Using Auxiliary Information in Statistical Function Estimation

By SERGEY TARIMA
Department of Statistics, University of Kentucky, Lexington, Kentucky, 40506-0027, U.S.A.
stari@ms.uky.edu

DMITRI PAVLOV
Biostatistics and Reporting, Pfizer Inc
50 Pequot Ave, New London, CT 06320, U.S.A.
dmitri.pavlov@groton.pfizer.com

Abstract

In many cases, auxiliary information is available from previous experiments. In the paper a general method of using auxiliary information is considered. It is shown that properties of standard empirical estimates can be improved by using additional knowledge. As an example, an improved empirical distribution estimate is given. It is demonstrated that it has faster rate of convergence. The method provides substantially more efficient confidence intervals than the standard procedures. It is shown by simulation that the theoretical results are valid even for moderate sample sizes.

Some key words: Auxiliary information; Non-parametric statistics

1. Introduction

Auxiliary information can be used in statistical function estimation in order to improve properties of standard procedures. Different methods of incorporating auxiliary information were suggested by a number of researchers (Dmitriev, Gal’chenko, Gurevich, Holt, Kuk, Pugachev, Zhang) and in most cases this information can be available in the form of prior knowledge about the estimated distributions, parameters, etc. Those methods deal mostly with auxiliary information of deterministic nature. For example, Chambers and Dunstan (1973) considered quantile estimation in presence of auxiliary variable when values of that auxiliary variable were known. The same “ultimate” knowledge on auxiliary variable was used by Holt (1991) to modify estimators derived on the data inflicted by non-responses. The above results are not applicable when the exact auxiliary information cannot be obtained. In most real problems auxiliary information come from experts or derived from previous experiments. But there are very few papers provide methods on how to use auxiliary information in a form of statistical estimates from the previous experiments. For example, Kuk and Mak (1989) used the median derived from an independent previous sample to improve the standard median estimator.

In 1973 Pugachev used effect of linear correlation between auxiliary and empirical information to derive modified estimators. His approach used exact knowledge of means of some statistical functions correlated to estimable one. Gal’chenko and Gurevich (1991) suggested modification of Pugachev’s method when auxiliary information is not known exactly but present in a form of statistical estimates from previous experiments. This paper suggests a new methodology of using auxiliary information in a form of statistical estimates derived from previous experiments which generalizes known approach of Gal’chenko and Gurevich. The method is applied to modify empirical cumulative distribution function and Nelson Cumulative Hazard Estimator.
2. Methodology

Consider problem of statistical estimation of $H(t) = \int_R \varphi(t,x)dF_X(x)$, where $\varphi(t,x)$ is a known function defined on $(R \times R)$ and $F_X(x)$ is an unknown cumulative distribution function of a random variable $X$ defined on real line. In order to estimate $H(t)$ a sample of size $n$ was obtained. For simplicity this sample will be called “current” sample. Let $\hat{H}_n(t)$ be a consistent and asymptotically normal estimator of $H(t)$ based on this sample.

Auxiliary information in form of vector of estimates $\tilde{H}_m(s) = (\tilde{H}_m(s_1), \ldots, \tilde{H}_m(s_k))$ was obtained independently from “current” sample, where subscript $m$ reflects the size of a sample used to derive these estimates and $-\infty < s_1 < \cdots < s_k < +\infty$. Again for simplicity this sample of size $m$ will be referred as "previous" sample.

Remark: In the notation for these estimators “hat” corresponds to one estimator’s type (e.g. empirical estimator), ”tilde” - to (may be, but not necessary) another (e.g. smoothed estimator).

Picking up the sum of squares as a loss function risk function becomes a mean square error, which is used in the paper as a measure of estimators’ optimality. Extracting an estimator with minimal mean square error may be conducted in the family of estimators

$$\hat{H}^\Lambda(t) = \hat{H}_n(t) + \Lambda(\Delta\tilde{H}_n(s) - \Delta\tilde{H}_m(s))^T,$$

where $\Lambda = (\lambda_1, \ldots, \lambda_k)$ is a vector of parameters corresponding to all possible estimators in family (1), $\Delta\tilde{H}_n(s) = (\tilde{H}_n(s_1), \ldots, \tilde{H}_n(s_k))$, $\Delta\tilde{H}_m(s) = (\tilde{H}_m(s_1), \ldots, \tilde{H}_m(s_k))$, $\Delta\hat{H}_n(s_i) = \hat{H}_n(s_i) - H_n(s_i - 1)$, $\Delta\hat{H}_m(s_i) = \hat{H}_m(s_i) - H_m(s_i - 1)$, $i = 2, \ldots, k$, $H_m(s_0) = H_n(s_0) = 0$, $T$ - transpose operator.

