View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by University of Richmond

University of Richmond
UR Scholarship Repository

Math and Computer Science Faculty Publications
Math and Computer Science

2002

Nonparametric Estimation of a Distribution Subject to a Stochastic Precedence Constraint

Miguel A. Arcones

Paul H. Kvam
University of Richmond, pkvam@richmond.edu

Francisco J. Samaniego

Follow this and additional works at: https://scholarship.richmond.edu/mathcs-faculty-publications

Part of the Applied Statistics Commons, and the Mathematics Commons

This is a pre-publication author manuscript of the final, published article.

Recommended Citation

Arcones, Miguel A.; Kvam, Paul H.; and Samaniego, Francisco J., "Nonparametric Estimation of a Distribution Subject to a Stochastic Precedence Constraint" (2002). *Math and Computer Science Faculty Publications*. 197.
https://scholarship.richmond.edu/mathcs-faculty-publications/197

This Post-print Article is brought to you for free and open access by the Math and Computer Science at UR Scholarship Repository. It has been accepted for inclusion in Math and Computer Science Faculty Publications by an authorized administrator of UR Scholarship Repository. For more information, please contact scholarshiprepository@richmond.edu.

Nonparametric Estimation of a Distribution Subject to a Stochastic Precedence Constraint

Miguel A. Arcones

Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902, USA

Paul H. Kvam

Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

Francisco J. Samaniego

Department of Statistics, University of California, Davis, CA 95616, USA

Abstract

For any two random variables X and Y with distributions F and G defined on $[0, \infty)$, X is said to stochastically precede Y if $P(X \leq Y) \geq 1/2$. For independent X and Y, stochastic precedence (denoted by $X \leq_{sp} Y$) is equivalent to $E[G(X-)] \leq 1/2$. The applicability of stochastic precedence in a variety of statistical contexts, including reliability modeling, tests for distributional equality vs. various alternatives and the relative performance of comparable tolerance bounds, is discussed. The problem of estimating the underlying distribution(s) of experimental data under the assumption that they obey a stochastic precedence (*sp*) constraint is treated in detail. Two estimation approaches, one based on data shrinkage and the other involving data translation, are used to construct estimators that conform to the *sp* constraint, and each is shown to lead to a root n-consistent estimator of the underlying distribution. The asymptotic behavior of each of the estimators is fully characterized. Conditions are given under which each estimator is asymptotically equivalent to the corresponding empirical distribution function, or, in the case of right censoring, the Kaplan Meier estimator. In the complementary cases, evidence is presented, both analytically and via simulation, which demonstrates that the new estimators tend to outperform the edf when sample sizes are sufficiently large.

Key Words: Empirical processes, Order statistics, Reliability, Stochastic order, U–statistics.

1 Introduction

The study of stochastic relationships between random variables or their distributions has been a fertile area of research in applied probability for some time. The notion that one random variable tends to be larger than another is one that can be quantified in many different ways. Among the best known stochastic relationships in the literature are stochastic ordering, uniform stochastic (or hazard rate) ordering and likelihood ratio ordering, denoted here by $X \leq_{st} Y$, $X \leq_{hr} Y$ and $X \leq_{lr} Y$, respectively. (When $X \sim F$ and $Y \sim G$, we will use the inequality $F \leq G$ as interchangeable with $X \leq Y$.)

Definitions and a comprehensive discussion of these and other orderings can be found in the recent monograph by Shaked and Shanthikumar (1994). It is well known that stochastic ordering is the weakest of these three notions, that is, $X \leq_{lr} Y$ implies $X \leq_{hr} Y$ which, in turn, implies $X \leq_{st} Y$. The fact that these concepts arise in a wide variety of statistical applications provides additional motivation for the study of such relationships. Early statistical applications of stochastic ordering include testing hypotheses for parametric families having monotone likelihood ratio and nonparametric tests for the equality of two distributions against a stochastic ordering alternative (see Chapter 3 of Lehmann (1986)).

Coincident with the studies alluded to above, there has been a growth of interest in the use of nonparametric statistical methods in the analysis of failure time data. Nonparametric reliability, for example, seeks to model life-testing data through known qualitative characteristics of the experiment in question; a nonparametric class such as the collection of distributions with increasing failure rate serves to describe experimental subjects tending to deteriorate over time without making restrictive parametric assumptions about the underlying probability distribution. For a good overview of nonparametric modeling in reliability, see Barlow and Proschan (1975). In the present article, we will be interested in the problem of estimating a distribution or survival function when it is assumed to be a member of a particular nonparametric class (to be described in detail below). Antecedents for the work presented here include Grenander's (1956) and Marshall and Proschan's (1965) studies on estimating a distribution with monotone failure rate, Boyles and Samaniego's (1984) study on estimating a survival curve under the "new better than used" constraint, the work of Brunk et al. (1966) and Dykstra et al. (1982) on estimation under a stochastic ordering constraint, and of Rojo and Samaniego (1991, 1993), Mukerjee (1996) and Arcones and Samaniego (2000) on estimation under a uniform stochastic ordering constraint.

The weakest of the orderings mentioned above, viz. $X \leq_{st} Y$, is still too strong an assumption in many problems in which one is inclined to believe that the X population is somehow smaller than the Y population. If F and G are the cumulative distribution functions (cdfs) of X and Y respectively, the stochastic ordering assumption prescribes uniform domination (i.e. $F(t) \geq G(t)$ for all t) of one distribution by the other. While that domination may well hold over an important part of the range of the relevant variables, it may be known to fail over another part of the range (due to infant mortality or planned obsolescence, for example), or may simply be unknown or unknowable over the entire range. Furthermore, stochastic ordering often fails when comparing distributions from different parametric families, and may be quite an unmanageable concept when the two cdfs of interest are not available in closed form. For these reasons, one might wish to entertain the possibility of alternative formulations of the relationship between two random variables. With this motivation, we introduce the following stochastic relationship as a way of comparing distributions:

Definition: Let X and Y be independent random variables with distributions F and G, respectively. Then the variable X is said to *stochastically precede* the variable Y if $P(X \leq Y) \geq 1/2$. This relationship will be denoted by $X \leq_{sp} Y$ or, equivalently, by $F \leq_{sp} G$.

Suppose X and Y are independent random variables, with $X \sim F$ and $Y \sim G$. It is then easily seen that stochastic ordering implies stochastic precedence. If $X \leq_{st} Y$, then $P(X \leq Y) = \int_X (1 - G(x-))dF(x) \geq \int_X (1 - F(x-))dF(x) = P(X \leq X') \geq 1/2$, where X' is an independent copy of X. It follows that stochastic precedence is a less restrictive assumption on the relationship between two random variables than stochastic ordering. If $X \sim N(\mu_1, {\sigma_1}^2)$ and $Y \sim N(\mu_2, {\sigma_2}^2)$, $X \leq_{st} Y$ iff $\mu_1 \leq \mu_2$ and ${\sigma_1}^2 = {\sigma_2}^2$, but $X \leq_{sp} Y$ iff $\mu_1 \leq \mu_2$. Stochastic ordering recognizes differences in normal random variables only if the variances are identical. Stochastic precedence, on the other hand, does not require equal variances to order X and Y.

The assumption $X \leq_{sp} Y$ is equivalent to the assertion that the median of the variable $Y - X$ is greater than or equal to zero. The relationship is thus seen to be different from, but of the same ilk as, the more familiar restriction $E(Y - X) \geq 0$. Statistical inference under the latter restriction and its natural generalizations has been studied extensively, and constitutes an important part of the field of order-restricted inference (see Robertson, Wright and Dykstra (1988)).

Interest in the probability $P(X \leq Y)$, where X and Y are independent random variables, has a fairly long history. Birnbaum (1956), for example, considered the problem of the estimating $P(X \leq Y)$ on the basis of two independent samples, advocating a scalar multiple of the Mann-Whitney statistic for this purpose, and deriving one-sided confidence intervals based on his estimator. Church and Harris (1970) studied a particular parametric version of this problem arising in reliability, pointing out the relevance of this probability in the modeling of stress-strength relationships. When Y represents the stress placed on a component under test and X represents its (breaking) strength, then $P(X \leq Y)$ is simply the probability that the component fails. Johnson (1988) provides a comprehensive review of work on modeling and inference related to stress-strength testing. While the probability $P(X \leq Y)$ has received a good deal of attention, its utility in ordering the variables X and Y, as in the relationship $X \leq_{sp} Y$ defined above, has not heretofore been carefully explored.

In the section that follows, we discuss the occurrence of stochastic precedence in several different statistical contexts, including accelerated life testing, test for contamination and the comparison of nonparametric tolerance intervals. From this discussion, we will see that interest in stochastic precedence extends well beyond the problems of stress-strength testing or mathematical comparisons against other stochastic orders.

Our main interest is in improving nonparametric estimation of the underlying distribution function F when that F is subject to a stochastic precedence constraint. Specifically, under the assumption that $F \leq_{sp} G$, we will consider both one and two-sample estimation problems. In the one sample case, the dominating distribution G is treated as known. Our goal is to develop estimators of F (and of G, when appropriate) which obey the postulated *sp* constraint. If the edf satisfies the *sp* constraint, then it will serve as a suitable estimator. The real challenge, of course, is to develop a good estimator in the more typical circumstance in which the edf violates the constraint.

In Section 3, we construct an estimator of F that satisfies the stochastic precedence constraint by *shrinking* the sample (generated from F) until the constraint holds. An alternative estimator for F is derived in Section 4, based on *shifting* the data, rather than shrinking it. Both estimators are shown to be consistent, and their asymptotic behavior is fully characterized. Similar results are obtained for estimators of F and G in the corresponding two sample problems. Finally, in Section 5, the two proposed estimators are compared to each other and to the edf, both on the basis of their asymptotic properties and in the context of a simulation study in which small sample comparisons are made using data generated from several well-known reliability models. Also discussed in the concluding section is constrained estimation based on censored data. All proofs of results in the body of the paper have been relegated to the appendices.

2 Stochastic Precedence in Statistics

In the introductory section, we motivated the concept of stochastic precedence as a useful weakening of various forms of stochastic ordering that arise in reliability applications. The origins of stochastic precedence can be traced to problems involving strength-stress testing in which the probability $P(X \leq Y)$ arises naturally. In this section, we consider a collection of additional applications of stochastic precedence. This discussion establishes that this ordering has a host of statistical ramifications, and constitutes a useful addition to existing approaches for quantifying the way in which two or more experiments might be related. We consider five distinct scenarios in which the concept of stochastic precedence stands to enhance current statistical theory and practice.

2.1 The analysis of data from ordered experiments.

The fields of isotonic regression (IR) and accelerated life testing (ALT) are both centered on data presumed to be derived from populations that are ordered in some way. As an example if IR analysis, Dykstra, Kochar and Robertson (1991) develop a test for uniform stochastic ordering and apply it to data on survival times for patients with carcinoma of the oropharynx in the presence of a covariate measuring the seriousness of their tumors. In ALT, materials are often tested at stress levels that are more severe than those at normal operating conditions. Failure time data from accelerated life tests are often treated using "linked" parametric models that assume a specific functional relationship between stress level and the parameters of the model (e.g., the Arrhenius model or the power law – see Nelson (1990)). Among the nonparametric approaches that have been taken to ALT is that of McNichols and Padgett (1984), who postulate that the distributions governing the experiments differ only by their respective scale factors.

The notion of stochastic precedence can be viewed as a new nonparametric version of traditional IR or ALT modeling. Because stochastic precedence is a weaker condition than that inherent in virtually all IR and ALT models in current use, it stands to be more broadly applicable to the analysis of data from ordered experiments. We shall describe below, in a concrete example, how the inference results developed in the sequel might influence such analysis. Indeed, treating the type of data available from separate phases of the military acquisitions process –developmental and operational testing – was one of the primary motivations for our study (see Cohen, Rolph and Steffey (1998)).

