NOTES

This section is devoted to brief research and expository articles and other short items.

THE SAMPLING DISTRIBUTION OF THE RATIO OF TWO RANGES FROM INDEPENDENT SAMPLES[1]

RICHARD F. LINK

University of Oregon

Let us consider a sample of n ordered observations $(x_1 < x_2 < \cdots < x_n)$ drawn from a population with variance σ^2. Let $w = (x_n - x_1)/\sigma$. Let us consider the joint sampling distribution of w_1 and w_2 for two samples, not necessarily the same size, drawn from populations with the same variance. If the two samples were drawn independently, then the joint sampling distributions of w_1 and w_2 may be written as the product of the sampling distributions of w_1 and w_2.

If we make the change of variable $r = w_1/w_2$, $w_2 = w$, and if w is integrated over its range of definition, the cumulative distribution of the ratio of two ranges remains. This may be written as

$$F(R) = \int_0^R dr \int_0^\infty dw \cdot w \cdot h_2(w) \cdot h_1(wr), \tag{1}$$

where h_1 is the pdf for w_1 and h_2 is the pdf for w_2.

To obtain more explicit results, specific distribution functions may be considered. The following table gives the sampling distribution of the ratio of two ranges from independent samples for the indicated density functions $f(x)$. Notice that for the normal distribution it was possible to obtain results only for some special cases.

In Table 1 for $F(R)$, w_1 and w_2 represent ranges computed from samples of size n_1 and n_2 respectively.

Notice that formula (1) for $F(R)$ is equivalent to the following expressions

$$Pr(w_1/w_2 < R) = F(R) = \int_0^\infty dw_2 \int_0^{Rw_2} dw_1\, h(w_1) \cdot h(w_2).$$

The region of integration for the last expression is simply the region in the w_2, w_1 plane to the right of the line $w_1 = Rw_2$.

This integration was done numerically. Table 2 gives values of R for all combinations of n_1 and $n_2 \leq 10$ and for $\alpha = .005, .01, .025, .05, .10$ such that

$$Pr(w_1/w_2 < R) = \alpha$$

where w_1 and w_2 are ranges computed from samples of size n_1 and n_2 drawn from

[1] This work was done under contract N6onr-218/IV with the Office of Naval Research.

normal populations with the same variance. It is believed that these values are correct to within one place in the last reported figure.

These tabled values may be used as critical values for testing the hypothesis that two independent samples were drawn from normal populations with the same variance. This test is therefore comparable to the F test. Some sort of

TABLE 1

$f(x)$	$F(R) = Pr\,(w_1/w_2 < R)$
$1 \quad 0 \le x \le 1$ $0 \quad$ all other x	$n_2(n_2-1)\,R^{n_1-1}\left[\frac{n_1}{n_1+n_2-2} - \frac{(n_1+R[n_1-1])}{n_1+n_2-1} + \frac{R(n_1-1)}{n_1+n_2}\right]$
$e^{-x} \quad 0 \le x < \infty$ $0 \quad x < 0$	$1-(n_1-1)(n_2-1)\sum_{i=0}^{n_1-2}\sum_{j=0}^{n_2-2}\binom{n_1-2}{i}\binom{n_2-2}{j}\cdot\frac{(-1)^{i+j}}{[1+j+(1+i)R]\,(1+i)}$
$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \quad -\infty < x < \infty$	$n_1 = 2, \quad n_2 = 2 \quad \frac{2}{\pi}\tan^{-1} R$ $n_1 = 2, \quad n_2 = 3 \quad \frac{6}{\pi}\tan^{-1}\frac{R}{\sqrt{4+3R^2}}$ $n_1 = 3, \quad n_2 = 2 \quad \frac{6}{\pi}\left(\tan^{-1}\sqrt{3+4R^2}-\frac{\pi}{3}\right)$ $n_1 = 3, \quad n_2 = 3$ $\int_0^R dr\left[\frac{27r}{2\pi^2}\left\{\frac{2}{r^2}(u\tan^{-1}u - v\tan^{-1}v) + \frac{1}{6r^2(1+r^2)}(w\tan^{-1}w - u\tan^{-1}u) + \frac{u^2y}{r}\tan^{-1}2ry\right\}\right]$ where $u = [3(r^2+1)]^{-\frac{1}{2}} \qquad w = (7r^2+3)^{-\frac{1}{2}}$ $v = (4r^2+3)^{-\frac{1}{2}} \qquad y = (3r^2+4)^{-\frac{1}{2}}$