Let also $K_{ts} = \|\text{cov}(\hat{H}_n(t), \Delta\tilde{H}_n(s_i) - \Delta\tilde{H}_m(s_i))\|_{i=1,\ldots,k}$ and $K_{ss} = \|\text{cov}(\Delta\tilde{H}_n(s_i), \Delta\tilde{H}_m(s_j) - \Delta\hat{H}_m(s_j))\|_{i,j=1,\ldots,k}$ then mean square error of $\hat{H}^\Lambda(t)$ is

$$MSE[\hat{H}^\Lambda(t)] = MSE[\hat{H}(t)] + 2\Lambda K_{ts} + \Lambda K_{ss} \Lambda^T.$$  

Condition for extremum is $\nabla_\Lambda MSE[\hat{H}^\Lambda(t)] \equiv 0$, where $\nabla_\Lambda$ is a gradient vector, then

$$\nabla_\Lambda MSE[\hat{H}^\Lambda(t)] = 2K_{ts} + 2\Lambda K_{ss} \equiv 0.$$  

It can be shown that

$$\Lambda_0 = -K_{ts}K_{ss}^{-1}$$

be the "optimal" vector $\Lambda_0$ corresponding to the estimator

$$\hat{H}^{\Lambda_0}(t) = \hat{H}(t) - K_{ts}K_{ss}^{-1}(\Delta\tilde{H}_n(s) - \Delta\tilde{H}_m(s))$$

with

$$MSE(\hat{H}^{\Lambda_0}(t)) = MSE(\hat{H}(t)) - K_{ts}K_{ss}^{-1}K_{ts}^T$$

minimal in family (1).

Covariance matrix $K_{ts}$ is non-negatively defined. But from (4) notice that $\text{det}(K_{ss}) \neq 0$ and then matrix $K_{ss}$ should be positively defined. In equation (6) $K_{ts}K_{ss}^{-1}K_{ts}^T$ is a quadratic form and therefore is always nonnegative. Moreover $K_{ts}K_{ss}^{-1}K_{ts}^T = 0$ if and only if $K_{ts} \equiv 0$.

In practice researcher usually doesn’t really know values of $K_{ts}$ and $K_{ss}$ and should use their estimators. Consider

$$K_{nts} = \|\text{cov}_n(\tilde{H}_n(t), \Delta\tilde{H}_n(s_i) - \Delta\tilde{H}_m(s_i))\|_{i=1,\ldots,k}$$

an estimator for $K_{ts}$ and

$$K_{nss} = \|\text{cov}(\Delta\tilde{H}_n(s_i) - \Delta\tilde{H}_m(s_i), \Delta\tilde{H}_n(s_j) - \Delta\tilde{H}_m(s_j))\|_{i,j=1,\ldots,k}$$
an estimator for $K_{ss}$.

Remark: Subscript $n$ in $\text{cov}_n$ defines only that this covariance estimator derived from "current" sample and does not define type of this estimator. Among desired properties of this estimator consistency will be pointed. Its consistency will be used later in Proposition 2.

Applying $K_{nts}$ and $K_{nss}$ instead of unknown $K_{ts}$ and $K_{ss}$ modifies (5) and (6) to the following forms

$$\hat{\Delta}_{\lambda n}(t) = \hat{\Delta}_{n}(t) - K_{nts}K_{nss}^{-1}(\Delta \hat{H}_n(s) - \Delta \hat{H}_n(s))$$

and

$$MSE_n(\hat{\Delta}_{\lambda n}(t)) = MSE_n(\hat{\Delta}_{n}(t)) - K_{nts}K_{nss}^{-1}K_{nts}^T.$$

The term (1 + $w^2$)$^{-1}$$_{C_{ts}C_{ss}^{-1}C_{ts}^T}$ is a non-negative one and its value depends on sample sizes only through $w$. If $w$ much larger than 1 then decrease of mean square error is very small and asymptotical properties of new estimator are almost the same as unmodified estimator has. If $w$ is close to 0 then auxiliary information derived from "previous" sample almost true and asymptotical properties of new estimator close to the case of incorporating information of exact knowledge.