In typical applications of the acquisitions process in the Department of Defense, a system under development (e.g., a vehicle, weapon or piece of hardware) is subjected to testing at various stages prior to "procurement". Developmental Testing (DT) occurs while prototypes are being built and refined. In the later stages of DT, tests are run on fairly mature prototypes, and the data set obtained at that juncture is a useful guide for predicting future performance. When the developmental process is complete, a set of prototypes is delivered to an independent agency that performs Operational Testing (OT) on the sample. It is generally the case that performance under OT is less impressive than that under DT. The main reason for this is that DT occurs under controlled and fairly restricted (laboratory-type) conditions, while OT is meant to investigate performance characteristics under real, anticipated operating conditions. Thus OT often takes place under stresses and strains that are not part of the DT environment.

The goal of a life-testing experiment during the OT phase is the estimation of the lifetime distribution F under OT conditions. This task is especially challenging due to the small sample sizes employed in OT (often smaller than DT sample sizes due to the extraordinary cost of testing under field conditions.) If G is the lifetime distribution under DT conditions, then the assumption $F \leq_{sp} G$ would typically be judged to be an eminently reasonable assumption. The imposition of this assumption could potentially have a substantial impact on the form and quality of the estimator of F. The example below is meant to illustrate this impact. We have applied the *sp*-constrained estimator developed in Section 3 to data generated from distributions F and G that satisfy a stochastic precedence constraint. The distributions involved do not satisfy the stronger stochastic ordering (st) constraint, so that inference results based on the latter ordering are not well suited for estimating F from these data.

Suppose that X is a prototype performance under OT conditions where $exp(\mathrm{X}) \sim \mathrm{N}(\mu, \sigma^2)$, or equivalently, $F \sim$ Lognormal with μ=0 and σ^2=1. Under DT conditions, $G \sim$ Lognormal with μ=0.18 and σ^2=0.8. The mean of both distributions is 1.6487, but $P(X \leq Y) = 0.5559$, so that $F \leq_{sp} G$. At x=2.46, the 0.816 quantile of F, the distribution functions cross, so that the stochastic ordering constraint fails to hold for this case. The DT and OT data shown in Table 1 below were generated from these respective distributions.

Table 1. Data from DT and OT, and the rescaled data OT*=(0.609)OT

DT		OT	OT*
0.3431	0.9727	0.4194	0.2554
0.4461	1.0487	0.8696	0.5296
0.4873	1.0516	0.8770	0.5341
0.5159	1.3262	1.1430	0.6961
0.5249	1.5254	1.3623	0.8297
0.5517	1.5805	1.3719	0.8355
0.5617	1.7979	1.7244	1.0502
0.6865	1.8327	2.7100	1.6504
0.8430	2.3811	3.9693	2.4173
0.8994	6.6239	6.8127	4.1490

Estimation under assumed constraints is characterized by "adjustments" made when the available data appear to violate the constraint. In the isotonic regression problem of estimating two ordered means, one would adjust the two sample means in constructing a pair of estimators consistent with the assumed ordering. In the case of interest here, an adjustment would be required when the empirical distribution functions F_n and G_n violate, relative to each other, the constraint assumed for the population as a whole, namely, $F \leq_{sp} G$. The estimator studied in the next section adjusts the estimator F_n by rescaling the data (the X sample, the Y sample or both) in a minimal way in order to achieve two new empirical distributions that do satisfy the *sp* constraint. Figure 1 below shows the two distribution functions F and G from which the data above was drawn, the two empirical distribution functions F_n and G_n and the *sp*-constrained estimator $\hat{F}_{1,t}$ discussed in Section 3.2, with t set equal to zero.

It is clear that, in the instance above, the adjustment made by imposing the *sp* constraint makes a huge difference in the accuracy of estimation. We do not wish to represent the picture above as typical. As with other constrained problems, the sampled data will often conform to the assumed constraint (just as sample means tend to be ordered in the same way as the means of the populations the samples were drawn from), and no adjustment is necessary. When is constrained estimation likely to be helpful? The example above is a good illustration of the answer: when the data violates the constraint, and especially when it violates it in a marked fashion. The application of the constraint in such situations stands to make a rather large difference in the resulting data analysis.

The discussion above is aimed at demonstrating that, in selected applications, *sp*-constrained estimation can have a strong effect on the practice of data analysis. In the

Figure 1: F, G, F_n, G_n and $\hat{F}_{1,0}$.

succeeding sections, we provide the theoretical justification for two specific types of *sp*-constrained estimators, demonstrating their consistency and discussing the comparative advantage that they have asymptotically over the standard unconstrained estimator when the *sp* constraint holds.

2.2 An embedding of the Behrens-Fisher problem.

The concept of stochastic precedence has, among its interesting applications, a natural connection to the Behrens-Fisher problem. While this famous testing problem admits to some reasonably satisfactory approximate solutions to due to Aspin (1949) and Trickett, et al. (1956), there continues to be some disagreement in the field about the best way to test the equality of two normal means in the presence of heteroscedasticity. The concepts and methods introduced in this paper provide a new way of treating these hypotheses. Consider testing the hypothesis of "*sp* equality" ($F =_{sp} G$) against the alternative of strict stochastic precedence (i.e., $F \leq_{sp} G$). In the heteroscedastic normal case, the hypothesis of equal means is nested within the null hypothesis above while the hypothesis that $\mu_F < \mu_G$ is nested within the alternative above. It follows that a size-α test in the nonparametric problem will have size no greater than α in the Behrens-Fisher problem.

Moreover the nonparametric test for stochastic precedence has reasonable power when the true distributions are normal with different means, and constitutes a robust alternative to the parametric procedures in common use. The nonparametric test to which we allude is specified in greater detail in Arcones, Kvam and Samaniego (2001).

2.3 Testing against a contaminated normal alternative.

In the literature on robust estimators of location (see Andrews et al. (1974)) and on the detection and treatment of outliers, the contaminated normal distribution plays a distinguished role. The model is meant to describe the potential that exists for a small fraction of the data available in a given experiment to be drawn from an extraneous source, perhaps simply from the distribution governing gross errors. When one considers the possibility of trying to detect the presence of contamination, the problem of testing that data came from a single normal population rather than from a mixture of two normals naturally arises. It can be shown that when a single normal distribution and a mixture of two normals are assumed to have the same mean, the two distributions enjoy a strict stochastic precedence relationship, provided that the mixing probability differs from 1/2. This suggests that a test of the null hypothesis of normality against the alternative hypothesis of a contaminated normal can be based on a test statistic that measures the extent to which standardized data appears to come from a distribution that stochastically precedes or is preceded by the standard normal distribution. In Arcones, et al. (2001), we show that, asymptotically, such a test achieves the nominal significance level under the null hypothesis and has limiting power 1 under contaminated normal alternatives.

2.4 Comparing complex coherent systems.

Methods for comparing competing system designs relative to either deterministic or stochastic criteria are of importance in reliability engineering. For an overview, see Kochar, Mukerjee and Samaniego (1999). Recent work by Boland et al. (1992) and by Singh and Misra (1994) focuses on the question of whether active or parallel redundancy produces better performance in particular systems of interest. If X and Y are the respective lifetimes of the systems under study, Singh and Misra suggest that the second system be judged better than the first if $P(X < Y)$ exceeds $P(X > Y)$, a condition that is essentially equivalent to $X \leq_{sp} Y$. Now, the demonstration that a given system is better than another in the sense above can be an imposing analytical problem. For complex systems, the comparison can be virtually intractable. Thus, statistical procedures for testing the condition $X \leq_{sp} Y$, and for estimating each system's lifetime distribution under an *sp* constraint, may be the only practical way to make the determination of superiority. In industrial applications in which competing prototypes can be constructed and tested under fixed conditions, the methods for estimation and testing developed in this paper can be applied to establish system superiority in the sense suggested by Singh and Misra.

2.5 Comparing fixed-level tolerance limits

Suppose that $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ are the order statistics from a random sample of size n from a continuous distribution F on the real line. We will refer to the order statistic $X_{i:n}$ as an approximate $100(1-\alpha)\%$ upper tolerance limit for $100(1-\gamma)\%$ of the population if $P(X_{i:n} \geq F^{-1}(1-\alpha)) \geq 1-\gamma$ and $P(X_{i-1:n} \geq F^{-1}(1-\alpha)) < 1-\gamma$. Thus, $X_{i:n}$ covers $100(1-\alpha)\%$ of the population with probability at least $1-\gamma$, no smaller order statistic attains such coverage at a probability level of at least $1-\gamma$ and $P(X_{i-1:n-1} \geq F^{-1}(1-\alpha) < 1-\gamma)$. Since the variable $F(X_{i:n})$ has the $Beta(i, n-i+1)$ distribution, the probabilities above can be easily evaluated as partial binomial sums.

Using the language above, the order statistics $X_{29:29}$ and $X_{45:46}$ are both approximate 95% upper tolerance limits for 90% of the population. The smallest sample size for which an order statistic can serve as a 95% upper tolerance limit for 90% of the population is $n = 29$, and the maximum order statistic $X_{n:n}$ serves in that capacity for $29 \leq n \leq 45$. When $n = 46$, the second largest order statistic provides 90% coverage at the desired probability level. While it seems natural to claim that $X_{45:46}$ is superior to $X_{29:29}$ as an upper tolerance limit for 90% of the population, the sense in which it is superior may not be obvious. One might guess, for example, that $X_{45:46}$ is stochastically smaller than $X_{29:29}$. If that were so, then $X_{45:46}$ could be declared as preferable on the grounds that it is an upper limit that tends to be smaller than the upper limit $X_{29:29}$. It is not difficult to check directly that the conjectured stochastic ordering does not hold between these two order statistics. A more general result is summarized in the theorem below. The proof is given in Arcones, Kvam and Samaniego (2000b).

Theorem 2.1. Let $X_{r_1:k_1}$, $X_{r_2:k_2}$ be order statistics from independent samples of sizes k_1 and k_2, respectively, drawn from a continuous distribution F. Then $X_{r_1:k_1} \leq_{st} X_{r_2:k_2}$ if and only if $r_2 \geq r_1$ and $k_1 - r_1 \geq k_2 - r_2$.

This result effectively eliminates the possibility of comparing upper tolerance limits via well known order relations like $\leq_{st}$, $\leq_{hr}$ or $\leq_{lr}$. The sense in which $X_{45:46}$ might be better than $X_{29:29}$ as an upper tolerance limit remains to be identified.

The comparisons of greatest interest are those involving order statistics which constitute approximate upper tolerance limits (UTLs) for fixed coverage and probability levels. For sample sizes up to 100, approximate 95% UTLs for 90% of the population are identified as $X_{29:29}, X_{45:46}, X_{59:61}, X_{73:76}$ and $X_{85:89}$. The comparison of such upper tolerance limits will be accomplished via the notion of stochastic precedence. Specifically, one can show that each order statistic in the above list is stochastically preceded by all the order statistics that follow it. The computation involved in such comparisons is displayed in the following result, which is a special case of Theorem 4 from Kvam and Samaniego (1993).

Theorem 2.2. If $X_{j:n}$ and $X_{i:k}$ are order statistics from independent samples of sizes n and k, respectively, from a continuous distribution F, then

$$P(X_{j:n} \le X_{i:k}) = \sum_{\ell=j}^{n} \frac{\binom{n}{\ell}\binom{k}{i}}{\binom{n+k}{\ell+i}} \frac{i}{i+\ell}. \tag{2.1}$$

From Theorem 2.2, we find that $P\left(X_{45:46} \le X_{29:29}\right) = .62703$, that is, we have that $X_{45:46} \le_{sp} X_{29:29}$. It follows that, among these two approximate 95% upper tolerance limits for 90% of the population, the order statistic $X_{45:46}$ is preferable, since it provides a smaller upper tolerance limit 62.7% of the time. Similar computations establish successive stochastic precedence relationships among the remaining upper tolerance limits in the list above.

