Asymptotic Joint Distribution of Sample Mean and a Sample Quantile

Thomas S. Ferguson
UCLA

1. Introduction. The joint asymptotic distribution of the sample mean and the sample median was found by Laplace almost 200 years ago. See Stigler [2] for an interesting historical discussion of this achievement. For a review of other work on this problem, see the Problem Corner of the IMS Bulletin, (1992) Vol. 21, p. 234, and the Problem Corner of Chance magazine, (2000) Vol. 13 No. 3, p. 51. In this talk, I derive the asymptotic joint distribution of the sample mean and an arbitrary quantile. It is hoped that the proof may be new and of interest.

2. The Main Theorem. Let $X_1, \ldots, X_n$ be i.i.d. with distribution function $F(x)$, density $f(x)$, mean μ and finite variance σ^2. Let $0 < p < 1$ and let x_p denote the pth quantile of F, so that $F(x_p) = p$. Assume that the density $f(x)$ is continuous and positive at x_p. Let $\overline{X}_n = (1/n)\sum_1^n X_i$ be the sample mean, and let $Y_n = X_{(n:\lceil np \rceil)}$ denote the sample pth quantile. As is well-known,

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2) \tag{1}$$

and

$$\sqrt{n}(Y_n - x_p) \xrightarrow{\mathcal{L}} \mathcal{N}(0, p(1-p)/f(x_p)^2). \tag{2}$$

In this note, we find the joint asymptotic distribution of $\overline{X}_n$ and Y_n. As is expected, this asymptotic distribution is bivariate normal, so the main interest is in the asymptotic covariance of $\overline{X}_n$ and Y_n. This asymptotic covariance is easy to describe in terms of the minimum pth deviation, a measure of spread for use in estimating the pth quantile. Let

$$L_p(x-a) = \begin{cases} p(x-a) & \text{if } x > a \\ (1-p)(a-x) & \text{if } x \le a. \end{cases} \tag{3}$$

The pth deviation about the point a, $\mathrm{E}(L_p(X-a))$, is minimized by the choice $a = x_p$. Therefore, we define the minimum pth deviation to be

$$\tau(p) = \mathrm{E}L_p(X - x_p). \tag{4}$$

When $p = 1/2$, the corresponding quantile is the median which we denote here by ν. In this case, the minimum pth deviation is half the mean deviation about the median, $\tau(.5) = (1/2)\mathrm{E}(|X-\nu|)$.

Theorem. *Under the assumptions of the first paragraph, we have*

$$\sqrt{n}\left(\begin{pmatrix} \overline{X}_n \\ Y_n \end{pmatrix} - \begin{pmatrix} \mu \\ x_p \end{pmatrix}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \tau(p)/f(x_p) \\ \tau(p)/f(x_p) & p(1-p)/f(x_p)^2 \end{pmatrix}\right). \tag{5}$$

3. Mean and Median. In the application of the main theorem to the asymptotic distribution of the mean and median, we find

$$\sqrt{n}(\begin{pmatrix} \overline{X}_n \\ Y_n \end{pmatrix} - \begin{pmatrix} \mu \\ \nu \end{pmatrix}) \xrightarrow{\mathcal{L}} \mathcal{N}(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \mathrm{E}|X-\nu|/(2f(\nu)) \\ \mathrm{E}|X-\nu|/(2f(\nu)) & 1/(2f(\nu))^2 \end{pmatrix}). \tag{6}$$

What linear combination of sample mean and median has smallest mean squared error? Let μ and ν denote the mean and median of the distribution, and let $d = \nu - \mu$. We consider unbiased estimates of the form $\hat{\nu} = \alpha \overline{X}_n + (1-\alpha) m_n + d$. Let σ_x^2, σ_{xy} and σ_y^2 denote the asymptotic variances and covariance of $\overline{X}_n$ and Y_n. The asymptotic distribution of $\sqrt{n}(\hat{\nu} - \nu)$ is $\mathcal{N}(0, v^2)$, where

$$v^2 = \alpha^2 \sigma_x^2 + 2\sigma_{xy} + (1-\alpha)^2 \sigma_y^2.$$

The choice of α that minimizes this quantity is

$$\alpha = \frac{\sigma_y^2 - \sigma_{xy}}{\sigma_x^2 - 2\sigma_{xy} + \sigma_y^2} = \frac{\frac{1}{4} - f(\nu)\tau}{f(\nu)^2 \sigma^2 - 2f(\nu)\tau + \frac{1}{4}}.$$

When $4f(\nu)\tau = 1$, then $\alpha = 0$ and we use the median only to estimate ν. When $\sigma^2 f(\nu) = \tau$, then $\alpha = 1$ and we use the mean only.

