Improved Cramer-Rao Inequality for Randomly Censored Data

B.L.S. Prakasa Rao


Received: 05/10/2017, Revision received: 05/02/2018, Published online: 10/08/2018

Abstract. As an application of the improved Cauchy-Schwartz inequality due to Walker (2017), we obtain an improved version of the Cramer-Rao inequality for randomly censored data derived by Abdushukurov and Kim (1987, pp. 2171-2185). We derive a lower bound of the Bhattacharya type for the mean square error of a parametric function based on randomly censored data.

Keywords. Cauchy-Schwartz inequality, Cramer-Rao inequality, Randomly censored data, Walker’s inequality.

MSC: 62F.

1 Introduction

Suppose that \(\psi\) and \(\eta\) are random variables defined on a probability space with distribution functions \(F(x, \theta)\) and \(G(x, \theta)\) respectively where \(\theta \in \Theta \subset R\). Further suppose that \(\psi_i, 1 \leq i \leq n\) are independent and identically distributed (i.i.d.) as the random variable \(\psi\) and \(\eta_i, 1 \leq i \leq n\) be i.i.d. random variables distributed as the random variable \(\eta\). Define \(\zeta_i = \min(\psi_i, \eta_i)\) and \(\delta_i = I(\psi_i \leq \eta_i), 1 \leq i \leq n\) where \(I(A)\) denotes the indicator function of a set \(A\). Suppose that \(\psi_i\) and \(\eta_i\) are not observable but \((\zeta_i, \delta_i)\) is observable for
1 ≤ i ≤ n. The data (\(\zeta_i, \delta_i\)), 1 ≤ i ≤ n is termed as \textit{randomly censored data} and is common in studies in survival analysis and reliability. It is assumed that the random variables \(\psi_{\eta_i}, 1 ≤ i ≤ ns\) are independent of the \textit{censoring} random variables \(\eta_i, 1 ≤ i ≤ n\).


2 Preliminaries

In the sequel, we assume that the distribution functions \(F(x, \theta)\) and \(G(x, \theta)\) are absolutely continuous with densities \(f(x, \theta)\) and \(g(x, \theta)\) respectively. It is easy to see that the random variables \(\zeta_i, 1 ≤ i ≤ n\) are independent and identically distributed random variables with the distribution function

\[
H(x, \theta) = 1 - \bar{F}(x, \theta)\bar{G}(x, \theta),
\]

where \(\bar{F} = 1 - F\) and \(\bar{G} = 1 - G\). Suppose that the densities \(f\) and \(g\) are differentiable with respect to \(\theta\) up to \(k\) times. Let \(f^{(i)}\) denote the \(i\)-th derivative \(f\) with respect to \(\theta\) and \(g^{(i)}\) denote the \(i\)-th derivative of \(g\) with respect to \(\theta\). Let

\[
\lambda^{(i)}(x, \theta) = \frac{f^{(i)}(x, \theta)}{f(x, \theta)} \quad \text{and} \quad \mu^{(i)}(x, \theta) = \frac{g^{(i)}(x, \theta)}{g(x, \theta)}.
\]

Suppose the following regularity conditions hold:
The derivatives \( f^{(i)}(x, \theta) \) and \( g^{(i)}(x, \theta) \) with respect to \( \theta \) exist for \( i = 1, \ldots, k \) and
\[
\int_{-\infty}^{\infty} |f^{(i)}(x, \theta)| dx < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} |g^{(i)}(x, \theta)| dx < \infty.
\]