From this discussion, it is clear that the *sp* relationship between random variables has some interesting statistical ramifications. The examples are by no means exhaustive. For instance, stochastic precedence has also arisen in the comparison of point estimators via Pitman's measure of closeness (see Mason, et al. (1990)). In this paper, our main interest is in estimating the underlying distributions of two variables which are subject to a stochastic precedence constraint. We now turn to the development and comparison of two distinct approaches to that inference problem.

3 Estimation via Data Rescaling

Let F and G be two continuous distributions on the positive real line, and assume that $F \le_{sp} G$. If the edfs based on respective samples from F and G fail to meet the *sp* constraint, we seek to modify one or both edfs until the constraint is achieved. In this section, we do this by minimally rescaling the observations to achieve the *sp* constraint. We assume that F and G are continuous with support in $(0, \infty)$. Here, we treat the one-sample case, where G is assumed known, and the two-sample case, where samples are available from both populations, using the "data-shrinking" strategy.

3.1 The One-Sample Case

Let us assume that G is known and that a random sample $X_1, ..., X_n$ is available from F. The results we derive here hold somewhat more generally than for estimation under a stochastic precedence constraint. We will obtain a consistent estimator of F under the assumption that F satisfies the constraint $E[\phi(X)] \leq 0$, where $\phi(.)$ is an arbitrary non-decreasing function on $[0, \infty)$.

When $\phi(x) = G(x-) - 1/2$, the inequality $F \leq_{sp} G$ is equivalent to $E[\phi(X)] \leq 0$. Define

$$\theta_n = \sup\{t \geq 0 : n^{-1} \sum_{j=1}^{n} \phi(tX_j) \leq 0\}. \quad (3.2)$$

We have that $n^{-1} \sum_{j=1}^{n} \phi(\theta_n X_j -) \leq 0 \leq n^{-1} \sum_{j=1}^{n} \phi(\theta_n X_j +)$. Let $\lambda_n = \min(\theta_n, 1)$, and define our estimator of F as a function of λ_n:

$$\hat{F}_1(x) = n^{-1} \sum_{j=1}^{n} I(\lambda_n X_j \leq x). \quad (3.3)$$

By construction, $\int \phi(x) d\hat{F}_1(x) \leq 0$, thus $\hat{F}_1$ stochastically precedes G. The statistic λ_n is the scale factor used to shrink the data. That is, when $\int \phi(x) dF_n(x) > 0$, we multiply the set $X_1, ..., X_n$ by λ_n, with $0 < \lambda_n < 1$, so that $\int \phi(x) d\hat{F}_1(x) = 0$.

It is well known that $\{n^{1/2}(F_n(x) - F(x)) : x \in \mathbb{R}\}$ converges weakly to $\{W(F(x)) : x \in \mathbb{R}\}$, where $\{W(u) : 0 \leq u \leq 1\}$ is a Brownian bridge. If $E[\phi(X)] < 0$, then $\Pr\{n^{-1} \sum_{j=1}^{n} \phi(X_j) \leq 0\} \to 1$, which implies $\Pr\{\hat{F}_1(x) = F(x), \text{ for each } x\} \to 1$. This fact implies that when the stochastic precedence is strict, (that is, $P(X \leq Y) > 1/2$), $\hat{F}_1$ has the same asymptotic limit as F_n. We record this result as

Theorem 3.1. If $E[\phi(X)] < 0$, then $\{n^{1/2}(\hat{F}_1(x) - F(x)) : x \in \mathbb{R}\} \xrightarrow{w} \{W(F(x)) : x \in \mathbb{R}\}$.

If $P(X \leq Y) = 1/2$, the *sp*-constraint can more strongly affect the asymptotic variance of $\hat{F}_1$. The difference is seen in part (ii) of the theorem that follows. For this theorem, the limiting condition must be satisfied:

(c3.1) $\lim_{h \to 1+} \sup_{x \geq 0} |F(hx) - F(x) - x(h-1)F'(x)|/(h-1) = 0$, $\sup_{x \geq 0} xF'(x) < \infty$.

Theorem 3.2. Let $U_n = n^{-1/2} \sum_{j=1}^{n} (\phi(X_j) - E[\phi(X_j)])$. If $E[\phi^2(X)] < \infty$, then U_n converges in distribution to $N(0, \text{Var}(\phi(X)))$. In addition, $\{n^{1/2}(F_n(x) - F(x)) : x \in \mathbb{R}\}$ and U_n converge jointly to $\{W(F(x)) : x \in \mathbb{R}\}$ and U with covariance function $\text{Cov}(W(F(x)), U) = \text{Cov}(I(X \leq x), \phi(X))$. Define $\zeta(t) = E[\phi(tX)]$. If $E[\phi(X)] = 0$ and $\zeta'(1)$ exists and is positive, then

(i) $n^{1/2}(\theta_n - 1) + (\zeta'(1))^{-1}U_n \xrightarrow{\text{Pr}} 0$ and $n^{1/2}(\lambda_n - 1) + (\zeta'(1))^{-1}U_n^+ \xrightarrow{\text{Pr}} 0$.

(ii) Under (c3.1) or (c3.2), $\{n^{1/2}(\hat{F}_1(x) - F(x)) : x \geq 0\} \xrightarrow{\text{w}} \{W(F(x)) + xF'(x)(\zeta'(1))^{-1}U^+ : x \geq 0\}$.

Theorem 3.2 holds for other nondecreasing functions $\phi(.)$ as long as $E[\phi^2(X)] < \infty$. We now turn our attention to the mean squared error (MSE) of the estimator $\hat{F}_1$ at an arbitrary value $x > 0$. More specifically, we compare F to $\hat{F}$ using the standardized limit

$$MSE(\hat{F}) = \lim_{n\to\infty} n^{1/2} E(\hat{F}(x) - F(x))^2. \tag{3.4}$$

By Lemma A.1, when $\phi(x) = G(x-) - 1/2$, MSE($\hat{F}_1$) simplifies to $E[(W(F(x))] + E[xF'(x)(\zeta'(1))^{-1}U^+)^2] = F(x)(1 - F(x)) + xF'(x)(\zeta'(1))^{-1}\text{Cov}(G(X-), I(X \leq x)) + 2^{-1}(xF'(x))^2(\zeta'(1))^{-2}\text{Var}(G(X-))$.

Note that the first term of the right hand side of this equation is the MSE of F_n. For example, if X and Y have a Uniform$(0,1)$ distribution, then MSE($\hat{F}_1(x)$) $= x(1 - x) + x^2(x - 5/6)$. It follows that this MSE is smaller than MSE($F_n(x)$) for $x < 5/6$. The integrated mean squared error, defined here as IMSE($\hat{F}_1$)$=\int$ MSE($\hat{F}_1(x)$)$dF(x) = 5/36$, is slightly smaller than IMSE(F_n)=1/6. If X and Y are distributed as exponential with mean $\mu = 1$, then $MSE(\hat{F}_1) = (1 - e^{-x})e^{-x} + 2xe^{-2x}(x/3 - 1 + e^{-x})$. In this case, $MSE(\hat{F}_1) \leq MSE(F_n)$ for all $x < 2.8214$, which is approximately $x_{0.94}$, the 0.94 quantile of F. The integrated mean squared error is 77/648=0.1183, again less than that of F_n.

These examples show that improvement gains with $\hat{F}_1$ depend on the underlying distributions of F and G. While uniform improvement over F_n cannot be guaranteed, we see from the above that $\hat{F}_1$ can offer improvement upon the MSE of F_n over a large portion of the effective support set of the distribution F. In Section 5, we further investigate the potential improvements made in reducing MSE from using $\hat{F}_1$ over F_n, and we compare $\hat{F}_1$ to the alternative estimators derived in Section 4.

3.2 The Two-Sample Case.

Here we consider the estimation of F in the case in which G is also unknown. We assume that an independent random sample $Y_1, ..., Y_m$ from G is available, along with the original sample $X_1, ..., X_n$ from F. In this case, we have to estimate two distribution functions simultaneously. Let F_n and let G_m be the empirical distributions based on $X_1, ..., X_n$ and on $Y_1, ..., Y_m$, respectively. For brevity, we shall repress further indexing of statistics $\xi_{n,m}$ that are based on both samples, and use $\hat{\xi}$ instead.

Let $H_1(t) = P(Y < tX)$, and define $\hat{H}_1(t) = (nm)^{-1} \sum_{i=1}^n \sum_{j=1}^m I(Y_j < tX_i)$. Analogous to (3.1), define

$$\hat{\theta}_1 = \sup\{t \geq 0 : \hat{H}_1(t) \leq 1/2\}, \tag{3.5}$$

and let $\hat{\lambda}_1 = \min(\hat{\theta}_1, 1)$. Given $0 \leq t \leq 1$, we define the two–sample estimators of F and G to be $\hat{F}_{1,t}(x) = n^{-1} \sum_{j=1}^n I(\hat{\lambda}_1^{1-t} X_j \leq x)$ and $\hat{G}_{1,t}(x) = m^{-1} \sum_{j=1}^m I(\hat{\lambda}_1^{-t} Y_j \leq x)$, respectively. Observe that $\hat{H}_1(t) = (nm)^{-1} \sum_{i=1}^n \sum_{j=1}^m I(\hat{\lambda}_1^{-t} Y_j < \hat{\lambda}_1^{1-t} X_j)$. Hence, for each $0 \leq t \leq 1$, $\hat{F}_{1,t}(x) \leq_{sp} \hat{G}_{1,t}(x)$. At $t = 0$, we achieve the *sp* constraint by rescaling only the sample from F. At $t = 1$, only the sample from G is rescaled, and for values $t \in (0, 1)$, both samples are simultaneously rescaled.

By the law of large numbers for U–statistics based on two samples (see Theorem 1 in McConnell (1987)) we have $(nm)^{-1} \sum_{i=1}^n \sum_{j=1}^m I(Y_j < X_j) \to H_1(1) < 1/2$ a.s. Then, with probability one, for n large enough, $\hat{\lambda}_1 = 1$, $\hat{F}_{1,t}(x) = F(x)$ and $\hat{G}_{1,t}(x) = G(x)$ for each x. By the Donsker theorem, $\{n^{1/2}(F_n(x) - F(x)) : x \in I\!R\}$ and $\{m^{1/2}(G_m(x) - G(x)) : x \in I\!R\}$ converge weakly to $\{W_1(F(x)) : x \in I\!R\}$ and to $\{W_2(F(x)) : x \in I\!R\}$, respectively, where W_1 and W_2 are two independent Brownian bridges. Similar results for $\hat{F}_{1,t}$ and $\hat{G}_{1,t}$ are summarized in the following theorem.

Theorem 3.3. If $H_1(1) < 1/2$ and $m, n \to \infty$, then, for each $0 \leq t \leq 1$, $\{n^{1/2}(\hat{F}_{1,t}(x) - F(x)), m^{1/2}(\hat{G}_{1,t}(y) - G(y)) : x, y \in I\!R\} \xrightarrow{w} \{W_1(F(x)), W_2(G(y)) : x, y \in I\!R\}$.