Proposition 1 showed asymptotical properties of estimator (5) but (5) cannot be used in most practical cases because $\Lambda_0$ usually is not known. In these situations estimator (7) should be applied and its asymptotical properties described in

Proposition 2. Suppose assumptions of Proposition 1 hold and

1) $a_n^2(\text{cov}_n(\hat{H}_n(t), \Delta \hat{H}_n(s_i)) - \text{cov}(\hat{H}_n(t), \Delta \hat{H}_n(s_i)))$ are asymptotically normal with mean 0 and finite variance (which can be zero), $i = 1, \ldots, k$;

2) $a_n^2(\text{cov}_n(\Delta \hat{H}_n(s_i), \Delta \hat{H}_n(s_j)) - \text{cov}(\Delta \hat{H}_n(s_i), \Delta \hat{H}_n(s_j)))$ are asymptotically normal with mean 0 and finite variance (which can be zero), $i, j = 1, \ldots, k$. 


Then \( a_n \left( \hat{H}_A(t) - H(t) \right) \rightarrow \eta(t) \), as \( n \rightarrow \infty \), where \( \eta(t) \) defined in Proposition 1.

Remarks:
1) In Proposition 2 there are two additional assumptions on convergence of covariances’ estimators. If these assumptions do not hold asymptotical properties of estimator (7) can not only differ from these of (5) but also be even worse (in sense of asymptotical variance) when of unmodified estimator.

2) For the cases when \( w = 0 \) (7) become the same as the estimator derived by method of correlated processes (Pugachev, 1973) and has the same asymptotical properties as empirical likelihood estimator in the presence of auxiliary information of the same type (Zhang, 1996).

3. ”Memoryless” case

Assume that \( H_n(t_2) - H_n(t_1) \) does not correlate with \( H_n(t_4) - H_n(t_3) \) for arbitrary \( t_1 < t_2 < t_3 < t_4 \), that is

\[
E((H_n(t_2) - H_n(t_1)) \cdot (H_n(t_4) - H_n(t_3))) = 0.
\]

Then

\[
E(\hat{H}_n(x) \cdot \hat{H}_n(y)) = E(\hat{H}_n(\min(x, y))\hat{H}_n(\min(x, y)))
\]

for any positive \( x \) and \( y \).

With these assumptions for \( H_n(t) \)

\[
K_{ts} = \left\| \text{cov} \left( \hat{H}_n(t), \Delta \hat{H}_n(s_i) - \Delta \hat{H}_m(s_i) \right) \right\|_{i=1,...,k} = \left\| \text{cov} \left( \hat{H}_n(t), \Delta \hat{H}_n(s_i) \right) \right\|_{i=1,...,k} = 0
\]

and \( K_{ss} \) becomes diagonal matrix

\[
K_{ss} = \left\| \text{cov} \left( \Delta \hat{H}_n(s_i), \Delta \hat{H}_n(s_j) - \Delta \hat{H}_m(s_j) \right) \right\|_{i,j=1,...,k} =
\]

\[
\left\| \text{cov} \left( \Delta \hat{H}_n(s_i), \Delta \hat{H}_n(s_j) \right) \right\|_{i,j=1,...,k} =
\]

\[
\left\| \text{cov} \left( \Delta \hat{H}_n(s_i), \Delta \hat{H}_n(s_j) \right) \right\|_{i=1,...,k}.
\]

In this case estimator (7) may be rewritten as

\[
\hat{H}_A(t) = \hat{H}(t) - \sum_{j=1}^{k} \left[ \Delta \hat{H}(s_j) - \Delta \hat{H}(s_j) \right] \times
\]

\[
\frac{\text{cov} \left( \hat{H}_n(\min(t, s_i)), \hat{H}_n(\min(t, s_i)) \right) - \text{cov} \left( \hat{H}_n(\min(t, s_{i-1})), \hat{H}_n(\min(t, s_{i-1})) \right)}{\text{cov} \left( \Delta \hat{H}_n(s_i), \Delta \hat{H}_n(s_j) \right) + \text{cov} \left( \Delta \hat{H}_m(s_i), \Delta \hat{H}_m(s_j) \right)}
\]

and mean square error (8) as

\[
\text{MSE}(\hat{H}_A(t)) = \text{MSE}(\hat{H}(t)) -
\]

\[
- \sum_{j=1}^{k} \left[ \frac{\text{cov} \left( \hat{H}_n(\min(t, s_i)), \hat{H}_n(\min(t, s_i)) \right) - \text{cov} \left( \hat{H}_n(\min(t, s_{i-1})), \hat{H}_n(\min(t, s_{i-1})) \right)}{\text{cov} \left( \Delta \hat{H}_n(s_i), \Delta \hat{H}_n(s_j) \right) + \text{cov} \left( \Delta \hat{H}_m(s_i), \Delta \hat{H}_m(s_j) \right)} \right]^2.
\]
Adaptive estimator (7) can be used in this situation which transform (15) to form cumulative distribution function. According to Proposition 2 estimator (17) has the same asymptotical properties as (15) but \( \lambda \) cannot be used in all possible cases because as only \( F_n(s) = 0 \) or \( F_n(s) = 1 \) denominator and estimator with minimal mean square error in (13) is

