_{L(i)} \end{bmatrix}, \tag{13}$$

where $Y_{k(i)}$, $\Delta + 1 \le k \le L$ are independent and identically distributed Bernoulli random variables with probability of success given by $\Pr(Y_i \le k \le X_i) = \Pr(Y \le k \le X \mid Y \le X) = C(k)$ for $k = \Delta + 1, \ldots, L$. Thus, $E[Y_{k(i)}] = C(k)$ and $Var[Y_{k(i)}] = C(k)(1 - C(k))$. Now, since

$$\mathbf{1}_{Y_i \le k' \le X_i} \mathbf{1}_{Y_i \le k \le X_i} = \mathbf{1}_{Y_i \le \min(k',k), X_i \ge \max(k'k)},$$

we have

$$E[Y_{k'(i)} Y_{k(i)}] = E[\mathbf{1}_{Y_i \le \min(k',k), X_i \ge \max(k'k)}] = c(k', k), \tag{14}$$

for $k', k = \Delta + 1, \ldots, L$. Thus,

$$\begin{aligned} Cov[Y_{k'(i)} Y_{k(i)}] &= E[Y_{k'(i)} Y_{k(i)}] - E[Y_{k'(i)}]E[Y_{k(i)}] \\ &= c(k', k) - C(k')C(k). \end{aligned}$$

Recall that (14) reduces to $C(k)$ when $k' = k$. The result then follows by Theorem 8.21 [Multivariate CLT] (pg. 61) of Lehmann and Casella (1998).

□

Corollary 3.0.1. *As* $n \to \infty$, $\hat{\mathbf{C}}_n \xrightarrow{\mathcal{P}} \mathbf{C}$.

Proof. Applying Theorem 8.2 [Weak Law of Large Numbers] (pg. 54-55) of Lehmann and Casella (1998) to (13) gives us the result.

□

The discrete nature of X and Y along with the finite sample space of trapezoid A simplifies the calculations considerably. The same is true for the Woodroofe-type estimator of the hazard rate λ, which we now examine.

Theorem 3.1 ($\hat{\mathbf{\Lambda}}_n$ Asymptotic Normality)**.** *Define* $\hat{\mathbf{\Lambda}}_n = (\hat{\lambda}_n(\Delta+1), \hat{\lambda}_n(\Delta+2), \ldots, \hat{\lambda}_n(L))^\top$. *Then,*

$$\sqrt{n}(\hat{\mathbf{\Lambda}}_n - \mathbf{\Lambda}) \xrightarrow{\mathcal{L}} N\big(\mathbf{0}, \mathbf{\Sigma}_f\big), \ \ \text{as } n \to \infty,$$

where $\mathbf{\Lambda} = (\lambda(\Delta+1), \lambda(\Delta+2), \ldots, \lambda(L))^\top$ *with* $\lambda(z) = f_*(z)/C(z)$ *and*

$$\mathbf{\Sigma}_f = \operatorname{diag}\bigg(\frac{f_*(\Delta+1)c(\Delta+1, \Delta+2)}{C(\Delta+1)^3}, \ldots, \frac{f_*(L-1)c(L-1, L)}{C(L-1)^3}, 0\bigg). \tag{15}$$

That is, the estimators $\hat{\Lambda}_n(\Delta+1)$, …, $\hat{\Lambda}_n(L)$ *are asymptotically normal and independent.*

Proof. Recall (6)–(7) and observe

$$\hat{\boldsymbol{\Lambda}}_n - \boldsymbol{\Lambda} = \begin{bmatrix} \hat{\lambda}_n(\Delta+1) \\ \vdots \\ \hat{\lambda}_n(L) \end{bmatrix} - \begin{bmatrix} \lambda(\Delta+1) \\ \vdots \\ \lambda(L) \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{1}_{X_i=\Delta+1}}{C_n(\Delta+1)} - \frac{f_*(\Delta+1)}{C(\Delta+1)} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{1}_{X_i=L}}{C_n(L)} - \frac{f_*(L)}{C(L)} \end{bmatrix}$$

$$= \mathbf{A}_n \times \frac{1}{n}\sum_{i=1}^{n} \begin{bmatrix} Z_{\Delta+1(i)} \\ \vdots \\ Z_{L(i)} \end{bmatrix},$$

where, for $\Delta + 1 \leq k \leq L$,

$$Z_{k(i)} = \mathbf{1}_{X_i=k}C(k) - \mathbf{1}_{Y_i \leq k \leq X_i} f_*(k),$$

and $\mathbf{A}_n = \text{diag}([C_n(\Delta+1)C(\Delta+1)]^{-1}, \ldots, [C_n(L)C(L)]^{-1})$. That is,

$$\hat{\boldsymbol{\Lambda}}_n - \boldsymbol{\Lambda} = \mathbf{A}_n \times \frac{1}{n}\sum_{i=1}^{n} \mathbf{Z}_{(i)},$$

where $\mathbf{Z}_{(i)} = (Z_{\Delta+1(i)}, \ldots, Z_{L(i)})^\top$, $1 \leq i \leq n$ are independent and identically distributed random vectors. We will also subsequently show that the components of random vector $\mathbf{Z}_{(i)}$ are uncorrelated.

More specifically, $\mathbf{1}_{X_i=x}$ is a Bernoulli random variable with probability of success $f_*(x)$ and, similarly, $\mathbf{1}_{Y_i \leq x \leq X_i}$ is a Bernoulli random variable with probability of success $C(x)$. Thus,

$$E[Z_{k(i)}] = f_*(k)C(k) - C(k)f_*(k) = 0.$$

Therefore,

$$\begin{aligned} & Cov[Z_{k(i)}Z_{k'(i)}] \qquad (16) \\ &= E\left[\left(\mathbf{1}_{X_i=k}C(k) - \mathbf{1}_{Y_i \leq k \leq X_i} f_*(k)\right)\left(\mathbf{1}_{X_i=k'}C(k') - \mathbf{1}_{Y_i \leq k' \leq X_i} f_*(k')\right)\right] \\ &= C(k)C(k')E[\mathbf{1}_{X_i=k}\mathbf{1}_{X_i=k'}] - f_*(k)C(k')E[\mathbf{1}_{X_i=k'}\mathbf{1}_{Y_i \leq k \leq X_i}] \\ &\quad - C(k)f_*(k')E[\mathbf{1}_{X_i=k}\mathbf{1}_{Y_i \leq k' \leq X_i}] + f_*(k)f_*(k')E[\mathbf{1}_{Y_i \leq k \leq X_i}\mathbf{1}_{Y_i \leq k' \leq X_i}]. \qquad (17) \end{aligned}$$

We proceed to calculate $Cov[Z_{k(i)}Z_{k'(i)}]$ by cases.

Case 1: $k = k'$.

