ON THE BIAS AND MEAN-SQUARE ERROR OF THE SAMPLE MINIMUM AND THE MAXIMUM LIKELIHOOD ESTIMATOR FOR TWO ORDERED NORMAL MEANS

SANJAY CHAUDHURI AND MATHIAS DRTON

Abstract. This note considers the estimation of ordered means of two independent normal random variables. When estimating the smaller one of the two means, two simple (non-trivial) estimators are available: the minimum of the two normal random variables and the maximum likelihood estimator. For these two estimators we provide simple elementary derivations of bias and mean square error.

1. Estimation of ordered means of two independent normal random variables

Let $X_1 \sim N(\mu_1, \sigma^2)$ and $X_2 \sim N(\mu_2, \sigma^2)$ be two independent normally distributed random variables with common variance $\sigma^2 > 0$. Assume the prior knowledge $\mu_1 \geq \mu_2$ which yields the model

$$\big(N(\mu_1, \sigma^2) \otimes N(\mu_2, \sigma^2) \mid \mu_1 \geq \mu_2,\ \sigma^2 \in (0, \infty)\big). \tag{1.1}$$

This note considers the estimation of the smaller mean μ_2 under quadratic loss, which means that different estimators are compared in terms of their mean square error (MSE). This problem is also considered in Cohen and Sackrowitz [1]. The following Lemma 1 states that the minimum $M \equiv \min(X_1, X_2)$ dominates the unbiased estimator X_2 in the estimation of μ_2.

Lemma 1 (Cohen and Sackrowitz [1], Lemma A.1)**.** *Suppose X_1 and X_2 are independent normal random variables distributed as*

$$(X_1, X_2) \sim N(\mu_1, \sigma^2) \otimes N(\mu_2, \sigma^2),$$

where $\mu_1 \geq \mu_2$. Then for estimating μ_2 the MSE of $M \equiv \min(X_1, X_2) \leq$ MSE of X_2.

The proof in Cohen and Sackrowitz [1] refers to the Theorem in Lee [2]. However, from this Theorem it follows only that MSE of $\hat{\mu}_2 \leq$ MSE of X_2, where $\hat{\mu}_2$ is the second component of the maximum likelihood estimator (MLE) $\hat{\mu}$ of μ in the model (1.1). The MLE $\hat{\mu}_2$ differs, however, from the minimum M, as shown in the following.

Date: July 15, 2003.

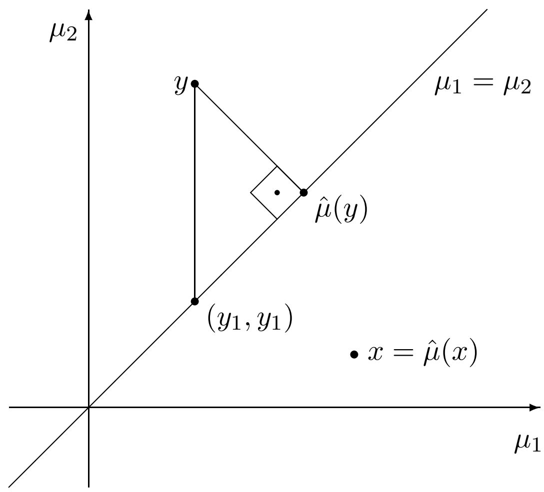

FIGURE 1. Illustration of the MLE. For the observation $x = (x_1, x_2)$ with $x_1 \geq x_2$ the MLE $\hat{\mu}(x)$ coincides with x. For the observation $y = (y_1, y_2)$ with $y_1 < y_2$ the MLE $\hat{\mu}(y)$ is the orthogonal projection on the line $\mu_1 = \mu_2$.

The log-likelihood $\ell(\mu \mid x)$ of model (1.1) based on the observation of $x = (x_1, x_2)$ is

$$\ell(\mu \mid x) = -\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{2}(x_i - \mu_i)^2.$$

Hence, to find $\hat{\mu}$ we need to find the minimum of the squared Euclidean distance between x and μ, i.e.

$$|x - \mu|^2 = (x_1 - \mu_1)^2 + (x_2 - \mu_2)^2,$$

where μ is restricted by $\mu_1 \geq \mu_2$.

We have to distinguish two cases.

