NONPARAMETRIC ESTIMATION OF $\mathbb{P}(X < Y)$ WITH LAPLACE ERROR DENSITIES

DANG DUC TRONG¹, TON THAT QUANG NGUYEN¹²

¹Faculty of Mathematics and Computer Science, VNUHCM-University of Science
²Faculty of Fundamental Science, Industrial University of Ho Chi Minh City
tonthatquangnguyen@iuh.edu.vn

Abstract. We survey the nonparametric estimation of the probability $\theta := \mathbb{P}(X < Y)$ when two random variables $X$ and $Y$ are observed with additional errors. Specifically, from the noise versions $X'_1, \ldots, X'_n$ of $X$ and $Y'_1, \ldots, Y'_m$ of $Y$, we introduce an estimator $\hat{\theta}$ of $\theta$ and then establish the mean consistency for the suggested estimator when the error random variables have the Laplace distribution. Next, using some further assumption about the condition of the densities $f_X$ of $X$ and $f_Y$ of $Y$, we then derive the convergence rate of the root mean square error for the estimator.

Keywords. Nonparametric, error density, estimator, convergence rate.

1. INTRODUCTION

Let $X_1, \ldots, X_n$ be i.i.d. random variables from an unknown density function $f_X$ of $X$ and $Y_1, \ldots, Y_m$ be i.i.d. random variables from an unknown density function $f_Y$ of $Y$. We concern the problem of estimating the quantity

$$\theta := \mathbb{P}(X < Y)$$

from given the two independent samples

$$X'_j = X_j + \xi_j, \quad Y'_k = Y_k + \eta_k, \quad j = 1, \ldots, n; \quad k = 1, \ldots, m.$$  \quad (2)

Here, one observes $X'_j$ from $f_{X'_j}, \quad j = 1, \ldots, n$ and $Y'_k$ from $f_{Y'_k}, \quad k = 1, \ldots, m$. The random variables $\xi_j$ and $\eta_k$ are known as error ones. The random variables $X_j, \quad \xi'_j, \quad Y_k, \quad \eta_k$ are assumed to be mutually independent for $1 \leq j, j' \leq n, \quad 1 \leq k, k' \leq m$. In addition, assume that each $\xi_j$ has its own known density $g_{\xi,j}$ and each $\eta_k$ has its own known density $g_{\eta,k}$. The densities $g_{\xi,j}$ and $g_{\eta,k}$ are also called error densities.

The quantity $\theta$ has many applicabilities in various fields. For instance, $\theta$ is equal to the area under ROC curve which is used as a graphical tool for evaluation of the performance of diagnostic tests (see Metz [1], Bamber [3], Hughes et al. [11], Kim-Gleser [17], Coffin-Sukhatme [20], Zhou [27]). Besides, the quantity $\theta$ plays an important role in biostatistics (see Pepe [21]) and in engineering (see Kotz et al. [24]). Additionally, the quantity $\theta$ is also applied in agriculture (see Dewdney et al. [22]).

In the context of error free data, i.e., $\xi_j = 0$ and $\eta_k = 0$, there are many papers researching in both parametric and nonparametric approaches (see Kundu-Gupta [7, 8], DeLong et al. [9], Wilcoxon [10], Mann-Whitney [12], Tong [13], Montoya-Rubio [16], Constantine et al. [18], Huang et al. [19], Kotz et al. [24], Woodward-Kelley [26], among others). However, for the problem of estimating the quantity $\theta$ from given contaminated observations as in (2), the problem has not been studied much. For a nonparametric framework, there are a few papers related to the problem. In Coffin-Sukhatme [20], with contaminated observations, the Wilcoxon-Mann-Whitney estimator was used to survey the bias of the estimator. In Kim-Gleser [17], the authors used the SIMEX method, proposed by Cook-Stefanski [15], to construct an estimator of $\theta$, in which the measurement errors have the standard normal distribution. Applying nonparametric deconvolution tools and basing on the contaminated samples, Dattner [14] developed an
optimal estimator of $\theta$ when error density functions $g_{\zeta,j}$ and $g_{\eta,k}$ are assumed to be supersmooth. Herein a density is called supersmooth if its Fourier transform decays with an exponential rate at infinity. Next, Trong et al. [4] considered the problem in the case where $g_{\zeta,j}$ and $g_{\eta,k}$ are compactly supported ones. Following the latter paper, Trong et al. [5] considered the problem with heteroscedastic errors. This means that $\zeta_j$ and $\eta_k$ have different distributions for $1 \leq j \leq n$, $1 \leq k \leq m$. Recently, Phuong-Thuy [2] concentrated on the case where the distribution of the random errors is unknown but symmetric around zero and can be estimated from some additional samples.