Notice $\mathbf{1}_{X_i=k}\mathbf{1}_{X_i=k'} = \mathbf{1}_{X_i=k}$ and $E[\mathbf{1}_{X_i=k}\mathbf{1}_{X_i=k'}] = f_*(k)$. Further,

$$\mathbf{1}_{X_i=k'}\mathbf{1}_{Y_i \leq k \leq X_i} = \mathbf{1}_{X_i=k, Y_i \leq k \leq X_i} = \mathbf{1}_{X_i=k}.$$

Hence, $E[\mathbf{1}_{X_i=k}\mathbf{1}_{Y_i \leq k' \leq X_i}] = f_*(k)$. Also note that

$$\mathbf{1}_{Y_i \leq k \leq X_i}\mathbf{1}_{Y_i \leq k' \leq X_i} = \mathbf{1}_{Y_i \leq k \leq X_i},$$

and thus $E[\mathbf{1}_{Y_i \leq k \leq X_i}] = C(k)$. Replacing the expectations in (17) yields

$$Cov[Z_{k(i)}Z_{k'(i)}] = C(k)C(k')f_*(k) - f_*(k)C(k')f_*(k)$$

$$
\begin{aligned}
&- C(k)f_*(k')f_*(k) + f_*(k)f_*(k')C(k) \\
&= C(k)^2 f_*(k) - 2f_*(k)^2 C(k) + f_*(k)^2 C(k) \\
&= f_*(k)C(k)[C(k) - f_*(k)]. \qquad (18)
\end{aligned}
$$

However,

$$
\begin{aligned}
C(k) - f_*(k) &= \sum_{y=\Delta+1}^{k} \sum_{x=k}^{L} h_*(x,y) - \sum_{y=\Delta+1}^{k} h_*(k,y) \\
&= \sum_{y=\Delta+1}^{k} \sum_{x=k+1}^{L} h_*(x,y) \\
&= c(k,k+1). \qquad (19)
\end{aligned}
$$

Replacing (19) in (18) and simplifying yields the diagonal matrix

$$
\mathbf{D} = \operatorname{diag}\big(f_*(\Delta+1)C(\Delta+1)c(\Delta+1,\Delta+2),\ldots,f_*(L)C(L)c(L,L+1)\big).
$$

We emphasize here that $c(L,L+1) = 0$.

Case 2: $k \neq k'$.

Certainly, $\mathbf{1}_{X_i=k}\mathbf{1}_{X_i=k'} = 0$ when $k \neq k'$. Therefore,

$$
E[\mathbf{1}_{X_i=k}\mathbf{1}_{X_i=k'}] = 0. \qquad (20)
$$

Assume $k < k'$ and notice $\mathbf{1}_{X_i=k'}\mathbf{1}_{Y_i\le k\le X_i} = \mathbf{1}_{X_i=k',Y_i\le k\le X_i}$. Thus, $E[\mathbf{1}_{X_i=k'}\mathbf{1}_{Y_i\le k\le X_i}] = r(k',k)$. On the other hand, $\mathbf{1}_{X_i=k}\mathbf{1}_{Y_i\le k'\le X_i} = \mathbf{1}_{X_i=k,Y_i\le k'\le X_i} = 0$ because $\{X_i = k \cap k' \le X_i\} = \emptyset$ when $k < k'$. Now observe the symmetry between $\mathbf{1}_{X_i=k}\mathbf{1}_{Y_i\le k'\le X_i}$ and $\mathbf{1}_{(X_i=k'}\mathbf{1}_{Y_i\le k\le X_i}$ to drop the assumption $k < k'$ and more generally claim

$$
\begin{aligned}
&- f_*(k)C(k')E[\mathbf{1}_{X_i=k'}\mathbf{1}_{Y_i\le k\le X_i}] - C(k)f_*(k')E[\mathbf{1}_{X_i=k}\mathbf{1}_{Y_i\le k'\le X_i}] \\
= \;&- r(k,k')f_*(\min(k,k'))C(\max(k,k')). \qquad (21)
\end{aligned}
$$

Further, $\mathbf{1}_{Y_i\le k\le X_i}\mathbf{1}_{Y_i\le k'\le X_i} = \mathbf{1}_{Y_i\le k\le X_i,Y_i\le k'\le X_i} = \mathbf{1}_{Y_i\le\min(k,k'),X_i\ge\max(k,k')}$. Hence,

$$
E[\mathbf{1}_{Y_i\le k\le X_i}\mathbf{1}_{Y_i\le k'\le X_i}] = c(k,k'). \qquad (22)
$$

Replacing the expectations (20), (21), and (22) in (17) and simplifying yields

$$
E[Z_{k(i)}Z_{k'(i)}] = f_*(\min(k,k')) \times \{f_*(\max(k,k'))c(k,k') - r(k,k')C(\max(k,k'))\}.
$$

But,

$$
\begin{aligned}
&f_*(\max(k,k'))c(k,k') \\
&= \frac{\Pr(X=\max(k,k'),Y\le X)}{\alpha}\,\frac{\Pr(Y\le\min(k,k'))\Pr(X\ge\max(k,k'))}{\alpha} \\
&= \frac{\Pr(X=\max(k,k'))\Pr(Y\le\max(k,k'))}{\alpha}\,\frac{\Pr(Y\le\min(k,k'))\Pr(X\ge\max(k,k'))}{\alpha} \\
&= \frac{\Pr(X=\max(k,k'))\Pr(Y\le\min(k,k'))}{\alpha}\,\frac{\Pr(Y\le\max(k,k'))\Pr(X\ge\max(k,k'))}{\alpha} \\
&= r(k,k')C(\max(k,k')),
\end{aligned}
$$

and so (17) is zero whenever $k \neq k'$. Now define

$$\bar{\mathbf{Z}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_{(i)},$$

and use Theorem 8.21 [Multivariate CLT] (pg. 61) of Lehmann and Casella (1998) to claim

$$\sqrt{n}[\bar{\mathbf{Z}}_n - \mathbf{0}] \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{D}), \text{ as } n \to \infty.$$

Further note by Corollary 3.0.1,

$$\mathbf{A}_n \xrightarrow{\mathcal{P}} \mathbf{V}, \text{ as } n \to \infty$$

where $\mathbf{V} = \operatorname{diag}(C(\Delta+1)^{-2}, \ldots, C(L)^{-2})$. Therefore, by Theorem 5.1.6 [Multivariate Slutsky's Theorem] (pg. 283) of Lehman (1998),

$$\sqrt{n}[\mathbf{A}_n\bar{\mathbf{Z}}_n] \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{V}\mathbf{D}\mathbf{V}^\top), \text{ as } n \to \infty.$$

Finally, observe $\mathbf{V}\mathbf{D}\mathbf{V}^\top = \mathbf{\Sigma}_f$ and $\mathbf{A}_n\bar{\mathbf{Z}}_n = \hat{\mathbf{\Lambda}}_n - \mathbf{\Lambda}$ to complete the proof.

□

Remark. *There is an alternative form of $\mathbf{\Sigma}_f$ that may be preferable. Observe for $x \in [\Delta+1, \ldots, L]$,*

$$\begin{aligned}
\frac{f_*(x)c(x,x+1)}{C(x)^3} &= \frac{f_*(x)}{C(x)}\frac{c(x,x+1)}{C(x)^2} = \frac{f_*(x)}{C(x)}\frac{\alpha^{-1}\Pr(Y \le x)\Pr(X \ge x+1)}{[\alpha^{-1}\Pr(Y \le x)\Pr(X \ge x)]^2} \\
&= \frac{f_*(x)}{C(x)}\frac{\alpha^{-1}\Pr(Y \le x)}{\alpha^{-1}\Pr(Y \le x)}\frac{\Pr(X \ge x+1)}{\alpha^{-1}\Pr(Y \le x)\Pr(X \ge x)^2} \\
&= \frac{f_*(x)}{C(x)}\frac{\Pr(X \ge x+1)}{\Pr(X \ge x)}\frac{1}{\alpha^{-1}\Pr(Y \le x)\Pr(X \ge x)} \\
&= \frac{f_*(x)}{C(x)}\frac{\Pr(X \ge x) - \Pr(X = x)}{\Pr(X \ge x)}\frac{1}{\alpha^{-1}\Pr(Y \le x)\Pr(X \ge x)} \\
&= \lambda(x)[1-\lambda(x)]\frac{1}{C(x)} = \frac{\lambda(x)^2[1-\lambda(x)]}{f_*(x)}.
\end{aligned}$$

Hence, we also claim

$$\mathbf{\Sigma}_f = \operatorname{diag}\left(\frac{\lambda(\Delta+1)^2[1-\lambda(\Delta+1)]}{f_*(\Delta+1)}, \ldots, \frac{\lambda(L-1)^2[1-\lambda(L-1)]}{f_*(L-1)}, 0\right). \tag{23}$$

Further, (15) and (23) are equivalent when the true quantities are replaced by their NPMLEs. That is, for $x \in [\Delta+1, \ldots, L]$ with

$$\hat{f}_{*,n}(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{X_i = x}, \quad \text{and} \quad \hat{c}_n(x,x+1) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{Y_i \le x, X_i \ge x+1},$$

it is easy to show

$$\frac{\hat{f}_{*,n}(x)\hat{c}_n(x,x+1)}{C_n(x)^3} = \frac{\hat{\lambda}_n(x)^2[1-\hat{\lambda}_n(x)]}{\hat{f}_{*,n}(x)}.$$

We now state the following corollary without proof for completeness.