\[
\text{MSE}(\hat{F}_n(s)) = \text{MSE}(\hat{F}_m(s)) = \text{MSE}(\hat{F}_n(s)) - \text{MSE}(\hat{F}_m(s) - \text{MSE}(\hat{F}_n(s-1))
\]

and in case of unbiasedness of involved estimators mean square errors become variances. Remarks: 1) Notice that denominators in (11) and (12) can be transformed in the following way

\[
cov(\Delta \hat{H}_n(s_i), \Delta \hat{H}_n(s_i)) + cov(\Delta \hat{H}_m(s_i), \Delta \hat{H}_m(s_i)) =
\]

\[
= \text{MSE}(\Delta \hat{H}_n(s_i)) + \text{MSE}(\Delta \hat{H}_m(s_i)) = \text{MSE}(\hat{F}_n(s)) - \text{MSE}(\hat{F}_m(s)) + \text{MSE}(\hat{F}_m(s)-\text{MSE}(\hat{F}_n(s-1))
\]

2) If “hat” and “tilde” correspond to the estimators of the same type then covariances in nominators of (11) and (12) become mean square errors (variances in unbiased case).

4. Empirical cumulative distribution function modification

Suppose \( X_1, \ldots, X_n \) are independent and identically distributed random variables with an unknown cumulative distribution function \( F(t) \) \( t \in (-\infty, \infty) \). Consider problem of non-parametric estimation of \( F(t) \) in presence of auxiliary information. In addition to “current” sample \( X_1, \ldots, X_n \) statistic \( F_n(s) \) was derived from “previous” sample \( Y_1, \ldots, Y_m, \) where \( F_n(s) = \frac{m}{n+m} \sum_{i=1}^m I_{(-\infty,s)}(Y_i) \) and \( Y_1, \ldots, Y_m \) have the same distribution as \( X_1, \ldots, X_n \). Indicator function \( I_A(X) \) equals to 1 if \( X \in A \) and 0 otherwise. Notice also here that only statistic \( F_n(s) \) is known not all “previous” sample.

All methodology described in first section can be applied in this section with following details

\[
H(t) = \int_R \varphi(t,x)dF(x) = \int_R I_{(-\infty,t)}(x)dF(x) = F(t), k = 1, \hat{H}_n(t) = F_n(t), \hat{H}_m(s) = F_m(s),
\]

In order to modify empirical cumulative distribution function \( F_n(t) \) consider family

\[
F_n^\lambda(t) = F_n(t) + \lambda(F_n(s) - F_m(s))
\]

which is a special case of (1).

Optimal \( \lambda_0 \) defined by (4) take a form

\[
\lambda_0 = \frac{m}{n+m} \frac{F(\min(s,t)) - F(s)F(t)}{F(s)(1-F(s))}
\]

and estimator with minimal mean square error in (13) is

\[
F_n^{\lambda_0}(t) = F_n(t) - \frac{m}{n+m} \frac{F(\min(s,t)) - F(s)F(t)}{F(s)(1-F(s))}(F_n(s) - F_m(s)).
\]

\( F_n(t) \) is an unbiased estimator and as a result all family (13) consists of unbiased estimators and their mean square errors become variances. Variance of (15) is

\[
var \left( F_n^{\lambda_0}(t) \right) = \frac{F(t)(1-F(t))}{n} - \frac{m}{n+m} \frac{(F(\min(s,t)) - F(s)F(t))^2}{F(s)(1-F(s))}.
\]

But in \( \lambda_0 \) in (14) is not known and (15) with (16) cannot be used without proper modification. Adaptive estimator (7) can be used in this situation which transform (15) to form

\[
F_n^{\lambda_\alpha}(t) = F_n(t) - \frac{m}{n+m} \frac{F_n(\min(s,t)) - F_n(s)F_n(t)}{F_n(s)(1-F_n(s))}(F_n(s) - F_m(s)).
\]

According to Proposition 2 estimator (17) has the same asymptotical properties as (15) but \( F_n^{\lambda_\alpha}(t) \) cannot be used in all possible cases because as only \( F_n(s) = 0 \) or \( F_n(s) = 1 \) denominator and nominator in (14) become 0. Different amendments can be used in this case, for example, applying
When \( \lambda_n = 0/0 \) resolves this problem.