To the best of our knowledge, so far the problem of estimating the quantity $\theta$ when error densities $g_{\zeta,j}$ and $g_{\eta,k}$ are ordinary smooth has not been considered in any research yet. This is a popular standard condition where the error densities have the Fourier transform decaying with polynomial rate at infinity. Therefore, in our current work, we fill partially the gap by considering the problem in the setting where error densities $g_{\zeta,j}$ and $g_{\eta,k}$ are the Laplace density, which is a specific case of ordinary smooth density. This is also the condition about the problem that has never been considered before. Moreover, it is also known that the Laplace distribution plays an important role in many scientific fields. It has attracted interesting applications in the modeling of detector relative efficiencies, measurement errors, extreme wind speeds, position errors in navigation, the Earth’s magnetic field, wind shear data and stock return. An in-depth survey of the Laplace distribution including various properties and applications is provided by Kotz et al. [23].

For convenience, we introduce some notations. The convolution of two functions $f$ and $g$ is denoted by $f \ast g$. The notation $\mathcal{F}^{-1}(h(t)) = \int_{-\infty}^{\infty} e^{itx}h(x)dx$ denotes the Fourier transform of a function $h(x)$, $i = \sqrt{-1}$. The notations $\mathbb{R}\{z\}$ and $\overline{z}$ denote the imaginary part and conjugate of $z$, respectively. The number $\lambda(A)$ is the Lebesgue measure of a measurable set $A \subset \mathbb{R}$. For two sequences of positive real numbers $\{a_{n,m}\}$ and $\{b_{n,m}\}$, the notation $a_{n,m} = \mathcal{O}(b_{n,m})$ means $a_{n,m} \leq \text{const} \cdot b_{n,m}$ for large $n,m$. The notation $\mathcal{O}(1)$ is a positive constant which is independent of $n,m$.

2. MAIN RESULTS

We know that, for a continuous distribution function $F$, one has

$$F(x) = \frac{1}{2} - \frac{1}{\pi} \int_{-\infty}^{\infty} e^{itx} \mathcal{F}(f(t)) dt, \quad x \in \mathbb{R},$$

where $f$ is the density function corresponding to $F$. Let $Z = X - Y$. Then $\theta = \mathbb{P}(Z < 0) = F_Z(0)$, where $F_Z$ is the distribution function of $Z$. In addition, since $f_Z = f_X \ast f_Y$, we get that

$$\theta = F_Z(0) = \frac{1}{2} - \frac{1}{\pi} \int_{-\infty}^{\infty} e^{itx} \mathcal{F}(f_Z(t)) dt = \frac{1}{2} - \frac{1}{\pi} \int_{0}^{\infty} e^{itx} \mathcal{F}(f_X(t)) \mathcal{F}(f_Y(t)) dt.$$  (3)

From (3), in the present paper, we suggest an estimator of $\theta$ in the form

$$\hat{\theta}_\gamma := \frac{1}{2} - \frac{1}{nm} \sum_{j=1}^{n} \sum_{k=1}^{m} \Delta_{j,k},$$  (4)

in which

$$\Delta_{j,k} = \frac{1}{\pi} \int_{0}^{\infty} e^{itx} \mathcal{F}(g_{\zeta,j}(t)g_{\eta,k}(t)) \left( (\gamma t^s + |g_{\zeta,j}(t)|^2)(\gamma t^s + |g_{\eta,k}(t)|^2) e^{\mu(\chi - \chi')} \right) dt,$$  (5)

where $s > 1$ and the number $\gamma \in (0,1)$ plays a role as a regularization parameter and must be selected
according to the sample sizes \( n, \ m \) later. The estimator \( \hat{\Theta}_s \) was also considered in Trong et al. [5]. Now, for error random variables, we assume that \( \zeta_j \) and \( \eta_k \), \( j = 1, \ldots, n; \ k = 1, \ldots, m \) have the Laplace distribution where the densities of \( \zeta_j \) and \( \eta_k \) have the form \( g_j(x) = (1/2) \cdot e^{-|x|} \) with \( g_j^x (t) = 1/(1+t^2) \). It is well-known that in the additive measurement error model, the class of the ordinary smooth error densities is a popular standard class where the error densities have the Fourier transform decaying with polynomial rate at infinity and the Laplace density is a famous example belonging to this class. In order to prove some below results, we need the following specific quality of the Laplace density,

\[
1 - \left| g_j^x (t) \right|^2 = \frac{2t^2 + t^4}{(1+t^2)} \leq 2t^2 + 4t^2, \ 0 \leq t \leq 1/2.
\]