Let $\hat{a}^2 = nm/(m\text{Var}(G(X-)) + n\text{Var}(F(Y)))$ and $\hat{U} = \hat{a}\hat{H}_1(1)$. By Theorem 4.5.1 of Koroljuk and Borovskich (1994), if $\text{Var}(G(X-)) > 0$, $\text{Var}(F(Y)) > 0$ and $m, n \to \infty$, then $\hat{U}$ converges in distribution to $N(0, 1)$. In fact, $\hat{Z} = (n^{1/2}(F_n(x) - F(x)), m^{1/2}(G_m(y) - G(y)), \hat{U})$ converges jointly to $Z = (W_1(F(x)), W_2(G(y)), U)$ if $m/n \to c$, with $0 \leq c \leq \infty$. Here, Z is distributed as trivariate normal with mean zero and covariance matrix Σ, where $\Sigma_{1,1} = F(x)(1 - F(x))$, $\Sigma_{2,2} = G(y)(1 - G(y))$, $\Sigma_{3,3}$=1, $\Sigma_{1,2}$=0,

$$\Sigma_{1,3} = \lim_{n\to\infty} \left(\text{Var}(G(X-)) + nm^{-1}\text{Var}(F(Y))\right)^{-1/2} \text{Cov}(I(X \leq x), G(X-)),$$

$$\Sigma_{2,3} = \lim_{n\to\infty} \left(mn^{-1}\text{Var}(G(X-)) + \text{Var}(F(Y))\right)^{-1/2} \text{Cov}(I(Y \leq x), 1 - G(Y)).$$

The asymptotic behavior of these estimators is substantially more complex when the stochastic precedence between F and G is not strict. In the theorem below, we partition this remaining problem into disjoint cases based on limiting conditions (c3.1) from Section 3.1, along with (c3.2) below:

(c3.2) For all $x \geq 0$, $\lim_{h\to 1+} \sup_{x\geq 0} |G(hx) - G(x) - x(h-1)G'(x)|/(h-1) = 0$, $\sup_{x\geq 0} xG'(x) < \infty$.

Theorem 3.4. Suppose that $H_1(1) = 1/2$, $H_1'(1) > 0$, $m, n \to \infty$, $\mathrm{Var}(G(X-)) > 0$ and $\mathrm{Var}(F(Y)) > 0$. Then,

(a) $\hat{a}(\hat{\theta}_1 - 1) + (H_1'(1))^{-1}\hat{U} \overset{\Pr}{\to} 0$ and $\hat{a}(\hat{\lambda}_1 - 1) + (H_1'(1))^{-1}\hat{U}^{+} \overset{\Pr}{\to} 0$.

(b) If conditions c3.1 or c3.2 are satisfied, and if $n/m \to c$ for some $0 \leq c < \infty$, then
$\{n^{1/2}(\hat{F}_{1,t}(x) - F(x)) : x \geq 0\} \overset{w}{\to}$
$\{W_1(F(x)) + (1-t)(\mathrm{Var}(G(X-)) + c\mathrm{Var}(F(Y)))^{1/2} xF'(x)(H_1'(1))^{-1}U^{+} : x \geq 0\}$
and $\{n^{1/2}(\hat{G}_{1,t}(x) - G(x)) : x \geq 0\} \overset{w}{\to}$
$\{c^{1/2}W_2(G(x)) - t(\mathrm{Var}(G(X-)) + c\mathrm{Var}(F(Y)))^{1/2} xG'(x)(H_1'(1))^{-1}U^{+} : x \geq 0\}$.

(c) If conditions c3.1 or c3.2 are satisfied, and if $n/m \to \infty$, then
$\{m^{1/2}(\hat{F}_{1,t}(x) - F(x)) : x \geq 0\} \overset{w}{\to}$
$\{(1-t)(\mathrm{Var}(F(Y)))^{1/2} xF'(x)(H_1'(1))^{-1}U^{+} : x \geq 0\}$, and
$\{m^{1/2}(\hat{G}_{1,t}(x) - G(x)) : x \geq 0\} \overset{w}{\to}$
$\{W_2(G(x)) - t(\mathrm{Var}(F(Y)))^{1/2} xG'(x)(H_1'(1))^{-1}U^{+} : x \geq 0\}$.

If $n/m \to 0$, we obtain infinitely more information on G, and in the case $t = 0$, the limits are identical to those in the one-sample case. If $n/m \to \infty$, our information about G is relatively sparse, and the rate of convergence is $m^{1/2}$, which is slower than that of the one-sample case. In case (b), we obtain the following two expressions for the MSEs:

$$\begin{aligned} &MSE(\hat{F}_{1,t}) \\ &= E[(W_1(F(x)) + (1-t)(\mathrm{Var}(G(X-)) + c\mathrm{Var}(F(Y)))^{1/2} xF'(x)(H_1'(1))^{-1}U^{+})^2] \\ &= F(x)(1-F(x)) + (1-t)\mathrm{Cov}(G(X-), I(X \leq x))xF'(x)(H_1'(1))^{-1} \\ &+2^{-1}(1-t)^2(\mathrm{Var}(G(X-)) + c\mathrm{Var}(F(Y)))(xF'(x))^2(H_1'(1))^{-2}, \end{aligned} \tag{3.6}$$

$$\begin{aligned} &MSE(\hat{G}_{1,t}) \\ &= E[(c^{1/2}W_2(G(x)) - t(\mathrm{Var}(G(X-)) + c\mathrm{Var}(F(Y)))^{1/2} xF'(x)(H_1'(1))^{-1}U^{+})^2] \\ &= cG(x)(1-G(x)) - c^{1/2}t\mathrm{Cov}(1-F(Y), I(Y \leq x))xG'(x)(H_1'(1))^{-1} \\ &+2^{-1}t^2(\mathrm{Var}(G(X-)) + c\mathrm{Var}(F(Y)))(xG'(x))^2(H_1'(1))^{-2}. \end{aligned} \tag{3.7}$$

The variable $t \in [0,1]$ determines which samples are rescaled to achieve the *sp* constraint, and by how much. From the properties of $\hat{F}_{1,t}$ and $\hat{G}_{1,t}$, it is not immediately clear what value of t should be used to construct the two-sample estimators. If we judge the estimators based on the MSE criterion, we might seek the value of t that minimizes $\rho MSE(\hat{F}_{1,t}) + (1-\rho) MSE(\hat{G}_{1,t})$, for some fixed value of $\rho \in (0,1)$ which depends on the precision required in estimating F relative to the precision in estimating G. In the case $\rho = 1/2$, the value of t which minimizes

$$\begin{aligned}
&MSE(\hat{F}_{1,t}) + MSE(\hat{G}_{1,t}) = F(x)(1-F(x)) + cG(x)(1-G(x)) \\
&+(1-t)\text{Cov}(G(X-), I(X \leq x))xF'(x)(H_1'(1))^{-1} \\
&+2^{-1}(1-t)^2(\text{Var}(G(X-)) + c\text{Var}(F(Y)))(xF'(x))^2(H_1'(1))^{-2} \\
&-c^{1/2}t\text{Cov}(1-F(Y), I(Y \leq x))xG'(x)(H_1'(1))^{-1} \\
&+2^{-1}t^2(\text{Var}(G(X-)) + c\text{Var}(F(Y)))(xG'(x))^2(H_1'(1))^{-2} \\
&= A + (1-t)B + 2^{-1}(1-t)^2C + Dt + 2^{-1}Et^2,
\end{aligned} \tag{3.8}$$

is t=(B+C-D)/(C+E). If the distributions F and G are continuous, this value can be estimated by $\hat{t}(x) = (\hat{B}+\hat{C}-\hat{D})/(\hat{C}+\hat{E})$, where $\hat{B} = \hat{\text{Cov}}(G(X-), I(X \leq x))x\hat{f}(x)(\hat{H}_1'(1))^{-1}$, $\hat{C} = (\hat{\text{Var}}(G(X-)) + c\text{Var}(F(Y)))(x\hat{f}(x))^2(\hat{H}_1'(1))^{-2}$, $\hat{D} = c^{1/2}\hat{\text{Cov}}(1-F(Y), I(Y \leq x))x\hat{g}(x)(\hat{H}_1'(1))^{-1}$, $\hat{E} = 2^{-1}(\hat{\text{Var}}(G(X-)) + c\hat{\text{Var}}(F(Y)))(x\hat{g}(x))^2\hat{H}_1'(1)^{-2}$, $\hat{\text{Cov}}(G(X-), I(X \leq x)) = n^{-1}\sum_{i=1}^n G_m(X_i)I(X_i \leq x)$ - $n^{-1}\sum_{i=1}^n G_m(X_i)F_n(x)$, $\hat{\text{Cov}}(1-F(Y), I(Y \leq x)) = m^{-1}\sum_{j=1}^m (1-F_n(Y_j))I(Y_j \leq x)$ -$n^{-1}\sum_{j=1}^m (1-F_m(Y_j))G_m(x)$, $\hat{H}_1'(1) = \int_0^\infty \hat{f}(t)\hat{g}(t)t\,dt$, $\hat{\text{Var}}(G(X-)) = n^{-1}\sum_{i=1}^n G_m^2(X_i)$ -$(n^{-1}\sum_{i=1}^n G_m(X_i))^2$, $\hat{\text{Var}}(F(Y)) = m^{-1}\sum_{j=1}^m F_n^2(Y_j)$ - $(m^{-1}\sum_{j=1}^m F_n(Y_j))^2$, and $\hat{f}(t)$ and $\hat{g}(t)$ are density estimators of $f(t) = F'(t)$ and $g(t) = G'(t)$, respectively.

4 Estimation via Data Translation

In this section, we present an alternative estimator for F (denoted by $\hat{F}_2$) based on transforming the data with a *location* rather than a scale change to achieve stochastic precedence. If needed, the data $X_1, ..., X_n$ are minimally shifted by some constant amount to the left until the edf based on the shifted data stochastically precedes G. In the two sample case where G is also unknown, we simultaneously shift the data $Y_1, ..., Y_m$ (from G) by a constant to the *right* until the *sp*-constraint holds.

While the methods employed in the present section can be applied to problems involving positive random variables (on which we focused in Section 3), they also apply more broadly. Here, we assume only that $F \leq_{sp} G$, with F and G being continuous cdfs on the real line. We treat one- and two-sample problems below. In Section 5, we compare these estimators with those developed in Section 3, showing that both offer potential improvement over F_n, but that neither uniformly dominates the other.

4.1 The One-sample case

Let us assume that G is known and that a random sample $X_1, ..., X_n$ is available from F. As before, the results we derive here hold somewhat more generally than for estimation under a stochastic precedence constraint. We will obtain a consistent estimator of F under the assumption that F satisfies the constraint $E[\phi(X)] \leq 0$, where $\phi(.)$ is an arbitrary non-decreasing function on $[0, \infty)$. When $\phi(x) = G(x-) - 1/2$, the inequality $F \leq_{sp} G$ is equivalent to $E[\phi(X)] \leq 0$. Define

$$\theta_n = \sup\{t \in I\!R : n^{-1} \sum_{j=1}^{n} \phi(t + X_j) \leq 0\}. \tag{4.9}$$

We have that $n^{-1} \sum_{j=1}^{n} \phi(\theta_n + X_j-) \leq 0 \leq n^{-1} \sum_{j=1}^{n} \phi(\theta_n + X_j+)$. Let $\lambda_n = \min(\theta_n, 0)$, and define our estimator of F as a function of λ_n:

$$\hat{F}_2(x) = n^{-1} \sum_{j=1}^{n} I(\lambda_n + X_j \leq x). \tag{4.10}$$

By shifting the data an amount λ_n, we have $\hat{F}_2$ stochastically preceding G. The location-shift statistic λ_n is analogous to the scale-shift statistic from Section 3.1. The properties of the estimators are similar, as well, but they are not identical in any case of interest. This fact is made clear in the theorems below.

The difference between $\hat{F}_1$ and $\hat{F}_2$ is best appreciated through an example. The two estimators, and the edf F_n, are graphed in Figure 2 against the true distribution ($F = G$) from which a sample of size ten was drawn from a Weibull distribution with shape parameter α=2 and scale parameter β=1, so that $F(x) = 1 - exp(-x^2)$, $x > 0$. For illustration, we choose $G = F$. The edf, graphed as a solid-line step function in Figure 2, clearly disagrees with the constraint of stochastic precedence for this sample. That is, while F_n clearly violates the *sp* constraint relative to G, the alternative estimators $\hat{F}_1$ and $\hat{F}_2$, which differentially shift F_n up and to the left, minimally satisfy the constraint within each of their respective classes. The scale transformation estimator, $\hat{F}_1$, is graphed as a dashed-line step function while $\hat{F}_2$, is graphed as a dotted-line step function. $\hat{F}_1$ is shifted left of F_n by multiplying all the observed data by 0.8253 to ensure G has stochastic precedence over $\hat{F}_1$. $\hat{F}_2$ is shifted to the left by subtracting 0.1732 from each observation. Naturally, the estimators are the same at $x = 0.1732/(1 - 0.8253) = 0.9914$, where the translations are identical. The disagreement between $\hat{F}_1$ and $\hat{F}_2$ is much more dramatic at values of x for which $F(x)$ is close to 1.