If \( t \leq s \) then (17) can be transformed to the form

\[
F_n^{\lambda_0}(t) = F_n(t) F_n(s)
\]

(18)

and if \( t > s \) (17) becomes

\[
F_n^{\lambda_0}(t) = F_n(s) + [1 - F_n(s)] \frac{F_n(t) - F_n(s)}{1 - F_n(s)}.
\]

(19)

From (18) and (19) derive representation of (17)

\[
F_n^{\lambda_0}(t) = F_n(s) \frac{F_n(\min(t, s))}{F_n(s)} + [1 - F_n(s)] \frac{F_n(\max(t, s)) - F_n(s)}{1 - F_n(s)}.
\]

(20)

Estimator (20) was also derived by Little and Rubin when they applied maximum likelihood ratio to monotone missing data with ignorable mechanism of missing data generation.

When the \( m(m+n)^{-1} \) goes to 0 estimator (17) converges to empirical one. When the \( m(m+n)^{-1} \) goes to 1, this case becomes similar to the case when the auxiliary information is known exactly. Applying the ultimate case when \( m = +\infty \) and \( n \) is finite \( m(m+n)^{-1} = 1 \) (17) and (20) become

\[
F_n^{\lambda_0}(t) = F_n(t) - F_n(s) F_n(s) (1 - F_n(s))
\]

(21)

and

\[
F_n^{\lambda_0}(t) = F(s) F_n(\min(t, s)) + [1 - F(s)] \frac{F_n(\max(t, s)) - F_n(s)}{1 - F_n(s)}.
\]

(22)

Notice that (22) is also a non-parametric maximum likelihood estimator constructed with a prior knowledge of \( F(s) \) (Owen, 2001).

In order to show graphically how modified estimator (17) differs from empirical one consider the following example.

**Example 4.1.** Suppose "current" and "previous" samples were obtained from \( F(t) = 1 - \exp(-t) \). Auxiliary information is available in the form of the following statistic: \( F_m(1) (s = 1, k = 1) \) based on "previous" sample. The following three experiments were conducted: 1) \( n=50, m=150 \) 2) \( n=50, m=10 \) 3) \( n=50, m=10000 \)
On Pictures 1,3,5 empirical estimators with their 95% confidence intervals are shown. Pictures 2,4,6 correspond to modified estimators (with their confidence intervals) based on the same samples as empirical ones on pictures 1,3,5 correspondingly.

Picture 2 shows the closer \( t \) to \( s \) (and the stronger correlation between \( H_n(t) \) and \( H_m(s) \)) the smaller confidence interval becomes and the biggest influence from auxiliary information came at point \( s \).

If the size of "previous" sample \( (m) \) greatly overwhelms the size of "current" sample \( (n) \) then picture 6 describes such a situation.

Pictures 1 - 6 show that the best improvement of confidence interval corresponds to \( t = 1 \) because in this time point the strongest correlation \( F_m(t) \) and \( F_n(t) \) shows up. Pictures 1 and 2 were plotted on the basis of only one simulation.

All methodology in the first section can be applied not only to empirical but also to the other estimators. Consider, for example, modification of Nelson cumulative hazard estimator.

5. Modification of Nelson Cumulative Hazard Estimator

Let \( Y_1, \ldots, Y_n \) be independent and identically distributed random variables with an unknown \( F_Y(t) \) \((t \in [0, \infty))\), \( C_1, \ldots, C_n \) - with unknown \( F_C(t) \) \((t \in [0, \infty))\). If \( T_j = \min(Y_j, C_j) \) and \( \delta_j = I(Y_j \leq C_j) \), for \( j = 1, 2, \ldots, n \), then paired observations \((T_1, \delta_1), \ldots, (T_n, \delta_n)\) are called right censored data. One
may take interest in the cumulative hazard function $H(t) = \int_0^t (1 - F_Y(x-))^{-1} dF_Y(x)$ estimation, where $F(x-) = \lim_{\varepsilon \to 0} F(x-\varepsilon)$, $\varepsilon > 0$. For this purpose well known estimators may be used. But it is possible to improve them when extra information is known. Consider Nelson estimator of cumulative hazard function.

$$\hat{H}_n(t) = H_{n, NA}(t) = \sum_{x \leq t} \frac{d_n(x)}{R_n(x)},$$

(23)

where $d_n(x) = \sum_{i=1}^n I(Y_i = x) \cdot \delta_i$ (the number of uncensored events $Y_i$ registered at $x$) and $R_n(x) = \sum_{i=1}^n I(Y_i \geq x)$ (the number of events $Y_i$ registered at and after $x$) and Aalen’s estimator for its mean square error is

$$MSE_n(\hat{H}_n(t)) = MSE_n(\hat{H}_{n, NA}(t)) = \sum_{x \leq t} \frac{d_n(x)}{R_n^2(x)}.$$

(24)

Fleming (1991) is a good reference concerning estimators (23) and (24).

