COMPARING THE SAMPLE MEAN AND THE SAMPLE MEDIAN: AN EXPLORATION IN THE EXPONENTIAL POWER FAMILY

Michael Sherman*
Texas A&M University

Abstract:

We compare the (asymptotic) efficiencies of the sample mean and the sample median in the exponential power family. This single family of densities is an ideal one to illustrate the effect of "heavy tails" on the two estimators. We find the density for which the mean and median are equally efficient, and observe what happens for limiting cases within the family. Only basic tools are needed in the analysis and the statistical development leads to natural discoveries of mathematical facts, e.g., the nonmonotonicity and existence of a minimum of the gamma function.

Keywords: asymptotic, efficiency, gamma function

* Michael Sherman is Assistant Professor in the Department of Statistics, Texas A&M University, College Station, TX 78143

1: INTRODUCTION

In introductory mathematical statistics courses the concept of (asymptotic) efficiency of an estimator is frequently discussed. The efficiency of the ubiquitous estimators the sample mean, $\overline{x}$, and the sample median, $\tilde{x}$, are often compared for the normal distribution, in which case the ratio of the variance of $\overline{x}$ to $\tilde{x}$ is eff($\overline{x}$,$\tilde{x}$):=$\lim_{n\to\infty}$Var($\overline{x}$)/Var($\tilde{x}$)=$2/\pi$. Thus, the mean is a more efficient estimator of location for normally distributed data. In addition to this isolated result, it is sometimes shown that for the Cauchy distribution eff($\overline{x}$,$\tilde{x}$)$\to\infty$, which shows that the mean is very inefficient for Cauchy data.

These isolated results give no indication of what happens for distributions that have tail behavior between these two extremes, nor of what happens for tails that are less heavy than those of the normal. In this note we consider the exponential power family of densities:

$$f(x,p) = c(p)exp(-|x-\theta|^p), \text{ for } x \in (-\infty,\infty), p \in (0,\infty), \text{ and } \theta \in (-\infty,\infty).$$

This family is seldom studied due to the difficulty of estimating θ and p simultaneously. For this reason it is not of much use in modelling data. The goal in this note, however, is to study the effect of marginal variability on the efficiencies of $\overline{x}$ and $\tilde{x}$. We seek to do this with minimal mathematical machinery, and for this purpose the exponential power family is particularly well suited.

In Section 2 we introduce this family of densities and discuss its properties as a function of the parameter p. Then we show, through direct computations, that we can calculate Var($\overline{x}$) exactly, and the asymptotic Var($\tilde{x}$) (and hence eff($\overline{x}$,$\tilde{x}$)) for all finite p, as well as for the limiting cases, p$\to$ 0 and p$\to\infty$. In the statistical development we discover an interesting property of the gamma function, namely its nonmonotonicity. Knowledge of only single variable calculus, some properties of the gamma function, and the use of Sterling's formula is required for all derivations. These factors make this a powerful example for an introductory course in mathematical statistics.

We note that the exponential power family has been discussed by, e.g., Box and Tiao (1973), although their formulation only allows for p $\in(1,\infty)$. The properties of various robust and adaptive estimators for data from this family have been studied by, e.g., D'Agostino and Lee (1974) and Hogg (1974). Also we note that the estimators, $\overline{x}$ and $\tilde{x}$, are considered here for ease of exposition. These are the maximum likelihood estimators only when p=2 and p=1, respectively. For other known values of p, the asymptotically best linear estimator can be obtained using the results of Chernoff, Gastwirth, and Johns (1967), see also D'Agostino and Lee (1974).

2: COMPARING THE EFFICIENCY OF $\overline{x}$ AND $\tilde{x}$ IN THE EXPONENTIAL POWER FAMILY

Consider observations generated from the density $f(x,p) = c(p)exp(-|x-\theta|^p)$, for $x \in$

$(-\infty,\infty), p \in (0,\infty)$, and $\theta \in (-\infty,\infty)$. θ is the location parameter of statistical interest, p determines the spread of the distribution, and c(p) is the constant that makes f(x,p) a density for each p. For each p, f(x,p) is symmetric about θ, and thus θ is both the mean and the median of the distribution. We will show that the mean is finite for all values of p.

As we vary the parameter p, we obtain densities with quite different tail behavior. For p=2, we obtain the normal density with σ^2=1/2, for p=1, the double exponential. As p$\to$ 0, f(x,p) has very heavy tails and very light tails as p$\to \infty$. In fact as p$\to \infty$, f(x,p) converges to the Uniform distribution on [$-$1,1]. Pictured in Figure 1 are six representative members of this family of densities.

