Estimation in Roberts’ Correlation Model for Twin Studies

Jose Almer T. Sanqui
Department of Mathematical Sciences
Appalachian State University
ASU Box 32092, Boone, NC 28608-2092, USA

Truc T. Nguyen
Department of Mathematics and Statistics
Bowling Green State University
Bowling Green, OH 43403-0221, USA

Arjun K. Gupta
Department of Mathematics and Statistics
Bowling Green State University
Bowling Green, OH 43403-0221, USA

Abstract
Estimation of the correlation coefficient in the model based on the minimum of $X$ and $Y$, where $(X, Y)$ has a bivariate normal distribution, has been studied. The truncated Roberts’ estimator is investigated and compared with the maximum likelihood estimator. Simulation study shows that neither estimator is uniformly superior.

Mathematics Subject Classification: Primary 62F12; Secondary 62F10

Keywords: bivariate normal distribution; skew-normal distribution; estimation; asymptotic normality; consistency; unbiasedness

1 Introduction
The correlation coefficient is one of the most studied statistical ideas in the literature. One obvious reason is its importance in studying association of

1sanquijat@appstate.edu
random variables. Another reason is its role as one of the parameters of the bivariate normal distribution which remains to be one of the most popular bivariate distributions both in statistical theory and applications.

Various estimators of the correlation coefficient of the bivariate normal distribution corresponding with the various configurations of the parameter space and possible data patterns have been studied. An extensive account is given in Section 8.1 of Kotz et al. (2000).

Roberts (1966) proposed an estimator of the correlation coefficient for a special configuration of the parameter space and a special data pattern for the bivariate normal distribution. This special configuration and data pattern was proposed by Roberts as a model that can be utilized in the study of twins. This model and Roberts’ estimator of the correlation coefficient will be reviewed in Section 2.

Asymptotic properties of a truncated version of Roberts’ estimator is investigated in Section 3. In particular, the consistency, asymptotic normality and the bias of this estimator are studied.

Maximum likelihood estimation is discussed in Section 4. Finally, the results of a small simulation study comparing the performance of the truncated Roberts’ estimator with the maximum likelihood estimator is presented in the last section.

2 The Roberts Correlation Model and the Roberts Estimator

The random vector \((X, Y)\) has a bivariate-normal distribution with parameters \(-\infty < \mu_1, \mu_2 < \infty, \sigma_1 > 0, \sigma_2 > 0, \) and \(\rho \in (-1, 1)\), denoted by \(\text{BN}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)\), if its p.d.f. is given by

\[
f(x, y; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi \sqrt{1-\rho^2} \sigma_1 \sigma_2} \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left\{ \left( \frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho \left( \frac{x-\mu_1}{\sigma_1} \right) \left( \frac{y-\mu_2}{\sigma_2} \right) + \left( \frac{y-\mu_2}{\sigma_2} \right)^2 \right\} \right\}
\]

Roberts (1966) studied the estimation of \(\rho\) in the special case where \(\mu_1 = \mu_2 = \mu, \sigma_1 = \sigma_2 = \sigma, \mu\) and \(\sigma\) are known, and the only available data are the independent observations \(z_1, z_2, \ldots, z_n\) on \(Z = \min(X, Y)\). We will call this special configuration of the parameter space and the special data pattern the Roberts’ correlation model.

For this model, it was shown that the density of \(W = (Z - \mu)/\sigma\) is the skew-normal density with parameter \(-\sqrt{(1-\rho)/(1+\rho)}\) of Azzalini (1985) although he did not call it as such. This model has also been studied by
Chiogna (1998), Henze (1986), Gupta et al. (2002), Arnold and Beaver (2002), Azzalini and Dalla Valle (1996), Azzalini and Capitanio (2003) and Branco and Dey (2001). An extensive review of literature related to this model is given in Genton (2004). After deriving the density of $W$, Roberts obtained the moments of the distribution of $W$ and used this result to show that the estimator

$$\bar{\rho} = 1 + \frac{\pi}{n - 1} - \frac{\pi}{n} \left( \frac{\sum_{i=1}^{n} \frac{Z_i - \mu}{\sigma}}{n(n - 1)} \right)^2$$

is an unbiased and consistent estimator of $\rho$. He also showed that the asymptotic distribution of $\bar{\rho}$ is

$$N\left( \rho, \frac{4(1-\rho)}{n(1-\rho) + \rho} \right).$$

Since $\mu$ and $\sigma$ are known, then without loss of generality, the estimator in (2.1) can be examined by assuming that $\mu = 0$ and $\sigma = 1$. We refer to the estimator in (2.1) with $\mu = 0$ and $\sigma = 1$ as the Roberts’ estimator.