Theorem 4.1 below states that if stochastic precedence is strict, $\hat{F}_2$ has the same asymptotic limit as F_n. Theorem 4.2 examines the asymptotic limit of $\hat{F}_2$ in the case that

Figure 2: Weibull cdf (gray line), F_n (solid line), $\hat{F}_1$ (dashed line), and $\hat{F}_2$ (dotted line) under *sp* constraint.

stochastic precedence is not strict, and relies on the following limiting condition:

(c4.2) For all $x \in I\!R$, $\lim_{h\to 0}\sup_{x\geq 0}(|F(x+h)-F(x)-hF'(x)|)/h = 0$, $\sup_{x\geq 0} F'(x) < \infty$.

Theorem 4.1. If $E[\phi(X)] < 0$, then $\{n^{1/2}(\hat{F}_2(x) - F(x)) : x \in I\!R\} \overset{w}{\to} \{W(F(x)) : x \in I\!R\}$.

Theorem 4.2. Define $\zeta(t) = E[\phi(t + X)]$. If $E[\phi(X)] = 0$ and $\zeta'(0) > 0$, then

(i) $n^{1/2}\theta_n + (\zeta'(0))^{-1}U_n \overset{\text{Pr}}{\to} 0$ and $n^{1/2}\lambda_n + (\zeta'(0))^{-1}U_n^+ \overset{\text{Pr}}{\to} 0$.

(ii) Under (c4.1) or (c4.2), $\{n^{1/2}(\hat{F}_2(x) - F_n(x)) : x \geq 0\} \overset{w}{\to} \{W(F(x)) + F'(x)(\zeta'(1))^{-1}U^+ : x \in I\!R\}$.

By Lemma A.1, when $\phi(x) = G(x-) - 1/2$, the MSE of $\hat{F}_2$ defined in (3.4) simplifies to

$$E[(W(F(x)) + F'(x)(\zeta'(0))^{-1}U^+)^2] = F(x)(1 - F(x))$$
$$+F'(x)(\zeta'(0))^{-1}\text{Cov}(G(X-), I(X \leq x)) + 2^{-1}(F'(x))^2(\zeta'(0))^{-2}\text{Var}(G(X-)). \quad (4.11)$$

The MSE in (4.11) neither dominates or is dominated by the MSE for the scale-translation estimator. For example, if X and Y have a Uniform$(0, 1)$ distribution, we have from Section 3 that $\text{MSE}(\hat{F}_1(x)) = x(1-x) + x^2(x - (5/6))$. For the location-translation estimator, $\text{MSE}(\hat{F}_2(x)) = x(1-x)/2 + 1/24$. The integrated MSE for $\hat{F}_2 = 1/8$, which is slightly smaller than the IMSE for $\hat{F}_1$. On the other hand, when X and Y have identical exponential distributions, the IMSE for $\hat{F}_2$ is slightly larger than that of $\hat{F}_1$. In the case $\mu = 1$, then $\text{MSE}(\hat{F}_1(x)) = e^{-x}(1-e^{-x}) + xe^{-2x}(2e^{-x} - 2 + x/6)$ and $\text{MSE}(\hat{F}_2(x)) = e^{-x}(1-e^{-x}) + e^{-2x}(e^{-x} - 5/6)$. Here, $\text{IMSE}(\hat{F}_1) = 0.1183 < \text{IMSE}(\hat{F}_2) = 0.1389$.

4.2 The Two-sample case

Next, we consider the estimation of F in the case in which G is also unknown. We assume that an independent random sample $Y_1, ..., Y_m$ from G is available, along with the original sample $X_1, ..., X_n$ from F. We proceed similarly to Section 3.2, but, in this case, we need not assume that the r.v.s are nonnegative. Let $H_2(t) = P(Y < t + X)$, and define $\hat{H}_2(t) = (nm)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} I(Y_j < t + X_j)$. Analogous to (3.6), define

$$\hat{\theta}_2 = \sup\{t \in I\!R : H_2(t) \leq 1/2\}, \tag{4.12}$$

and define $\hat{\lambda}_2 = \min(\hat{\theta}_2, 0)$. We define the two-sample estimator of F and G to be $\hat{F}_{2,t}(x) = n^{-1} \sum_{i=1}^{n} I((1-t)\hat{\lambda}_2 + X_i \leq x)$ and $\hat{G}_{2,t}(x) = m^{-1} \sum_{j=1}^{m} I(-t\hat{\lambda}_2 + Y_j \leq x)$. Note that, by definition, $\hat{F}_{2,t}(x) \leq_{sp} \hat{G}_{2,t}(x)$. At $t = 0$, only data from F are shifted (to the left), and at $t = 1$, only data from G are shifted (to the right). For values of $t \in (0,1)$, both samples are shifted. Theorem 4.3 below follows by the law of the large numbers for U–statistics: $\hat{H}_2(0-) \to E[G(X-)] < 1/2$, so $\hat{\lambda}_2 = 0$ for n large enough. Along with limiting conditions (c4.1) from Section 4.1, we have

(c4.2) For all$x \in I\!R$, $\lim_{h\to 0} \sup_{x \in I\!R} |G(x+h) - G(x) - hG'(x)|/h{=}0$, $\sup_{x\in I\!R} G'(x) < \infty$.

Theorem 4.3. If $H_2(0) < 1/2$, and $m, n \to \infty$, then $\{n^{1/2}(\hat{F_{2,t}}(x) - F(x)) : x \in I\!R\} \stackrel{w}{\to} \{W(F(x)) : x \in I\!R\}$.

Theorem 4.4. Suppose that $H_2(0) = 1/2$, $H_2'(0) > 0$, $m, n \to \infty$, $\text{Var}(G(X-)) > 0$ and $\text{Var}(F(Y)) > 0$. Then,

(a) $\hat{a}\hat{\theta}_2 + (H_2'(0))^{-1}\hat{U} \stackrel{\text{Pr}}{\to} 0$ and $\hat{a}\hat{\lambda}_2 + (H_2'(0))^{-1}\hat{U}^+ \stackrel{\text{Pr}}{\to} 0$.

(b) If conditions (c4.1) or (c4.2) hold, and if $n/m \to c$ for some $0 \leq c < \infty$, then

$$\{n^{1/2}(\hat{F}_{2,t}(x) - F(x)) : x \in I\!R\} \stackrel{\text{w}}{\to}$$

$$\{W_1(F(x)) + (1-t)(\text{Var}(G(X-)) + c\text{Var}(F(Y)))^{1/2} F'(x)(H_2'(0))^{-1} U^+ : x \in I\!R\}$$

and $\{n^{1/2}(\hat{G}_{2,t}(x) - G(x)) : x \in I\!R\} \stackrel{w}{\to}$
$\{c^{1/2}W_2(G(x)) + t(\mathrm{Var}(G(X-)) + c\mathrm{Var}(F(Y)))^{1/2}G'(x)(H_2'(0))^{-1}U^+ : x \in I\!R\}$.

(c) If conditions (c4.1) or (c4.2) hold, and if $n/m \to \infty$, then
$\{m^{1/2}(\hat{F}_{2,t}(x) - F(x)) : x \in I\!R\} \stackrel{w}{\to}$
$\{(1-t)(\mathrm{Var}(F(Y)))^{1/2}F'(x)(H_2'(0))^{-1}U^+ : x \in I\!R\}$ and
$\{m^{1/2}(\hat{G}_{2,t}(x) - G(x)) : x \in I\!R\} \stackrel{w}{\to}$
$\{W_2(G(x)) + t(\mathrm{Var}(F(Y)))^{1/2}G'(x)(H_2'(0))^{-1}U^+ : x \in I\!R\}$.

As we discussed in the last section, the MSE criteria can be used to find an optimal value of $t \in [0,1]$. The derivation and expression for $\hat{t}$ is similar to that in Section 3.

5 Discussion

In this section, we discuss the properties of the nonparametric estimators derived in Sections 3 and 4. Examples from those sections suggest that there is no strict ranking of the estimators according to the integrated mean squared error criterion. When $E(G(X)) = 1/2$, the two estimators have different asymptotic variances, and these both differ from that of F_n. The examples show that each of the *sp* estimators can improve upon, but does not dominate F_n with regard to the mean squared error criterion in (3.4). To further examine the relationship between respective IMSEs, we consider comparisons based on distributions that are commonly applied to reliability and life-testing problems: the Gamma, Weibull and Lognormal distributions.

For the Gamma distribution, the IMSE can be computed directly for each of the estimators. To compare IMSE of $\hat{F}_1$ versus $\hat{F}_2$, we let $X \sim$ Gamma(r,λ), where r is the shape parameter and $1/\lambda$ is the scale parameter. Let $Y \sim$ Exponential(1), the special case for which $(r,\lambda) = (1,1)$. Note that for $X \leq_{sp} Y$, $E[G(X)] \leq 1/2$ implies that $\lambda \geq 1/(2^{1/r} - 1)$. Because the IMSE is based on asymptotic variances for which $E[G(X)] = 1/2$, we compare the estimators at values of (r,λ) for which $\lambda = 1/(2^{1/r} - 1)$. Relative Error (R.E.) for $\hat{F}$ is defined as IMSE$(\hat{F})$/IMSE(F_n), and is plotted in Figure 3 as a function of the Gamma scale parameter λ. Both estimators improve significantly on F_n, with $\hat{F}_1$ exhibiting a greater amount of improvement than $\hat{F}_2$.

Analytic solutions for IMSE are not possible in the case that data have Weibull or Lognormal distributions. Figures 4 and 5 show the computed IMSE for simulations based on the Weibull and Lognormal distributions, respectively. The results of each figure are based on simulations of 250,000. For Figure 4, samples of n=20 were generated from a Weibull(a, b) distribution, where a is the shape parameter and b is the scale parameter that is set to $b = 1$ here. For the generated samples $X_1, ..., X_{20}$, a is set to 3, and we assume Y is distributed Weibull(a,1), with $a \in (1, 3)$ so that $X \leq_{sp} Y$. For the simulated Weibull data, both estimators outperform F_n, and $\hat{F}_2$ has smaller IMSE than $\hat{F}_1$

In Figure 5, we compute the IMSE based on samples of size n=20 generated from a Lognormal distribution. We examine the IMSE in the case X is distributed Lognormal($\mu = 0$, $\sigma = 1$) and Y is distributed Lognormal($\mu = a$, $\sigma = 1$), with $a \in (0, 0.4)$. From the Example in Section 1, we see that stochastic precedence holds for $a \geq 0$. In contrast to the simulated Weibull data, $\hat{F}_1$ has smaller IMSE than $\hat{F}_2$ in this comparison.

These results confirm the fact that neither estimator dominates the other, and that both estimators can offer substantial improvement over F_n if stochastic precedence is known to exist between two distributions. However, the amount of improvement depends on the underlying distribution of the data, and is not easily characterized analytically.

The approach we have taken to the estimation of F, given $F \leq_{sp} G$, is unabashedly ad hoc. It is an approach that has considerable intuitive appeal when F and G are continuous. Both of the approaches we have considered can be applied to failure time (i.e., nonnegative) data, though we consider this to be the natural domain of applicability of $\hat{F}_1$. For measurement (i.e., real valued) data, $\hat{F}_2$, based on a change in location, is clearly the more suitable. Since most applications in reliability involve positive random variables, both approaches constitute new and usable techniques for analyzing reliability data when stochastic precedence is a reasonable assumption. Because there is no universally accepted approach to constrained nonparametric estimation, it is common to seek to exploit the specific structure of the constrained class one is working with. Our approach has been to transform the data in a minimal way so that the empirical distribution of the transformed data will satisfy the *sp* constraint. The theoretical results we have derived show that this approach is quite efficacious, yielding estimators which satisfy the assumed constraint for any sample size and which inherit many of the good properties of F_n (and G_m) asymptotically. We note that both of these approaches to estimation have limitations. The first is for models for which $P(X < 0) > 0$, and neither is recommended for discrete data. Due to these observations, and because there are a number of other approaches which might yield competitive estimators, we plan to report on the relative performance of alternative approaches in sequel to the present paper.