Corollary 3.1.1. *As* $n \to \infty$, $\hat{\mathbf{\Lambda}}_n \xrightarrow{\mathcal{P}} \mathbf{\Lambda}$.

When estimating G is of interest, we may also obtain the sister statement for the reverse hazard rate β as follows.

Theorem 3.2 ($\hat{\mathbf{B}}_n$ Asymptotic Normality)**.** *Define* $\hat{\mathbf{B}}_n = (\hat{\beta}_n(\Delta+1), \hat{\beta}_n(\Delta+2), \ldots, \hat{\beta}_n(\Delta+M))^\top$. *Then,*

$$\sqrt{n}(\hat{\mathbf{B}}_n - \mathbf{B}) \xrightarrow{\mathcal{L}} N\big(\mathbf{0}, \mathbf{\Sigma}_g\big), \text{ as } n \to \infty,$$

where $\mathbf{B} = (\beta(\Delta+1), \beta(\Delta+2), \ldots, \beta(L))^\top$ *with* $\beta(z) = g_*(z)/C(z)$ *and*

$$\mathbf{\Sigma}_g = \operatorname{diag}\bigg(0, \frac{g_*(\Delta+2)c(\Delta+1, \Delta+2)}{C(\Delta+2)^3}, \ldots, \frac{g_*(\Delta+M)c(\Delta+M-1, \Delta+M)}{C(\Delta+M)^3}\bigg).$$

That is, the estimators $\hat{\beta}_n(\Delta+1), \ldots, \hat{\beta}_n(\Delta+M)$ *are asymptotically normal and independent.*

Proof. See the proof of Theorem 3.1, substituting g_* for f_* and adjusting the indicator logic as appropriate. It is useful to introduce similar notation to (12). That is,

$$\begin{aligned} s(u,v) &= \Pr(Y_i = \min(u,v), X_i \geq \max(u,v)) \\ &= \frac{1}{\alpha}\Pr(Y = \min(u,v))\Pr(X \geq \max(u,v)). \end{aligned} \tag{24}$$

□

Remark. *One may also write*

$$\mathbf{\Sigma}_g = \operatorname{diag}\bigg(0, \frac{\beta(\Delta+2)^2[1-\beta(\Delta+2)]}{g_*(\Delta+2)}, \ldots, \frac{\beta(\Delta+M)^2[1-\beta(\Delta+M)]}{g_*(\Delta+M)}\bigg). \tag{25}$$

We again state the following corollary without proof for completeness.

Corollary 3.2.1. *As* $n \to \infty$, $\hat{\mathbf{B}}_n \xrightarrow{\mathcal{P}} \mathbf{B}$.

3.2 Survival and Distribution Function

For most analysts of survival data, the key quantity of interest is the survival function, $S(x) = 1 - F(x)$. From (6) and (8), we have the estimator

$$\hat{S}_n(x) = \prod_{\Delta+1 \leq k \leq x} [1 - \hat{\lambda}_n(k)]. \tag{26}$$

Asymptotic normality also extends to (26), which we now show.

Theorem 3.3 ($\hat{\mathbf{S}}_n$ Asymptotic Normality)**.** *For the estimator* $\hat{\mathbf{S}}_n = (\hat{S}_n(\Delta+1), \hat{S}_n(\Delta+2), \ldots, \hat{S}_n(L))^\top$,

$$\sqrt{n}(\hat{\mathbf{S}}_n - \mathbf{S}) \xrightarrow{\mathcal{L}} N\big(0, \mathbf{RK\Sigma}_f[\mathbf{RK}]^\top\big), \text{ as } n \to \infty,$$

where $\mathbf{S} = (S(\Delta+1), S(\Delta+2), \ldots, S(L))^\top$, $\mathbf{\Sigma}_f$ *follows from Theorem 3.1,*

$$\mathbf{K} = \begin{bmatrix} -[1-\lambda(\Delta+1)]^{-1} & 0 & \ldots & 0 \\ -[1-\lambda(\Delta+1)]^{-1} & -[1-\lambda(\Delta+2)]^{-1} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -[1-\lambda(\Delta+1)]^{-1} & -[1-\lambda(\Delta+2)]^{-1} & \ldots & -[1-\lambda(L)]^{-1} \end{bmatrix},$$

and $\mathbf{R} = \operatorname{diag}(S(\Delta+1), S(\Delta+2), \ldots, S(L))$.

Proof. To motivate the demonstration, let $x \in [\Delta+1, \dots, L]$ and recall (4) to write,

$$S(x) = \prod_{z=\Delta+1}^{x} [1 - \lambda(z)].$$

Now consider the natural log,

$$\ln S(x) = \sum_{z=\Delta+1}^{x} \ln[1 - \lambda(z)].$$

Hence,

$$\begin{aligned}
\sqrt{n}[\ln S_n(x) - \ln S(x)] &= \sqrt{n}\Bigg[\sum_{z=\Delta+1}^{x} \ln\left(\frac{1-\lambda_n(z)}{1-\lambda(z)}\right)\Bigg] \\
&= \sqrt{n}\Bigg[\sum_{z=\Delta+1}^{x} \ln\left(1 + \frac{\lambda(z)-\lambda_n(z)}{1-\lambda(z)}\right)\Bigg].
\end{aligned}$$

But $\ln(1+x) = \sum_{n\geq 1}(-1)^{n+1}x^n/n$ and so

$$\begin{aligned}
\sqrt{n}[\ln S_n(x) - \ln S(x)] &= \sqrt{n}\Bigg[\sum_{z=\Delta+1}^{x} \left\{\frac{\lambda(z)-\lambda_n(z)}{1-\lambda(z)} - \frac{1}{2}\left[\frac{(\lambda(z)-\lambda_n(z))^2}{(1-\lambda(z))^2}\right] + \cdots\right\}\Bigg] \\
&= \sqrt{n}\Bigg[\sum_{z=\Delta+1}^{x} \frac{\lambda(z)-\lambda_n(z)}{1-\lambda(z)} + O_p(|\lambda(z)-\lambda_n(z)|^2)\Bigg] \\
&= \sqrt{n}\left(-\sum_{z=\Delta+1}^{x} \frac{\lambda_n(z)-\lambda(z)}{1-\lambda(z)}\right) + o_p(1), \qquad (27)
\end{aligned}$$

where (27) follows by Corollary 3.1.1 and Theorem 8.10 [Slutsky's Theorem] (pg. 58) of Lehmann and Casella (1998). Now consider all $x \in [\Delta+1, \dots, L]$ to write,

$$\sqrt{n}\begin{bmatrix} \{\ln S_n(\Delta+1) - \ln S(\Delta+1)\} \\ \vdots \\ \{\ln S_n(L) - \ln S(L)\} \end{bmatrix} = \mathbf{K} \times \sqrt{n}(\hat{\mathbf{\Lambda}}_n - \mathbf{\Lambda}) + o_p(1).$$

Thus, by Theorem 3.1 and Theorem 5.1.6 [Multivariate Slutsky's Theorem] (pg. 283) of Lehman (1998)

$$\mathbf{D} \times \sqrt{n}(\hat{\mathbf{\Lambda}}_n - \mathbf{\Lambda}) + o_p(1) \xrightarrow{\mathcal{L}} N(0, \mathbf{K}\mathbf{\Sigma}_f\mathbf{K}^\top), \text{ as } n \to \infty.$$

Finally, note $S(x) = \exp\{\ln S(x)\}$ and apply the Theorem 8.22 [Multivariate Delta Method] (pg. 61) of Lehmann and Casella (1998) to complete the proof.