Let \( \phi(\theta) \) be a real-valued function of the parameter \( \theta \) and suppose that \( \phi(\theta) \) is differentiable \( k \) times with respect to \( \theta \). Let \( \phi^{(i)}(\theta) \) denote the \( i \)-th derivative of \( \phi(\theta) \) with respect to \( \theta \). Suppose that \( \hat{\phi}(\zeta, \delta) \) is an estimator of the function \( \phi(\theta) \) based on the observation \((\zeta, \delta)\). Let \( E_\theta[\hat{\phi}(\zeta, \delta)] = \gamma(\theta) \). Observe that
\[
\int_{\mathbb{R} \times [0,1]} \hat{\phi}(x, y) h(x, y, \theta) d\nu(x, y) = \gamma(\theta),
\]
where
\[
h(x, y, \theta) = [\hat{G}(x, \theta)f(x, \theta)]^{y}[\hat{F}(x, \theta)g(x, \theta)]^{1-y}.
\]
is the joint density function of the random vector \((\zeta, \delta)\) and \( \nu(x, y) \) is appropriate \( \sigma \)-finite measure on \( \mathbb{R} \times [0,1] \).

Suppose the integral on the left side of the equation (2.1) is differentiable \( k \)-times with respect to \( \theta \) under the integral sign.

In particular, it follows that
\[
\int_{\mathbb{R} \times [0,1]} \hat{\phi}(x, y) h^{(i)}(x, y, \theta) d\nu(x, y) = \gamma^{(i)}(\theta), i = 1, \ldots, k,
\]
where
\[
h^{(i)}(x, y, \theta) = \frac{\partial^i}{\partial \theta^i} h(x, y, \theta).
\]

Suppose that
\[
E_\theta[|\lambda^{(i)}(\psi, \theta)|^2 < \infty \quad \text{and} \quad E_\theta[|\mu^{(i)}(\eta, \theta)|^2 < \infty.
\]
Consider the random variable
\[
S_i(\zeta, \delta, \theta) = \sum_{j=0}^{i} \binom{i}{j} \delta^{i-j} \left[ \frac{f^{(i)}(\zeta, \delta)}{f(\zeta, \delta)G(\zeta, \delta)} + (1-\delta) \frac{g^{(i-j)}(\zeta, \delta)\hat{G}^{(i)}(\zeta, \delta)}{g(\zeta, \delta)\hat{F}(\zeta, \delta)} \right],
\]
where
\[
\hat{F}^{(i)}(x, \theta) = \frac{\partial^i}{\partial \theta^i} F(x, \theta) \quad \text{and} \quad \hat{G}^{(i)}(x, \theta) = \frac{\partial^i}{\partial \theta^i} G(x, \theta).
\]
Define

\[ S^T(\zeta, \delta, \theta) = (S_1(\zeta, \delta, \theta), \ldots, S_k(\zeta, \delta, \theta)), \]

where \( S^T \) denotes the transpose of the vector \( S \).

(C4) Suppose the matrix

\[ I(\theta) = E_\theta[S(\zeta, \delta, \theta)S^T(\zeta, \delta, \theta)] = ((I_{ij}(\theta)))_{k \times k}, \]

is positive definite where

\[ I_{ij}(\theta) = \sum_{\ell=0}^{i} \sum_{m=0}^{j} \binom{i}{\ell} \binom{j}{m} \int_{-\infty}^{\infty} \frac{f^{(\ell)}(x, \theta)f^{(m)}(x, \theta)\tilde{G}^{(i-\ell)}(x, \theta)\tilde{G}^{(j-m)}(x, \theta)}{f(x, \theta)\tilde{G}(x, \theta)} \, dx + \]

\[ + \int_{-\infty}^{\infty} \frac{g^{(i-\ell)}(x, \theta)g^{(j-m)}(x, \theta)\tilde{F}^{(\ell)}(x, \theta)\tilde{F}^{(m)}(x, \theta)}{g(x, \theta)\tilde{F}(x, \theta)} \, dx. \]

(C5) The set \( \{ x : \tilde{G}(x, \theta)f(x, \theta) = 0 \} \cup \{ x : \tilde{F}(x, \theta)g(x, \theta) = 0 \} \) does not depend on the parameter \( \theta \).

Let us first consider the case \( k = 1 \). The following theorem is due to Abdushukurov and Kim (1987) generalizing the Cramer-Rao inequality for the randomly censored data.