One of the important extensions of our results that merits some commentary is the applicability of the approaches we have studied to censored data. While we are not in a position to present comprehensive results in this case, we have obtained preliminary results that demonstrate the feasibility and efficacy of our approach in censored data problems.

We have limited our investigation to the one sample problem using the rescaling approach of Section 3.1. We find that the approach generalizes quite easily, producing an estimator of the underlying distribution F based on a censored X sample when F satisfies an sp constraint relative to a known G. Under explicit smoothness conditions on the distribution G, we show that the estimator $\tilde{F}$ based on rescaled censored data, is asymptotically equivalent to the Kaplan-Meier estimator when the sp constraint is strict. In the remaining case, we identify the weak limit of the estimating process and compare its performance to the standard, unconstrained estimator $\tilde{F}$.

As in the uncensored case, evidence is presented supporting the superiority of the constrained estimator. For example, if X and $Y \sim$ U$(0,1)$, then $\text{MSE}(F_n(x)) = 2^{-1}x(2-x)$, $\text{MSE}(\tilde{F}(x)) = 2^{-1}x(2-x) + x(1-x)\ln(1-x) + 2^{-2}x^2$, $\text{IMSE}(F_n(x)) = 1/3$ and $\text{IMSE}(\tilde{F}(x)) = 41/180 < 1/3$. If X and $Y \sim$ exponential$(\mu = 1)$, then $\text{MSE}(F_n(x)) = 2^{-1}(1-e^{-2x})$, $\text{MSE}(\tilde{F}(x)) = 2^{-1}(1-e^{-2x})-x^2e^{-2x}$, $\text{IMSE}(F_n(x)) = 1/3$ and $\text{IMSE}(\tilde{F}(x)) = 7/27$. In both of these cases, the constrained estimator $\tilde{F}$ is seen to outperform the KME in terms of the global IMSE criterion. Detailed results of the derivation for the constrained estimator with censored data can be found in Appendix B.

Appendices

A Asymptotic Results

Lemma A.1. Let (Z_1, Z_2) be a bivariate normal random vector with zero means. Then, $E[Z_1 \max(Z_2, 0)] = 2^{-1}E[Z_1Z_2]$ and $E[(\max(Z_2, 0))^2] = 2^{-1}E[Z_2^2]$.

Proof: Since (Z_1, Z_2) and $(-Z_1, -Z_2)$ have the same distribution, $E[Z_1 \max(Z_2, 0)] = E[-Z_1 \max(-Z_2, 0)] = E[Z_1 \min(Z_2, 0)]$. From this and the fact that $x = \max(x, 0) + \min(x, 0)$, the first formula follows. The second formula follows from the fact that $x^2 = (\max(x, 0))^2 + (\min(x, 0))^2$. □

Now, we apply limit theorems for M–estimators. Asymptotic properties of M–estimators when $h(x, \theta)$ is nondecreasing in θ are in several references, for example Lemma 4 of Huber (1964). We will need generalizations of this theorem.

Theorem A.2. Let $\{Z_n(\theta) : \theta \in \mathbb{R}\}$ be a sequence of stochastic processes. Let $\theta_0 \in \mathbb{R}$.

Let $\{a_n\}$ be a sequence of real numbers converging to infinity. Let $\theta_n = \sup\{t : Z_n(t) \leq 0\}$. Assume that:

(i) As a function on θ, $Z_n(\theta)$ is non increasing.

(ii) There exists a positive constant b such that for each $\tau \in I\!R$, $a_n E[Z_n(\theta_0 + a_n^{-1}\tau) - Z_n(\theta_0)] \to b\tau$.

(iii) $a_n Z_n(\theta_0) = O_P(1)$.

(iv) For each $\tau \in I\!R$, $a_n(Z_n(\theta_0 + a_n^{-1}\tau) - Z_n(\theta_0) - E[Z_n(\theta_0 + a_n^{-1}\tau) - Z_n(\theta_0)]) \stackrel{\Pr}{\to} 0$.

Then, $a_n(\theta_n - \theta_0) + b^{-1} a_n Z_n(\theta_0) \stackrel{\Pr}{\to} 0$.

Proof: Given $\tau > 0$, we prove that

$$(A.1) \qquad \Pr\{b a_n(\theta_n - \theta_0) + a_n Z_n(\theta_0) < -\tau\} \to 0 \quad \text{and}$$

$$(A.2) \qquad \Pr\{b a_n(\theta_n - \theta_0) + a_n Z_n(\theta_0) \leq \tau\} \to 1.$$

This implies the claim.

It is well known that if sequence of nondecreasing functions converges to a continuous function, then it does so uniformly on compact sets. This is also true for a sequence of nondecreasing random functions converging in probability to a continuous function. This implies that for each $0 < M < \infty$,

$$(A.3) \qquad \sup_{|\tau| \leq M} |a_n(Z_n(\theta_0 + a_n^{-1}\tau) - Z_n(\theta_0)) - \tau b| \stackrel{\Pr}{\to} 0.$$

Given t, we have that $\{\theta_n < t\} \subset \{Z_n(t) > 0\}$. So,

$$(A.4) \qquad \Pr\{b a_n(\theta_n - \theta_0) + a_n Z_n(\theta_0) < -\tau\}$$

$$= \Pr\{\theta_n < \theta_0 - a_n^{-1} b^{-1}(\tau + a_n Z_n(\theta_0))\} \leq \Pr\{Z_n(\theta_0 - a_n^{-1} b^{-1}(\tau + a_n Z_n(\theta_0))) > 0\}$$

By condition (iii) and (A.3)

$$|a_n(Z_n(\theta_0 - a_n^{-1} b^{-1}(\tau + a_n Z_n(\theta_0))) - Z_n(\theta_0)) + (\tau + a_n Z_n(\theta_0))| \stackrel{\Pr}{\to} 0.$$

Thus, $a_n Z_n(\theta_0 - a_n^{-1} b^{-1}(\tau + a_n Z_n(\theta_0))) \stackrel{\Pr}{\to} -\tau$. This and (A.4) imply (A.1). Finally, given t, we have that $\{Z_n(t) > 0\} \subset \{\theta_n \leq t\}$. So, $\Pr\{b a_n(\theta_n - \theta_0) + a_n Z_n(\theta_0) \leq \tau\} \geq$

$\Pr\{Z_n(\theta_0 - a_n^{-1}b^{-1}(-\tau + a_n Z_n(\theta_0))) > 0\}$, and as before we have $a_n Z_n(\theta_0 - a_n^{-1}b^{-1}(-\tau + a_n Z_n(\theta_0))) \stackrel{\Pr}{\to} \tau$, which implies (A.2). □

In the previous theorem, condition (ii) is implied by $E[Z_n(\theta)] = Z(\theta)$ does not depend on n and $Z'(\theta_0) > 0$.

The next corollary is similar to Lemma 4 in Huber (1964). In that paper it is shown that the M–estimator is asymptotically normal. We show that the M–estimator minus a linear approximation goes to zero in probability.

Corollary A.3. Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. random variables. Let $h : I\!\!R \times I\!\!R \to I\!\!R$ be function such that $h(\cdot, \theta) : I\!\!R \to I\!\!R$ is measurable for each θ and $h(x, \cdot) : I\!\!R \to I\!\!R$ is nondecreasing for each x. Let $\theta_0 \in I\!\!R$. Let $\theta_n = \sup\{t : n^{-1}\sum_{j=1}^n h(X_j, t) \le 0\}$. Assume that:

(i) $Z(\theta_0) = 0$ and $Z'(\theta_0) > 0$, where $Z(\theta) := E[h(X, \theta)]$.

(ii) $E[h^2(X, \theta_0)] < \infty$.

(iii) $\lim_{\theta \to \theta_0} E[(h(X, \theta) - h(X, \theta_0))^2] = 0$.

Then, $n^{1/2}(\theta_n - \theta_0) + (Z'(\theta_0))^{-1} n^{-1/2} \sum_{j=1}^n (h(X_j, \theta_0) - E[h(X_j, \theta_0)]) \stackrel{\Pr}{\to} 0$.

Proof: We apply Theorem A.2 with $Z_n(\theta) = n^{-1}\sum_{j=1}^n h(X_j, \theta)$. We have that $\mathrm{Var}(n^{1/2}(Z_n(\theta_0 + n^{-1/2}\tau) - Z_n(\theta_0))) = E[(h(X, \theta_0 + \tau n^{-1/2}) - h(X, \theta_0))^2] - (E[h(X, \theta_0 + \tau n^{-1/2}) - h(X, \theta_0)])^2$, which converges to zero. □

To obtain asymptotic properties of $\hat{F}$ in the two-sample case, we apply formulas for U–statistics. If $h : I\!\!R^2 \to I\!\!R$, then the Hoeffding decomposition can be written as

$$(A.5) \qquad n^{-1}m^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}(h(X_i, Y_j) - E[h(X_i, Y_j)])$$

$$= n^{-1}\sum_{i=1}^{n}(h_1(X_i) - E[h_1(X_i)]) + m^{-1}\sum_{j=1}^{m}(h_2(Y_j) - E[h_2(Y_j)])$$

$$+ n^{-1}m^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}(h(X_i, Y_j) - h_1(X_i) - h_2(Y_j) + E[h(X_i, Y_j)]),$$

where $h_1(x) = E[h(x, Y)]$ and $h_2(y) = E[h(X, y)]$. Since this is a decomposition into orthogonal components, we have that

$$(A.6) \qquad \mathrm{Var}(n^{-1}m^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}h(X_i,Y_j))$$

$$= n^{-1}\mathrm{Var}(h_1(X)) + m^{-1}\mathrm{Var}(h_2(Y)) + n^{-1}m^{-1}\mathrm{Var}(h(X,Y) - h_1(X) - h_2(Y))$$

$$\leq (3m^{-1}n^{-1} + m^{-1} + n^{-1})\mathrm{Var}(h(X,Y)).$$

We also have that if $\mathrm{Var}(h_1(X))$, $\mathrm{Var}(h_2(Y)) > 0$ and $\min(n,m) \to \infty$, then

$$(A.7) \qquad b_n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}(h(X_i,Y_j) - E[h(X_i,Y_j)]) \xrightarrow{\mathrm{d}} N(0,1),$$

where $b_n^2 = \mathrm{Var}(\sum_{i=1}^{n}\sum_{j=1}^{m}h(X_i,Y_j))$ (see Theorem 4.5.1 of Koroljuk and Borovskich (1994)). We will use the fact that

$$\lim_{\min(n,m)\to\infty}\frac{b_n^2}{n^{-1}\mathrm{Var}(h_1(X)) + m^{-1}\mathrm{Var}(h_2(Y))} = 1$$

.

Corollary A.4. Let $\{X_i\}_{i=1}^{\infty}$ and let $\{Y_j\}_{j=1}^{\infty}$ be two independent sequences of i.i.d.r.v.'s with possibly different distributions. Let $h : I\!R^3 \to I\!R$ be function such that $h(\cdot,\cdot,\theta) : I\!R^2 \to I\!R$ is measurable for each θ and $h(x,y,\cdot) : I\!R \to I\!R$ is nondecreasing for each x, y. Let $\theta_0 \in I\!R$. Let $\{m\}$ be a sequence of positive integers converging to infinity. Let $\theta_n = \sup\{t : n^{-1}m^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}h(X_i,Y_j,t) \leq 0\}$, and assume that:

(i) $Z(\theta_0) = 0$ and $Z'(\theta_0) > 0$, where $Z(\theta) := E[h(X_1,Y_1,\theta)]$.