□

The sister theorem for analysts working instead to estimate G is as follows.

Theorem 3.4 ($\hat{\mathbf{G}}_n$ Asymptotic Normality)**.** *For the estimator* $\hat{\mathbf{G}}_n = (\hat{G}_n(\Delta+1), \hat{G}_n(\Delta+2), \dots, \hat{G}_n(\Delta+M))^\top$

$$\sqrt{n}(\hat{\mathbf{G}}_n - \mathbf{G}) \xrightarrow{\mathcal{L}} N\big(0, \mathbf{W}\mathbf{M}\mathbf{\Sigma}_g[\mathbf{W}\mathbf{M}]^\top\big), \text{ as } n \to \infty,$$

where $\mathbf{G} = (G(\Delta+1), G(\Delta+2), \ldots, G(\Delta+M))^\top$*,* $\boldsymbol{\Sigma}_g$ *follows from Theorem 3.2,*

$$\mathbf{M} = \begin{bmatrix} -[1-\beta(\Delta+1)]^{-1} & -[1-\beta(\Delta+2)]^{-1} & \ldots & -[1-\beta(\Delta+M)]^{-1} \\ 0 & -[1-\beta(\Delta+2)]^{-1} & \ldots & -[1-\beta(\Delta+M)]^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & -[1-\beta(M+\Delta)]^{-1} \end{bmatrix},$$

and $\mathbf{W} = \text{diag}(G(\Delta+1), G(\Delta+2), \ldots, G(\Delta+M))$*.*

Proof. Recall (10) and see the proof of Theorem 3.1.

□

3.3 Hypothesis Test

In many applications it is desirable to test if the distribution of F or G corresponds to a known distribution. As one may anticipate, there is some history in the literature. For a starting point, we encourage the reader to review Hyde (1977) and the associated citations. For our purposes, we review a few notable examples. To begin, Guilbaud (1988) generalizes the ordinary Kolmogorov-Smirov one-sample tests based on the product-limit estimator. The test we develop from Theorem 3.5 is more akin to a goodness-of-fit test, however. Mandel and Betensky (2007) is related, though they assume continuous F and G to introduce several goodness-of-fit tests for the truncation distribution. Similarly, Hwang and Wang (2008) assume the lifetime, truncation, and censoring random variables are continuous in proposing a chi-square test to test the hypothesis that the truncation distribution follows a parametric family. Further, the asymptotic properties of the nonparametric test of Ning et al. (2010) were derived assuming a continuous survival function. See also Moreira et al. (2014), in which goodness-of-fit tests are proposed for a semiparametric model under random double truncation. As there is no clear application to discrete F or discrete G, we extend our results to propose a simple hypothesis testing procedure using a chi-square random variable. We state our results for the truncation distribution G.

Theorem 3.5. *Assume that* G *follows a known distribution over the discrete points* $[\Delta+1, \ldots, \Delta+M]$*. Then the test statistic*

$$\mathbb{Q}_G = [\sqrt{n}(\hat{\mathbf{B}}_n^* - \mathbf{B}^*)]^\top [\boldsymbol{\Sigma} g^*]^{-1} [\sqrt{n}(\hat{\mathbf{B}}_n^* - \mathbf{B}^*)] \xrightarrow{\mathcal{L}} \chi_q^2,$$

where $\hat{\mathbf{B}}_n^* = \big(\hat{\beta}_n(\Delta+2), \ldots, \hat{\beta}_n(\Delta+M)\big)^\top$*,* $\mathbf{B}^* = \big(\beta(\Delta+2), \ldots, \beta(\Delta+M)\big)^\top$*,*

$$\boldsymbol{\Sigma}_g^* = \text{diag}\left(\frac{\beta(\Delta+2)^2[1-\beta(\Delta+2)]}{g_*(\Delta+2)}, \ldots, \frac{\beta(\Delta+M)^2[1-\beta(\Delta+M)]}{g_*(\Delta+M)}\right),$$

and $q = \mathbf{card}\{\Delta+2, \ldots, \Delta+M\}$*. We emphasize here that the point* $\Delta+1$ *with the degenerate estimator* $\hat{\beta}_n(\Delta+1) = 1$ *and* $Var[\hat{\beta}_n(\Delta+1)] = 0$ *is omitted from* $\mathbb{Q}_G$*.*

Proof. Begin with Theorem 3.2 along with (25) and use the well-known multivariate normal Results 5.2.8 (all subsets of multivariate normal random vectors themselves have normal distributions) and 5.3.3 (a centered and scaled quadratic form of a p dimensional multivariate normal random vector is a chi-squared random variable with p degrees of freedom) from Ravishanker and Dey (2002). The result then follows by Corollary 8.11 [Continuous Mapping Theorem] (pg. 58) of Lehmann and Casella (1998).

□

Specifically, it is often of interest to test if G follows a uniform distribution. This is an important assumption in the case length-biased sampling, see for instance Asgharian et al. (2002), and De Uña-Álvarez (2004). There is a long history of proposed methods for checking this assumption in the literature. For example, one stated use of the NPMLE for the truncated distribution derived by Wang (1991) is to informally check the validity of the stationarity assumption (i.e., G coincides with a uniform distribution). Similarly, Asgharian et al. (2006) propose a graphical method to check the stationarity of the underlying incidence times. Addona and Wolfson (2006) propose a formal test for stationarity of the incidence rate, but they require a continuous truncation density (via Theorem 1 of Asgharian et al. (2006)). Our test, however, allows for exact p-value calculations and considers discrete G.

Corollary 3.5.1. *Assuming the conditions and notation of Theorem 3.5, under the null hypothesis that G is a uniform distribution over $[\Delta+1,\ldots,\Delta+M]$, the test statistic*

$$\mathbb{Q}_U = \big|[\sqrt{n}(\hat{\mathbf{B}}_n^* - \mathbf{B}_U^*)]^\top [\boldsymbol{\Sigma}_{g,U}^*]^{-1} [\sqrt{n}(\hat{\mathbf{B}}_n^* - \mathbf{B}_U^*)] \xrightarrow{\mathcal{L}} \chi_q^2,$$

where $\mathbf{B}_U^* = (1/2, 1/3, \ldots, 1/M)$, and

$$\boldsymbol{\Sigma}_{g,U}^* = \operatorname{diag}\left(\frac{[1/2]^2[1-1/2]}{\hat{g}_{*,n}(\Delta+2)}, \ldots, \frac{[1/M]^2[1-1/M]}{\hat{g}_{*,n}(\Delta+M)}\right).$$

Proof. By Theorem 8.2 [Weak Law of Large Numbers] (pg. 54-55) of Lehmann and Casella (1998), $\hat{g}_{*,n} \xrightarrow{\mathcal{P}} g_*$. Further, if G is uniform over the discrete set of points $[\Delta+1,\ldots,\Delta+M]$, then for $y \in [\Delta+1,\ldots,\Delta+M]$

$$\beta(y) = \frac{\Pr(Y=y)}{\Pr(Y \le y)} = \frac{1}{M}\frac{M}{y-(\Delta+1)+1} = \frac{1}{y-\Delta}.$$

Finally, use the results of Theorem 3.5 substituting $\beta(y)$ for $y \in \{\Delta+2,\ldots,\Delta+M\}$ as appropriate along with Theorem 5.1.6 [Multivariate Slutsky's Theorem] (pg. 283) of Lehman (1998) to complete the proof.

