A NOTE ON THE EXACT DISTRIBUTION OF λ_n.

BY K. RAGHAVAN NAIR

STATISTICAL LABORATORY, CALCUTTA.

1. In *Biometrika*, Vol. XXV. pp. 379—410, Karl Pearson had developed what he called the $P(\lambda_n)$ test, for "determining whether a sample of size n, supposed to have been drawn from a parent population having a known probability integral, has probably been drawn at random."

If $x_1, x_2, \ldots\ldots x_n$ are the variates of the sample and $p_1, p_2 \ldots\ldots p_n$ the corresponding probability integrals, then λ_n is defined by

$$\lambda_n = p_1 . p_2 . \ldots\ldots p_n . \qquad \ldots (1)$$

Karl Pearson had shown that the p's follow the rectangular law of distribution (*Ibid*, p. 380). For the application of the test he evaluated the probability integral of λ_n, *viz.* $P(\lambda_n)$, and showed its connection with the incomplete Γ-function.

The expression for $P(\lambda_n)$ had been obtained with the help of hyperspace geometry. It is easy to derive the frequency distribution of λ_n from this expression for $P(\lambda_n)$.

Suppose $f(\lambda_n)$ to be the distribution law of λ_n. Then

$$P(\lambda_n) = \int_0^{\lambda_n} f(\lambda_n) \, . \, d\lambda_n \qquad \ldots (1.1)$$

or, in another form,

$$\frac{dP(\lambda_n)}{d\lambda_n} = f(\lambda_n) \qquad \ldots (1.2)$$

Karl Pearson had shown that

$$P(\lambda_n) = 1 - \lambda_n \left[1 - \frac{\log_e \lambda_n}{1!} + \frac{(\log_e \lambda_n)^2}{2!} - \ldots\ldots + (-1)^{n-1} \frac{(\log_e \lambda_n)^{n-1}}{(n-1)!} \right] \ldots (2.1)$$

Therefore, $$\frac{dP(\lambda_n)}{d\lambda_n} = -\left[1-\frac{\log_e\lambda_n}{1!}+\frac{(\log_e\lambda_n)^2}{2!}-\ldots+(-1)^{n-1}\frac{(\log_e\lambda_n)^{n-1}}{(n-1)!}\right]$$

$$-\lambda_n\left[-\frac{1}{1!\,\lambda_n}+\frac{2\log_e\lambda_n}{2!\,\lambda_n}-\ldots+(-1)^{n-1}\frac{(n-1)(\log_e\lambda_n)^{n-2}}{(n-1)!\,\lambda_n}\right] \quad \ldots \quad (2\cdot2)$$

which, on simplification, gives

$$\frac{d\,P(\lambda_n)}{d\,\lambda_n} = -\frac{1}{(n-1)!}(-\log_e\lambda_n)^{n-1} \quad \ldots \quad (2\cdot3)$$

Since λ_n lies between 0 and 1, $-\log_e\lambda_n$ is positive. We see therefore that $dP(\lambda_n)/d\lambda_n$ is negative, which simply means that the probability integral of λ_n has been measured from that end of its distribution which makes $P(\lambda_n)$ decrease as λ_n increases. We can, therefore, write the distribution of λ_n as

$$f(\lambda_n) = \frac{1}{(n-1)!}(-\log_e\lambda_n)^{n-1} \quad \ldots \quad (3.0)$$

2. This result can be obtained in another way without using hyperspace geometry The p's follow the rectangular distribution $\theta(p) = 1$, so that the chance of a variate falling between (p) and $(p+dp)$ is dp. Since $\lambda_n = p_1 . p_2 \ldots\ldots p_n$, we have

$$(-\log_e\lambda_n) = (-\log_e p_1)+(-\log_e p_2)+\ldots\ldots+(-\log_e p_n) \quad \ldots \quad (4\cdot1).$$

Let $u = -\log_e p$, and let $\phi(u)$ be its distribution. Then

$$\phi(u)\,.\,du = \theta(p)\,.\,dp = dp = d(e^{-u}) = -e^{-u}\,.\,du \quad \ldots \quad (4\cdot2)$$

Thus, $$\phi(u) = -e^{-u} \quad \ldots \quad (4\cdot3)$$

and $$-\log_e\lambda_n = u_1+u_2+\ldots\ldots+u_n = (n\,.\,U),\ \text{say} \quad \ldots \quad (4\cdot4)$$

The distribution law of U, the mean of a sample of n from the population $y=-e^{-u}$ has been given by J. Neyman and E. S. Pearson*

$$-\frac{n^n}{(n-1)!}(U)^{n-1}\,.\,e^{-nU} \quad \ldots \quad (5\cdot1)$$

From (4·1), the distribution law of $(-\log_e\lambda_n/n)$ is found to be

$$-\frac{n^n}{(n-1)!}\left(-\frac{\log_e\lambda_n}{n}\right)^{n-1}.\,\lambda_n \quad \ldots \quad (5\cdot2)$$

* On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference, (*Biometrika*, Vol. XXA, pp 175—240).

If $f(\lambda_n)$ be the frequency distribution of λ_n,

$$f(\lambda_n)\,.\,d\lambda_n = -\frac{n^n}{(n-1)!}\left(-\frac{\log_e \lambda_n}{n}\right)^{n-1}.\,\lambda_n\,.\,d\left(-\frac{\log_e \lambda_n}{n}\right) = \frac{(-\log_e \lambda_n)^{n-1}}{(n-1)!}\,.\,d\lambda_n \quad (5\cdot3)$$

whence

$$f(\lambda_n) = \frac{(-\log_e \lambda_n)^{n-1}}{(n-1)!} \quad \cdots \quad (3a)$$

For practical work the distribution of $(-\log_e \lambda_n)$ is more convenient than $f(\lambda_n)$ because it makes possible the direct use of the Incomplete Γ-function Tables. Thus if $t = -\log_e\lambda_n$, the distribution of t is given by

$$-\frac{1}{(n-1)!}\,t^{n-1}\,.\,e^{-t} \quad \cdots \quad (6)$$

This is a Type III curve, which is also the distribution curve for the Pearsonian χ^2. This is why the probability integrals of both χ^2 and λ_n can be obtained from Tables of the Incomplete Γ-function. In fact if we take

$$\chi^2 = -2\log_e \lambda_n \quad \ldots \quad (7\cdot1)$$

with the degrees of freedom $= 2n$, then the corresponding value of $P(\chi^2)$ will be equal to $Q(\lambda_n)$, that is

$$P(\chi^2) = Q(\lambda_n) = 1 - P(\lambda_n) \quad \ldots \quad (7\cdot2)$$

3. A third proof for the exact distribution of λ_n is available in Section 21. 1 of R. A. Fisher's *Statistical Methods for Research Workers*, 6th edition (1936) where the author briefly explains how to make use of the $P(\chi^2)$-Table when a number of probability integrals derived from independent samples for testing separately the significance of each sample estimate of a statistic (as for example, t, z, r etc.) have to be combined to yield a single comprehensive test of significance. A detailed exposition of Fisher's method is attempted below.

Let $p_1, p_2 \ldots\ldots p_n$ be the probability integrals, one or more, for each sample. It is not a condition of the problem that the p's derived for the various samples refer to the same statistic, a point emphasized by Karl Pearson in his paper on $P(\lambda_n)$ test. Thus for n_1 of the samples we might derive p's for t-distribution, for n_2 of the samples the p's may be for the sample means and so on.

In virtue of this property that the distribution of p is unaffected by the particular distribution of the statistic of which it is the probability integral, there is no objection in making the assumption that each p is got as the probability integral of any continuous statistical population. Fisher selects for the latter the distribution of χ^2 with 2 degrees of freedom, which choice brings out in a simple way the relationship between the $P(\lambda_n)$ and $P(\chi^2)$ tests.

Now the probability integral of the χ^2 distribution with $n=2$ is

$$\int_{\chi}^{\infty} e^{-\frac{1}{2}\chi^2} \cdot \chi d\chi = e^{-\frac{1}{2}\chi^2} \quad \ldots \quad (8\cdot1)$$

Thus if $p = e^{-\frac{1}{2}\chi^2}$, $-2 \log_e p$ follows the same distribution as χ^2 with 2 degrees of freedom.

Therefore

$$\log_e \lambda_n = \log_e p_1 + \log_e p_2 \ldots . + \log_e p_n = -\tfrac{1}{2}(\chi_1^2 + \chi_2^2 + \ldots\ldots + \chi_n^2) \quad \ldots \quad (8\cdot2)$$

Because of the additive property of χ^2, the right hand side of (8·2) follows the χ^2-distribution with 2 n degrees of freedom. The distribution of $-2 \log_e \lambda_n$ is therefore the same as that of χ^2 with $2n$ degrees of freedom.

This property has been derived earlier in (7·1) and (7·2) by a different route; without assuming the distribution of χ^2 we deduced that of λ_n and hence of $-2 \log_e \lambda_n$ which latter turned out to be identical in form with the distribution of χ^2 with $2n$ degrees of freedom.

SUMMARY

The paper describes three different methods of getting the exact distribution of λ_n, the product of the probability integrals of n independent values of a variable following a known continuous distribution.

7th November, 1936.