(ii) $\mathrm{Var}(h_1(X_1)), \mathrm{Var}(h_2(Y_1)) > 0$, where $h_1(x) = E[h(x,Y_1,\theta_0)]$ and $h_2(y) = E[h(X_1,y,\theta_0)]$.

(iii) $\lim_{\theta\to\theta_0} E[(h(X_1,Y_1,\theta) - h(X_1,Y_1,\theta_0))^2] = 0$.

Then,

$$a_n(\theta_n - \theta_0) + (Z'(\theta_0))^{-1}a_n n^{-1}m^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}(h(X_i,Y_j,\theta_0) - E[h(X_i,Y_j,\theta_0)]) \xrightarrow{\mathrm{Pr}} 0,$$

where $a_n^2 = nm/(m\mathrm{Var}(h_1(X)) + n\mathrm{Var}(h_2(Y)))$.

Proof: We apply Theorem A.2 with $Z_n(\theta) = n^{-1}m^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}h(X_i,Y_j,\theta_0)$. By (A.6), $\mathrm{Var}(a_n(Z_n(\theta_0+n^{-1/2}\tau) - Z_n(\theta_0))) \leq (3m^{-1}n^{-1}+m^{-1}+n^{-1})a_n^2\mathrm{Var}(h(X_1,Y_1,\theta_0+n^{-1/2}\tau)$-$h(X_1,Y_1,\theta_0))$, which converges to zero. □

B Treating Censored Data

We restrict attention to the one sample problem. Suppose we observe $(Z_1, \delta_1), ..., (Z_n, \delta_n)$ where $Z_i = \min(X_i, Y_i)$, $\delta_i = I\{X_i \leq Y_i\}$, and $X_1, ..., X_n \overset{iid}{\sim} F$ and $Y_1, ..., Y_n \overset{iid}{\sim} K$ are independent samples from a lifetime distribution F and a censoring distribution K, respectively. Suppose further that $F \leq_{sp} G$. Let F_n represent the Kaplan–Meier estimator of F.

In the rescaling case, we define

$$\theta_n = \sup\{t \geq 0 : \int_{-\infty}^{\infty} G(tx)\, dF_n(x) \leq 1/2\}.$$

Let $\lambda_n = \min(\theta_n, 1)$. We define our estimator of F as $\tilde{F}_n(x) = F_n(x/\lambda_n)$.It is well known (see, e.g., Breslow and Crowley (1974), Gill (1981)) that $\sup_{t>0} |F_n(t) - F(t)| \to 0 \quad a.s.$, and that $\{n^{1/2}(F_n(t) - F(t)) : t \geq 0\} \overset{w}{\to} \{Z(t) : t \geq 0\}$, where $\{Z(t) : t \geq 0\}$ is a Gaussian process with mean zero and covariance $E\,(Z(s)Z(t)) = C(s)(1 - F(s))(1 - F(t))$, and where $s < t$ and $C(s) = \int_0^s ((1 - F(t))^2(1 - K(t)))^{-1} dF(t)$. Assuming that G is absolutely continuous, we have that, (using integration by parts)

$$Z_n(t) \equiv \int_{-\infty}^{\infty} G(tx)\, dF_n(x) = \int_{-\infty}^{\infty} G(tx)\, d(F_n(x) - 1)$$

$$= G(tx)(F_n(x) - 1))|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} tg(tx)(F_n(x) - 1)dx = \int_{-\infty}^{\infty} tg(tx)(1 - F_n(x))dx,$$

where $g(x) = G'(x)$. Hence, we have that

$$|\int_{-\infty}^{\infty} G(x)\, dF_n(x) - E[G(X)]| \leq \int_{-\infty}^{\infty} tg(tx)|F_n(x) - F(x)|dx \to 0 \;\; a.s.$$

This implies, by the argument in the proof of Theorem 3.1, the following theorems.

Theorem B.1. Assume $E[G(X)] = 1/2$ and that G is absolutely continuous with $g(x) = G'(x)$, then $\{n^{1/2}(\tilde{F}(x) - F(x)) : x \geq 0\} \overset{w}{\to} \{Z(x) : x \geq 0\}$, where $\{Z(x) : x \geq 0\}$ is the limit distribution of the normalized Kaplan–Meier estimator.

Theorem B.2. Assume $E[G(X)] = 1/2$ and that G is twice differentiable with bounded first and second derivatives and $\lim_{\epsilon \to 0} \int_{-\infty}^{\infty} |g((1 + \epsilon)x) - g(x))|dx = 0$. Define $b = \int_{-\infty}^{\infty} xf(x)g(x)\, dx$, $U = -n^{1/2} \int_{-\infty}^{\infty} g(x)Z(x)\, dx$. Then

(i) $n^{1/2}(\theta_n - 1) + b^{-1}U_n \overset{\Pr}{\to} 0$ and $n^{1/2}(\lambda_n - 1) + b^{-1}U_n^+ \overset{\Pr}{\to} 0$, and $\hat{U} = n^{1/2}(\int_{-\infty}^{\infty} G(x)\, dF_n(x) - E[\int_{-\infty}^{\infty} G(x)\, dF_n(x)]) = -n^{1/2} \int_{-\infty}^{\infty} g(x)(F_n(x) - E[F_n(x)])\, dx.$

(ii) If $\lim_{h\to 1+}(|F(hx)-F(x)-x(h-1)F'(x)|)/(h-1)=0$, then $n^{1/2}(\tilde{F}(x)-F(x)) \overset{w}{\to} Z(x)+xF'(x)b^{-1}U^{+}$

(iii) If $\lim_{h\to 1+}\sup_{x\geq 0}(|F(hx)-F(x)-x(h-1)F'(x)|)/(h-1)=0$, and $\sup_{x\geq 0} xF'(x)<\infty$, then $\{n^{1/2}(\tilde{F}(x)-F(x)) : x\geq 0\} \overset{w}{\to} \{Z(x)+xF'(x)b^{-1}U^{+} : x\geq 0\}$.

PROOF. We apply Theorem A.2. Hypothesis (i) in Theorem A.2 is trivially satisfied. As to condition (ii),

$$n^{1/2}E[Z_n(1+n^{-1/2}\tau)-Z_n(1)]$$

$$= n^{1/2}\int_{-\infty}^{\infty}((1+n^{-1/2}\tau)g((1+n^{-1/2}\tau)x)-g(x))(1-E[F_n(x)])dx$$

$$= \int_{-\infty}^{\infty}\tau g((1+n^{-1/2}\tau)x)(1-E[F_n(x)])dx + n^{1/2}\int_{-\infty}^{\infty}(g((1+n^{-1/2}\tau)x)-g(x))(1-E[F_n(x)])dx$$

$$\to \int_{-\infty}^{\infty}\tau g(x)(1-F(x))dx+\int_{-\infty}^{\infty}\tau x g'(x)(1-F(x))dx = \tau\int_{-\infty}^{\infty}\tau g(x)f(x)dx.$$

As to condition (ii),

$$n^{1/2}|Z_n(1+n^{-1/2}\tau)-Z_n(1)-E[Z_n(1+n^{-1/2}\tau)-Z_n(1)]|$$

$$\leq n^{1/2}\int_{-\infty}^{\infty}((1+n^{-1/2}\tau)g((1+n^{-1/2}\tau)x)-g(x))|F_n(x)-E[F_n(x)]|dx$$

$$\leq O_P(1)\int_{-\infty}^{\infty}|(1+n^{-1/2}\tau)g((1+n^{-1/2}\tau)x)-g(x))|dx = o_p(1).$$

Theorem A.2 applies because

$$\int_{-\infty}^{\infty}|g((1+n^{-1/2}\tau)x)-g(x))|dx\to 0, \quad \text{and}$$

$$\int_{-\infty}^{\infty}n^{-1/2}g((1+n^{-1/2}\tau)x)dx = n^{-1/2}((1+n^{-1/2}\tau)^{-1}G(((1+n^{-1/2}\tau)x)|_{-\infty}^{\infty}$$

$$= n^{-1/2}(1+n^{-1/2}\tau)^{-1}\to 0.$$

By Lemma A.1, $MSE(\tilde{F}) = E[(Z(x)+xF'(x)b^{-1}U^{+})^2] = C(x)(1-F(x))^2$ - $b^{-1}xF'(x)\times \int_0^{\infty} g(t)C(t\wedge x)(1-F(x))(1-F(t))\,dt$ $+2^{-1}(b^{-1}xF'(x))^2\int_0^{\infty}\int_0^{\infty} g(s)g(t)C(s\wedge t)(1-F(s))(1-F(t))\,ds\,dt$. $\square$

C Remaining Proofs

Proof of Theorem 3.2: To prove (i), we apply Corollary A.3 with $h(x,t) = \phi(tx)$. We need to prove that $\lim_{\theta\to 1} E[(h(X,\theta) - h(X,1))^2] = 0$. We have that $\lim_{\theta\to 1-} E[(\phi(\theta X) - \phi(X))^2] = E[(\phi(X) - \phi(X-))^2]$ and $\lim_{\theta\to 1+} E[(\phi(\theta X) - \phi(X))^2] = E[(\phi(X+) - \phi(X))^2]$. Since ζ is continuous, $E[\phi(X+) - \phi(X-)] = 0$, and $\phi(X+) - \phi(X-)$ is nonnegative, thus, $E[(\phi(X+) - \phi(X-))^2] = 0$. From this, condition (iii) in Corollary A.3 is satisfied. Therefore, by Corollary A.3, $n^{1/2}(\theta_n - 1) + (\zeta'(1))^{-1}U_n \overset{\Pr}{\to} 0$.

We have that $n^{1/2}(\lambda_n - 1) + (\zeta'(1))^{-1}U_n^+ = n^{1/2}(\min(\theta_n, 1) - 1) + \max(0, (\zeta'(1))^{-1}U_n)$ $= -\max(-n^{1/2}(\theta_n - 1), 0) + \max(0, (\zeta'(1))^{-1}U_n)$. From the inequality $|x^+ - y^+| \le |x - y|$, we have $|n^{1/2}(\lambda_n - 1) + (\zeta'(1))^{-1}U_n^+| \le |n^{1/2}(\theta_n - 1) + (\zeta'(1))^{-1}U_n| \overset{\Pr}{\to} 0$.

To prove (ii), we can partition $n^{1/2}(\hat{F}_1(x) - F(x))$ into four distinct elements: $n^{1/2}(\hat{F}_1(x) - F(x))$=$[n^{1/2}(F_n(x) - F(x))]$+ $[n^{1/2}(\lambda_n^{-1} - 1)xF'(x)]$+ $[n^{1/2}(F_n(\lambda_n^{-1}x) - F_n(x) - F(\lambda_n^{-1}x) + F(x))]$+ $[n^{1/2}(F(\lambda_n^{-1}x) - F(x) - x(\lambda_n^{-1} - 1)F'(x))]$= $[I] + [II] + [III] + [IV]$.