**Proposition 2.1** Let the observations be given by model (2). Let the quantity \( \theta \) be defined as in (1) and the estimator \( \hat{\Theta}_s \) be as in (4) with \( \gamma \in (0, 1), \ s > 1 \). Suppose that \( f_x^j f_y^j \in L^1(\mathbb{R}) \). Besides, suppose that \( g_{\zeta,j} \) and \( g_{\eta,k} \) are the Laplace density, \( j = 1, \ldots, n; \ k = 1, \ldots, m \). Then, we have

\[
\left| \mathbb{E}(\hat{\Theta}_s) - \Theta \right| \leq C_0 \left\{ \gamma + \frac{1}{nm} \sum_{j=1}^{n} \sum_{k=1}^{m} \int_{1/2}^{\infty} \frac{\gamma t^{-1} \left| f_x^j (t) \right| \left| f_y^j (t) \right|}{\gamma t + \left| g_{\zeta,j}^x (t) \right|^2} dt + \frac{1}{nm} \sum_{j=1}^{n} \sum_{k=1}^{m} \int_{1/2}^{\infty} \frac{\gamma t^{-1} \left| f_x^j (t) \right| \left| f_y^j (t) \right|}{\gamma t + \left| g_{\eta,k}^x (t) \right|^2} dt \right\},
\]

where the constant \( C_0 \) only depends on \( s \).

**Proposition 2.2** Let the observations be given by model (2). Let the quantity \( \theta \) be defined as in (1) and the estimator \( \hat{\Theta}_s \) be given by (4) with \( \gamma \in (0, 1), \ s > 1 \). Suppose that \( g_{\zeta,j} \) and \( g_{\eta,k} \) are the Laplace density, \( j = 1, \ldots, n; \ k = 1, \ldots, m \) along with \( f_x^j f_y^j \in L^1(\mathbb{R}) \). Then, we get

\[
\mathbb{E} \left| \hat{\Theta}_s - \Theta \right|^2 \leq C_1 \times \left( \gamma + \frac{1}{nm} \sum_{j=1}^{n} \sum_{k=1}^{m} \int_{1/2}^{\infty} \frac{\gamma t^{-1} \left| f_x^j (t) \right| \left| f_y^j (t) \right|}{\gamma t + \left| g_{\zeta,j}^x (t) \right|^2} dt + \frac{1}{nm} \sum_{j=1}^{n} \sum_{k=1}^{m} \int_{1/2}^{\infty} \frac{\gamma t^{-1} \left| f_x^j (t) \right| \left| f_y^j (t) \right|}{\gamma t + \left| g_{\eta,k}^x (t) \right|^2} dt \right)^2 + \left( \frac{1}{n} + \frac{1}{m} \right)^{1/2},
\]

where the constant \( C_1 \) only depends on \( s \).

Next, the following theorem represents the mean consistency of the estimator \( \hat{\Theta}_s \).

**Theorem 2.3** The assumptions are the same as in Proposition 2.2. Besides, suppose that \( \gamma > 0 \) is a parameter depending on the sample sizes \( n, \ m \) such that \( \gamma \to 0, \ n\gamma^2 \to \infty, \ m\gamma^2 \to \infty \) as \( n, m \to \infty \). Then, \( \mathbb{E} \left| \hat{\Theta}_s - \Theta \right|^2 \to 0 \) as \( n, m \to \infty \). Now, in order to obtain the rate of the convergence of the estimator \( \hat{\Theta}_s \), we need the following definition.

For \( \beta > \frac{1}{2} \) and \( C > 0 \), we consider the class
From (3) and (4), combining the latter equality, we obtain

\[ \mathbb{E} \left( \text{sign} \left( \hat{\theta}_x - \theta \right) \right) = 0. \]

Proof of Proposition 2.1. Using the Fubini theorem, we get

\[ \mathbb{E} \left( \Delta_{j,k} \right) = \frac{1}{\pi} \int_0^{\infty} \frac{1}{t^3} \left[ \frac{g_{X_j}^f(t) g_{Y_n}^f(t) f_{X_j}^t(t) f_{Y_n}^t(t)}{\left( \gamma t^4 + \left| g_{X_j}^f(t) \right|^2 \left| g_{Y_n}^f(t) \right|^2 \right)^{1/2}} dt \right. \]

\[ \left. - \frac{1}{\pi} \int_0^{\infty} \frac{1}{t^3} \left[ \frac{g_{X_j}^f(t) g_{Y_n}^f(t) f_{X_j}^t(t) f_{Y_n}^t(t)}{\left( \gamma t^4 + \left| g_{X_j}^f(t) \right|^2 \left| g_{Y_n}^f(t) \right|^2 \right)^{1/2}} dt \right. \]