4. Examples. (1) The standard normal: $\begin{pmatrix} 1 & 1 \\ 1 & \pi/2 \end{pmatrix}$. The sample mean has smaller variance. In fact, since the sample mean is a sufficient statistic for the mean of the distribution, no further reduction of the variance can be obtained by considering also the sample median.

(2) The logistic: $\begin{pmatrix} \pi^2/3 & 4\log 2 \\ 4\log 2 & 4 \end{pmatrix}$. Again the mean has smaller asymptotic variance. But here some asymptotic improvement can be obtained by considering also the sample median. The linear combination of the form $\alpha \overline{X}_n + (1-\alpha) Y_n$ with the smallest asymptotic variance occurs at $\alpha = (1 - \log 2)/(1 - 2\log 2 + \pi^2/12) = .7035$.

(3) The uniform on $(0,1)$: $\begin{pmatrix} 1/12 & 1/8 \\ 1/8 & 1/4 \end{pmatrix}$. Again the sample mean has smaller variance. The optimal linear combination is $(3/2)\overline{X}_n - (1/2)Y_n$, with minimum variance 1/16. Notice an interesting feature. If we first learn the mean, then after learning the median, the estimate does not move toward the median, but rather in the opposite direction. This example was also worked out by Samuel-Cahn [1].

(4) The double exponential: $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. Here the sample median has smaller asymptotic variance. The maximum likelihood estimate is the sample median, and it is efficient, so that no linear combination involving the sample mean can improve on it.

(5) As an example of a distribution with differing mean and median, consider the extreme value distribution, with density $f(x|\theta) = \exp\{-e^{-(x-\theta)} - (x-\theta)\}$. The mean of the distribution is $\mu = \theta + \gamma$, where $\gamma = 0.5772\ldots$ is Euler's constant. The median of the distribution is $\nu = \theta - \log\log 2 = \theta + 0.3665\ldots$. We have $\sigma_x^2 = \pi^2/6 = 1.6449\ldots$, $\mathrm{E}|X-\nu| = 0.5466\ldots$, $f(\nu|\theta) = 0.3466$, from which we have the asymptotic covariance matrix,

$$\begin{pmatrix} 1.645 & 0.789 \\ 0.789 & 2.081 \end{pmatrix}$$

The asymptotically best linear combination of the mean and median to estimate θ is then $\hat\theta = \alpha(\overline{X}_n - \gamma) + (1-\alpha)(Y_n + \log\log 2)$, where $\alpha = .601\ldots$.

5. Proof. Let Z_i be i.i.d. exponential random variables with mean 1, and let $S_n = Z_1 + \cdots + Z_n$ for all n. For fixed n let

$$U_j = S_j/S_{n+1}. \tag{7}$$

Then $(U_1, \ldots, U_n)$ has the same distribution as the order statistics of a sample of size n from the uniform distribution on $[0,1]$. Let $F(x)$ be the distribution function of interest, and let $g(x) = F^{-1}(x)$ be the inverse function, so that $(g(U_1), \ldots, g(U_n))$ are distributed as the order statistics of a sample of size n from F. Then the sample pth quantile and the sample mean are

$$Y_n = g(U_{\lceil np \rceil}) \quad \text{and} \quad \overline{X}_n = \frac{1}{n}\sum_1^n g(U_j) \tag{8}$$

We are interested in the joint asymptotic distribution of these two quantities.

Expand the first about p:

$$Y_n = g(p) + g'(p)(U_{\lceil np \rceil} - p) + \text{ higher order terms} \tag{9}$$

and expand the jth term of the second about $j/(n+1)$:

$$\overline{X}_n = \frac{1}{n}\sum_1^n \left(g(\frac{j}{n+1}) + g'(\frac{j}{n+1})(U_j - \frac{j}{n+1}) + \text{ higher order terms} \right) \tag{10}$$

The higher order terms become negligible even when multiplied by $\sqrt{n}$. Let $\mu_n = (1/n)\sum_1^n g(j/(n+1))$. We are to find the joint asymptotic distribution of

$$\begin{aligned} \sqrt{n}(Y_n - g(p)) &\sim \sqrt{n}g'(p)(\frac{S_{\lceil np \rceil}}{S_{n+1}} - p) \\ &= \sqrt{n}g'(p)\frac{(1-p)(S_{\lceil np \rceil} - np) - p(S_{n+1} - S_{\lceil np \rceil} - n(1-p))}{S_{n+1}} \end{aligned} \tag{11}$$

and

$$\sqrt{n}(\overline{X}_n - \mu_n) \sim \frac{1}{\sqrt{n}} \sum_{j=1}^{n} g'(\frac{j}{n+1})(\frac{S_j}{S_{n+1}} - \frac{j}{n+1})$$
$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} g'(\frac{j}{n+1}) \frac{(n+1-j)(S_j - j) - j(S_{n+1} - S_j - (n+1-j))}{(n+1)S_{n+1}} \tag{12}$$

Since $S_{n+1}/n \xrightarrow{a.s.} 1$ as $n \to \infty$, we may replace the S_{n+1} in the denominators of (11) and (12) by n without changing the limiting distributions.

