NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

ESTIMATING THE PARAMETERS OF A RECTANGULAR DISTRIBUTION

By A. George Carlton

Columbia University

1. Introduction. In this note, the range and midrange of the sample are shown to be a pair of sufficient statistics, and maximum likelihood estimates, for the true range and true mean of a rectangular distribution; exact and limiting distribution of midrange, range, and their ratio are derived; the "efficiencies" of the sample mean and median as estimates of the true mean are calculated; and the limiting distribution of the difference between two sample midranges is derived. All the limiting distributions are non-normal, and the error of estimate is of order n^{-1} rather than the customary order $n^{-\frac{1}{2}}$. The limiting distribution of midrange, and the limiting ratio of variances of the midrange and sample mean were given by Fisher [1].

$f(x)$ and $F(x)$ are used throughout to designate the probability density function of x and the distribution function (cumulative probability function) of x; the argument will also indicate the random variable being considered.

2. Exact distribution of midrange, range, and their ratio. Let $x_1, \cdots, x_n$ be a set of n independent observations on a random variable having the rectangular distribution $f(x) = 1/L$, $(\theta - L/2 \leq x \leq \theta + L/2)$, where θ is the true mean, and L the true range. The minimum observation u and the maximum observation v are a pair of sufficient statistics for θ and L, as the conditional distribution of the remaining observations for given u and v is independent of θ and L:

$$f(x_1, \cdots, x_n \mid u, v) = (v - u)^{-(n-2)}$$

The midrange $\bar{\theta} = \frac{1}{2}(u + v)$ and the range $\bar{L} = v - u$ are maximum likelihood estimates of θ and L, respectively, as they are the parameter values which uniquely maximize $f(x_1, \cdots, x_n)$ for the given set of observations. We shall assume that the random variable is normalized by change of origin and change of scale so that $\theta = 0$ and $L = 1$. The joint probability density function of u and v is

$$f(u, v) = \frac{d^2 F(u, v)}{dv\, d(-u)} = \frac{d^2 (v - u)^n}{dv\, d(-u)} = n(n - 1)(v - u)^{n-2}, \qquad (-\tfrac{1}{2} \leq u \leq v \leq \tfrac{1}{2}). \tag{1}$$

Making the transformation $\bar{\theta} = \frac{1}{2}(u + v)$, $\bar{L} = v - u$ in (1),

$$(2) \qquad f(\bar{\theta}, \bar{L}) = n(n-1)\bar{L}^{n-2} \qquad (0 \leq 2\,|\,\bar{\theta}\,| \leq 1 - \bar{L} \leq 1).$$

Integrating out $\bar{L}$ from 0 to $(1 - 2\,|\bar{\theta}\,|)$,

$$(3) \qquad \begin{aligned} f(\bar{\theta}) &= n(1 - 2\,|\,\bar{\theta}\,|)^{n-1}, & (|\,\bar{\theta}\,| \leq \tfrac{1}{2}). \\ |\,F(\bar{\theta}) - F(0)\,| &= \tfrac{1}{2} - \tfrac{1}{2}(1 - 2\,|\,\bar{\theta}\,|)^n, & (|\,\bar{\theta}\,| \leq \tfrac{1}{2}). \end{aligned}$$

Odd moments vanish by symmetry; even order moments are

$$(4) \qquad \mu_{2k}(\bar{\theta}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} n\bar{\theta}^{2k}(1 - 2\,|\,\bar{\theta}\,|)^{n-1}\,d\bar{\theta} = 2^{-2k} \Big/ \binom{2k+n}{2k}.$$

In (2), integrating out $\bar{\theta}$ from $\frac{1}{2}(\bar{L} - 1)$ to $\frac{1}{2}(1 - \bar{L})$,

$$(5) \qquad \begin{aligned} f(\bar{L}) &= n(n-1)\bar{L}^{n-2}(1 - \bar{L}), \qquad (0 \leq \bar{L} \leq 1). \\ F(\bar{L}) &= n(n-1) \int_0^{\bar{L}} \bar{L}^{n-2}(1 - \bar{L})\,d\bar{L} = n(n-1)B_{\bar{L}}(n-1, 2), \qquad (0 \leq \bar{L} \leq 1). \\ \mu_k(\bar{L}) &= n(n-1) \int_0^1 \bar{L}^{n-2+k}(1 - \bar{L})\,d\bar{L} = \frac{n(n-1)}{(n+k)(n+k-1)}. \end{aligned}$$

Thus $\mu_1(\bar{L}) = (n-1)/(n+1)$; hence the bias of $\bar{L}$ can be removed by multiplying $\bar{L}$ by $(n+1)/(n-1)$.