From (3) and (4), combining the latter equality, we obtain

\[ \mathbb{E} \left( \text{sign} \left( \hat{\theta}_x - \theta \right) \right) \leq \frac{1}{nm} \sum_{j=1}^{m} \sum_{k=1}^{n} \int_0^{\infty} \left[ 1 - \frac{\left| g_{X_j}^f(t) \right|^2 \left| g_{Y_n}^f(t) \right|^2}{\left( \gamma t^4 + \left| g_{X_j}^f(t) \right|^2 \left| g_{Y_n}^f(t) \right|^2 \right)^{1/2}} \right] \delta \left( f_{X_j}^t(t) f_{Y_n}^t(t) \right) dt \]

\[ \leq \frac{1}{nm} \sum_{j=1}^{m} \sum_{k=1}^{n} \int_0^{\infty} \left[ 1 - \frac{\left| g_{X_j}^f(t) \right|^2 \left| g_{Y_n}^f(t) \right|^2}{\left( \gamma t^4 + \left| g_{X_j}^f(t) \right|^2 \left| g_{Y_n}^f(t) \right|^2 \right)^{1/2}} \right] \delta \left( f_{X_j}^t(t) f_{Y_n}^t(t) \right) dt \]

\[ \leq \frac{1}{nm} \sum_{j=1}^{m} \sum_{k=1}^{n} \int_0^{\infty} \left[ 1 - \frac{\left| g_{X_j}^f(t) \right|^2 \left| g_{Y_n}^f(t) \right|^2}{\left( \gamma t^4 + \left| g_{X_j}^f(t) \right|^2 \left| g_{Y_n}^f(t) \right|^2 \right)^{1/2}} \right] \delta \left( f_{X_j}^t(t) f_{Y_n}^t(t) \right) dt \]

\[ \leq \frac{1}{nm} \sum_{j=1}^{m} \sum_{k=1}^{n} \int_0^{\infty} \left[ 1 - \frac{\left| g_{X_j}^f(t) \right|^2 \left| g_{Y_n}^f(t) \right|^2}{\left( \gamma t^4 + \left| g_{X_j}^f(t) \right|^2 \left| g_{Y_n}^f(t) \right|^2 \right)^{1/2}} \right] \delta \left( f_{X_j}^t(t) f_{Y_n}^t(t) \right) dt \]
Since \(1 - |g_{x,j}^F(t)|^2 < 4t^2\) and \(1 - |g_{n,k}^F(t)|^2 < 4t^2\), for all \(t \in [0, 1/2\sqrt{2}]\). Therefore, we get

\[
\int_0^{1/2\sqrt{2}} \frac{1}{t} \left| \frac{\gamma^2 t^2 + 2t}{\gamma t + |g_{x,j}^F(t)|^2} \right| \frac{1}{t} \left| \frac{\gamma t + |g_{x,j}^F(t)|^2}{\gamma t + |g_{n,k}^F(t)|^2} \right| dt \\
\leq 4 \left( \frac{\gamma^2 (1/2\sqrt{2})^2}{2s} + \frac{2\gamma(1/2\sqrt{2})^2}{s} \right) \gamma.
\]

Also, we get

\[
\int_{1/2\sqrt{2}}^{1/2\sqrt{2}} \frac{1}{t} \left| \frac{\gamma^2 t^2 + 2t}{\gamma t + |g_{x,j}^F(t)|^2} \right| \frac{1}{t} \left| \frac{\gamma t + |g_{x,j}^F(t)|^2}{\gamma t + |g_{n,k}^F(t)|^2} \right| dt \\
\leq 2 \int_{1/2\sqrt{2}}^{1/2\sqrt{2}} \frac{\gamma t + |g_{n,k}^F(t)|^2}{\gamma t + |g_{x,j}^F(t)|^2} dt + \int_{1/2\sqrt{2}}^{1/2\sqrt{2}} \frac{\gamma t + |g_{n,k}^F(t)|^2}{\gamma t + |g_{x,j}^F(t)|^2} dt.
\]

From (6), (7), and (8), we get the conclusion of the proposition.

To prove Proposition 2.2, we need to prove Lemma 1 and Lemma 2.

**Lemma 1** Let the observations be given by model (2). Suppose that \(g_{x,j}^F\) and \(g_{n,k}^F\) are the Laplace density, \(j = 1, \ldots, n, k = 1, \ldots, m\). Let \(\Delta_{j,k}\) be given as in (5) with \(s > 1\) and \(\gamma \in (0, 1)\). Then, we have

\[
\max \left\{ \mathbb{E}\left( \Delta_{j,k}^2 \right); \mathbb{E}\left( \Delta_{j,k} \Delta_{j,k'} \right); \mathbb{E}\Delta_{j,k} \mathbb{E}\Delta_{j,k'} \right\} \leq \frac{1}{\pi^2} \left( \frac{11}{4} + \frac{3}{2s} + \frac{\omega s}{s-1} \right) \frac{1}{\gamma^2},
\]

where \(\omega := \min \left\{ \frac{1}{6\min[2,s]}, 1/4 \right\} \), \(j = 1, \ldots, n; k = 1, \ldots, m\).