By (i), $I + II$ converges weakly to $W(F(x)) + xF'(x)(\zeta'(1))^{-1}U^+$. Since F is continuous at x, $|III| \overset{\Pr}{\to} 0$. By hypothesis, $|IV| \le n^{1/2}|\lambda_n^{-1} - 1|o(1) \overset{\Pr}{\to} 0$; hence, part (ii) follows. Observe that

$$\sup_{x\ge 0} \frac{|F(\lambda_n^{-1}x) - F(x) - (\lambda_n^{-1} - 1)xF'(x)|}{\lambda_n^{-1} - 1} \overset{\Pr}{\to} 0$$

and uniform convergence holds in this case. □

Proof of Theorem 3.4: We apply Corollary A.4 with $h(x,y,t) = I(y < tx)$. We have that $h_1(x) = G(tx-)$ and $h_2(y) = 1 - F(t^{-1}y)$. Since $H_1'(1) > 0$, $\Pr\{Y = X\} = 0$ and $\lim_{t\to 0} E[(h(x,y,t) - h(x,y,1))^2] = 0$, which implies condition (iii) in Corollary A.4 holds. Therefore, claim (a) follows. The rest of the claims follow by doing a decomposition similar to that in the one sample case. Observe that if $n/m \to c < \infty$, then $\hat{a}^{-1}n^{1/2} = (\mathrm{Var}(G(X-)) + m^{-1}n\mathrm{Var}(F(Y)))^{1/2} \to (\mathrm{Var}(G(X-)) + c\mathrm{Var}(F(Y)))^{1/2}$ and $n^{1/2}(\lambda_n - 1)xF'(x) \overset{d}{\to} -(\mathrm{Var}(G(X-)) + c\mathrm{Var}(F(Y)))^{1/2}xF'(x)(H_1'(1))^{-1}U^+$.

If $n/m \to \infty$, then $\hat{a}^{-1}m^{1/2} = (\mathrm{Var}(F(Y)))^{1/2}$. Thus, $m^{1/2}(F_n(x) - F(x)) \overset{Pr}{\to} 0$ and $m^{1/2}(\lambda_n - 1)xF'(x) \overset{d}{\to} -(\mathrm{Var}(F(Y)))^{1/2}xF'(x)(H_1'(1))^{-1}U^+$. In case (c) we have the following decomposition:

$$\begin{aligned}
&n^{1/2}(\hat{F}_{1,t}(x) - F(x)) \\
&= n^{1/2}(F_n(x) - F(x)) + n^{1/2}(\lambda_n^{t-1} - 1)xF'(x) \\
&+ n^{1/2}(F_n(\lambda_n^{t-1}x) - F_n(x) - F(\lambda_n^{t-1}x) + F(x)) \\
&+ n^{1/2}(F(\lambda_n^{t-1}x) - F(x) - x(\lambda_n^{t-1} - 1)F'(x)) \\
&\simeq n^{1/2}(F_n(x) - F(x)) + n^{1/2}(t-1)(\lambda_n - 1)xF'(x).
\end{aligned}$$

This decomposition and previous limits imply (c). The proof for (b) follows by repeating previous arguments. □

Proof of Theorem 4.1: By the law of the large numbers, with probability one, $n^{-1}\sum_{j=1}^{n}\phi(X_j) \to E[\phi(X)] < 0$. So, for n large enough, $\theta_n \geq 0$, $\lambda_n = 0$ and $\hat{F}_n = F_n$. □

Proof of Theorem 4.2: By Theorem A.2, $n^{1/2}\theta_n + (\zeta'(0))^{-1}U_n \stackrel{\Pr}{\to} 0$, and $n^{1/2}\lambda_n + (\zeta'(0))^{-1}U_n^+ = n^{1/2}(\min(\theta_n, 0) - 1) + \max(0, (\zeta'(0))^{-1}U_n) = -\max(-n^{1/2}\theta_n, 0) + \max(0, (\zeta'(0))^{-1}U_n)$. From the inequality $|x^+ - y^+| \leq |x - y|$, we have $|n^{1/2}\lambda_n + U_n^+| \leq |n^{1/2}\theta_n + U_n| \stackrel{\Pr}{\to} 0$. This implies (i). As to (ii), we have that

$$
\begin{aligned}
n^{1/2}(\hat{F}_2(x) - F(x)) &= n^{1/2}(F_n(x) - F(x)) + (H_2'(0))^{-1}F'(x)U_n^+ \\
&+n^{1/2}(F_n(x - \lambda_n) - F_n(x) - F(x - \lambda_n) + F(x)) \\
&+n^{1/2}(F(x - \lambda_n) - F(x) + \lambda_n F'(x)) \\
&-F'(x)\left(H'(0))^{-1}U_n^+ + n^{1/2}\lambda_n\right) = I + II + III + IV + V.
\end{aligned}
$$

We have that I and II converge jointly and $III, IV, V \stackrel{\Pr}{\to} 0$. □

Proof of Theorem 4.4: By Theorem A.2, $\hat{a}\theta_n + (H_2'(0))^{-1}U_n \stackrel{\Pr}{\to} 0$. This implies (a). For (b), we have that $n^{1/2}(\hat{F}_{2,t}(x) - F(x))= n^{1/2}(F_n(x) - F(x)) + (H_2'(0))^{-1}U_n^+ + n^{1/2}(F_n(x - (1-t)\lambda_n)$ -$F_n(x) - F(x - (1-t)\lambda_n) + F(x)) + n^{1/2}(F(x - (1-t)\lambda_n) - F(x) - (1-t)\lambda_n F'(x))$ -$\left((H_2'(0))^{-1}U_n^+ - n^{1/2}(1-t)\lambda_n F'(x)\right) \equiv I + II + III + IV + V$. We have that I and II converge jointly and $III, IV, V \stackrel{\Pr}{\to} 0$. The rest of the proof follows similarly. □

References

[1] Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rodgers, W. H., Tukey, J. W. (1972). *Robust Estimates of Location: Survey and Advances*, Princeton University Press, New Jersey.

[2] Arcones, M. A. and Samaniego, F. J. (2000). "On the asymptotic distribution theory of a class of consistent estimators of a distribution satisfying a uniform stochastic ordering constraint", *Annals of Statistics*, 28, 116-150.

[3] Arcones, M. A., Kvam, P.H. and Samaniego, F. J. (2000a). "On combining information from ordered experiments", *Proceedings of the Fifth Conference on Applied Statistics*, to appear.

[4] Arcones, M. A., Kvam, P. H. and Samaniego, F. J. (2000b). "Nonparametric estimation of a distribution subject to a stochastic precedence constraint", Department of Statistics, University of California, Technical Report, No. 364.

[5] Arcones, M. A., Kvam, P. H. and Samaniego, F. J. (2001). “A nonparametric test to detect stochastic precedence between two distributions”, Department of Statistics, University of California, Technical Report, No. 399.

[6] Andrews, D. F. (1974). “A robust method for multiple linear regression”. *Technometrics*, 16, 523–531.

[7] Aspin, A. A. (1949). “An examination and further development of a formula arising in the problem of comparing two mean values”. *Biometrika*, 35, 88–96.

[8] Barlow, R. E. and Proschan, F. (1981) *Statistical Theory of Reliability and Life Testing*, Silver Springs, MD: To Begin With.

[9] Birnbaum, Z. W. (1965) “On a use of the Mann-Whitney statistic”. *Proc. of the Third Berkeley Symposium on Prob. and Stat.*, Vol. 1, U.C. Press, 13–17.

[10] Boland, P. J., El-Neweihi, E. and Proschan, F. (1992) “Stochastic order for redundancy allocations in series and parallel systems”. *Adv. Appl. Prob.*, 24, 161–171.

[11] Boyles, R. A. and Samaniego, F. J. (1984) “Estimating a survival curve when new is better than used”, *Operations Research*, 32, 732–40.

[12] Breslow, N. and Crowly, J. (1974). “A large sample study of the life table and product limit estimates under random censorship”. *Annals of Statistics*, 2, 437–453.

[13] Brunk, D. R., Franck, W. E., Hanson, D.L. and Hogg, R. V. (1966) “Maximum likelihood estimation of two distributions of two stochastically ordered random variables”, *Jou. Am. Stat. Assoc.*, 61, 1067–1080.

[14] Church, J. D. and Harris, B. (1970) “The estimation of reliability from stress-strength relationships”, *Technometrics*, 12, 49–54.

[15] Dykstra, R. L. (1982) “Maximum likelihood estimation of the survival functions of stochastically ordered random variables”, *Jou. Am. Stat. Assoc.*, 77, 627–627.

[16] Dykstra, R. L., Kochar, S. and Robertson, T. “Statistical inference for uniform stochastic ordering in several populations”, (1991), *Annals of Statistics*, 19, 870–888.

[17] Gill, R. D. (1981). “Testing replacement and the product limit estimator”. *Annals of Statistics*, 9, 853–860.

[18] Johnson, R. A. (1988). “Stress-strength models for reliability”. In *Handbook of Statistics, Vol. 7: Quality Control and Reliability*, Editors: P.R. Krishniah and C. R. Rao, 27–54. Elsevier, New York.

[19] Grenander, U. (1956). “On the theory of mortality measurement, part 2”, *Scand. Akt.*, 39, 125–53.

[20] Huber, P. J. (1964). “Robust estimation of a location parameter”, *Annals of Mathematical Statistics*, 35, 73–101.

[21] Kochar, S., Mukerjee, H. and Samaniego, F. J. (1999). ”On the Signature of a Coherent System and its application to comparisons among systems” *Naval Research Logistics*, 49, 507–523.

[22] Koroljuk, V. S. and Borovskich, Y. V. (1994). *Theory of U–statistics*. Kluwer Academic Publishers. Dordrecht.

[23] Kvam, P. H. and Samaniego, F. J. (1993). "On the inadmissibility of standard estimators based on ranked set sampling", *Jou. Stat. Plan. & Inf.*, 36, 39–55.

[24] Lehmann, E. L.(1955). "Ordered families of distributions", *The Annals of Mathematical Statistics*, 26, 399–419.

[25] Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd Edition, New York: Wiley and Sons.

[26] Marshall, A. W. and Proschan, F. (1965). "Maximum likelihood estimation for distributions with monotone failure rate", *Annals of Mathematical Statistics*, 36, 69–77.

[27] Mason, R. L., Keating, J. P., Sen, P. K., and Blaylock, N. W. (1990) "Comparison of linear estimators using Pitman's measure of closeness", *Jou. Am. Stat. Assoc.*, 85, 579–581.

[28] McConnel, T. R. (1987). "A two–parameter maximal ergodic theorem with dependence", *Annals of Probability*, 15, 1569–1585.

[29] McNichols, D. T. and Padgett, W. J. (1984). "Nonparametric estimation from accelerated life tests with random censorship", in *Reliability Theory and Models*, eds. M.S. Abdel-Hameed, E. Cinlar, and J. Quinn, Orlando: Academic Press.

[30] Mukerjee, H. (1996). "Estimation of survival functions under uniform stochastic ordering", *Jou. Am. Stat. Assoc.*, 91, 1684–89.

[31] Nelson, W. (1990). *Accelerated Testing: Statistical Models, Test Plans and Data Analysis*, New York: Wiley and Sons.

[32] Robertson, T., Wright, F.T. and Dykstra, R.L. (1988). *Order Restricted Statistical Inference*, New York: John Wiley and Sons.

[33] Rojo, J. and Samaniego, F. J. (1991). "On nonparametric maximum likelihood estimation of a distribution uniformly stochastically smaller than a standard", *Statistics and Probability Letters*, 11, 267–271.

[34] Rojo, J. and Samaniego, F. J. (1993). "On estimating a survival curve subject to a uniform stochastic ordering constraint", *Jou. Am. Stat. Assoc.*, 88, no. 422, 566–572.

[35] Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic Orders and their Applications*, Academic Press, Inc., Boston.

[36] Singh, H. and Misra, N. (1994). "On redundancy allocations in systems". *Journal of Applied Probability*, 31, 1004–1014.

[37] Trickett, W. H., Welch, B. L. and James, G. S. (1956). "Further critical values for the two-means problem". *Biometrika*, 43, 203–205.

Figure 3: Relative Error of $\hat{F}_1$, $\hat{F}_2$ with respect to F_n for $X \sim$ Gamma$(r, \hat{\lambda})$ and $Y \sim$ Exponential(1).

Figure 4: IMSE for $\hat{F}_1$ (solid line), $\hat{F}_2$ (curved dotted line) and R_n (dashed line) based on Weibull data.

Figure 5: IMSE for $\hat{F}_1$ (solid line), $\hat{F}_2$ (curved dotted line) and R_n (dashed line) based on Lognormal data with n=20.