**Theorem 2.1.** Suppose the conditions (C1) – (C5) hold and the estimator \( \hat{\phi}(\zeta, \delta) \) is an unbiased estimator of the function \( \phi(\theta) \). Then

\[ \text{Var}(\hat{\phi}(\zeta, \delta)) \geq \frac{[\phi^{(1)}(\theta)]^2}{I(\theta)}. \quad (2.3) \]

**Remark 1.** A Bayesian version of the result stated above was obtained in Prakasa Rao (1995). An improved version of this result will be derived later as an application of the improved Cauchy-Schwartz inequality due to Walker (2017).

The following theorem, again due to Abdushukurov and Kim (1987), generalizes Theorem 2.1 to the case \( k > 1 \) and it is an analogue of the Bhattacharyya bound in the uncensored case. They have also characterized the class of distributions for which equality is attained in the equation (2.3) and in the inequality (2.4) given below.
Theorem 2.2. Suppose the conditions \((C_1) - (C_3)\) hold and the estimator \(\hat{\phi}(\zeta, \delta)\) is an unbiased estimator of the function \(\phi(\theta)\). Then

\[
\text{Var}(\hat{\phi}(\zeta, \delta)) \geq \sum_{i,j=1}^{k} \phi^{(i)}(\theta)\phi^{(j)}(\theta)\tilde{I}_{ij}(\theta),
\]  

where

\[
((\tilde{I}_{ij}(\theta))_{k\times k} = ((I_{ij}(\theta))^{-1}_{k\times k}).
\]

3 Improved Cauchy-Schwartz Inequality

Walker (2017) obtained an improved version of the Cauchy-Schwartz inequality which implies the following probabilistic version.

Theorem 3.1. (Walker’s inequality) If \(X\) and \(Y\) are random variables defined on a probability space \((\Omega, \mathcal{F}, P)\) with finite second moments, then

\[
|E(XY)|^2 \leq E(X^2)E(Y^2) - |E(X)|\sqrt{\text{Var}(Y)} - |E(Y)|\sqrt{\text{Var}(X)})^2.
\]  

(3.1)

In particular, if the random variable \(Y\) has mean zero but positive variance, then it follows that

\[
|E(XY)|^2 \leq E(X^2)E(Y^2) - |E(X)|^2E(Y^2) = \text{Var}(X)E(Y^2).
\]  

(3.2)

Hence

\[
E(X^2) \geq (E(X))^2 + \frac{|E(XY)|^2}{E(Y^2)},
\]

and we have the following corollary.

Corollary 3.1. If \(X\) and \(Y\) are random variables defined on a probability space \((\Omega, \mathcal{F}, P)\) with finite second moments and if \(E(Y) = 0\), then

\[
E(X^2) \geq (E(X))^2 + \frac{|E(XY)|^2}{E(Y^2)}.
\]  

(3.3)
Walker (2017) indicated that the inequality derived in Theorem 3.1 is an improved inequality over the Cauchy-Schwarz inequality. For instance, if \( Y \) is a standard normal random variable and \( X \) is a random variable with mean \( \mu \) and variance \( \sigma^2 \), then the Cauchy-Schwarz inequality shows that
\[
\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2) = \mu^2 + \sigma^2,
\]
where as the improved Cauchy-Schwarz inequality implies that
\[
\mathbb{E}(XY)^2 \leq \sigma^2,
\]
which is a strict improvement over the earlier inequality.

We now discuss some applications of the inequality derived in Corollary 3.1.

### 4 Main Results

The following lemma is proved in Abdushukurov and Kim (1987).