**Proof.** For all \(t \in (0, \omega)\), since \(1 - |g_{x,j}^F(t)|^2 < 4t^2\) and \(1 - |g_{n,k}^F(t)|^2 < 4t^2\), \(j = 1, \ldots, n, k = 1, \ldots, m\) and \(\gamma \in (0, 1)\), we have

\[
1 - |g_{x,j}^F(t)|^2 - \gamma t < 1 - |g_{x,j}^F(t)|^2 + t^2 < 4t^2 + t^2 < 5t^{\min[2,s]} < 1.
\]

Therefore,

\[
\frac{1}{\gamma t^2 + |g_{x,j}^F(t)|^2} = 1 + \sum_{k=1}^{\infty} \left( 1 - |g_{x,j}^F(t)|^2 - \gamma t \right)^k = 1 + Z_{x,j}(t).
\]

Likewise, \(\frac{1}{\gamma t^2 + |g_{n,k}^F(t)|^2} = 1 + \sum_{k=1}^{\infty} \left( 1 - |g_{n,k}^F(t)|^2 - \gamma t \right)^k = 1 + Z_{n,k}(t)\) for all \(t \in (0, \omega)\). Hence, for all \(r \in \mathbb{R}\), we obtain
\[ S(r) := \left| \int_0^\infty \frac{\sin(\omega t)}{t} \left( \gamma t^2 + \left| g_{\epsilon,j}^\omega(t) \right|^2 \right)^{2r} \right| dt \leq \int_0^\infty \frac{\sin(\omega t)}{t} dt + \int_0^\infty \frac{Z_{\epsilon,j}(t)}{t} dt + \int_0^\infty \frac{Z_{\eta,k}(t)}{t} dt + \int_0^\infty \frac{Z_{\zeta,j}(t)Z_{\eta,k}(t)}{t} dt =: S_1 + S_2 + S_3 + S_4. \]

From Lemma 2.6.2, Section 2.6 in Kawata [25], we obtain
\[ S_1 = \left| \int_0^\infty \frac{\sin(u)}{u} du \right| \leq 2. \]
Using the inequality
\[ (a_1 + a_2)^k \leq 2^{k-1} (a_1^k + a_2^k) \]
for all \( a_1, a_2 \geq 0, \ k \geq 1 \), we get that
\[ |Z_{\zeta,j}(t)| \leq \sum_{k=1}^\infty \left( 1 - |g_{\epsilon,j}^\omega(t)|^2 \right)^k \leq \sum_{k=1}^\infty 2^{k-1} \left( 1 - |g_{\epsilon,j}^\omega(t)|^2 \right)^k + \sum_{k=1}^\infty 2^{k-1} (\gamma t)^k \]
for any \( t \in (0, \omega_a) \). With the same argument as above, we also have
\[ |Z_{\eta,k}(t)| \leq \frac{1}{2} \sum_{k=1}^\infty (8t^2)^k + \frac{1}{2} \sum_{k=1}^\infty 2^k t^{sk}, \ \forall t \in (0, \omega_a). \] (9)

From (9), we obtain
\[ S_2 \leq \int_0^\infty \frac{1}{t} \left( \frac{1}{2} \sum_{k=1}^\infty (8t^2)^k + \frac{1}{2} \sum_{k=1}^\infty 2^k t^{sk} \right) dt \]
\[ = \frac{1}{2} \int_0^\infty \left( \sum_{k=1}^\infty (8k^2)^{\frac{k}{2}} + \sum_{k=1}^\infty 2^{k-1} t^{sk-1} \right) dt \]
\[ = \frac{1}{2} \left( \frac{1}{2} \sum_{k=1}^\infty \left( \frac{8\omega_a^2}{k} \right)^k + \sum_{k=1}^\infty 2^k \frac{\omega_a^k}{sk} \right) \leq \frac{1}{2} \left( \frac{1}{2} \sum_{k=1}^\infty (8\omega_a^2)^k + \frac{1}{s} \sum_{k=1}^\infty (2\omega_a^k)^k \right) \]
\[ \leq \frac{1}{4} + \frac{1}{2s}. \]