## 3 Truncated Roberts’ Estimator

Roberts’ estimator has one major weakness. Its value is not always in the range $[-1, 1]$ which is the only range of possible values for $\rho$. This observation is easily verified by simulation or by Lemma 3.3.

One obvious way of eliminating this weakness is to modify $\bar{\rho}$ by assigning the value 1 when it exceeds unity and by assigning the value $-1$ when its value is less than $-1$. That is, the modified estimator is given by

$$\tilde{\rho} = \bar{\rho} \cdot 1_{\{\bar{\rho} \in [-1,1]\}} + 1_{\{\bar{\rho} > 1\}} - 1_{\{\bar{\rho} < -1\}}.$$

(3.1)

It can be shown that when $-1 < \rho < 1$, this estimator, which we will call the truncated Roberts’ estimator is asymptotically equivalent to Roberts’ estimator. To see this, we need the following lemmas:

**Lemma 3.1.** If $n > \frac{\pi}{1-\rho}$ and $-1 < \rho < 1$ then $P(\tilde{\rho} > 1) \leq \frac{1}{\sqrt{n(1-\rho)}/\pi - 1}}$.

**Proof.**

$$P(\tilde{\rho} > 1) = P\left( \left( \sum_{i=1}^{n} Z_i \right)^2 / n \leq n \right).$$
\[
\begin{align*}
&= P \left( -\sqrt{n} + n \sqrt{\frac{1-\rho}{\pi}} \leq \sum_{i=1}^{n} Z_i + n \sqrt{\frac{1-\rho}{\pi}} \leq \sqrt{n} + n \sqrt{\frac{1-\rho}{\pi}} \right) \\
&\leq P \left( \sum_{i=1}^{n} Z_i + n \sqrt{\frac{1-\rho}{\pi}} \geq -\sqrt{n} + n \sqrt{\frac{1-\rho}{\pi}} \right) \\
&\leq \frac{n \left[ 1 - \frac{(1-\rho)}{\pi} \right]}{\left[ -\sqrt{n} + n \sqrt{\frac{1-\rho}{\pi}} \right]^2} \\
&= \frac{\left[ 1 - \frac{(1-\rho)}{\pi} \right]}{\left[ \sqrt{n \left(1-\rho\right)}/\pi - 1 \right]^2} \\
&\leq \frac{1}{\left[ \sqrt{n \left(1-\rho\right)}/\pi - 1 \right]^2}
\end{align*}
\]

The first inequality holds if the hypothesis is true. The second inequality follows from Chebyshev’s Inequality. \(\square\)

**Lemma 3.2.** If \(n > \frac{\pi-2}{1+\rho}\) and \(-1 < \rho < 1\) then \(P(\bar{\rho} < -1) \leq \frac{16 \pi}{\pi(1+\rho)^2}\).

**Proof.**

\[
\begin{align*}
P(\bar{\rho} < -1) &= P \left( \left[ \sum_{i=1}^{n} Z_i \right]^2 > \frac{n \left( 2n - 2 + \pi \right)}{\pi} \right) \\
&= P \left( \sum_{i=1}^{n} Z_i + n \sqrt{\frac{1-\rho}{\pi}} > \sqrt{\frac{n \left( 2n - 2 + \pi \right)}{\pi}} + n \sqrt{\frac{1-\rho}{\pi}} \right) \\
&\quad + P \left( \sum_{i=1}^{n} Z_i + n \sqrt{\frac{1-\rho}{\pi}} < -\sqrt{\frac{n \left( 2n - 2 + \pi \right)}{\pi}} + n \sqrt{\frac{1-\rho}{\pi}} \right) \\
&\leq P \left( \left| \sum_{i=1}^{n} Z_i + n \sqrt{\frac{1-\rho}{\pi}} \right| > \sqrt{n \left( 2n - 2 + \pi \right)}/\pi - n \sqrt{\frac{1-\rho}{\pi}} \right) \\
&\leq \frac{n \left[ 1 - \frac{(1-\rho)}{\pi} \right]}{\left[ \sqrt{n \left(2n - 2 + \pi\right)}/\pi - n \sqrt{\frac{1-\rho}{\pi}} \right]^2}
\end{align*}
\]
\[
\frac{n [1 - (1 - \rho)/\pi] \left( \sqrt{n (2n - 2 + \pi)/\pi} + n \sqrt{(1 - \rho)/\pi} \right)^2}{\left( n^2 (1 + \rho) + n (\pi - 2) \right)/\pi^2}
\leq \frac{n \left( \sqrt{n (2n + 2n)/\pi} + n \sqrt{2/\pi} \right)^2}{n^4 (1 + \rho)^2/\pi^2}
\leq \frac{n \pi (4n)^2}{n^4 (1 + \rho)^2/\pi^2}
= \frac{16 \pi}{n (1 + \rho)^2}
\]