□

Consequently, for H_0 that Y is uniformly distributed and significance level $0 \le \alpha \le 1$, one can reject H_0 if $\mathbb{Q}_U \le \chi^2_{q,\alpha/2}$ or $\mathbb{Q}_U \ge \chi^2_{q,1-\alpha/2}$, where $\chi^2_{q,\theta}$ is the $(100\times\theta)$th $(0<\theta<1)$ percentile of a chi-square distribution with q degrees of freedom.

The accuracy of the asymptotic chi-square distribution was investigated for a uniform G in our simulation study. The empirical distribution of the test statistics matches closely to the limiting chi-square distribution, which we validated down to a sample size of $n = 500$.

4 Simulation Study

In this section, we examine the finite sample behavior of the $\hat{\lambda}_n(x)$ estimator for $x \in [\Delta+1,\ldots,L]$ and $\hat{\beta}_n(y)$ for $y \in [\Delta+1,\ldots,\Delta+M]$. In addition to serving as an experimental verification of Theorem 3.1, our intention is to help practitioners evaluate an appropriate sample size of truncated data to achieve a desired level of estimation accuracy. In particular, we assume that Y follows a discrete uniform distribution over $\mathcal{Y} = [1,2,\ldots,10]$, and that X follows a truncated geometric distribution over $\mathcal{X} = [1,2,\ldots,24]$. The pmf of X is

$$\Pr(X=x) = \begin{cases} p(1-p)^{x-1}, & x = 1,2,\ldots,23, \\ \sum_{x=24}^{\infty} p(1-p)^{x-1}, & x = 24, \\ 0, & \text{otherwise}, \end{cases} \tag{28}$$

where $0 < p < 1$. From this, we may calculate many key quantities of interest. For example, with $p = 0.20$:

$$\alpha = \Pr(Y \leq X) = \sum_{y=1}^{10} \Pr(Y = y) \Pr(X \geq y) = 0.4463,$$

as well as the useful quantities (3), (11), (12), and (24). Notice here that $\Delta = 0$, $M = 10$, and $L = 24$. All results in this section used 1,000 replicates.

We remark here on the behavior of (15) across various values of p. For larger values of p, the variance of $\hat{\lambda}_n$ for values of X closer to 23 quickly explodes. This is not unexpected because, as p increases, it becomes more and more unlikely to observe large values of X. On the other hand, for very small values of p close to zero, the variance of $\hat{\lambda}_n$ for values of X close to 1 is the largest and rapidly decreases until $X = 10$, the truncation point. This suggests what we can already glean from a careful examination of (23): estimation accuracy of $\hat{\lambda}_n(x)$ is dependent on the quantities $\lambda(x)$ and $f_*(x)$.

We also compare the empirical covariance against the asymptotic covariance suggested by Theorems 3.1 and 3.2 by way of examining the resulting confidence interval estimates. As is standard practice, we construct the confidence interval estimates on a log scale with the delta method and then transform them back to the original scale. Specifically, the 95% confidence intervals for $\lambda(x)$, $x \in [1, \ldots, 24]$, are

$$\exp\left\{ \ln \hat{\lambda}_n(x) \pm 1.96 \times \sqrt{\frac{1 - \hat{\lambda}_n(x)}{\hat{f}_{*,n}(x) \times n}} \right\}, \tag{29}$$

where

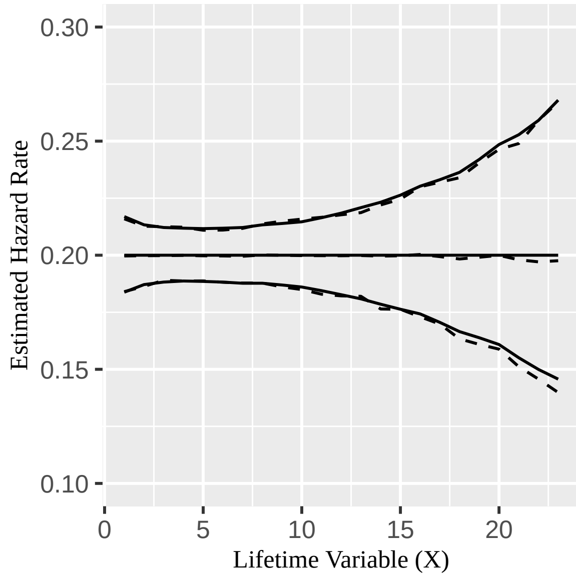
$$\hat{f}_{*,n}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{x_i = x}. \tag{30}$$

Throughout this section, we assume $p = 0.20$. In our analysis, we compared two quantities. The first quantity is the average of the estimated confidence intervals derived from the simulated data, i.e., (29), over the 1,000 replicates. We represent this quantity by the solid lines in Figure 3. To validate the variance in Theorems 3.1 and 3.2 we also report the 2.5th and 97.5th empirical quantiles of the 1,000 replicates of the estimated quantity $\hat{\lambda}_n(x)$, for $x \in [1, \ldots, 24]$. This is the second quantity and is represented by a dashed line in Figure 3. Naturally, if the asymptotic relationship is appropriate, the expected confidence interval should match closely to the empirical confidence interval for large n. We see that this is indeed the case in examining Figure 3. Similar results are also available for $\hat{\beta}_n$. Readers interested in the full covariance matrix comparison may contact the corresponding author for more details, though we note here that the empirical off-diagonal elements are indeed very close to zero, validating independence.

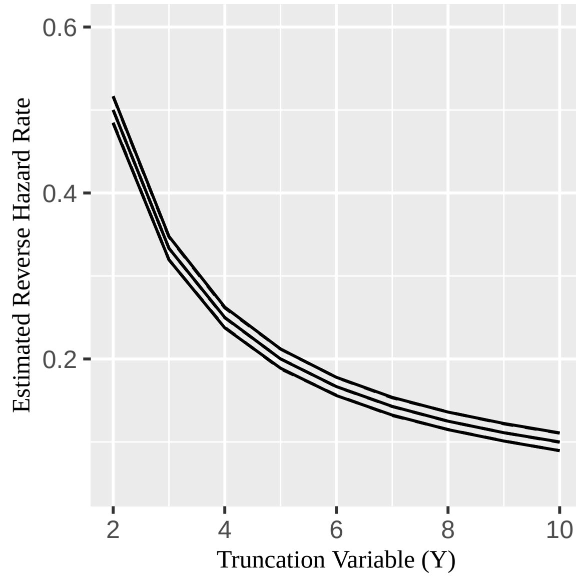
4.1 Estimation Accuracy

We next explore the approximation accuracy across various sample sizes. It is interesting to observe that the approximation accuracy is a function of the underlying distribution. This is again clear from (23); the variance of the estimator $\hat{\lambda}_n(x)$ is a function of the distributions of X and Y. Hence, we see that a larger sample size is necessary to control the approximation accuracy towards the right tail of the distribution of X, values of which occur with much smaller probability. We see tail failures of the approximation begin to materialize when n is as large as 1,000. On the other hand, the approximation for $\hat{\beta}_n(x)$ still works well for $n = 1{,}000$. Again, see Figure 3.