Similarly, the inequality (10) results in
\[ S_3 \leq \frac{1}{4} + \frac{1}{2s}. \]
Next, we have
\[ S_4 \leq \int_0^\infty \frac{1}{t} \left( \frac{1}{2} \sum_{k=1}^\infty (8t^2)^k + \frac{1}{2} \sum_{k=1}^\infty 2^k t^{sk} \right) \left( \frac{1}{2} \sum_{k=1}^\infty (8t^2)^k + \frac{1}{2} \sum_{k=1}^\infty 2^k t^{sk} \right) dt \]
\[ \leq \frac{1}{4} \left( \sum_{k=1}^\infty (8\omega_a^2)^k + \sum_{k=1}^\infty 2^{k} \omega_a^k \right) \]
\[ \int_0^\infty \left( \sum_{k=1}^\infty (8k^2)^{\frac{k}{2}} + \sum_{k=1}^\infty 2^{k-1} t^{sk-1} \right) dt \leq \frac{1}{4} + \frac{1}{2s}. \]

From the bounds of \( S_1, S_2, S_3 \) and \( S_4 \), these imply \( S(r) \leq \frac{11}{4} + \frac{3}{2s} \) for all \( r \in \mathbb{R} \).

Next, for convenience, we denote
\[ S_{a,b,j,k}^{t,h} = \int_0^\infty \frac{1}{t} \left( \left( \gamma t^2 + \left| g_{\epsilon,j}^\omega(t) \right|^2 \right)^{2r} \left( \gamma t^2 + \left| g_{\eta,k}^\omega(t) \right|^2 \right)^{2r} \right) dt, \quad 0 \leq a \leq b \leq \infty, \ j \in \mathbb{N}, \ k \in \mathbb{N}. \]
By the Fubini theorem and the inequality $S(r) \leq \frac{11}{4} + \frac{3}{2s}$, we obtain

$$
\left| S_{0,m}^{l,k} \right| \leq \frac{11}{4} + \frac{3}{2s} 
$$

Additionally, using the inequalities $\left| S \{ z \} \right| \leq |z|$ for $z \in \mathbb{C}$ and $(a_1 + a_2 + a_3)^2 \leq (a_1^2)(a_1^2)$ for $a_1, a_2, a_3 \geq 0$, we have

$$
\left| S_{0,m}^{l,k} \right| \leq \frac{11}{4} + \frac{3}{2s}. 
$$

Therefore,

$$
E(\Delta_{j,k}) \leq \frac{1}{\pi^2} E \left[ \left( |S_{0,m}^{l,k}| + |S_{0,m}^{l,k'}| \right) \right] \leq \frac{1}{\pi^2} \left( \frac{11}{4} + \frac{3}{2s} \right) \frac{1}{\gamma^2},
$$

$$
E(\Delta_{j,k})^2 = \frac{1}{\pi^2} E \left[ \left( |S_{0,m}^{l,k}| + |S_{0,m}^{l,k'}| \right)^2 \right] \leq \frac{1}{\pi^2} \left( |S_{0,m}^{l,k}| + |S_{0,m}^{l,k'}| \right)^2 \leq \frac{1}{\pi^2} \left( \frac{11}{4} + \frac{3}{2s} \right) \frac{1}{\gamma^2},
$$

$$
E(\Delta_{j,k})E(\Delta_{j,k'}) = \frac{1}{\pi^2} E \left[ \left( |S_{0,m}^{l,k}| + |S_{0,m}^{l,k'}| \right) \right] E \left[ \left( |S_{0,m}^{l,k'}| + |S_{0,m}^{l,k''}| \right) \right] \leq \frac{1}{\pi^2} \left( |S_{0,m}^{l,k}| + |S_{0,m}^{l,k'}| \right)^2 \leq \frac{1}{\pi^2} \left( \frac{11}{4} + \frac{3}{2s} \right) \frac{1}{\gamma^2}.
$$

Finally, from the bounds of $E(\Delta_{j,k})^2$, $E(\Delta_{j,k})$; $E(\Delta_{j,k})E(\Delta_{j,k'})$, we get the conclusion of the lemma.

Lemma 2 Let the observations be given by model (2). Suppose that $g_{\xi,j}$ and $g_{\eta,k}$ are the Laplace density, $j = 1, \ldots, n$; $k = 1, \ldots, m$. Let the estimator $\hat{\Theta}_j$ be given by (4) with $\gamma \in (0,1)$ and $s > 1$. Then, we have

$$
\text{Var}(\hat{\Theta}_j) \leq O \left( \left( \frac{1}{n} + \frac{1}{m} \right) \frac{1}{\gamma^2} \right).
$$

Proof. Since $E(\Delta_{j,k}) = E(\Delta_{j,k})E(\Delta_{j,k'})$ for any $j \neq j'$, $k \neq k'$, we have

$$
\text{Var}(\hat{\Theta}_j) = \frac{1}{n^2m^2} \sum_{j=1}^{n} \sum_{k=1}^{m} E(\Delta_{j,k})^2 + \frac{2}{n^2m^2} \sum_{j=1}^{n} \sum_{k=2}^{m} E(\Delta_{j,k})^2 - \frac{2}{n^2m^2} \sum_{j=1}^{n} \sum_{k=2}^{m} \sum_{j'=1}^{n} \sum_{k'=2}^{m} E(\Delta_{j,k})E(\Delta_{j,k'})
$$

© 2021 Industrial University of Ho Chi Minh City
Using (12), we infer
\[
\frac{1}{n^m} \sum_{j=2}^{n} \sum_{k=1}^{m-1} \mathbb{E}(\Delta_{j,k}\Delta_{j,k}) - \frac{1}{n^m} \sum_{j=2}^{n} \sum_{k=1}^{m-1} \mathbb{E}\Delta_{j,k}\mathbb{E}\Delta_{j,k}.
\]

Hence, we obtain \( \text{Var}(\hat{\theta}_j) \leq \mathcal{O}\left( \frac{1}{n^m} \right) \) by using Lemma 1 and (11). This completes the proof of the lemma.

**Proof of Proposition 2.2.** Using Proposition 2.1, Lemma 2, and the common variance-bias decomposition
\[
\mathbb{E}|\hat{\theta}_j - \theta|^2 = \mathbb{E}(\hat{\theta}_j - \theta)^2 + \text{Var}(\hat{\theta}_j),
\]
we get the result of the proposition.

**Proof of Theorem 2.3.** From the assumptions of the theorem and from Proposition 2.2, we only need to prove that
\[
\int_{\mathcal{E}} \gamma \gamma^{-1} \frac{f_X^2(t)}{f_X^2(t)} \frac{f_Y^2(t)}{f_Y^2(t)} \, dt \to 0, \quad \int_{\mathcal{E}} \gamma \gamma^{-1} \frac{f_X^2(t)}{f_X^2(t)} \frac{f_Y^2(t)}{f_Y^2(t)} \, dt \to 0 \quad \text{as } n, m \to \infty.
\]
Indeed, using the Lebesgue dominated convergence theorem, we get the result of the theorem.

To prove Theorem 2.4, we use the following lemma, the statement and the proof for the general case of which are presented in Trong-Phuong [6].

**Lemma 3** Suppose that the error density \( g \) satisfies the condition
\[
\frac{\sin t}{(1+t^2)} \leq |g^e(t)|.
\]
For \( R > 1 \) and \( \varepsilon > 0 \), put
\[
\mathcal{B}_{g_e,R,\varepsilon} = \{0 < t < R : |g^e(t)| \leq \varepsilon\}.
\]
Then there exists a constant \( K(g) > 0 \) depending on \( g \) such that
\[
\lambda(\mathcal{B}_{g_e,R,\varepsilon}) \leq K(g) \varepsilon R^3.
\]

**Proof of Theorem 2.4.** Given \( (f_X, f_Y) \in \mathcal{Z}_{\beta,C} \), we set
\[
Q_{\varepsilon,j} := \int_{\mathcal{E}} \gamma \gamma^{-1} \frac{f_X^2(t)}{f_X^2(t)} \frac{f_Y^2(t)}{f_Y^2(t)} \, dt, \quad Q_{\eta,k} := \int_{\mathcal{E}} \gamma \gamma^{-1} \frac{f_X^2(t)}{f_X^2(t)} \frac{f_Y^2(t)}{f_Y^2(t)} \, dt.
\]
For \( \varepsilon_1 > 0 \) small enough and let \( R_1 > 1 \), we write \( Q_{\varepsilon,j} = Q_{\varepsilon,j,1} + Q_{\varepsilon,j,2} + Q_{\varepsilon,j,3} \), where
\[
Q_{\varepsilon,j,1} := \int_{t < R_1} \gamma \gamma^{-1} \frac{f_X^2(t)}{f_X^2(t)} \frac{f_Y^2(t)}{f_Y^2(t)} \, dt,
\]
\[
Q_{\varepsilon,j,2} := \int_{1/2, \mathcal{E} : x \leq R_1} \gamma \gamma^{-1} \frac{f_X^2(t)}{f_X^2(t)} \frac{f_Y^2(t)}{f_Y^2(t)} \, dt,
\]
\[
Q_{\varepsilon,j,3} := \int_{1/2, \mathcal{E} : x \leq R_1} \gamma \gamma^{-1} \frac{f_X^2(t)}{f_X^2(t)} \frac{f_Y^2(t)}{f_Y^2(t)} \, dt.
\]
We have
\[
Q_{\varepsilon,j,1} \leq \frac{1}{R_1} \int_{t > R_1} \left| f_X^2(t) \right| \left| f_Y^2(t) \right| (1+t)^{\beta} (1+t^2)^{\beta} \, dt \leq \frac{C}{2 R_1^{2\beta+1}}.
\]
Using (12), we infer \( Q_{\varepsilon,j,2} \leq 2\sqrt{2\lambda(\mathcal{B}_{g_e,R_1,\varepsilon})} \leq 2\sqrt{2K(g_{\varepsilon,j})} \varepsilon_1 R_1^3 \). In addition, applying the Cauchy
inequality, we get that
\[
Q_{s,j,3} \leq \frac{\sqrt{2\gamma R_i^*}}{e_i} \int_{1/2}^e \left| f_X^F(t) \right| \left| f_Y^F(t) \right| dt \leq \frac{C}{\sqrt{2}} \cdot \frac{\sqrt{\gamma R_i^*}}{e_i}.
\]
From these upper bounds of \( Q_{s,j,1}, Q_{s,j,2}, \) and \( Q_{s,j,3} \), we obtain
\[
Q_{s,j} \leq \frac{C}{2R_i^{2\beta+1}} + 2\sqrt{2K(g_{s,j})} e_i R_i^3 + \frac{C\sqrt{\gamma R_i^*}}{\sqrt{2e_i}}.
\]  
(13)

With the same argument, we also have
\[
Q_{s,k} \leq \frac{C}{2R_j^{2\beta+1}} + 2\sqrt{2K(g_{s,j})} e_i R_i^3 + \frac{C\sqrt{\gamma R_i^*}}{\sqrt{2e_i}}.
\]  
(14)

Using Proposition 2.2, (13), (14), and the inequality
\[
(b_1 + b_2 + b_3 + b_4)^2 \leq 4(b_1^2 + b_2^2 + b_3^2 + b_4^2),
\]
we deduce
\[
\mathbb{E} \left| \hat{\theta}_v - \theta \right|^2 \leq O(1) \cdot \left\{ \gamma^2 + \frac{R_i^{-2(2\beta+1)}}{e_i} + \frac{\gamma R_i^*}{e_i} + \left( \frac{1 + 1}{m} \right) \frac{1}{\gamma^2} \right\}.
\]
Choosing \( \gamma = \left( \frac{1}{n} + \frac{1}{m} \right)^3 e_i^{-2/3} R_i^{-1/3} \) and \( e_i = R_i^{-12(2\beta+1)+3} \) results in
\[
\mathbb{E} \left| \hat{\theta}_v - \theta \right|^2 \leq O(1) \cdot \left\{ \beta^2 \left( \frac{1 + 1}{m} \right)^{2/3} R_i^{-2(2\beta+1)\beta} + \left( \frac{1 + 1}{m} \right)^{1/3} R_i^{2(2\beta+1)+2\beta+3} \right\}.
\]
Choosing \( R_i = \left( \frac{1}{n} + \frac{1}{m} \right)^{-d} \), we obtain
\[
\mathbb{E} \left| \hat{\theta}_v - \theta \right|^2 \leq O(1) \cdot \left\{ \left( \frac{1 + 1}{m} \right)^{2(2\beta+1)d} \right\} \leq O(1) \cdot \left\{ n^{-2(2\beta+1)d} + m^{-2(2\beta+1)d} \right\}.
\]
Finally, applying the inequality \( \sqrt{c_1 + c_2} \leq \sqrt{c_1} + \sqrt{c_2} \) for all \( c_1, c_2 > 0 \), we get the conclusion of the theorem.

**ACKNOWLEDGMENT**
The authors would like to thank the reviewers for careful reading, helpful comments and suggestions leading to the improved version of the paper.

**REFERENCES**


ƯỚC LƯỢNG PHI THAM SỐ CỦA $\mathbb{P}(X < Y)$ VỚI CÁC HÀM MẬT ĐỘ SAI SỐ LAPLACE

Tóm tắt. Chúng tôi khảo sát sự ước lượng phi tham số của xác suất $\theta := \mathbb{P}(X < Y)$ khi hai biến ngẫu nhiên $X$ và $Y$ được quan trắc có tính đến sai số. Cụ thể, từ các phiên bản nhiễu $X'_1, \ldots, X'_n$ của $X$ và $Y'_1, \ldots, Y'_m$ của $Y$, chúng tôi giới thiệu một ước lượng $\hat{\theta}$ của $\theta$ và sau đó, thiết lập tính vững theo trung bình của ước lượng được đề xuất khi các biến ngẫu nhiên sai số có phân phối Laplace. Tiếp theo, sử dụng giả thiết thêm vào về điều kiện của các hàm mật độ $f_X$ của $X$ và $f_Y$ của $Y$, chúng tôi rút ra được tốc độ hội tụ của căn bậc hai trung bình sai số bình phương của ước lượng $\hat{\theta}$.

Từ khóa. Phi tham số, mật độ sai số, ước lượng, tốc độ hội tụ.

Received on: 26/05/2021
Accepted on: 16/08/2021