Stat 582
Emerson, Win 12
Homework #4 Key
February 6, 2012

Written problems to be handed in Monday, February 13, 2012.

1. For the following estimators, find consistent estimators of the mean based on the sample median, and provide their asymptotic distributions. In each case, find the asymptotic relative efficiency of median based estimator to the sample mean.

Ans: We consider continuous distributions with density f that is nonzero at the median. The approximate distribution for an estimator of population mean μ based on the sample mean $\overline{X}_n$ will be

$$\sqrt{n}(\overline{X} - \mu) \rightarrow_d \mathcal{N}(0, \sigma^2),$$

where σ^2 is the population variance. The approximate distribution for a sample median $X_n^{(m)}$ as an estimator of the population median ϕ will be

$$\sqrt{n}(X_n^{(m)} - \phi) \rightarrow_d \mathcal{N}\left(0, \frac{1}{4f^2(\phi)}\right).$$

Supposing that $g(\phi) = \mu$ with derivative $g'(\phi)$ at ϕ, we then use the delta method to find for $\tilde{\mu}_n = g(X_n^{(m)})$

$$\sqrt{n}(\tilde{\mu}_n - \mu) \rightarrow_d \mathcal{N}\left(0, \frac{[g'(\phi)]^2}{4f^2(\phi)}\right).$$

The asymptotic relative efficiency of the median based estimator to the sample mean is thus

$$e_{\tilde{\mu}_n, \overline{X}_n} = \frac{4\sigma^2 f^2(\phi)}{[g'(\phi)]^2}.$$

a. The normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$.

Ans: The median as a function of μ is $\phi = \mu$, so $g(u) = u$, $\tilde{\mu} = X_n^{(m)}$, and $g'(\phi) = 1$.

The density at the median is $f(\phi) = 1/(\sqrt{2\pi}\sigma)$, so the asymptotic distribution of $\tilde{\mu}$ is

$$\sqrt{n}(\tilde{\mu}_n - \mu) \rightarrow_d \mathcal{N}\left(0, \frac{\pi\sigma^2}{2}\right),$$

and the desired asymptotic relatively efficiency is

$$e_{\tilde{\mu}_n, \overline{X}_n} = \frac{2}{\pi} = 0.637.$$

b. The exponential distribution $f(x) = \lambda e^{-x\lambda} 1_{(0,\infty)}(x)$.

Ans: The mean is $\mu = 1/\lambda$, the variance is $\sigma^2 = 1/\lambda^2 = \mu^2$, and the median as a function of μ is $\phi = \mu \log(2)$, so $g(u) = u/\log(2)$, $\tilde{\mu} = X_n^{(m)}/\log(2)$, and $g'(\phi) = 1/\log(2)$.

The density at the median is $f(\phi) = 1/(2\mu)$, so the asymptotic distribution of $\tilde{\mu}$ is

$$\sqrt{n}(\tilde{\mu}_n - \mu) \rightarrow_d \mathcal{N}\left(0, \frac{\mu^2}{[\log(2)]^2}\right),$$

and the desired asymptotic relatively efficiency is

$$e_{\tilde{\mu}_n,\overline{X}_n} = [\log(2)]^2 = 0.480.$$

c. The Laplace distribution $f(x) = \lambda e^{-|x-\mu|\lambda}/2$.

Ans: The mean is μ, the variance is $\sigma^2 = 2/\lambda^2$, and the median as a function of μ is $\phi = \mu$, so $g(u) = u$, $\tilde{\mu} = X_n^{(m)}$, and $g'(\phi) = 1$.

The density at the median is $f(\phi) = \lambda/2$, so the asymptotic distribution of $\tilde{\mu}$ is

$$\sqrt{n}(\tilde{\mu}_n - \mu) \to_d \mathcal{N}\left(0, \frac{1}{\lambda^2}\right),$$

and the desired asymptotic relatively efficiency is

$$e_{\tilde{\mu}_n,\overline{X}_n} = \frac{1}{3} = 0.333.$$

d. The uniform distribution $\mathcal{U}(0,\theta)$.

Ans: The mean is $\mu = \theta/2$, the variance is $\sigma^2 = \theta^2/12$, and the median as a function of μ is $\phi = \mu$, so $g(u) = u$, $\tilde{\mu} = X_n^{(m)}$, and $g'(\phi) = 1$.

The density at the median is $f(\phi) = 1/theta$, so the asymptotic distribution of $\tilde{\mu}$ is

$$\sqrt{n}(\tilde{\mu}_n - \mu) \to_d \mathcal{N}\left(0, \frac{\theta^2}{4}\right),$$

and the desired asymptotic relatively efficiency is

$$e_{\tilde{\mu}_n,\overline{X}_n} = 2.$$

e. The lognormal distribution X such that $\log X \sim \mathcal{N}(\omega, \tau^2)$. (Note the change in notation on the key to facilitate the same notation in my general problem.)

Ans: Note that in this problem, we have to deal with the bivariate parameter $\theta = (\omega, \tau^2)$. We can work the problem as if τ^2 is known, though there would be more variability of $\tilde{\mu}$ if τ^2 has to be estimated.