Suppose asymptotically normal statistical estimates

$$\tilde{H}_m(s_i) = H_{m, NA}(s_i) = \sum_{x \leq t} \frac{d_m(x)}{R_m(x)}$$

of $H(t)$ were derived from ”previous” sample and covariances’ estimates (satisfying to proposition 2)

$$cov_n(\hat{H}_m(s_i), \hat{H}_m(s_j)) = MSE_n(\hat{H}_{m, NA}(\min(s_i, s_j)))$$

and

$$cov_n(H_n(s_i), H_n(s_j)) = MSE_n(\hat{H}_{n, NA}(\min(s_i, s_j)))$$

can be calculated on the basis of ”current” sample, where $i, j = 1, \ldots, k$ (see equation (31)). It is reasonable to propose that this extra data is independent from $(T_1, \delta_1), \ldots, (T_n, \delta_n)$.

It can be shown that assumptions for ”memoryless” situation (9) hold when $H_n = H_{NA}$. Applying (23) and (24) to (11) receive modification of Nelson estimator

$$H^*_{NA}(t) = \sum_{x \leq t} \frac{d_n(x)}{R_n(x)} - \frac{m}{n + m} \sum_{j=1}^k \left( \sum_{\min(t, s_{j-1}) \leq x \leq \min(t, s_j)} \frac{d_n(x)}{R_n^2(x)} \left( \sum_{s_{j-1} \leq x \leq s_j} \frac{d_n(x)}{R_n^2(x)} \right)^{-1} \right) \times$$

$$\times \left( \sum_{s_{j-1} \leq x \leq s_j} \frac{d_n(x)}{R_n(x)} - \sum_{s_{j-1} \leq x \leq s_j} \frac{d_m(x)}{R_m(x)} \right)$$

(25)

and applying (23) and (24) to (11)

$$MSE_n(H^*_{NA}(t)) = \sum_{x \leq t} \frac{d_n(x)}{R^2_n(x)} - \frac{m}{n + m} \sum_{j=1}^k \left( \sum_{\min(t, s_{j-1}) \leq x \leq \min(t, s_j)} \frac{d_n(x)}{R_n^2(x)} \right)^2 \left( \sum_{s_{j-1} \leq x \leq s_j} \frac{d_n(x)}{R_n^2(x)} \right)^{-1}$$

(26)
6. Conclusion

A problem of using auxiliary information in statistical function estimation was considered. It is assumed that auxiliary information is presented in the form of statistics obtained from previous experiments. New estimators have been developed which are generalizations of Gal’chenko and Gurevich results for a wider (arbitrary convergence rate) class of auxiliary information. The given theorems provide asymptotic properties of these estimators such as consistency, normality, etc. The obtained results were used to build modified empirical estimators of cumulative distribution function and cumulative hazard function. It is shown that the suggested estimators are consistent, asymptotically normal. Also, they way of estimation produces better results in terms of variance. The discussed methodology does not require strong assumptions and it can be used to improve properties of other widely-used procedures.

7. Appendix

Proof of Proposition 1. From the first assumption \( \hat{H}_n(t) \longrightarrow H(t) \) in probability, as \( n \to \infty \), from the second \( \hat{H}_n(s_j) \longrightarrow H(s_j) \) , as \( n \to \infty \), for all \( j \), and from second and fourth \( \hat{H}_{f(n)}(s_j) \longrightarrow H(s_j) \), as \( n \to \infty \), for all \( j \). So \( \hat{H}_n(t) \) is consistent.

Notice that

\[
\frac{b_n}{f(n)} \left( \Delta \hat{H}_n(s_j) - \Delta H(s_j) \right) = \zeta_n(s_j) - \zeta_n(s_j - 1) = \Delta \zeta_n(s_j)
\]

and

\[
\frac{b_n}{f(n)} \left( \Delta \hat{H}_{f(n)}(s_j) - \Delta H(s_j) \right) = \frac{b_n}{f(n)} \left( \zeta_{f(n)}(s_j) - \zeta_{f(n)}(s_j - 1) \right) = \frac{b_n}{f(n)} \Delta \zeta_{f(n)}(s_j)
\]

then \( \eta_n(t) \) may be rewritten as

\[
\eta_n(t) = \xi_n(t) - \frac{a_n}{b_n} K_{ts} K_{ss}^{-1} \left[ \Delta \zeta_n(s_j) - \frac{b_n}{f(n)} \Delta \zeta_{f(n)}(s_j) \right]_{j=1, \ldots, k}.
\]

In condition when \( K_{ss} \) is positively defined

\[
K_{ts} K_{ss}^{-1} = \left\| \text{cov}(\hat{H}_n(t), \Delta \hat{H}_n(s_i) - \Delta \hat{H}_m(s_i)) \right\|_{i=1, \ldots, k} \times
\]

\[
\times \left\| \text{cov}(\Delta \hat{H}_n(s_i) - \Delta \hat{H}_m(s_i), \Delta \hat{H}_n(s_j) - \Delta \hat{H}_m(s_j)) \right\|_{i,j=1, \ldots, k}^{-1}.
\]

First consider the following representation of elements of matrix \( K_{ts} \)

\[
\text{cov}(\hat{H}_n(t), \Delta \hat{H}_n(s_i) - \Delta \hat{H}_m(s_i)) = \text{cov}(\hat{H}_n(t), \Delta \hat{H}_n(s_i)) - \text{cov}(\hat{H}_n(t), \hat{H}_n(s_i - 1)) - \text{cov}(\hat{H}_n(t), \hat{H}_m(s_i)) + \text{cov}(\hat{H}_n(t), \hat{H}_m(s_i - 1)),
\]

where covariances may be represented as

\[
\text{cov}(\hat{H}_n(t), \hat{H}_n(s_i)) = \text{cov}(\hat{H}_n(t) - H(t), \hat{H}_n(s_i) - H(s_i)) = \frac{1}{a_n b_n} \text{cov}(\xi_n(t), \zeta_n(s_i)).
\]

So

\[
\text{cov}(\hat{H}_n(t), \hat{H}_n(s_i)) = \frac{1}{a_n b_n} \text{cov}(\xi_n(t), \zeta_n(s_i)).
\]

By analogy find

\[
\text{cov}(\hat{H}_n(t), \hat{H}_n(s_i - 1)) = \frac{1}{a_n b_n} \text{cov}(\xi_n(t), \zeta_n(s_i - 1)).
\]
\begin{align}
\text{cov}(\hat{H}_n(t), \hat{H}_m(s_i)) &= \frac{1}{a_nb_n} \text{cov}(\xi_n(t), \zeta_m(s_i)) = 0, \\
\text{cov}(\hat{H}_n(t), \hat{H}_m(s_i-1)) &= \frac{1}{a_nb_n} \text{cov}(\xi_n(t), \zeta_m(s_i-1)) = 0.
\end{align}

(30)\hspace{1cm}(31)

Zeroes in (30) and (31) came from independence of “current” and “previous” samples.

Then the elements of $K_{ts}$ may be represented as

\begin{align}
\text{cov}(\hat{H}_n(t), \Delta \hat{H}_n(s_i) - \Delta \hat{H}_m(s_i)) &= \frac{1}{a_nb_n} \text{cov}(\xi_n(t), \Delta \zeta_m(s_i)).
\end{align}

Use similar technique to find the following representation of elements of $K_{ss}$

\begin{align}
\text{cov}(\Delta \hat{H}_n(s_i) - \Delta \hat{H}_m(s_i), \Delta \hat{H}_n(s_j) - \Delta \hat{H}_m(s_j)) &= \frac{1}{b_n} \text{cov}(\Delta \zeta_n(s_i), \Delta \zeta_n(s_j)) + \frac{1}{b_m} \text{cov}(\Delta \zeta_m(s_i), \Delta \zeta_m(s_j)).
\end{align}

Now consider asymptotical situation, where $\frac{b_n}{b_m} = \frac{b_n}{\sigma^2(s_i)} \rightarrow w \in [0, +\infty) , \text{ as } n \rightarrow \infty$.

Then

\begin{align}
\eta_n(t) &\rightarrow \eta(t) = \xi(t) - ||\text{cov}(\xi(t), \Delta \zeta(s_i))||_{i=1, \ldots, k} \\
&\times ||\text{cov}(\Delta \zeta(s_i), \Delta \zeta(s_j)) + w^2 \cdot \text{cov}(\Delta \zeta(s_i), \Delta \zeta(s_j))||_{i,j=1, \ldots, k} \\
&\times ||\text{cov}(\Delta \zeta(s_i), \Delta \zeta(s_j))||_{i,j=1, \ldots, k} = \\
&\xi(t) - ||\text{cov}(\xi(t), \Delta \zeta(s_i))||_{i=1, \ldots, k} (1 + w^2) \cdot ||\text{cov}(\Delta \zeta(s_i), \Delta \zeta(s_j))||_{i,j=1, \ldots, k} \\
&\times ||\text{cov}(\xi(t), \Delta \zeta(s_i))||_{i=1, \ldots, k}.
\end{align}

Notice here that $\Delta \zeta_1(\cdot)$ and $\Delta \zeta_2(\cdot)$ are independent random vectors with the same $k$-dimensional distribution function as $\Delta \zeta(\cdot)$.