1. If $x_1 \geq x_2$ then the Euclidean distance is obviously minimized by $\hat{\mu} = x$.
2. If $x_1 < x_2$ then the MLE is given by the orthogonal projection of x onto the line $\{\mu \mid \mu_1 = \mu_2\}$. This projection is given by $\hat{\mu} = (\bar{x}, \bar{x})$ where $\bar{x} = (x_1 + x_2)/2$ is the mean of x. This is easily verified by checking that

$$x - \hat{\mu} = (x_2 - \bar{x})\begin{pmatrix} -1 \\ 1 \end{pmatrix} \perp \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Figure 1 illustrates the two cases and the resulting MLE. Note also that $\hat{\mu}_2$ equals the mean of M and X_2, i.e. $\hat{\mu}_2 = (M + X_2)/2$, and hence, $M \leq \hat{\mu}_2 \leq X_2$.

2. THE MEAN SQUARE ERROR OF THE MINIMUM

Next, we provide a simple proof of Lemma 1.

Proof of Lemma 1. Consider the difference in MSEs

$$\begin{aligned} \text{MSE}(M) - \text{MSE}(X_2) &= E[(M-\mu_2)^2 - (X_2-\mu_2)^2] \\ &= E[\big((X_1-\mu_2)^2 - (X_2-\mu_2)^2\big)1_{\{X_1 \le X_2\}}] \\ &= E[\big((X_1-X_2)(X_1+X_2-2\mu_2)\big)1_{\{X_1 \le X_2\}}]. \end{aligned} \tag{2.1}$$

Now, transform the variables as

$$\begin{pmatrix} Z \\ W \end{pmatrix} \equiv \begin{pmatrix} X_1 - X_2 \\ X_1 + X_2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}. \tag{2.2}$$

The new variables' joint distribution is

$$\begin{pmatrix} Z \\ W \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 - \mu_2 \\ \mu_1 + \mu_2 \end{pmatrix}, \begin{pmatrix} 2\sigma^2 & 0 \\ 0 & 2\sigma^2 \end{pmatrix} \right). \tag{2.3}$$

In particular, Z and W are independent. Hence,

$$\begin{aligned} \text{MSE}(M) - \text{MSE}(X_2) &= E[Z(W-2\mu_2)1_{\{Z \le 0\}}] \\ &= E[Z1_{\{Z \le 0\}}]E[W-2\mu_2] \\ &= E[Z1_{\{Z \le 0\}}](\mu_1-\mu_2) \\ &\le 0 \times (\mu_1-\mu_2) \le 0, \end{aligned} \tag{2.4}$$

which yields our claim. □

This proof can be readily extended to calculate the exact value of the MSE of the minimum M. We use the notation φ and Φ for the density and the cdf of $N(0,1)$.

Theorem 2 (Mean square error of minimum)**.** *The MSE of the minimum M equals*

$$\text{MSE}(M) = \sigma^2 \left\{ 1 - 2\gamma\varphi(\gamma) + 2\gamma^2 (1 - \Phi(\gamma)) \right\}, \tag{2.5}$$

where $\gamma \equiv \frac{\mu_1 - \mu_2}{\sqrt{2}\sigma} \ge 0$.

Proof. Picking up the proof of Lemma 1, it follows by Lemma A.1 that

$$E[Z1_{\{Z \le 0\}}] = -\sqrt{2}\sigma\varphi(\gamma) + (1-\Phi(\gamma))(\mu_1-\mu_2), \tag{2.6}$$

with

$$\gamma = \frac{E[Z]}{\sqrt{Var[Z]}} = \frac{\mu_1-\mu_2}{\sqrt{2}\sigma}. \tag{2.7}$$

Multiplying with the factor $(\mu_1 - \mu_2)$ from (2.4) and using that $\text{MSE}(X_2) = \sigma^2$, we obtain

$$\begin{aligned} \text{MSE}(M) &= \sigma^2 - \sqrt{2}\sigma(\mu_1-\mu_2)\varphi(\gamma) + (1-\Phi(\gamma))(\mu_1-\mu_2)^2 \\ &= \sigma^2 \left\{ 1 - 2\gamma\varphi(\gamma) + 2\gamma^2(1-\Phi(\gamma)) \right\}. \end{aligned} \tag{2.8}$$