The statistic $t = \bar{\theta}/\bar{L}$ can be used to test the hypothesis that the mean of a rectangular distribution of unknown range is 0. To obtain the distribution of t when the hypothesis is true, set $t = \bar{\theta}/\bar{L}$ and $\bar{L} = \bar{L}$ in (2):

$$(6) \qquad \begin{aligned} f(t, \bar{L}) &= n(n-1)\bar{L}^{n-1}, \qquad (\bar{L} \leq (1 + 2\,|\,t\,|)^{-1}). \\ f(t) &= (n-1)(1 + 2\,|\,t\,|)^{-n}. \\ |\,F(t) - F(0)\,| &= \tfrac{1}{2} - \tfrac{1}{2}(1 + 2\,|\,t\,|)^{1-n}. \end{aligned}$$

Moments of t do not exist for order greater than $(n - 2)$; for $k \leq n - 2$, odd moments vanish by symmetry and

$$\mu_{2k}(t) = 2(n-1) \int_0^{\infty} t^{2k}(1 + 2t)^{-n}\,dt = 2^{2k} \Big/ \binom{n-2}{2k}.$$

3. Limiting distributions. $\bar{\theta}$, $\bar{L}$, and t have non-normal limiting distributions, although $\bar{\theta}$ and $\bar{L}$ are maximum likelihood estimates; this is explained by the discontinuity of $f(x, \bar{\theta})$ at $x = \bar{\theta} \pm \frac{1}{2}$. We obtain the limiting distributions of $q = n\bar{\theta}$ and $r = n(1 - \bar{L})$. Substituting q and r in (2), and proceeding to the limit for increasing n,

$$\lim f(q, r) = \lim \frac{n-1}{n}\left(1 - \frac{r}{n}\right)^{n-2} = e^{-r}, \qquad (0 \leq 2\,|\,q\,| \leq r < \infty).$$

The necessary simple integrations yield the following limiting distributions:

$$
(7)\qquad
\begin{aligned}
f(q) &= e^{-2|q|}. \\
|F(q) - F(0)| &= \tfrac{1}{2} - \tfrac{1}{2}e^{-2|q|}. \\
\mu_{2k}(q) &= (2k)!/2^{2k};\ \mu_{2k+1} = 0. \\
f(r) &= re^{-r}, && (r \geq 0) \\
F(r) &= 1 - (1 + r)e^{-r}, && (r \geq 0) \\
\mu_k(r) &= (k + 1)!
\end{aligned}
$$

The limiting distribution of $s = nt$ is the same as that of $n\bar{\theta}$, as is seen by comparing (3) and (6).

4. Comparison of $\bar{\theta}$ with $\bar{x}$ and $\tilde{x}$ as estimates of θ. The sample mean $\bar{x}$ and median $\tilde{x}$ are unbiased estimates of θ.

$$
(8)\qquad
\begin{aligned}
\mu_2(\bar{x}) &= \frac{1}{n}\int_{-\frac{1}{2}}^{\frac{1}{2}} x^2\,dx = 1/(12n). \\
\mu_2(\tilde{x}) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \tilde{x}^2 f(\tilde{x})\,d\tilde{x} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \tilde{x}^2 \frac{(2m+1)!}{m!\,m!}\left(\tfrac{1}{2} - \tilde{x}\right)^m\left(\tilde{x} + \tfrac{1}{2}\right)^m d\tilde{x},
\end{aligned}
$$

for $n = 2m + 1$, m an integer. Substituting $z = 1 - 4\tilde{x}^2$, then simplifying the Beta function obtained on integration,

$$
(9)\qquad \mu_2(\tilde{x}) = \frac{(2m+1)!}{m!\,m!\,2^{2m+3}}\int_0^1 z^m(1 - z)^{\frac{1}{2}}\,dz = \frac{1}{4(2m+3)} = \frac{1}{4(n+2)}.
$$