**Lemma 4.1.** Suppose the conditions \((C_1)\) and \((C_5)\) hold. Then, almost everywhere on \( R \times \{0, 1\} \), there exist derivatives \( h^{(1)}(x, y, \theta) \) for all \( \theta \in \Theta \) and for \( i = 1, \ldots, k \)

\[
h^{(1)}(x, y, \theta) = h(x, y, \theta) \sum_{j=0}^{i} \binom{i}{j} y^j f^{(j)}(x, \theta) \mathbb{G}^{(i-j)}(x, \theta) G(x, \theta) + (1 - y) S^{(i-j)}(x, \theta) F^{(i)}(\zeta, \theta).
\]

Let
\[
S_{1i}(\zeta, \delta, \theta) = \sum_{j=0}^{i} \binom{i}{j} \mathbb{G}^{(i-j)}(\zeta, \theta) \frac{\delta^{(j)}(\zeta, \theta)}{f(\zeta, \theta) G(\zeta, \theta)}, i = 1, \ldots, k,
\]
and
\[
S_{2i}(\zeta, \delta, \theta) = \sum_{j=0}^{i} \binom{i}{j} \frac{(1 - \delta)^{i-j} \mathbb{G}^{(i-j)}(\zeta, \theta) F^{(j)}(\zeta, \theta)}{g(\zeta, \theta) F(\zeta, \theta)}], i = 1, \ldots, k.
\]

Define
\[
S^{(1)}(\zeta, \delta, \theta) = (S_{11}(\zeta, \delta, \theta), \ldots, S_{1k}(\zeta, \delta, \theta)),
\]
and
\[
S^{(2)}(\zeta, \delta, \theta) = (S_{21}(\zeta, \delta, \theta), \ldots, S_{2k}(\zeta, \delta, \theta)).
\]
Note that
\[ S(\zeta, \delta, \theta) = S^{(1)}(\zeta, \delta, \theta) + S^{(2)}(\zeta, \delta, \theta). \]

We will first obtain an improved version of the inequality derived in Theorem 2.1 due to Abdushukurov and Kim (1987). Let
\[ F'(x, \theta) = P_\theta(\zeta \leq x, \delta = 1) = \int_{-\infty}^{x} \bar{G}(t, \theta) f(t, \theta) dt, \]
and
\[ G'(x, \theta) = P_\theta(\zeta \leq x, \delta = 0) = \int_{-\infty}^{x} \bar{f}(t, \theta) g(t, \theta) dt. \]

Observe that
\[ F'(x, \theta) + G'(x, \theta) = H(x, \theta) = 1 - \bar{F}(x, \theta)\bar{G}(x, \theta), x \in R, \]
and
\[ F'(\infty, \theta) + G'(\infty, \theta) = 1, \]
where
\[ F'(\infty, \theta) = \lim_{x \to \infty} F(x, \theta) = P_\theta(\delta = 1) = E_\theta[\delta], \]
and
\[ G'(\infty, \theta) = \lim_{x \to \infty} G(x, \theta) = P_\theta(\delta = 0) = E_\theta[1 - \delta]. \]

Let us first consider the case \( k = 1 \). Suppose that \( \hat{\phi}(\zeta, \delta) \) is an estimator of the parametric function \( \phi(\theta) \) and suppose that
\[ E_\theta[\hat{\phi}(\zeta, \delta)] = \gamma(\theta), \theta \in \Theta. \]

Note that
\[ \int_{R \times [0,1]} \hat{\phi}(x, y) h(x, y, \theta) d\nu(x, y) = \gamma(\theta). \]

Then, by Lemma 4.1, it follows that
\[ S(\zeta, \delta, \theta) = S_{11}(\zeta, \delta, \theta) + S_{21}(\zeta, \delta, \theta), \]
where
\[ S^{(1)} = S_{11} = \frac{h^{(1)}(\zeta, \delta, \theta)}{h(\zeta, \delta, \theta)} = \delta [\frac{f^{(1)}}{f} + \frac{\bar{G}^{(1)}}{\bar{G}}]. \]
and

\[ S^{(2)} = S_{21} = (1 - \delta) \left[ \frac{g^{(1)}}{g} + \frac{f^{(1)}}{F} \right]. \]