measure of the relative performance of these two tests seems desirable. An attempt to measure the performance of this test relative to the F test was made by comparing the tolerance intervals of the distribution of this ratio with those of the F test.

The length of the interval containing the central $1 - 2\alpha$ proportion of the distribution of F was compared with a similar length for the distribution of w_1/w_2 for $n_1 = n_2 = n$. The square of the ratio of these lengths will be called δ_α^2.

TABLE 2

$$Pr\left(\frac{w_1}{w_2} < R\right) = .005$$

n_1 \ $n_2 =$	2	3	4	5	6	7	8	9	10
2	.0078	.0052	.0043	.0039	.0038	.0037	.0036	.0035	.0034
3	.096	.071	.059	.054	.051	.048	.045	.042	.041
4	.21	.16	.14	.13	.12	.12	.11	.11	.10
5	.30	.24	.22	.20	.19	.18	.18	.17	.16
6	.38	.32	.28	.26	.25	.24	.23	.22	.22
7	.44	.38	.34	.32	.30	.29	.28	.27	.26
8	.49	.43	.39	.36	.35	.33	.32	.31	.30
9	.54	.47	.43	.40	.38	.37	.36	.35	.34
10	.57	.50	.46	.44	.42	.40	.39	.38	.37

$$Pr\left(\frac{w_1}{w_2} < R\right) = .01$$

n_1 \ $n_2 =$	2	3	4	5	6	7	8	9	10
2	.0157	.0105	.0080	.0070	.0068	.0066	.0063	.0062	.0061
3	.136	.100	.084	.079	.073	.069	.065	.062	.060
4	.26	.20	.18	.17	.16	.15	.14	.14	.13
5	.38	.30	.26	.24	.23	.22	.21	.21	.20
6	.46	.37	.33	.31	.29	.28	.27	.26	.26
7	.53	.43	.39	.36	.34	.33	.32	.31	.30
8	.59	.49	.44	.41	.39	.37	.36	.35	.34
9	.64	.53	.48	.45	.43	.41	.40	.39	.38
10	.68	.57	.52	.49	.46	.45	.43	.42	.41

$$Pr\left(\frac{w_1}{w_2} < R\right) = .025$$

n_1 \ $n_2 =$	2	3	4	5	6	7	8	9	10
2	.039	.026	.019	.018	.017	.016	.016	.015	.015
3	.217	.160	.137	.124	.115	.107	.102	.098	.095
4	.37	.28	.25	.23	.21	.20	.19	.18	.18
5	.50	.39	.34	.32	.30	.28	.27	.26	.25
6	.60	.47	.42	.38	.36	.34	.33	.32	.31
7	.68	.54	.48	.44	.42	.40	.38	.37	.36
8	.74	.59	.53	.49	.46	.44	.43	.42	.41
9	.79	.64	.57	.53	.50	.48	.47	.46	.44
10	.83	.68	.61	.57	.54	.52	.50	.49	.48