(4), with $k = 1$, gives $\mu_2(\bar{\theta}) = \dfrac{1}{2(n+1)(n+2)}$. Comparison of this with (8) and (9) shows that $\mu_2(\bar{\theta})/\mu_2(\bar{x}) = \dfrac{6n}{(n+1)(n+2)}$ and $\mu_2(\tilde{x})/\mu_2(\bar{x}) = 3n/(n+2)$. As n increases, $\mu_2(\bar{\theta})/\mu_2(\bar{x}) \to 6/n \to 0$; and $\mu_2(\tilde{x})/\mu_2(\bar{x}) \to 3$. Thus the "efficiency" of the mean is zero, and the median is only one-third as "efficient" as the mean. (The concept of efficiency is not strictly applicable as $\bar{\theta}$ does not have a normal limiting distribution.)

5. Limiting distribution of difference between two midranges. Let $\bar{\theta}_1$ and $\bar{\theta}_2$ be the midranges of samples of n_1 and n_2 observations, respectively, from two normalized rectangular populations, and let $z = q_1 - q_2 = n_1\bar{\theta}_1 - n_2\bar{\theta}_2$. Applying the formula for composition of random variables, one obtains from (7),

$$
(10)\qquad
\begin{aligned}
f(z) &= \int_{-\infty}^{\infty} f(z - q)f(q)\,dq = \int_{-\infty}^{\infty} e^{-2|z-q|}e^{-2|q|}\,dq \\
&= \int_{-\infty}^{0} e^{-2|z|}e^{-4q}\,dq + \int_0^{|z|} e^{-2|z|}\,dq + \int_{|z|}^{\infty} e^{2|z|}e^{-4q}\,dq \\
&= \tfrac{1}{4}e^{-2|z|} + |z|e^{-2|z|} + \tfrac{1}{4}e^{-2|z|} = \left(|z| + \tfrac{1}{2}\right)e^{-2|z|}
\end{aligned}
$$

$$
|F(z) - F(0)| = \tfrac{1}{2} - \frac{|z| + 1}{2}e^{-2|z|}.
$$

$$
\mu_{2k}(z) = (k + 1)(2k)!/2^{2k}.
$$

$z = \frac{n_1}{2} \frac{v_1 + u_1}{v_1 - u_1} - \frac{n_2(v_2 + u_2)}{2(v_2 - u_2)}$ can be used to test the hypothesis of equality of means of any two rectangular populations, and has in the limit the distribution (10), if the means of the populations are equal.

6. The one-parameter rectangular distribution. If $f(x) = 1/\lambda$, $(0 \leq x \leq \lambda)$, then $f(x_1, \cdots, x_n \mid v) = v^{1-n}$. Thus v is a sufficient statistic and is evidently the maximum likelihood estimate of λ. Here $F(v) = (v/\lambda)^n$; $f(v) = nv^{n-1}\lambda^{-n}$; and $\mu_k(v) = \lambda^k n/(n + k)$. The normalized error $y = n(\lambda - v)/\lambda$ has the probability density function $f(y) = (1 - y/n)^{n-1}$, which tends to e^{-y} as n increases.

REFERENCE

[1] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Phil. Trans. Roy. Soc. London*, Series A, Vol. 222 (1921), pp. 309–368.

ON THE POWER FUNCTION OF THE SIGN TEST FOR SLIPPAGE OF MEANS

By John E. Walsh

Princeton University

1. Summary. This note compares the power functions of the sign test for slippage with the power functions of the most powerful test for the case of normal populations. The sign test is found to be approximately 95% efficient for small samples.

2. Introduction. Let us consider a univariate population whose mean equals its median and whose cumulative distribution function is continuous at the mean. A sampling method of testing the supposition that the mean of this population exceeds a given constant value μ_0 (slippage to the right) is furnished by considering how many values of the sample are less than μ_0. An analogous method applies for testing whether the mean is less than μ_0 (slippage to the left). A particular class of populations for which the sign test is valid are the normal populations. This note compares the power functions of the sign test with the power functions of the most powerful test for slippage for the case in which the population is normal (Table I). It is shown that the sign test is approximately 95% as efficient as the most powerful test (the Student t-test) for samples of size 4, 5 and 6, and that although the relative efficiency of the sign test decreases as the sample size increases, its efficiency is approximately 75% for samples of size 13. This supports the idea that for normal populations little efficiency is lost by using attributes instead of continuous variables if the sample size is small.

In choosing between the sign and Student t-tests for slippage the following considerations may be of interest: