PITMAN'S CLOSENESS CRITERION AND SHRINKAGE ESTIMATES OF THE VARIANCE AND THE S.D.

Yannis G. Yatracos

June 13, 2011

Address: School of Management and Economics, Cyprus U. of Technology, P.O. Box 50329, Lemesos 3603, Cyprus

e-mail: yannis.yatracos@cut.ac.cy

Summary

Pitman's closeness criterion (PCC) became a controversial topic since some statisticians expressed their wish to exclude it from the evaluation criteria of estimates. Herein, PCC is studied for an estimate t and its shrinkage ct, when the unknown parameter of interest θ is positive; $0 < c < 1$. PCC is transitive for shrinkage estimates with decreasing shrinkage coefficients and only t's distribution is needed to compute its value. When θ is the variance σ^2 or the standard deviation σ, exact calculations and simulations confirm that ct, which improves t's mean square error, may not improve *often* t's distance from θ and PCC takes large values. Consequently, some statisticians, their clients and some statistics' users will not use shrinkage estimates of σ^2 and of σ. For this group, PCC is a useful information tool to be used along with other evaluation criteria, as suggested by Rao (1993).

Some key words: Error reduction, Pitman's closeness criterion, Shrinkage estimates

1 Introduction

A common statistical problem is the estimation of an unknown, real-valued parameter θ ($\in \Theta$) of a density f, using n independent, identically f-distributed observations. Assume that potential estimates of θ are t_1 and t_2, and that calculations show that for all $\theta \in \Theta$

$$P[|t_1 - \theta| < |t_2 - \theta|] \tag{1}$$

equals .7. Then, at least some statisticians, their clients and statistics' users, including the author, would choose t_1 over t_2. This decision will remain unchanged for most of them if, in addition, for the mean square error (MSE) of t_1 and of t_2 it holds

$$E(t_2 - \theta)^2 < E(t_1 - \theta)^2 \tag{2}$$

and the difference $E(t_1 - \theta)^2 - E(t_2 - \theta)^2$ is small. However, the decision will be different if the value of (1) is not known: t_2 will be preferred over t_1 because of (2)[1]. Do you believe (1) provides helpful information in this problem?

Probability (1) is Pitman's closeness criterion (PCC, Pitman, 1937) and is vital for the choice of the estimate t_1 in the problem described above. When this probability is larger than .5, t_1 is "Pitman closer" (to θ) than t_2. Rao

[1]High value in (1) and (2) hold for various models when estimating the variance and the standard deviation, $t_2 = ct_1,\ 0 < c < 1$; see Table 1 and Figure 1.

(1980, 1981) stimulated interest for the use of (1) as an information tool.

Robert et al. (1993) pointed some PCC's drawbacks: (1) is difficult to compute because it depends on the joint distribution of t_1 and t_2; with three estimates t_1, t_2 and t_3 of θ, PCC may not be transitive, namely, it may occur that t_1 is Pitman closer than t_2 and t_2 is Pitman closer than t_3 but t_1 is not Pitman closer than t_3; PCC may "take into account a subset of the sample space that occurs only with probability slightly greater than 50%." In the examples presented, PCC's values were less than .62 (p. 76).

Other views, in PCC's favor or against it, were presented in the paper's discussion by Blyth, Casella and Wells, Ghosh, Keating and Sen, Peddada, and Rao. In the rejoinder, the authors agreed with Casella and Wells (1993), hoping that "the use of Pitman closeness as an evaluation criterion will finally cease". If their wish was binding, we would not have the information given by (1) !

In his comments to the paper by Robert et al. (1993), Professor Rao (1993) altered the probabilities in Example 2.1 and obtained the value $1-10^{-23}$ for (1) (the PCC). He asked whether t_2 should be used, having smaller risk than t_1 for the L^p loss, $p > 1$. Professor Rao answered his question: "The authors would perhaps recommend t_2, but a shrewd businessman may prefer t_1." Professor Rao also wrote: "I believe that the performance of an estimator should be examined under different criteria to understand the nature of the estimator and possibly to provide information to the decision maker. I would include PCC in my list of criteria, except perhaps in the rare case where the customer has a definite loss function."[2]

In this work, PCC is studied for shrinkage estimates of a positive parameter θ. When θ is the variance or the standard deviation, PCC takes values between .55 and .9 for various distributions, thus providing useful information for *some* decision makers, at least when PCC is high. The interested reader may consult Nayak (1994) and references therein for different PCC studies related with σ^2.

2 PCC for shrinkage estimates of $\theta > 0$

We use PCC to obtain information for shrinkage estimates of θ. For this problem, a competing estimate of t_1 has the form $t_2 = ct_1,\ c > 0$. Hence, the first consequence is that (1) is calculated from the distribution of t_1 only. Observe also that for $c > d > 0,\ \theta > 0$ and $t_1 > 0$,

$$P(|ct_1 - \theta| < |dt_1 - \theta|) = P(t_1 < \frac{2\theta}{c+d}). \tag{3}$$

Then, for $1 > c > d > 0$, if t_1 is Pitman closer than ct_1 and ct_1 is Pitman closer than dt_1, it follows from (3) that t_1 is Pitman closer than dt_1. Thus,

[2]No comments on these issues raised by Professor Rao appear in the rejoinder.

for this estimation problem, PCC is transitive for shrinkage estimates with decreasing shrinkage coefficients.

3 Shrinkage estimates of σ^2 and σ

For a sample $X_1, \ldots, X_n$ from the normal model with mean μ and variance σ^2 both unknown, the shrinkage estimate $\frac{1}{n+1}\sum_{i=1}^n (X_i - \bar{X})^2$ improves the MSE of the unbiased estimate of σ^2 for every n and all parameter values (see Lehmann 1983, p. 113). Under different models and assumptions, inadmissibility results in variance and standard deviation estimation were proved using shrinkage estimates, among others, by Stein (1964), Brown (1968) and Arnold (1970). Estimates that improve the MSE of the unbiased estimate of the variance and the covariance for every probability model and every n were recently presented (Yatracos, 2005).

Despite these and other similar results, for many non-academic statisticians it is not evident one should use shrinkage estimate ct, when it improves t's MSE. The reason is that the MSE can be drastically affected by extreme t-values having low probability, whereas in practice, an estimates' value is used and some decision makers, like Professor Rao's"shrewd businessman", want to know the chance t is closer to its target than its competitor ct. This information is obtained by computing (1).

4 The Examples

When $\theta = \sigma^2,\ \sigma$, (1) is obtained for shrinkage estimates under various models and for $n \leq 50$ takes values between .55 and .9.

Example 4.1. Let $X_1, ..., X_n$ be i.i.d. normal random variables with unknown mean μ and variance σ^2. The parameter of interest $\theta = \sigma^2$, and estimates of the form $b\sum_{k=1}^{n}(X_k - \bar{X}_n)^2$ are considered: the MLE $\hat{t}_n$ with $b = \frac{1}{n}$, the unbiased (jacknife) estimate $t_{n,J}$ with $b = \frac{1}{n-1}$, and the minimum risk equivariant estimate $\tilde{t}_n$ minimizing the MSE uniformly in n and σ with $b = \frac{1}{n+1}$. $\tilde{t}_n$ is a shrinkage estimate of $\hat{t}_n$ and of $t_{n,J}$, $\tilde{t}_n = \frac{n}{n+1}\hat{t}_n = \frac{n-1}{n+1}t_{n,J}$, and from (3) it holds

$$P\left[(\tilde{t}_n - \sigma^2)^2 > (\hat{t}_n - \sigma^2)^2\right] = P\left[\chi^2_{n-1} < \frac{2n(n+1)}{2n+1}\right], \tag{4}$$

$$P\left[(\tilde{t}_n - \sigma^2)^2 > (t_{n,J} - \sigma^2)^2\right] = P\left[\chi^2_{n-1} < \frac{(n^2-1)}{n}\right], \tag{5}$$

since $\sum_{i=1}^{n}(X_i - \bar{X})^2/\sigma^2$ follows a chi-square distribution with $(n-1)$ degrees of freedom, denoted χ^2_{n-1}. Tables 1 and 2 provide probabilities (4) and (5) for some n-values, $2 \leq n \leq 50$. The MLE $\hat{t}_n$ has more often smaller actual error than $\tilde{t}_n$.

The same phenomenon is observed for T-distributions, for which $\tilde{t}_n$ improves the MSE of $\hat{t}_n$ and $t_{n,J}$ when estimating σ^2 (Yatracos, 2005, *Remark 7*, p. 1174). Using 1000 simulations from T_m-distributions with degrees of freedom $m = 3,\ 10,\ 20,\ 30$, probability (4) is estimated for $n \leq 30$ and the results are in Figure 1 (after the references). The probabilities decrease slower than those in Table 1. The largest probability is .908, obtained for a T-distribution with 3 degrees of freedom and $n = 2$.