□

Remark 3. By the well-known inequality for Mills' ratio (see e.g. Shorack [3, p. 237]), $x\{1-\Phi(x)\} \le \varphi(x)$ for $x \ge 0$, we can bound the MSE of M by

$$\text{MSE}(M) \le \sigma^2 \{1 - 2\gamma\varphi(\gamma) + 2\gamma\varphi(\gamma)\} = \sigma^2 = \text{MSE}(X_2). \tag{2.9}$$

This reconfirms the result from Lemma 1. □

3. The mean square error of the MLE

As shown in Section 1, the MLE of μ_2 in the model (1.1) is given by

$$\hat{\mu}_2 = \begin{cases} x_2 & : \quad \text{if } x_1 \ge x_2, \\ \bar{x} & : \quad \text{else.} \end{cases} \tag{3.1}$$

By a similar approach as for the minimum, we can compute the exact value of the MSE of $\hat{\mu}_2$.

Theorem 4 (Mean square error of MLE)**.** *The MSE of the MLE $\hat{\mu}_2$ equals*

$$\text{MSE}(\hat{\mu}_2) = \sigma^2 \left\{ 1 - \frac{\gamma}{2}\varphi(\gamma) + \frac{\gamma^2 - 1}{2}\left(1 - \Phi(\gamma)\right) \right\}, \tag{3.2}$$

where $\gamma \equiv \frac{\mu_1 - \mu_2}{\sqrt{2}\sigma} \ge 0$.

Proof. Consider the difference in MSEs,

$$\begin{aligned} &\text{MSE}(\hat{\mu}_2) - \text{MSE}(X_2) \\ &= E\left[\left((\bar{X}-\mu_2)^2 - (X_2-\mu_2)^2\right) 1_{\{X_1 \le X_2\}}\right] \\ &= E\left[\left(\bar{X}^2 - 2\bar{X}\mu_2 - X_2^2 + 2X_2\mu_2\right) 1_{\{X_1 \le X_2\}}\right] \\ &= E\left[\left(\frac{1}{4}X_1^2 - \frac{3}{4}X_2^2 + \frac{1}{2}X_1X_2 - X_1\mu_2 + X_2\mu_2\right) 1_{\{X_1 \le X_2\}}\right] \\ &= E\left[\frac{1}{4}(X_1 - X_2)(X_1 + 3X_2 - 4\mu_2) 1_{\{X_1 \le X_2\}}\right]. \end{aligned} \tag{3.3}$$

Now, transform the variables as

$$\begin{pmatrix} Z \\ Y \end{pmatrix} \equiv \begin{pmatrix} X_1 - X_2 \\ X_1 + 3X_2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}. \tag{3.4}$$

The new variables' joint distribution is

$$\begin{pmatrix} Z \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 - \mu_2 \\ \mu_1 + 3\mu_2 \end{pmatrix}, \begin{pmatrix} 2\sigma^2 & -2\sigma^2 \\ -2\sigma^2 & 10\sigma^2 \end{pmatrix} \right). \tag{3.5}$$

It follows that

$$\begin{aligned}
&\mathrm{MSE}(\hat{\mu}_2) - \mathrm{MSE}(X_2) \\
&= \frac{1}{4}E[Z(Y-4\mu_2)1_{\{Z\le 0\}}] \\
&= \frac{1}{4}E\left[Z1_{\{Z\le 0\}}E[Y-4\mu_2 \mid Z]\right] \\
&= \frac{1}{4}E\left[Z1_{\{Z\le 0\}}\left\{(\mu_1+3\mu_2) - \frac{2\sigma^2}{2\sigma^2}(Z-(\mu_1-\mu_2)) - 4\mu_2\right\}\right] \\
&= \frac{1}{4}E\left[Z1_{\{Z\le 0\}}\left(2(\mu_1-\mu_2)-Z\right)\right] \\
&= \frac{1}{2}(\mu_1-\mu_2)E\left[Z1_{\{Z\le 0\}}\right] - \frac{1}{4}E\left[Z^2 1_{\{Z\le 0\}}\right].
\end{aligned} \tag{3.6}$$