Note that

\[ E_\theta[S_{11}(\zeta, \delta, \theta)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f^{(1)}(x, \theta) G(x, \theta) dF(x, \theta) \]

\[ = \int_{-\infty}^{\infty} f^{(1)}(x, \theta) G(x, \theta) dx \]

\[ = \frac{\partial}{\partial \theta} E_\theta[\delta]. \]

Similarly, we get that

\[ E_\theta[S_{21}(\zeta, \delta, \theta)] = \frac{\partial}{\partial \theta} E_\theta[1 - \delta]. \]

Combining the above equations, it follows that

\[ E_\theta[S(\zeta, \delta, \theta)] = \frac{\partial}{\partial \theta} E_\theta[\delta] + \frac{\partial}{\partial \theta} E_\theta[1 - \delta] = 0. \]

This equation was derived in Abdushukurov and Kim (1987). We include it here for completeness. Applying Corollary 3.1. with \( Y = S(\zeta, \delta, \theta) \) and \( X = \hat{\phi}(\zeta, \delta) - \phi(\theta) \), it follows that

\[ E_\theta[(\hat{\phi}(\zeta, \delta) - \phi(\theta))^2] \geq (E_\theta[\hat{\phi}(\zeta, \delta) - \phi(\theta)])^2 + \frac{\left| E(XY) \right|^2}{E(Y^2)}. \]

Observe that

\[ E(XY) = E_\theta[(\hat{\phi}(\zeta, \delta) - \phi(\theta))S(\zeta, \delta, \theta)] = E_\theta[\hat{\phi}(\zeta, \delta)S(\zeta, \delta, \theta)] \]

\[ = \gamma^{(1)}(\theta). \]

Hence we obtain the following inequality for randomly censored data improving the inequality proved by Abdushukurov and Kim (1987).

**Theorem 4.1.** Suppose the conditions (C1) to (C5) hold. Then, for all \( \theta \in \Theta \),

\[ E_\theta[(\hat{\phi}(\zeta, \delta) - \phi(\theta))^2] \geq (E_\theta[\hat{\phi}(\zeta, \delta) - \phi(\theta)])^2 + \frac{\left| \gamma^{(1)}(\theta) \right|^2}{E_\theta[S^2(\zeta, \delta, \theta)]}. \]
We will now consider the case \( k > 1 \). In addition to the conditions \((C_1) - (C_5)\), suppose the following condition holds.

\((C_6)\) Suppose the random functions \( S_i(\zeta, \delta, \theta) \) are linearly independent. Define the function

\[ Q(\zeta, \delta, \theta) = \hat{\phi}(\zeta, \delta) - \gamma(\theta) - \sum_{i=1}^{k} q_i S_i(\zeta, \delta, \theta), \]

where \( q_i, 1 \leq i \leq k \) are to be chosen. From the arguments given earlier, it follows that

\[ E_\theta[S_i(\zeta, \delta, \theta)] = 0, i = 1, \ldots, k. \]

Since \( E_\theta[\hat{\phi}(\zeta, \delta)] = \gamma(\theta) \), it follows that \( E_\theta[Q(\zeta, \delta, \theta)] = 0 \). Observe that

\[ Var_\theta[Q(\zeta, \delta, \theta)] = E_\theta[\{Q(\zeta, \delta, \theta)\}^2] \]

\[ = \int_{R \times [0,1]} \{\hat{\phi}(x, y) - \gamma(\theta) - \sum_{i=1}^{k} q_i S_i(x, y, \theta)\}^2 h(x, y, \theta) dx dy. \]

Suppose we choose \( q_i \) such that \( Var_\theta[Q(\zeta, \delta, \theta)] \) is minimum. This will imply that \( q_i, i = 1, \ldots, k \) should satisfy the equations

\[ \frac{\partial Var_\theta[Q]}{\partial q_i} = 0, i = 1, \ldots, k. \]

(4.11)

Under the conditions \((C_1) - (C_5)\), the system of equations given above reduce to the system

\[ \int_{R \times [0,1]} \{\hat{\phi}(x, y) - \gamma(\theta) - \sum_{i=1}^{k} q_i S_i(x, y, \theta)\} h^{(j)}(x, y, \theta) dx dy = 0, j = 1, \ldots, k. \] (4.12)

Solving these equations, by applying integration by parts, it can be checked that

\[ \gamma^{(j)}(\theta) = \sum_{i=1}^{k} q_i \int_{R \times [0,1]} S_i(x, y, \theta) S_j(x, y, \theta) h(x, y, \theta) dx dy \] (4.13)

\[ = \sum_{i=1}^{k} q_i I_{ij}(\theta), \]
using the conditions \((C_2)\) and \((C_5)\) and the relations \(h^{(j)} = S_j h, 1 \leq j \leq k\) from Lemma 4.1. Since the matrix \(((I_{ij}(\theta)))_{k \times k}\) is positive definite by assumption, there exists an inverse matrix \(((\tilde{I}_{ij}(\theta)))_{k \times k}\) and the linear system of equations (4.12) has the solution
\[
q_i^* = \sum_{j=1}^{k} \gamma^{(j)}(\theta) \tilde{I}_{ij}(\theta), 1 = 1, \ldots, k.
\]
Substituting these values of \(q_i, 1 = 1, \ldots, k,\) into the expression for the function \(Q,\) we obtain that
\[
Q'(\zeta, \delta, \theta) = \hat{\phi}(\zeta, \delta, \theta) - \gamma(\theta) - \sum_{ij=1}^{k} S_i(\zeta, \delta, \theta) \gamma^{(j)}(\theta) \tilde{I}_{ij}(\theta). \tag{4.14}
\]
It can be checked that
\[
\text{Var}_{\theta}[Q'(\zeta, \delta, \theta)] = E_{\theta}[\{Q'(\zeta, \delta, \theta)\}^2] = E_{\theta}[\{\hat{\phi}(\zeta, \delta, \theta) - \gamma(\theta)\}^2] - \sum_{ij=1}^{k} \gamma^{(ij)}(\theta) \tilde{I}_{ij}(\theta) \geq 0.
\]
Hence
\[
E_{\theta}[\{\hat{\phi}(\zeta, \delta, \theta) - \gamma(\theta)\}^2] \geq \sum_{ij=1}^{k} \gamma^{(ij)}(\theta) \tilde{I}_{ij}(\theta), \tag{4.15}
\]
and we obtain the following result.

**Theorem 4.2.** Suppose the conditions \((C_1)\) to \((C_6)\) hold for all \(\theta \in \Theta.\) Then, for any estimator \(\hat{\phi}(\zeta, \delta)\) of the parametric function \(\phi(\theta)\) with expectation \(\gamma(\theta),\)
\[
E_{\theta}[\{\hat{\phi}(\zeta, \delta, \theta) - \phi(\theta)\}^2] \geq [\phi(\theta) - \gamma(\theta)]^2 + \sum_{ij=1}^{k} \gamma^{(ij)}(\theta) \tilde{I}_{ij}(\theta), \tag{4.16}
\]
where
\[
((\tilde{I}_{ij}(\theta)))_{k \times k} = ((I_{ij}(\theta)))^{-1}_{k \times k}.
\]
This result generalizes the result of Abdushukurov and Kim (1987) from unbiased estimators to arbitrary estimators by the obvious relation
\[
E_{\theta}[\{\hat{\phi}(\zeta, \delta) - \phi(\theta)\}^2] = (E_{\theta}[\{\hat{\phi}(\zeta, \delta) - \phi(\theta)\}]^2 + \text{Var}_{\theta}(\hat{\phi}(\zeta, \delta)).
\]
Remark 2. It would be interesting to investigate whether Walker’s inequality can be extended to higher dimensions.

Acknowledgements

This work was done under the "Indian National Science Academy Senior Scientist" scheme at the CR Rao Advanced Institute of Mathematics, Statistics and Computer Science, Hyderabad 500046, India.

References