Our goal is to study the efficiency of the two estimators of θ: the sample mean, $\overline{x}$, and the sample median, $\tilde{x}$. Noting that the variance of these estimators does not depend on the value of θ, we take θ=0, without loss of generality. We take it to be known that nVar($\tilde{x}$)$\to$ $(4f^2(0,p))^{-1}$ as n$\to \infty$. This fact can be found in many textbooks, e.g., Devore (1995, p.263, problem 18).

In order to compute the variance of the two estimators for different values of p we need the following definitions: First we define the gamma function: for $t > 0, \Gamma(t) = \int_0^\infty y^{t-1}exp(-y)dy$. This function has the following three properties which we will use: 1) For $t > 0, \Gamma(t+1) = t\Gamma(t)$, 2) for t an integer, we have $\Gamma(t+1) = t!$ (1) and 2) can be shown using integration by parts), and 3) $\Gamma(1/2) = \pi^{1/2}$ (which follows by making the substitution $z = y^{1/2}$ and using the fact that the normal density integrates to 1). We will also use Stirling's formula to approximate factorials: for integer s, $s! \sim (2\pi)^{1/2}exp(-s)s^{s+1/2}$ ($\sim$ means that the ratio of the two sides tends to 1 as $s \to \infty$, see e.g., Feller (1968), II.9). We first derive the value of c(p), and the value of all positive moments of f(x,p). In particular this will give us the variance of f(x,p).

For even values of r$\geq$ 0 we have:

$$E(X^r) = \int_{-\infty}^{\infty} x^r c(p)exp(-|x|^p)dx = 2c(p)\int_0^\infty x^r exp(-x^p)dx$$
$$= 2(c(p)/p)\int_0^\infty y^{(r+1)/p-1}exp(-y)dy = 2c(p)\Gamma((r+1)/p)/p.$$

This follows by making the substitution $y = x^p$, and from the definition of the gamma function. Setting r=0, and setting the final expression equal to 1 shows that $c(p) = p/(2\Gamma(1/p)) = (2\Gamma(1 + (1/p)))^{-1}$, and thus that $E(X^r) = \Gamma((r+1)/p)/\Gamma(1/p)$. This coupled with the fact that $E(X) = 0$ implies that

$$nVar(\overline{x}) = Var(X) = \Gamma(3/p)/\Gamma(1/p) =: \sigma^2, \text{ say.}$$

Also noting that f(0,p)=c(p) we have

$$nVar(\tilde{x}) \to (4c^2(p))^{-1} = (\Gamma(1/p)/p)^2 = \Gamma^2(1+1/p)) =: \nu^2, \text{ say.}$$

To get some feel for these expressions, from the properties of the gamma function we can verify that for the normal distribution (p=2), we have $\sigma^2 = 1/2$, $\nu^2 = \pi/4$, and thus eff($\overline{x}$,$\tilde{x}$)=$2/\pi$. For the double exponential distribution (p=1), we have $\sigma^2 = 2$, $\nu^2 = 1$, and thus eff($\overline{x}$,$\tilde{x}$)=2. These are the well known results that for the normal distribution $\overline{x}$ is more efficient than $\tilde{x}$, with the reverse holding for the heavier tailed double exponential distribution. This raises the natural question:

Is there a value of p between 1 and 2 such that eff($\overline{x}$,$\tilde{x}$)=1?

This is the state of nature where we can rest easily that $\overline{x}$ and $\tilde{x}$ are equally efficient (or we could agonize over which one to employ). We can see that the answer to the above question is yes, and that the value is the one such that $p^2\Gamma(3/p) = \Gamma^3(1/p)$). It turns out that this value is approximately p^* =1.407426 (see Figure 1 for a picture of this density). Some other values of σ^2, ν^2, and the ratio, eff($\overline{x}$,$\tilde{x}$), are given in Table 1:

Table 1: The Aymptotic Variances of the Sample Mean and the Sample Median for the Family f(x,p)

p	σ^2	ν^2	eff($\overline{x}$,$\tilde{x}$)
.2	3.632X10^9	1.44X10^4	2.522X10^5
.5	120	4	30
1	2	1	2
1.407426	.8293	.8293	1
2	.5	.7854	.6366
6	.3184	.8607	.3700

We see from Table 1 that for p≤1.407426 the sample mean is worse than the sample median, but better for p≥1.407426. Table 1 raises some interesting further questions: What happens to σ^2, ν^2, and the ratio eff($\overline{x}$,$\tilde{x}$) as p→ 0, p→ ∞? Also, we see that ν^2 is not monotone in p. Is the same true for σ^2? When is eff($\overline{x}$,$\tilde{x}$) a minimum in p?