The mean is $\mu = \exp(\omega + \tau^2/2)$, the variance is $\sigma^2 = (\exp(\tau^2) - 1)(\exp(2\omega + \tau^2) = (\exp(\tau^2) - 1)\mu^2$, and the median as a function of μ is $\phi = \mu \exp(-\tau^2/2)$. Note that in this problem, we have to deal with the bivariate parameter , so $g(u) = u \exp(\tau^2/2)$, $\tilde{\mu} = X_n^{(m)} \exp(\tau^2/2)$, and $g'(\phi) = \exp(\tau^2/2)$.

The density at the median is $f(\phi) = \exp(\tau^2/2)/(\mu\sqrt{2\pi}\sigma)$, so the asymptotic distribution of $\tilde{\mu}$ is

$$\sqrt{n}(\tilde{\mu}_n - \mu) \to_d \mathcal{N}\left(0, \frac{\mu^2\tau^2\pi}{2\exp(\tau^2)}\right),$$

and the desired asymptotic relatively efficiency is

$$e_{\tilde{\mu}_n,\overline{X}_n} = \frac{2e^{\tau^2}(e^{\tau^2} - 1)}{\pi\tau^2},$$

which is 0.928, 1.36, 2.01, 2.97, 4.43, 6.62, 9.95, and 15.0 as τ^2 ranges from 0.25 to 2 (by 0.25). (Of course, the MLE is related to the sample geometric mean, which could be used more efficiently to estimate the mean.)

2. Find an approximate distribution for the interquartile range of a continuous random variable, where the interquartile range is defined as the difference between the sample 75th and 25th percentiles. Provide explicit expressions for the cases of a normal, exponential, and uniform random variable.

Ans: Let $\theta^{(25)}$ and $\theta^{(75)}$ be the 25th and 75th percentiles of the distribution, respectively, and let $\hat{\theta}_n^{(25)}$ and $\hat{\theta}_n^{(75)}$ be the respective estimates based on the sample quantiles. Then for continuous random variables having density f that is positive at $\theta^{(25)}$ and $\theta^{(75)}$, the "asymptotic covariance" of two sample estimators is found by

$$Cov(\hat{\theta}_n^{(100s)}, \hat{\theta}_n^{(100t)}) = \frac{\min(s,t) - st}{f(\hat{\theta}_n^{(100s)})f(\hat{\theta}_n^{(100t)})},$$

so for the 25th and 75th percentiles, we have joint asymptotic distribution

$$\sqrt{n}\left[\begin{pmatrix}\hat{\theta}_n^{(25)} \\ \\ \hat{\theta}_n^{(75)}\end{pmatrix} - \begin{pmatrix}\theta^{(25)} \\ \\ \theta^{(75)}\end{pmatrix}\right] \rightarrow_d \mathcal{N}_2\left(\begin{pmatrix}0 \\ \\ 0\end{pmatrix}, \begin{pmatrix}\frac{3}{16f^2(\theta^{(25)})} & \frac{1}{16f(\theta^{(25)})f(\theta^{(75)})} \\ \\ \frac{1}{16f(\theta^{(25)})f(\theta^{(75)})} & \frac{3}{16f^2(\theta^{(75)})}\end{pmatrix}\right).$$

Then using the continuous mapping theorem on the linear combination representing the difference ($\vec{a} = (-1 \;\; 1)^T$), we find

$$\sqrt{n}\left[\left(\hat{\theta}_n^{(75)} - \hat{\theta}_n^{(25)}\right) - \left(\theta^{(75)} - \theta^{(25)}\right)\right] \rightarrow_d \mathcal{N}_2\left(0, \frac{3f^2(\theta^{(25)}) - 2f(\theta^{(25)})f(\theta^{(75)}) + 3f^2(\theta^{(75)})}{16f^2(\theta^{(25)})f^2(\theta^{(75)})}\right).$$

3. Consider a Bayesian analysis that assumes $X_1, X_2, \ldots$ are independently distributed according to $X_i|\mu \sim \mathcal{N}(\mu, \sigma^2)$ with σ^2 known and $\mu \sim \mathcal{N}(\zeta, \tau^2)$.

   a. Find an expression for the $Pr(\mu \leq \mu_0 | X_1, \ldots, X_n)$ as would be derived for such an analysis.

   b. Suppose that the data are not truly normally distributed, but do have conditional mean μ and variance σ^2. Show that the expression you derived in part a is a valid approximation for a posterior probability and describe the sense in which it is valid, i.e., how might it differ from the posterior distribution we would have obtained if we knew the true distribution of the data.

4. Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with $X_i \sim \mathcal{N}(\mu, 1)$. For some $0 < \epsilon < 1$

$$T_n = \begin{cases} \overline{X}_n & \text{if } |\overline{X}_n| \geq n^{-1/4} \\ \\ \epsilon \overline{X}_n & \text{if } |\overline{X}_n| < n^{-1/4}. \end{cases}$$

   a. What is the distribution of T_n? Is it unbiased for μ? Is it asymptotically unbiased?

   b. Show that T_n is consistent for μ.

   c. Find the asymptotic distribution of $\sqrt{n}(T_n - \mu)$ as a function of μ.

   d. Comment on the variance of the limiting distribution derived in part a relative to the Cramér-Rao lower bound specified for finite samples for T_n.