COMPARING $\tilde{t}_n$ **WITH** $\hat{t}_n$						
PROBABILITIES (4) **FOR** $2 \leq n \leq 50$						
n	2	5	8	15	36	50
	0.8787	0.7562	0.7071	0.6541	0.6008	0.5858

Table 1

COMPARING $\tilde{t}_n$ **WITH** $t_{n,J}$						
PROBABILITIES (5) **FOR** $2 \leq n \leq 50$						
n	2	4	8	21	37	50
	0.7793	0.7102	0.6563	0.6001	0.5762	0.5658

Table 2

Example 4.2. Let $X_1, ..., X_n$ be i.i.d. random variables with density $f(x;\mu,\sigma) = \sigma^{-1}e^{-(x-\mu)/\sigma},\ x > \mu$, with $-\infty < \mu < +\infty,\ \sigma > 0$ and $X_{(1)}, \ldots, X_{(n)}$ the order statistics. The parameter of interest $\theta = \sigma$ and estimates of the form $b\sum_{k=2}^{n}(X_{(k)} - X_{(1)})$ are considered: the MLE $\hat{t}_n$ with $b = \frac{1}{n-1}$ that is also minimum variance unbiased and the estimate $\tilde{t}_n$ with $b = \frac{1}{n}$ that improves $\hat{t}_n$

uniformly. $\tilde{t}_n$ is a shrinkage estimate of $\hat{t}_n$, $\tilde{t}_n = \frac{n-1}{n}\hat{t}_n$, and from (3) it holds

$$P\left[(\tilde{t}_n - \sigma)^2 > (\hat{t}_n - \sigma)^2\right] = P\left[\Gamma_{n-1} < \frac{2n(n-1)}{2n-1}\right], \qquad (6)$$

since $\sum_{k=2}^{n}(X_{(k)} - X_{(1)})/\sigma$ follows a *Gamma* distribution with $n-1$ degrees of freedom, denoted Γ_{n-1}. Table 3 provides probability (6) for some n-values, $2 \le n \le 50$.

COMPARING $\tilde{t}_n$ WITH $\hat{t}_n$						
PROBABILITIES (6) **FOR** $2 \le n \le 50$						
n	2	3	11	30	44	50
	0.7364	0.6916	0.6001	0.5607	0.5501	0.5470

Table 3

Remark 4.1 *From the findings for the models in this section, it seems that at least for $n \le 10$, some statisticians, their clients and some statistics' users will not use shrinkage estimates of the variance or of the standard deviation.*

5 Concluding Remarks

A controversy started when some statistical decision theorists took the extreme position that PCC should not be used as an estimates' evaluation criterion. Examples have been presented herein showing that, for small and moderate sample size, PCC provides useful information for deciding whether to use or not shrinkage estimates of σ^2 and of σ. The message is twofold: PCC should be used along with all available evaluation criteria and shrinkage estimates of σ^2 and of σ should not be used unless proved successful for additional evaluation criteria. Some decision theorists will not be pleased on both counts.

References

[1] Arnold, B. (1970) Inadmissibility of the usual scale estimate for a shifted exponential distribution. *JASA*, **65**, 1260-1264.

[2] Blyth, C. R. (1993) Is Pitman Closeness a Reasonable Criterion? Comment *JASA* **88**, 72-74.

[3] Brown, L.(1968) Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Ann. Math. Statist.* **39**, 29-48.

[4] Casella, G. and Wells, M. T. (1993) Is Pitman Closeness a Reasonable Criterion? Comment *JASA* **88**, 70-71.

[5] Ghosh, M., Keating, J. P. and Sen, P. K. (1993) Is Pitman Closeness a Reasonable Criterion? Comment *JASA* **88**, 63-66.

[6] Lehmann, E. L. (1983) *Theory of Point Estimation.* Wiley, New York.

[7] Nayak, T. K. (1994) Pitman Nearness Comparison of Some Estimators of Population Variance. *The American Statistician* **48**, 99-102.

[8] Peddada, S. D. (1993) Is Pitman Closeness a Reasonable Criterion? Comment *JASA* **88**, 67-69.

[9] Rao, C. R. (1993) Is Pitman Closeness a Reasonable Criterion? Comment *JASA* **88**, 69-70.

[10] Rao, C. R. (1981) Some Comments on the Minimum Mean Square Error as a Criterion of Estimation. In *Statistics and Related Topics*, eds. M. Csorgo, D. Dawson, J. Rao, and A. Saleh, North-Holland, Elsevier, New York, 123-143.

[11] Rao, C. R. (1980) Discussion of 'Minimum Chi-square, Not Maximum Likelihood' by J. Berkson, *Ann. Stat.* **8**, 482-485.