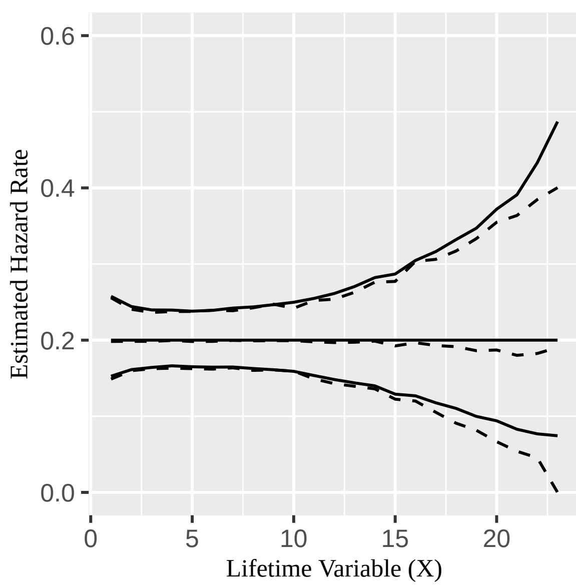
Finally, Table 1 summarizes the observed coverage probability over the 1,000 replicates for various sample sizes of n. That is, the percentage of the 1,000 replicates of confidence intervals



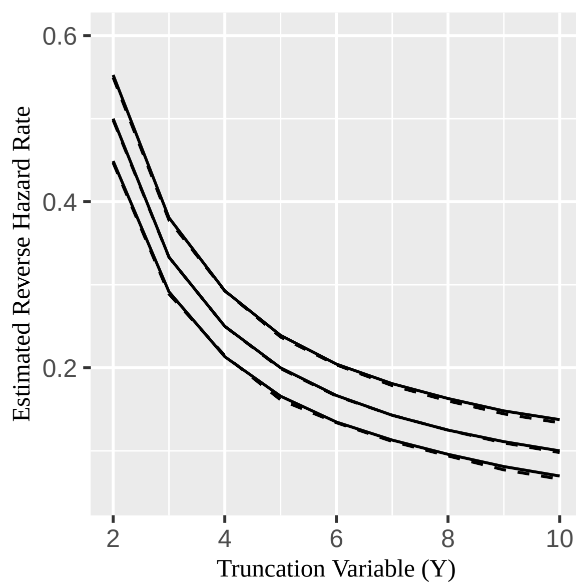
(a) $\hat{\lambda}_n(x)$ for $n = 10{,}000$



(b) $\hat{\beta}_n(y)$ for $n = 10{,}000$



(c) $\hat{\lambda}_n(x)$ for $n = 1{,}000$



(d) $\hat{\beta}_n(y)$ for $n = 1{,}000$

Figure 3: X follows a truncated geometric distribution at $x = 24$, as defined in (28). Y follows a discrete uniform distribution for the integers $\{1, 2, \ldots, 10\}$. The horizontal axis is the range of $X : [1, 2, \ldots, 24]$ (a), (c) or $Y : [1, 2, \ldots, 10]$ (b), (d). The solid line is the estimated confidence interval, as defined in (29). The dashed line is the empirical 2.5 and 97.5 quantiles based on the 1,000 estimates. For completeness, we also show the true values of $\lambda(x)$ and $\beta(y)$ as defined in (2) and (5), respectively, (solid line) against their estimates, $\hat{\lambda}_n(x)$ and $\hat{\beta}_n(y)$, defined in (6) and (9), respectively (dashed line).

that contained the true value of $\lambda(x)$ and $\beta(y)$. We also track the number of replicates that did not return a valid estimate (i.e., we did not observe any samples of X or Y at a particular value). Given these results, we recommend that a practitioner use judgment and available references to estimate the probability of less frequent observations. The smaller these probabilities, generally speaking, the larger the sample to ensure the approximation works well. Alternatively, a practitioner may instead identify the values of X or Y that are of most interest. For example, the confidence interval approximation for $\hat{\lambda}_n$ still works very well for $X \leq 10$ when $n = 1,000$. Once again, more details may be found in Table 1. Alternatively, if a known accurate estimate of $f_*(x)$ and $\lambda(x)$ is available, then determining the appropriate sample size is only a matter of selecting an approach to handle a simultaneous confidence region.

5 Application

Recall the motivating example in the Introduction. Here we apply the estimation and asymptotic results of earlier sections to auto lease securitization trust data. Specifically, we examine the Mercedes–Benz Auto Lease Trust (MBALT) 2017-A financial transaction. Detailed data and performance records are available at the individualized contract level from the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system, which is freely available to the public through the Securities and Exchange Commission. The MBALT 2017-A transaction had 56,402 lease contracts with original terms ranging from 24 to 60 months. For simplicity, we only considered ongoing lease contracts with an original termination schedule of 24 months. This reduced the sample to 866 lease contracts.

The MBALT 2017-A bond was placed in April of 2017. The transaction was paid in full and closed in August of 2019. Therefore, the observation window consisted of 28 months. Monthly loan performance information is available on EDGAR. Lease contracts must be delinquent no more than 30 days to be included in the securitization trust Mercedes-Benz (2017). Hence, the lease contracts are all active as of the onset of the transaction. At initialization, the oldest lease in the trust was 21 months old, and the youngest lease was 3 months old. Thus, to use our notation, $\Delta = 3$ and $M = 18$.

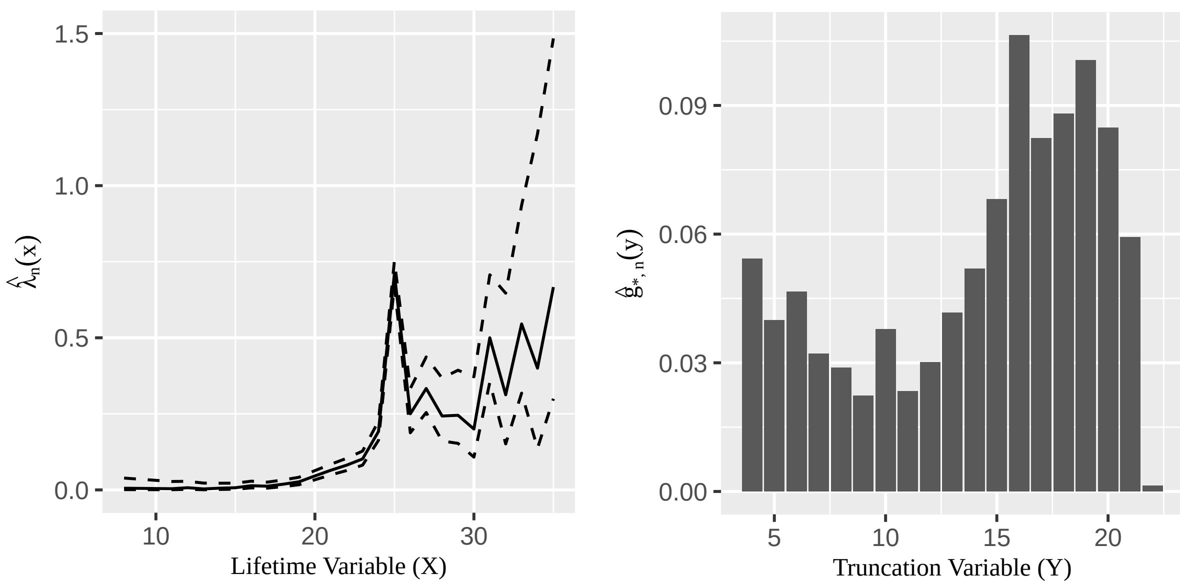
Though each lease is scheduled to terminated after 36 months, lease contracts may terminated early through default or consumer option. Additionally, lease contracts may extend beyond 36 months due to missed payments or various extension clauses. Therefore, to calculate the time of lease termination, we searched the data for three consecutive monthly reports of zero payment. Once three consecutive zero payments were found, the month of lease termination was assigned to be the month of the first zero. For example, if a lease contract recorded a zero payment for months 11, 12, and 13, then month 11 was assumed to be the lease termination age. After performing this search, we identified eight contracts that did not terminate during the observation window and were thus censored. However, for simplicity, we assumed these eight leases all terminated as of the last observation month. The termination time of the oldest lease was 37 months, and so $L = 37$. Formally, then, $Y \in [4, 22]$ and $X \in [4, 37]$.

In Figure 4, we plot the estimated hazard rate for lease terminations within the MBALT 2017-A transaction. Most leases have terminated at month 25, which we would expect for a pool of leases contractually designed to terminate after 24 payments. However, there are a few interesting observations. First, there is notable early lease termination activity beginning around lease age 20 months. Second, we have sporadic hazard rate behavior beyond lease age 25. Third and finally, the width of the 95% confidence band increases markedly beyond 25 months. However, the bands are quite narrow for leases that terminate prior to the original termination schedule of 24 months.

Table 1: Coverage percentages of the 95% confidence intervals in the simulation study for $\lambda(x)$ and $\beta(y)$ by sample size, for $x \in [1, \dots, 23]$ and $y \in [2, 10]$. Top: $\hat{\lambda}_n(x)$; Bottom: $\hat{\beta}_n(y)$. The coverage probability is the percentage of the 1,000 replicates of confidence intervals calculated using the sample data that contain the true value of $\lambda(x)$ for each $x \in X$. The None column denotes the number of the 1,000 replicates that did not observe any values of $x \in X$. For example, when $n = 250$, there were 431 occurrences in which we did not observe a value of $x = 23$ in the sample of 250 observations. The coverage probability adjusts for missing observations. For example, with $n = 250$ at $X = 23$, we divide the number of replicates with confidence intervals that contained $\lambda(x)$ by 569 instead of 1,000 to obtain 87.5. Probabilities displayed as percentages. Note that $X = 24$ and $Y = 1$ are not reported because they are degenerate values.

	$n = 250$		$n = 500$		$n = 750$		$n = 1{,}000$		$n = 10{,}000$	
x	Prob.	None	Prob.	None	Prob.	None	Prob.	None	Prob.	None
1	95.9	0	94.4	0	93.8	0	93.8	0	95.6	0
2	96.2	0	94.5	0	94.5	0	95.4	0	94.9	0
3	95.3	0	94.5	0	95.0	0	95.7	0	95.8	0
4	95.3	0	94.6	0	95.0	0	94.3	0	93.9	0
5	95.4	0	94.1	0	95.7	0	94.5	0	95.8	0
6	93.8	0	94.5	0	93.9	0	94.1	0	95.8	0
7	93.7	0	94.4	0	95.6	0	95.6	0	95.4	0
8	94.5	0	95.3	0	95.7	0	95.2	0	95.2	0
9	95.3	0	95.1	0	95.8	0	94.4	0	93.0	0
10	95.1	0	95.5	0	95.9	0	96.7	0	93.3	0
11	95.2	0	95.4	0	94.4	0	94.4	0	93.8	0
12	95.9	0	95.6	0	95.3	0	95.7	0	95.1	0
13	94.6	1	94.9	0	94.5	0	95.7	0	96.2	0
14	96.0	0	96.3	0	95.7	0	94.6	0	94.6	0
15	95.5	6	94.6	0	94.0	0	95.3	0	95.7	0
16	95.1	23	96.8	2	93.7	0	94.5	0	94.4	0
17	95.5	37	96.6	1	95.7	0	95.2	0	95.1	0
18	94.5	77	96.0	5	96.8	1	95.0	0	94.8	0
19	93.1	131	94.7	21	96.1	3	95.2	0	94.8	0
20	92.6	204	95.0	43	95.3	4	95.3	2	95.8	0
21	91.1	296	95.3	69	95.1	20	95.6	1	95.0	0
22	90.0	347	93.6	146	94.3	49	95.7	20	94.7	0
23	87.5	431	91.3	206	93.8	84	95.5	35	94.8	0

	$n = 100$		$n = 250$		$n = 500$		$n = 1{,}000$		$n = 10{,}000$	
y	Prob.	None	Prob.	None	Prob.	None	Prob.	None	Prob.	None
2	93.6	0	95.2	0	94.3	0	95.5	0	96.0	0
3	95.6	0	94.5	0	94.3	0	95.3	0	95.5	0
4	96.0	0	96.4	0	95.8	0	94.8	0	95.7	0
5	94.7	0	95.7	0	95.4	0	95.4	0	94.2	0
6	95.9	1	94.9	0	95.1	0	95.6	0	95.5	0
7	97.0	2	95.3	0	94.6	0	95.2	0	94.8	0
8	95.3	11	95.3	0	95.6	0	96.3	0	95.1	0
9	96.1	20	96.0	0	94.8	0	94.9	0	96.2	0
10	93.6	47	95.6	2	95.4	0	94.6	0	95.4	0



(a) $\hat{\lambda}_n(x)$ for MBALT 2017-A

(b) $\hat{g}_{*,n}(y)$ for MBALT 2017-A

Figure 4: (a) A plot of $\hat{\lambda}_n(x)$ for the MBALT 2017-A securitization (solid line). The dashed lines represent upper and lower 95% confidence bands. (b) The estimated pmf for Y, the truncation random variable, for the MBALT 2017-A securitization.

Table 2 presents complete results for the estimated quantities $\hat{f}_{*,n}$, $\hat{\lambda}_n$, $\hat{g}_{*,n}$, and $\hat{\beta}_n$, along with the standard errors for $\hat{\lambda}_n$ and $\hat{\beta}_n$.

Additionally, some practitioners may be more interested in estimating the truncation random variable, Y. To this end, we present the estimated probability mass function for Y in Figure 4. An interested investor could use this information to recover T, the distribution of lease origination times. Information about T may be compared with economic trends or the "Selection of the Leases" section of Mercedes-Benz (2017), for example. Finally, it may be of interest to determine if the distribution of Y is uniform, particularly if one wishes to extend the analysis to consider right censoring, such as Asgharian et al. (2002) and De Uña-Álvarez (2004). Though it may be obvious from Figure 4 that Y is not uniform, we may also use Corollary 3.5.1 to calculate $\mathbb{Q}_U = 1{,}530.6$, which corresponds to a p-value of effectively zero. Hence, we reject the null hypothesis of a uniform distribution for Y. Rejecting the null in this case implies utilizing a method to estimate a distribution function for X that relies on the assumption that the truncation random variable is uniform (i.e., stationarity), such as length-biased sampling, would be invalid for this application.

6 Discussion

We examined in detail the asymptotic properties of the hazard and reverse hazard rate estimators and the survival and distribution function estimators in the case of data subject to random truncation with the additional conditions that X and Y are discrete random variables with finite support. We proved that the random estimation vectors $\hat{\mathbf{\Lambda}}_n$ and $\hat{\mathbf{B}}_n$ are asymptotically normal with independent components (i.e., a diagonal covariance matrix). We also proved asymptotic normality extends to the survival function estimator $\hat{S}_n$ and the distribution function estimator $\hat{G}_n$. The last

Table 2: Estimated distributions for lease terminations (F_0) and the truncation random variable Y (G_0) and standard errors of the hazard rate and reverse hazard rate estimators for MBALT 2017-A.

	F_0			G_0		
Age	$\hat{f}_{*,n}$	$\hat{\lambda}_n$	$s.e.[\hat{\lambda}_n]$	$\hat{g}_{*,n}$	$\hat{\beta}_n$	$s.e.[\hat{\beta}_n]$
4	0	0	0	0.057	1	NA
5	0	0	0	0.042	0.424	1.577
6	0	0	0	0.048	0.331	1.229
7	0	0	0	0.033	0.186	0.917
8	0.001	0.005	0.161	0.030	0.143	0.763
9	0	0	0	0.023	0.100	0.621
10	0	0	0	0.039	0.145	0.675
11	0.001	0.004	0.115	0.024	0.082	0.505
12	0.002	0.007	0.147	0.031	0.096	0.516
13	0.001	0.003	0.093	0.043	0.117	0.531
14	0.002	0.006	0.115	0.053	0.127	0.515
15	0.003	0.007	0.121	0.069	0.143	0.502
16	0.008	0.014	0.152	0.107	0.182	0.503
17	0.008	0.012	0.135	0.082	0.124	0.404
18	0.014	0.019	0.157	0.087	0.117	0.373
19	0.022	0.027	0.177	0.097	0.118	0.355
20	0.040	0.046	0.223	0.080	0.090	0.305
21	0.058	0.065	0.260	0.053	0.059	0.250
22	0.068	0.081	0.298	0.001	0.001	0.041
23	0.079	0.102	0.345			
24	0.133	0.192	0.474			
25	0.397	0.711	0.607			
26	0.040	0.250	1.077			
27	0.040	0.333	1.354			
28	0.020	0.243	1.508			
29	0.015	0.245	1.739			
30	0.009	0.200	1.861			
31	0.018	0.500	2.601			
32	0.006	0.313	3.410			
33	0.007	0.545	4.418			
34	0.002	0.400	6.447			
35	0.002	0.667	8.009			
36	0	0	0			
37	0.001	1	NA			

main result of this work was to establish a hypothesis test to examine the shape of the distribution of G, which has clear applications to formally test the stationarity assumption of the truncation distribution in length-biased sampling.

At first glance, it may seem too restrictive to require discrete X and Y to be limited to a finite sample space for our work to have widespread applications. Indeed, this belief may be why the literature has historically given such little attention to the case of discrete, truncated survival data. However, when we consider various applications of survival analysis to model human lifetimes, in finance to model fixed length contracts, or the pragmatic limitations of data collection methods to the practicing statistician, we see what appears to be quite restrictive is actually not so. Furthermore, we feel any loss in generalization is overcome by the simplification of the results: a multivariate normal distribution with a diagonal covariance structure possesses many attractive statistical properties. Indeed, we hope one ancillary benefit of this work is to make the conclusions of Woodroofe (1985) and Wang et al. (1986) more accessible to a wider group of users in practice. For example, though our application was financial in nature, it is not difficult to see how our results may enjoy applications to event times that are practically discrete (e.g., in months) outside of finance, such as insurance, telecommunication, or epidemiology.

A natural extension for this research is to consider data that has also been censored in addition to truncated. Such an extension will allow for much wider applications. This problem will attract our attention in the sequel.

References

O. O. Aalen and S. Johansen (1978). "An empirical transition matrix for non-homogeneous markov chains based on censored observations." *Scandinavian Journal of Statistics* **5**, 141–150.

V. Addona and D. B. Wolfson (2006). "A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up." *Lifetime Data Analysis* **12**, 267–284.

M. Asgharian, C. E. M'Lan and D. B. Wolfson (2002). "Length-biased sampling with right censoring." *Journal of the American Statistical Association* **97**, 201–209.

M. Asgharian and D. B. Wolfson (2005). "Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data." *The Annals of Statistics* **33**, 2109 – 2131.

M. Asgharian, D. B. Wolfson and X. Zhang (2006). "Checking stationarity of the incidence rate using prevalent cohort survival data." *Statistics in Medicine* **25**, 1751–1767.

H. W. Block, T. H. Savits and H. Singh (1998). "The reversed hazard rate function." *Probability in the Engineering and Informational Sciences* **12**, 69–90.

M.-T. Chao and S.-H. Lo (1988). "Some representations of the nonparametric maximum likelihood estimators with truncated data." *The Annals of Statistics* **16**, 661–668.

K. Chen, M.-T. Chao and S.-H. Lo (1995). "On strong uniform consistency of the Lynden–Bell estimator for truncated data." *The Annals of Statistics* **23**, 440–449.

V. H. de la Peña and E. Giné (1999). *Decoupling: From Dependence to Independence*. Springer.

J. De Uña-Álvarez (2004). "Nonparametric estimation under length-biased sampling and type i censoring: A moment based approach." *Annals of the Institute of Statistical Mathematics* **56**, 667–681.

I. Gijbels and J. Wang (1993). "Strong representations of the survival function estimator for truncated and censored data with applications." *Journal of Multivariate Analysis* **47**, 210–229.

O. Guilbaud (1988). "Exact kolmogorov-type tests for left-truncated and/or right-censored data." *Journal of the American Statistical Association* **83**, 213–221.

Ü. Gürler (1996). "Bivariate estimation with right-truncated data." *Journal of the American Statistical Association* **91**, 1152–1165.

Ü. Gürler and J.-L. Wang (1993). "Nonparametric estimation of hazard functions and their derivatives under truncation model." *Annals of the Institute of Statistical Mathematics* **45**, 249–264.

S. He and G. L. Yang (1998a). "Estimation of the truncation probability in the random truncation model." *The Annals of Statistics* **26**, 1011–1027.

S. He and G. L. Yang (1998b). "The strong law under random truncation." *The Annals of Statistics* **26**, 992 – 1010.

C. Hu (2013). *Smoothing Spline ANOVA Models*. Springer.

C.-Y. Huang and J. Qin (2011). "Nonparametric estimation for length-biased and right-censored data." *Biometrika* **98**, 177–186.

Y.-T. Hwang and C.-c. Wang (2008). "A goodness of fit test for left-truncated and right-censored data." *Statistics & Probability Letters* **78**, 2420–2425.

J. Hyde (1977). "Testing survival under right censoring and left truncation." *Biometrika* **64**, 225–230.

A. F. Karr (1991). *Point Processess and Their Statistical Inference*. Marcel Dekker, Inc.

N. Keiding and R. D. Gill (1990). "Random truncation models and markov processes." *The Annals of Statistics* **18**, 582 – 602.

T. L. Lai and Z. Ying (1991). "Estimating a distribution function with truncated and censored data." *The Annals of Statistics* **19**, 417–442.

E. Lehman (1998). *Elements of Large-Sample Theory*. Springer.

E. Lehmann and G. Casella (1998). *Theory of Point Estimation, 2nd Edition*. Springer.

D. Lynden-Bell (1971). "A method of allowing for known observational selection in small samples applied to 3CR quasars." *Monthly Notices of the Royal Astronomical Society* **155**, 95–118.

M. Mandel and R. A. Betensky (2007). "Testing goodness of fit of a uniform truncation model." *Biometrics* **63**, 405–412.

Mercedes-Benz (2017). "Prospectus." *Mercedes-Benz Auto Lease Trust 2017-A* .

C. Moreira, J. De Uña-álvarez and I. Van Keilegom (2014). "Goodness-of-fit tests for a semiparametric model under random double truncation." *Computational Statistics* **29**, 1365–1379.

J. Ning, J. Qin and Y. Shen (2010). "Non-parametric tests for right-censored data with biased sampling." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**, 609–630.

A. B. Owen (2001). *Empirical Likelihood*. Chapman & Hall / CRC.

N. Ravishanker and D. Dey (2002). *A First Course in Linear Model Theory*. Chapman & Hall (CRC).

W. Stute (1993). "Almost sure representations of the product-limit estimator for truncated data." *The Annals of Statistics* **21**, 146 – 156.

W.-Y. Tsai, N. P. Jewell and M.-C. Wang (1987). "A note on the product-limit estimator under right censoring and left truncation." *Biometrika* **74**, 883–886.

Ü. Uzonḡullari and J.-L. Wang (1992). "A comparison of hazard rate estimators for left truncated and right censored data." *Biometrika* **79**, 297–310.

M.-C. Wang (1991). "Nonparametric estimation from cross-sectional survival data." *Journal of the American Statistical Association* **86**, 130–143.

M.-C. Wang, N. P. Jewell and W.-Y. Tsai (1986). "Asymptotic properties of the product limit estimate under random truncation." *The Annals of Statistics* **14**, 1597–1605.

M. Woodroofe (1985). "Estimating a distribution function with truncated data." *The Annals of Statistics* **13**, 163 – 177.

Y. Zhou (1996). "A note on the TJW product-limit estimator for truncated and censored data." *Statistics & Probability Letters* **26**, 381–387.

Y. Zhou and P. S. F. Yip (1999). "A strong representation of the product-limit estimator for left truncated and right censored data." *Journal of Multivariate Analysis* **69**, 261–280.