The first inequality holds if the hypothesis is true. The second inequality follows from Chebyshev’s Inequality.

It is also interesting to note the following results:

**Lemma 3.3.**

1. If \( \rho = 1 \), then \( P(\bar{\rho} > 1) = \Phi(1) - \Phi(-1) \) for all \( n \) and \( P(\bar{\rho} < -1) \to 0 \) as \( n \to \infty \).

2. If \( \rho = -1 \), then \( P(\bar{\rho} < -1) = 2 - 2 \Phi(1) \) for all \( n \) and \( P(\bar{\rho} > 1) \to 0 \) as \( n \to \infty \).

**Proof.** To prove part (1), we note that when \( \rho = 1 \), \( Z_i \sim SN(0) = N(0, 1) \). So, we have

\[
P(\bar{\rho} > 1) = P \left( \left( \sum_{i=1}^{n} Z_i \right)^2 < n \right)
= P \left( -\sqrt{n} < \sum_{i=1}^{n} Z_i < \sqrt{n} \right)
= P \left( -\frac{\sqrt{n}}{\sqrt{n}} < \frac{\sum_{i=1}^{n} Z_i}{\sqrt{n}} < \frac{\sqrt{n}}{\sqrt{n}} \right)
= P(-1 < Z < 1)\text{ where } Z \sim N(0, 1)
= \Phi(1) - \Phi(-1)
\]
Similarly, we have

\[
P(\bar{\rho} < -1) = P \left( \sum_{i=1}^{n} Z_i \right)^2 > \frac{n (2n - 2 + \pi)}{\pi} \]

\[
+ P \left( \sum_{i=1}^{n} Z_i \right) > -\sqrt{\frac{n (2n - 2 + \pi)}{\pi}} \]

\[
= P \left( \frac{\sum_{i=1}^{n} Z_i}{\sqrt{n}} > \sqrt{\frac{2n - 2 + \pi}{\pi}} \right) \]

\[
+ P \left( \frac{\sum_{i=1}^{n} Z_i}{\sqrt{n}} < -\sqrt{\frac{2n - 2 + \pi}{\pi}} \right) \]

\[
= 2 \Phi \left( -\sqrt{\frac{(2n - 2 + \pi)}{\pi}} \right) \]

\[
\rightarrow 0 \text{ as } n \to \infty \]

The proof of part (2) follows along similar line, that is, by using the fact that when \(\rho = -1\), \(Z_i \sim SN(-\infty)\) which is the distribution of \(-|Z|\) where \(Z \sim N(0,1)\).

Lemma 3.1 and Lemma 3.2 imply that when \(-1 < \rho < 1\), \(P(\bar{\rho} \in [-1,1])\) tends to 1 as \(n \to \infty\) and thus implying that \(\hat{\rho}\) and \(\bar{\rho}\) are equivalent asymptotically. As a consequence, we have

**Theorem 3.1.** When \(-1 < \rho < 1\), the asymptotic distribution of \(\bar{\rho}\) is \(N \left( \rho, \frac{4(1-\rho)(\pi-1+\rho)}{n} \right)\).

**Proof.** From (3.1) we have \(\bar{\rho} = \rho 1_{\{\rho \in [-1,1]\}} + 1_{\{\rho > 1\}} - 1_{\{\rho < -1\}}\). The asymptotic distribution of \(\bar{\rho}\) is \(N \left( \rho, \frac{4(1-\rho)(\pi-1+\rho)}{n} \right)\). Also it follows from Lemma 3.1 and Lemma 3.2 that as \(n \to \infty\), \(1_{\{\rho \in [-1,1]\}}\) tends to 1 in probability , \(1_{\{\rho > 1\}}\) tends to 0 in probability and \(1_{\{\rho < -1\}}\) tends to 0 in probability. The desired result then follows from Slutsky’s Theorem.

From Theorem 3.1, we immediately get the following corollaries:

**Corollary 3.1.** The estimator \(\bar{\rho}\) is asymptotically unbiased.
Corollary 3.2. The estimator $\tilde{\rho}$ is a consistent estimator of $\rho$.

Theorem 3.1 and Corollary 3.2 carry over all the asymptotic properties of $\bar{\rho}$ to $\tilde{\rho}$. That is, just like $\bar{\rho}$, $\tilde{\rho}$ is also consistent and asymptotically normally distributed. We already mentioned that $\bar{\rho}$ is unbiased. It is interesting to know whether this property of $\bar{\rho}$ also carry over to the modified estimator $\tilde{\rho}$, which is a truncated version of $\bar{\rho}$. In the last section, we study the bias of this estimator and compare its performance with the maximum likelihood estimator.

4 Maximum Likelihood Estimation of $\rho$

Consider a random sample $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ from $SN(\lambda)$ where $\lambda = -\sqrt{(1 - \rho)/(1 + \rho)}$. Note that $\lambda \to -1$ as $\rho \to 0$, $\lambda \to -\infty$ as $\rho \to -1$ and $\lambda \to 0$ as $\rho \to 1$. The likelihood function is given by

$$L(\lambda|\mathbf{z}) = 2^n \prod_{k=1}^{n} \Phi(\lambda z_k)\phi(z_k), \lambda \epsilon (-\infty, 0] \quad (4.1)$$

from which the log-likelihood equation

$$\sum_{k=1}^{n} \frac{z_k \phi(\lambda z_k)}{\Phi(\lambda z_k)} = 0. \quad (4.2)$$

is obtained. The maximum likelihood estimator $\hat{\lambda}$ of $\lambda$ can be obtained by numerically computing the solution of (4.2) or by numerically maximizing (4.1). The invariance property of maximum likelihood estimators can then be invoked to obtain the maximum likelihood estimator $\hat{\rho}$ of $\rho$ using

$$\hat{\rho} = \frac{1 - \hat{\lambda}^2}{1 + \hat{\lambda}^2} \quad (4.3)$$

Some results regarding the estimators $\hat{\lambda}$ and $\hat{\rho}$ are given in the following theorem.

Theorem 4.1. Let $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ be a random sample from $SN(\lambda)$ where $\lambda = -\sqrt{(1 - \rho)/(1 + \rho)}$.

(1) If all $z_i \geq 0$ and at least one $z_i$ is positive then $\hat{\lambda} = -\infty$ and $\hat{\rho} = -1$.

(2) If all $z_i \leq 0$ and at least one $z_i$ is negative then $\hat{\lambda} = 0$ and $\hat{\rho} = 1$. 

(3) There is a finite MLE \( \hat{\lambda} \) (and hence \( \hat{\rho} \) in \((-1,1)) \) if

\[
\sum_{z_i < 0} z_i + \sum_{z_i > 0} z_i < 0, \tag{4.4}
\]

at least one \( z_i \) is positive and at least one \( z_i \) is negative.

Proof. The derivative of the log-likelihood function is given by

\[
\frac{dl}{d\lambda} = \sum_{k=1}^{n} \frac{z_k \phi(\lambda z_k)}{\Phi(\lambda z_k)} = \sum_{z_k < 0} \frac{z_k \phi(\lambda z_k)}{\Phi(\lambda z_k)} + \sum_{z_k > 0} \frac{z_k \phi(\lambda z_k)}{\Phi(\lambda z_k)}
\]

(1) and (2) immediately follow since if all \( z_i \geq 0 \) and at least one \( z_i \) is positive, the likelihood function would be strictly increasing and if all \( z_i \leq 0 \) and at least one \( z_i \) is negative, the likelihood function would be strictly decreasing. To show (3), let \( l^* = \frac{dl}{d\lambda} \). Clearly, \( l^* \) is a continuous function of \( \lambda \). Since \( l^*(-\infty) = +\infty \), the log-likelihood equation \( l^*(\lambda^*) = 0 \) will have a finite solution in \((-\infty, 0]\) if and only if there exists a \( \lambda^* \) such that \( l^*(\lambda^*) < 0 \). But if \( \sum_{z_i < 0} z_i + \sum_{z_i > 0} z_i < 0 \), at least one \( z_i \) is positive and at least one \( z_i \) is negative, then we can take \( \lambda^* = 0 \).

\[\square\]

5 Simulation Results

In this section we present the results of a small simulation study of the bias and mean square errors of the estimators \( \tilde{\rho} \) and \( \hat{\rho} \). The results are given in Table 1.

We simulated observations from \( SN(-\sqrt{1-\rho}/(1+\rho)) \) for varying values of \( \rho \) and varying sample sizes using the S-Plus function rsn which can be obtained from Azzalini’s website (http://tango.stat.unipd.it/SN/). Each combination of \( \rho \) and sample size \( n \) was replicated 1000 times using the computer software S-Plus 6 implemented in a Windows PC. Estimates of the bias and mean square error for each estimator are then obtained by averaging the bias and the squared error of the 1000 estimates.

Table 1 shows that when \( \rho \) is close to \(-1\), \( \hat{\rho} \) generally outperforms \( \tilde{\rho} \) but when \( \rho \) is close to \( 1 \) \( \tilde{\rho} \) generally outperforms \( \hat{\rho} \).

Table 1 also shows that the bias of the truncated Roberts’ estimator could be substantial when the sample size is small and when \( \rho \) is close to \(-1\). Similarly the bias of the maximum likelihood estimator could also be substantial when the sample size is small and when \( \rho \) is close to \( 1 \). As the sample size
Robert’s correlation model

gets large, both the bias and mean square errors of the two estimators approach zero (which is as predicted by Corollary 3.1 for the truncated Roberts’ estimator).

Finally we note that if Roberts’ model will be utilized in the analysis of a twin study, it is more likely that $\rho$ is closer to 1 than to -1 since it is reasonable to expect that variables or traits measured on twins are highly positively correlated. In this case, the practitioner is advised to use the truncated Roberts’ estimator in the estimation of the correlation coefficient specially if there is a small set of twins available for the study.

Table 1: Comparison of the performance of the Truncated Roberts’ Estimator and the Maximum Likelihood Estimator of the Correlation Coefficient in Roberts’ Model

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\rho$</th>
<th>$\hat{\rho}$ Bias</th>
<th>$\tilde{\rho}$ Bias</th>
<th>$\hat{\rho}$ MSE</th>
<th>$\tilde{\rho}$ MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>-0.99</td>
<td>0.44355</td>
<td>-0.00121</td>
<td>0.46282</td>
<td>0.00067</td>
</tr>
<tr>
<td></td>
<td>-0.9</td>
<td>0.38470</td>
<td>-0.01506</td>
<td>0.46315</td>
<td>0.01744</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.12844</td>
<td>-0.10327</td>
<td>0.55017</td>
<td>0.30571</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>-0.16453</td>
<td>-0.32231</td>
<td>0.27049</td>
<td>0.32009</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>-0.16045</td>
<td>-0.33188</td>
<td>0.17694</td>
<td>0.28936</td>
</tr>
<tr>
<td>50</td>
<td>-0.99</td>
<td>0.16752</td>
<td>-0.00217</td>
<td>0.08823</td>
<td>0.00011</td>
</tr>
<tr>
<td></td>
<td>-0.9</td>
<td>0.12932</td>
<td>-0.00883</td>
<td>0.09391</td>
<td>0.00374</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.02210</td>
<td>-0.02023</td>
<td>0.16281</td>
<td>0.06931</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>-0.00781</td>
<td>-0.05332</td>
<td>0.02437</td>
<td>0.02825</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>-0.02880</td>
<td>-0.06343</td>
<td>0.00928</td>
<td>0.01413</td>
</tr>
<tr>
<td>100</td>
<td>-0.99</td>
<td>0.12287</td>
<td>-0.00159</td>
<td>0.04522</td>
<td>0.00007</td>
</tr>
<tr>
<td></td>
<td>-0.9</td>
<td>0.07249</td>
<td>-0.00539</td>
<td>0.04597</td>
<td>0.00201</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>-0.00974</td>
<td>-0.01608</td>
<td>0.08703</td>
<td>0.03648</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>-0.00191</td>
<td>-0.03029</td>
<td>0.01353</td>
<td>0.01434</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>-0.01277</td>
<td>-0.03180</td>
<td>0.00240</td>
<td>0.00411</td>
</tr>
</tbody>
</table>

References


Received: October 3, 2007