We first consider the heavy tail case, p→ 0. In this case, let s=(1−p)/p so that for integer s: $\sigma^2 = \Gamma(3(s+1))/\Gamma(s+1) = (3s+2)!/s! \to \infty$ as s→ ∞ (i.e., p→ 0). Similarly, $\nu^2 = ((s+1)!)^2 \to \infty$. Note, however, that eff($\overline{x}$,$\tilde{x}$)= $(2\pi)^{-1}((3s+2)^{3s+3/2}s^{-s-1/2}(s+1)^{-2s-3} \to \infty$ as $s \to \infty$. This follows from Stirling's formula. Thus, for very heavy tail distributions within the exponential power family we see that the variances of both $\overline{x}$ and $\tilde{x}$ become unbounded, but that the mean is still very inefficient relative to the median for small values of p.

As p→ ∞, by the continuity of the gamma function we have $\nu^2 \to \Gamma(1)$ =1. This agrees with ν^2 for the Uniform distribution on $[-1, 1]$. This suggests an alternative proof that ν^2 =1 using the fact that the densities converge (although formal details are slightly more complicated). Further, $\Gamma(3/p)/\Gamma(1/p) = (1/3)\Gamma(1+3/p)/\Gamma(1+1/p) \to 1/3$ as p→ ∞. I.e., $\sigma^2 \to 1/3$ and eff($\overline{x}$,$\tilde{x}$) $\to 1/3$. Thus, we see, using the entry for σ^2 when p=6 in Table 1 (.3184<1/3), that σ^2, like ν^2, is not monotone in p.

The nonmonotonicity of the two variances is due to the nonmonotonicity of the gamma function. The gamma function has a minimum when its argument is approximately 1.4615. Using the expression for σ^2 and ν^2 we see that ν^2 attains its minimum value when p is approximately 2.167 while σ^2 attains its minimum when p is approximately 9.10. These give "the most favorable" densities in the family for the two estimators. Finally, eff($\overline{x}$,$\tilde{x}$)=$(1/3)\Gamma(1+3/p)/\Gamma^3(1+1/p)$ is minimized as p$\rightarrow\infty$ at 1/3. One way to see this is to use Jensen's inequality on the convex function $h(x) = x^3$. Thus, we see that eff($\overline{x}$,$\tilde{x}$) approaches its minimum value of 1/3 for the Uniform[-1,1] distribution. In fact, it is known that eff($\overline{x}$,$\tilde{x}$) has an absolute minimum value of 1/3 for all symmetric, unimodal densities, and that the Uniform is the only distribution for which this minimum is attained (see, e.g., Lehmann (1983), Chapter 5, or Samuel-Cahn (1994)). Plotted in Figure 2 is eff($\overline{x}$,$\tilde{x}$) as a function of p.

We note that although Var(X), $\sigma^2 \rightarrow \infty$ as p$\rightarrow$ 0, no member of the exponential power family has tails as heavy as those of the Cauchy distribution. To see this note that the Cauchy distribution has density $g(x) = (\pi(1+x^2))^{-1}$, and f(x,p)/g(x) $\rightarrow$ 0 for any fixed p$\in (0,\infty)$. It is well known that $\sigma^2 = \infty$ for the Cauchy distribution. Note, however, that $\nu^2 > \pi^2/4$ for p<.581 in the exponential power model. Thus, $\tilde{x}$ can be more variable for lighter tail distributions. This has been seen earlier when we saw that the finite support Uniform distribution on [-1,1] has a larger value of ν^2 than any member of the exponential power family with p>1. In fact, the general expression for Var($\tilde{x}$) shows that the variability of the median depends only on the value of the density at the true population median, f(θ,p). Thus, for bimodal distributions with low density at the population median (e.g., "bathtub distributions"), $\tilde{x}$ will perform even more poorly.

REFERENCES

Box, G. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*, Addison Wesley, Reading.

Chernoff, H., Gastwirth, J.L., and Johns, M.V. (1967). Asymptotic Distribution of Linear Combinations of Functions of Order Statistics with Applications to Estimation, *Annals of Mathematical Statistics*, 38, 52-72.

D'Agostino, R.B. and Lee, A.F.S. (1977). Robustness of Location Estimators Under Changes of Population Kurtosis, *Journal of the American Statistical Association*, 72, 393-396.

Devore, J.L. (1995). *Probability and Statistics for Engineering and the Sciences (fourth edition)*, Duxbury Press, Belmont.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, Wiley, N.Y.

Hogg, R.V. (1974). Adaptive Robust Procedures: A Partial Review and Some Suggestions for the Future, *Journal of the American Statistical Association*, 69, 909-923.

Lehmann, E.L. (1983). *Theory of Point Estimation*, Wiley, N.Y.

Samuel-Cahn, E. (1994). Combining Unbiased Estimators, *The American Statistician*, 48, 34-36.