$\eta(t)$ is a normal random variable because it is a linear combination of normal random variables. $E(\eta(t)) = 0$ and after routine algebraic transformations find

\begin{align}
\text{var}(\eta(t)) &= \text{var}(\xi(t)) - ||\text{cov}(\xi(t), \Delta \zeta(s_i))||_{i=1, \ldots, k} \\
&\times ||\text{cov}(\Delta \zeta(s_i), \Delta \zeta(s_j)) + w^2 \cdot \text{cov}(\Delta \zeta(s_i), \Delta \zeta(s_j))||_{i,j=1, \ldots, k} \\
&\times ||\text{cov}(\Delta \zeta(s_i), \Delta \zeta(s_j))||_{i,j=1, \ldots, k} = \\
&\text{var}(\xi(t)) - \frac{1}{1 + w^2} C_{ts} C_{ts}^{-1} C_{ts}^T.
\end{align}

Q.E.D.

**Proof of Proposition 2.** Notice that $K_{nts} = K_{nts}^{-1}$ is a continuous function of $\text{cov}_n(\hat{H}_n(t), \Delta \hat{H}_n(s_i) - \Delta \hat{H}_m(s_i))$, and $\text{cov}_n(\Delta \hat{H}_n(s_i) - \Delta \hat{H}_m(s_i), \Delta \hat{H}_n(s_j) - \Delta \hat{H}_m(s_j))$, $i, j = 1, \ldots, k$, when $\text{det}(K_{nts}) > 0$. Probability of the cases when $\text{det}(K_{nts}) = 0$ goes to zero as $n$ goes to infinity and hence does not change asymptotical properties.

Consider Taylor expansion of $K_{nts} = K_{nts}^{-1}$ with respect to $\text{cov}_n(\cdot, \cdot)$ in a neighborhood of the true $\text{cov}(\cdot, \cdot)$, where $\text{cov}(\cdot, \cdot)$ is a $k + k^2$-dimensional vector with elements $\text{cov}(\hat{H}_n(t), \Delta \hat{H}_m(s_i) - \Delta \hat{H}_m(s_i))$, $i = 1, \ldots, k$, and $\text{cov}(\Delta \hat{H}_n(s_i) - \Delta \hat{H}_m(s_i), \Delta \hat{H}_n(s_j) - \Delta \hat{H}_m(s_j))$, $i, j = 1, \ldots, k$.

Then $K_{nts} = K_{ts}K_{ss}^{-1} + O_n(\text{cov}_n(\cdot, \cdot) - \text{cov}(\cdot, \cdot))$ or $K_{nts} - K_{ts}K_{ss}^{-1} = O_n(\text{cov}_n(\cdot, \cdot) - \text{cov}(\cdot, \cdot))$ which goes to normal random variable with mean 0 and variance 0, or to zero in probability.

Now

\begin{align}
&\frac{a_n(\hat{H}_n(t) - \hat{H}(t))}{\text{var}(\eta(t))} = \frac{a_n(\hat{H}_n(t) - \hat{H}_n(t))}{\text{var}(\eta(t))} + \frac{a_n(\hat{H}_n(t) - \hat{H}_n(t))}{\text{var}(\eta(t))} \\
&= (K_{ts}K_{ss}^{-1} - K_{nts}K_{nts}^{-1}) (\Delta \hat{H}_n(s) - \Delta \hat{H}_m(s)) a_n + a_n (\hat{H}_n(t) - \hat{H}(t)).
\end{align}

(33)
In (33) $K_{ts}K_{s}^{-1} - K_{nts}K_{nss}^{-1}$ goes to zero in probability and $a_n \left( \Delta \hat{H}_n(s) - \Delta \hat{H}_m(s) \right)$ is a consistent estimator of 0.

Hence $a_n \left( K_{ts}K_{s}^{-1} - K_{nts}K_{nss}^{-1} \right) \left( \Delta \hat{H}_n(s) - \Delta \hat{H}_m(s) \right)$ goes to zero in probability and asymptotical distributions for $a_n \left( \hat{H}^{\lambda_0}(t) - H(t) \right)$ and $a_n \left( \hat{H}^{\lambda_0}(t) - H(t) \right)$ are the same. Q.E.D.

8. References