TABLE 2—*Continued*

$$Pr\left(\frac{w_1}{w_2} < R\right) = .05$$

n_1 \ n_2 =	2	3	4	5	6	7	8	9	10
2	.079	.052	.039	.036	.034	.032	.031	.030	.028
3	.31	.23	.20	.18	.16	.15	.14	.14	.13
4	.50	.37	.32	.29	.27	.26	.25	.24	.23
5	.62	.49	.42	.40	.36	.35	.33	.32	.31
6	.74	.57	.50	.46	.43	.41	.40	.38	.37
7	.80	.64	.57	.52	.49	.47	.45	.44	.43
8	.86	.70	.62	.57	.54	.51	.50	.48	.47
9	.91	.75	.67	.61	.58	.55	.53	.52	.51
10	.95	.80	.70	.65	.61	.59	.57	.55	.54

$$Pr\left(\frac{w_1}{w_2} < R\right) = .10$$

n_1 \ n_2 =	2	3	4	5	6	7	8	9	10
2	.158	.105	.077	.074	.069	.066	.062	.059	.056
3	.46	.33	.28	.25	.23	.22	.21	.20	.19
4	.67	.49	.42	.38	.36	.34	.32	.31	.30
5	.84	.62	.53	.48	.45	.43	.41	.39	.38
6	.97	.72	.62	.56	.52	.50	.48	.46	.45
7	1.07	.80	.69	.63	.59	.56	.54	.52	.50
8	1.15	.87	.75	.68	.64	.61	.58	.56	.54
9	1.21	.92	.80	.73	.68	.65	.62	.60	.58
10	1.26	.98	.85	.77	.72	.68	.66	.64	.62

TABLE 3

n	*Relative precision of the range as an estimate of* σ	δ^2
2	1.00	1.00
3	.99	.99
4	.98	.97
5	.96	.95
6	.93	.92
7	.91	.90
8	.89	.89
9	.87	.88
10	.85	.86

For statistics having normal sampling distributions such a ratio would be independent of α and would be equivalent to the ratio of the variances of these sampling distributions. It was found that δ_α^2 is independent of α except for a maximum change of 1 in the second decimal for the values of $\alpha = .005, .01, .025, .05, .10$. These values of δ^2 are presented in Table 3 along with the relative precision of the range as an estimate of σ as given by Mosteller [1].

It is interesting to note that δ^2 corresponds very closely to the relative precision of the range as an estimate of σ.

REFERENCE

[1] F. Mosteller, "On some useful 'inefficient' statistics," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 377–408.

A NOTE ON THE ESTIMATION OF A DISTRIBUTION FUNCTION BY CONFIDENCE LIMITS

By Frank J. Massey, Jr.

University of Oregon

Let $F(x)$ be the continuous, cumulative distribution function of a random variable X, and let $x_1 < x_2 < x_3 < \cdots < x_n$ be the results of n independent observations on X arranged in order of size. We wish to estimate $F(x)$ by means of the band $S_n(x) \pm \lambda/\sqrt{n}$ where $S_n(x)$ is defined by

$$S_n(x) = \begin{cases} 0 & \text{if } x < x_1, \\ k/n & \text{if } x_k \leq x < x_{k+1}, \\ 1 & \text{if } x \geq x_n. \end{cases}$$

Thus we wish to know the probability, say $P_n(\lambda)$, that the band is such that $S_n(x) - \dfrac{\lambda}{\sqrt{n}} < F(x) < S_n(x) + \dfrac{\lambda}{\sqrt{n}}$ for all x. This problem has been previously studied [1] [2] [3] [4] [5] and a limiting distribution has been obtained [1] [4] [5] and tabled [3] [4]. However apparently no error terms for the limiting distribution, or practical methods of obtaining $P_n(\lambda)$ have been given. Such a method is given here.

It has been shown [2] that $P_n(\lambda)$ is independent of $F(x)$ provided only that $F(x)$ is continuous, and thus it is sufficient to consider only the case

$$F(x) = \begin{cases} 0 \text{ if } x < 0, \\ x \text{ if } 0 \leq x \leq 1, \\ 1 \text{ if } x \geq 1. \end{cases}$$

We will find the probability that $S_n(x)$ falls wholly in the band $F(x) \pm k/n$ (here $\lambda = k/\sqrt{n}$) where k is an integer or a rational number, and intermediate values may be obtained by interpolation. To illustrate the method we shall assume that k is an integer.