We can apply Lemma A.1 with

$$\gamma = \frac{E[Z]}{\sqrt{Var[Z]}} = \frac{\mu_1-\mu_2}{\sqrt{2}\sigma} \tag{3.7}$$

to find

$$\begin{aligned}
&\mathrm{MSE}(\hat{\mu}_2) - \mathrm{MSE}(X_2) \\
&= \frac{1}{2}(\mu_1-\mu_2)\left\{-\sqrt{2}\sigma\varphi(\gamma) + (\mu_1-\mu_2)\left(1-\Phi(\gamma)\right)\right\} \\
&\quad - \frac{1}{4}\left\{-\sqrt{2}\sigma(\mu_1-\mu_2)\varphi(\gamma) + \left((\mu_1-\mu_2)^2+2\sigma^2\right)\left(1-\Phi(\gamma)\right)\right\} \\
&= \frac{1}{4}(\mu_1-\mu_2)\left\{-\sqrt{2}\sigma\varphi(\gamma) + (\mu_1-\mu_2)\left(1-\Phi(\gamma)\right)\right\} - \frac{\sigma^2}{2}\left(1-\Phi(\gamma)\right) \\
&= -\frac{\sigma^2}{2}\gamma\varphi(\gamma) + \frac{\sigma^2}{2}(\gamma^2-1)\left(1-\Phi(\gamma)\right).
\end{aligned} \tag{3.8}$$

Adding $\mathrm{MSE}(X_2)=\sigma^2$ to this equation establishes our claim. □

Remark 5. By the inequality for Mills' ratio, we can bound the MSE of $\hat{\mu}_2$ by

$$\mathrm{MSE}(\hat{\mu}_2) \le \sigma^2\left\{1 - \frac{\gamma}{2}\varphi(\gamma) + \frac{\gamma^2}{2}\left(1-\Phi(\gamma)\right)\right\} \le \sigma^2 = \mathrm{MSE}(X_2). \tag{3.9}$$

This reconfirms the result in the Theorem by Lee [2] in the case of two samples. □

4. Illustration of the mean square errors

In Figure 2, we illustrate the formulas we derived. From Theorem 2 and 4, it can be easily seen that

$$\lim_{\gamma\to 0}\mathrm{MSE}(M) = \sigma^2, \quad \lim_{\gamma\to\infty}\mathrm{MSE}(M) = \sigma^2, \tag{4.1}$$

and

$$\lim_{\gamma\to 0}\mathrm{MSE}(\hat{\mu}_2) = \frac{3}{4}\sigma^2, \quad \lim_{\gamma\to\infty}\mathrm{MSE}(\hat{\mu}_2) = \sigma^2. \tag{4.2}$$

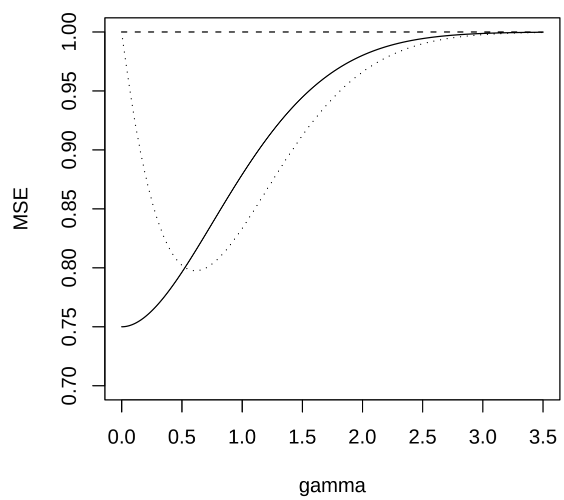

FIGURE 2. MSE of X_2 (dashed), M (dotted), and $\hat{\mu}_2$ (solid) for $\sigma^2 = 1$.

The MSE of the MLE is strictly increasing with γ since

$$\frac{\partial}{\partial \gamma}\mathrm{MSE}(\hat{\mu}_2) = -\sigma^2\gamma(1-\Phi(\gamma)) < 0. \tag{4.3}$$

The minimum of MSE(M) occurs at $\gamma \approx 0.612$ with a value of the MSE of about $0.798\sigma^2$. The MSE values of $\hat{\mu}_2$ and M coincide for $\gamma \approx 0.526$ with an MSE value of about $0.800\sigma^2$. For $\gamma < 0.526$, we have that $\mathrm{MSE}(\hat{\mu}_2) < \mathrm{MSE}(M)$, for $\gamma > 0.526$, $\mathrm{MSE}(\hat{\mu}_2) > \mathrm{MSE}(M)$.

5. The bias of the minimum and the MLE

As shown in the previous sections, the trivial estimator X_2, which is unbiased for μ_2, is dominated in MSE by both the minimum M and the MLE $\hat{\mu}_2$. However, this domination comes at the price of a bias. In the following Theorem 6, we show that both M and $\hat{\mu}_2$ underestimate μ_2 on average but that the bias of the MLE is only half of the bias of the minimum. Figure 3 illustrates the bias formula given in Theorem 6.

Theorem 6 (Bias of minimum and MLE)**.** *The bias of the MLE $\hat{\mu}_2$ equals half the bias of the minimum M. More precisely,*

$$2\big(E[\hat{\mu}_2] - \mu_2\big) = E[M] - \mu_2 = -\sqrt{2}\sigma\big\{\varphi(\gamma) - \gamma(1-\Phi(\gamma))\big\} \le 0, \tag{5.1}$$

where $\gamma \equiv \frac{\mu_1-\mu_2}{\sqrt{2}\sigma} \ge 0$.

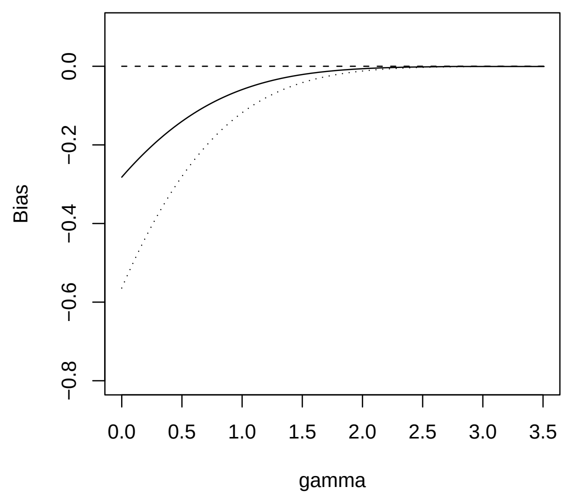

FIGURE 3. Bias of X_2 (dashed), M (dotted), and $\hat{\mu}_2$ (solid) for $\sigma^2 = 1$.

Proof. By (3.1), $\hat{\mu}_2 = (M + X_2)/2$, which yields that

$$E\left[\hat{\mu}_2\right] - \mu_2 = \frac{1}{2} E\left[M + X_2\right] - \mu_2 = \frac{1}{2}(E[M] - \mu_2). \tag{5.2}$$

In order to evaluate the bias, note that

$$E\left[M\right] - \mu_2 = E\left[M - X_2\right] = \frac{1}{2} E\left[(X_1 - X_2) 1_{\{X_1 - X_2 > 0\}}\right] = E\left[Y 1_{\{Y \leq 0\}}\right], \tag{5.3}$$

where $Y = (X_1 - X_2) \sim N\left((\mu_1 - \mu_2), 2\sigma^2\right)$. Then by Lemma A.1,

$$\begin{aligned} E\left[M\right] - \mu_2 &= -\sqrt{2}\sigma\varphi\left(\frac{\mu_1 - \mu_2}{\sqrt{2}\sigma}\right) + (\mu_1 - \mu_2)\left\{1 - \Phi\left(\frac{\mu_1 - \mu_2}{\sqrt{2}\sigma}\right)\right\} \\ &= -\sqrt{2}\sigma\left\{\varphi\left(\gamma\right) - \gamma\left(1 - \Phi\left(\gamma\right)\right)\right\} \leq 0, \end{aligned} \tag{5.4}$$

where we let $\gamma = \frac{\mu_1 - \mu_2}{\sqrt{2}\sigma} \geq 0$ and use the inequality for Mills' ratio. □

ACKNOWLEDGMENTS

We wish to thank Michael Perlman for directing our attention to the topic of this note.

APPENDIX A. A LEMMA FOR NORMAL RANDOM VARIABLES

Lemma A.1. *If* $X \sim N(\mu, \sigma^2)$ *then*

$$E\left[X 1_{\{X \leq 0\}}\right] = -\sigma\varphi\left(\frac{\mu}{\sigma}\right) + \mu\left(1 - \Phi\left(\frac{\mu}{\sigma}\right)\right), \tag{A.1}$$

and

$$E\left[X^2 1_{\{X\le 0\}}\right] = -\mu\sigma\varphi\left(\frac{\mu}{\sigma}\right) + (\mu^2+\sigma^2)\left(1-\Phi\left(\frac{\mu}{\sigma}\right)\right), \tag{A.2}$$

where φ and Φ are the density and the cdf of $N(0,1)$.

Proof. Let $S \equiv (X-\mu)/\sigma \sim N(0,1)$. Then

$$\begin{aligned} E\left[X 1_{\{X\le 0\}}\right] &= \\ &= \sigma E\left[\left(S+\frac{\mu}{\sigma}\right)1_{\{S\le -\mu/\sigma\}}\right] = \sigma E\left[S 1_{\{S\le -\mu/\sigma\}}\right] + \mu\Phi\left(-\frac{\mu}{\sigma}\right) \\ &= \sigma\int_{-\infty}^{-\mu/\sigma} s\varphi(s)\,ds + \mu\Phi\left(-\frac{\mu}{\sigma}\right) = -\sigma\varphi\left(-\frac{\mu}{\sigma}\right) + \mu\Phi\left(-\frac{\mu}{\sigma}\right) \\ &= -\sigma\varphi\left(\frac{\mu}{\sigma}\right) + \mu\left(1-\Phi\left(\frac{\mu}{\sigma}\right)\right), \end{aligned} \tag{A.3}$$

where we use $\varphi'(s) = -s\varphi(s)$.

Moreover,

$$\begin{aligned} E\left[X^2 1_{\{X\le 0\}}\right] &= \sigma^2 E\left[\left(S+\frac{\mu}{\sigma}\right)^2 1_{\{S\le -\mu/\sigma\}}\right] \\ &= \sigma^2\left\{E\left[S^2 1_{S\le -\mu/\sigma}\right] + 2\frac{\mu}{\sigma}E\left[S 1_{S\le -\mu/\sigma}\right] + \left(\frac{\mu}{\sigma}\right)^2\Phi\left(-\frac{\mu}{\sigma}\right)\right\} \\ &= \sigma^2\left\{\int_{-\infty}^{-\mu/\sigma} s^2\varphi(s)\,ds + 2\frac{\mu}{\sigma}\left(-\varphi\left(\frac{\mu}{\sigma}\right)\right) + \left(\frac{\mu}{\sigma}\right)^2\Phi\left(-\frac{\mu}{\sigma}\right)\right\} \\ &= \sigma^2\left\{\Phi\left(-\frac{\mu}{\sigma}\right) + \frac{\mu}{\sigma}\varphi\left(-\frac{\mu}{\sigma}\right) - 2\frac{\mu}{\sigma}\varphi\left(\frac{\mu}{\sigma}\right) + \left(\frac{\mu}{\sigma}\right)^2\Phi\left(-\frac{\mu}{\sigma}\right)\right\} \\ &= (\mu^2+\sigma^2)\Phi\left(-\frac{\mu}{\sigma}\right) - \mu\sigma\varphi\left(\frac{\mu}{\sigma}\right) \\ &= (\mu^2+\sigma^2)\left(1-\Phi\left(\frac{\mu}{\sigma}\right)\right) - \mu\sigma\varphi\left(\frac{\mu}{\sigma}\right), \end{aligned} \tag{A.4}$$

where we use

$$\frac{\partial}{\partial s}\left(\Phi(s) - s\varphi(s)\right) = \varphi(s) - \varphi(s) - s(-s\varphi(s)) = s^2\varphi(s).$$

□

References

1. A. Cohen and H.B. Sackrowitz, *Inference for the model of several treatments and a control*, J. Stat. Plann. Inference **107** (2002), 89–101.
2. C.I.C. Lee, *The quadratic loss of isotonic regression under normality*, Ann. Stat. **9** (1981), 686–688.
3. G.R. Shorack, *Probability for statisticians*, Springer Texts in Statistics, Springer-Verlag, New York, 2000.

Department of Statistics, University of Washington, Seattle, Washington, U.S.A.
E-mail address: {sanjay, drton}@stat.washington.edu