[12] Robert, C. P., Hwang J. T. G. and Strawderman, W. (1993) Is Pitman Closeness a Reasonable Criterion? *JASA* **88**, 57-63; Rejoinder, 74-76.

[13] Stein, C. (1964) Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155-160.

[14] Yatracos, Y. G. (2005) Artificially augmented samples, shrinkage and MSE reduction. *JASA* **100**, 1168-1175.

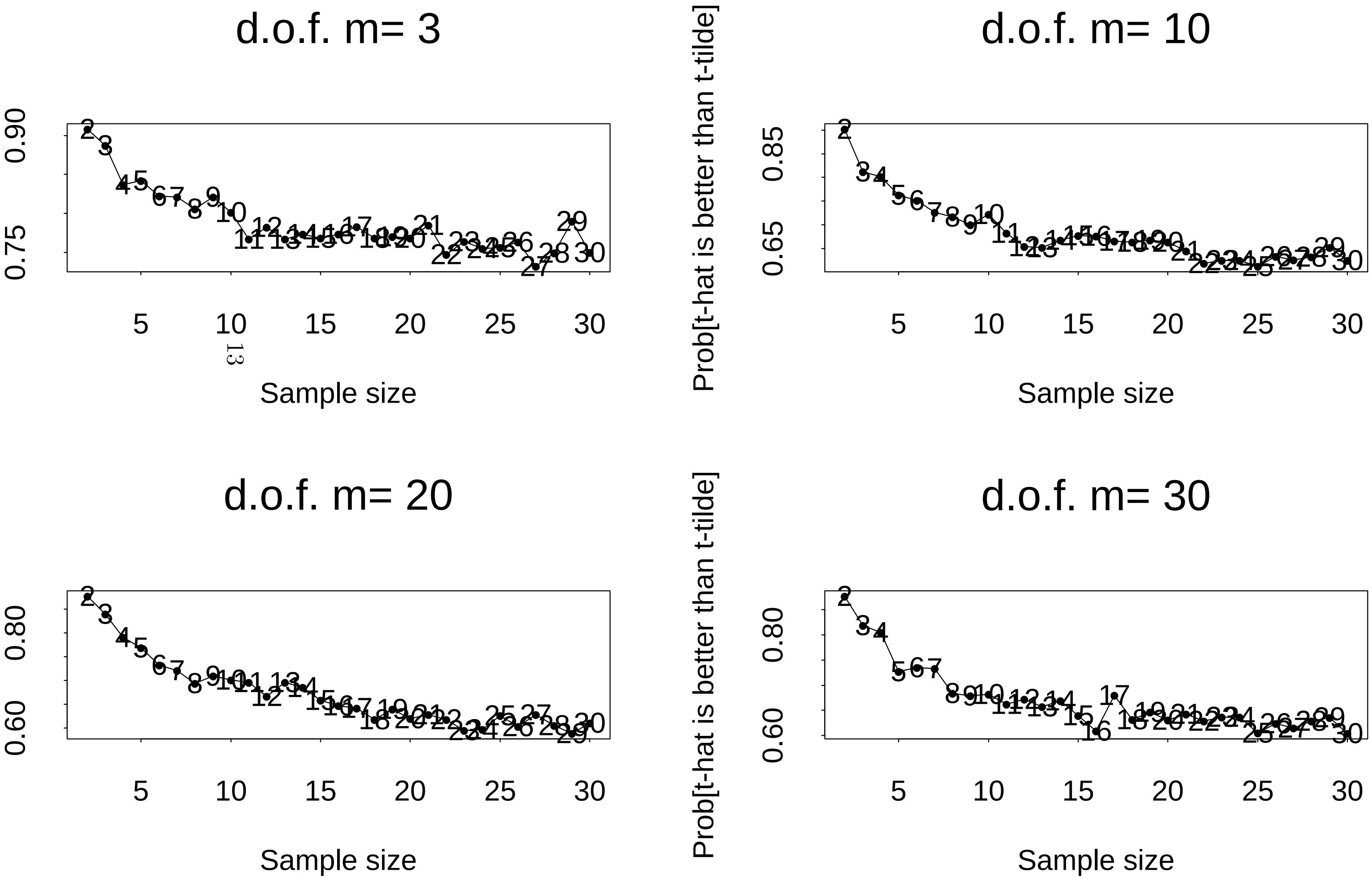
d.o.f. m= 3
d.o.f. m= 10
d.o.f. m= 20
d.o.f. m= 30
Prob[t-hat is better than t-tilde]
Prob[t-hat is better than t-tilde]
Sample size
Sample size
Sample size
Sample size

Figure 1