Now define $W_n(t)$ to be $W_n(i/(n+1)) = (S_i - i)/\sqrt{n+1}$ for $i = 0, 1, \ldots, n+1$ with linear interpolation between the points. Then by Donsker's Theorem, $W_n \xrightarrow{\mathcal{L}} W$ in $C([0,1])$ (the space of continuous functions on $[0,1]$ with sup-norm topology), where $W(t)$ is Brownian motion. The variables of interest now become

$$\begin{aligned}\sqrt{n}(Y_n - g(p)) &\sim g'(p)\left[(1-p)W_n(\frac{\lceil np \rceil}{n+1}) - p(W_n(1) - W_n(\frac{\lceil np \rceil}{n+1}))\right] \\ &\sim g'(p)\left[(1-p)W_n(p) - p(W_n(1) - W_n(p))\right]\end{aligned} \tag{13}$$

and

$$\begin{aligned}\sqrt{n}(\overline{X}_n - \mu_n) &\sim \sum_{j=1}^{n} g'(\frac{j}{n+1}) \frac{(n+1-j)W_n(\frac{j}{n+1}) - j(W_n(1) - W_n(\frac{j}{n+1}))}{n(n+1)} \\ &\sim \int_0^1 g'(t)[(1-t)W_n(t) - t(W_n(1) - W_n(t))]\, dt\end{aligned} \tag{14}$$

Now since the maps are continuous in the uniform topology, we conclude

$$\sqrt{n}(Y_n - g(p)) \xrightarrow{\mathcal{L}} g'(p)(W(p) - pW(1)) \tag{15}$$

and

$$\sqrt{n}(\overline{X}_n - \mu_n) \xrightarrow{\mathcal{L}} \int_0^1 g'(t)(W(t) - tW(1))\, dt. \tag{16}$$

as $n \to \infty$. Then the Cramér-Wold device shows that the limiting distribution is jointly normal. There remains for us to calculate the covariance since we already know what the variances must be. Since $\mathrm{E}W(s)W(t) = s \wedge t$,

$$\begin{aligned}\mathrm{Cov}(g'(p)(W(p) - pW(1)), &\int_0^1 g'(t)(W(t) - tW(1))\, dt \\ &= g'(p) \int_0^1 g'(t)\mathrm{E}(W(p) - pW(1))(W(t) - tW(1))\, dt \\ &= g'(p) \int_0^1 g'(t)(p \wedge t - pt)\, dt\end{aligned} \tag{17}$$

To simplfy this, note first that $g(p) = F^{-1}(p)$ implies that $g'(p) = 1/f(x_p)$. Also,

$$\begin{aligned}
\int_0^1 g'(t)(p \wedge t - pt)\,dt &= \int_0^1 (p \wedge t - pt)\,dg(t) \\
&= \int_0^p (1-p)t\,dg(t) + \int_p^1 p(1-t)\,dg(t) \\
&= -(1-p)\int_0^p (g(t)-g(p))\,dt + p\int_p^1 (g(t)-g(p))\,dt \\
&= p\int_{x_p}^{\infty} (x-x_p)\,dF(x) - (1-p)\int_{-\infty}^{x_p} (x-x_p)\,dF(x) \\
&= \tau(p)
\end{aligned} \tag{18}$$

so (17) reduces to $\tau(p)/f(x_p)$ as claimed.

One other thing may be checked. Namely that $\mu_n = (1/n)\sum_1^n g(j/(n+1)) \to \int_0^1 g(t)\,dt = \int_{-\infty}^{\infty} x\,dF(x) = \mu$. Furthermore, the convergence is rapid enough ($\sqrt{n}(\mu_n - \mu) \to 0$) so that we may replace μ_n by μ in (16). ■

Thanks are due to Rao Chaganti for pointing out the references.

References.

[1] Ester Samuel-Cahn (1994) Combining unbiased estimators, *The American Statistician*, **48**, 34-36.

[2] Stephen M. Stigler (1973) Laplace, Fisher, and the discovery of the concept of sufficiency, *Biometrika* **60**, 439-445.