TIPPETT'S FORMULAS AND OTHER RESULTS ON SAMPLE RANGE AND EXTREMES

K. V. MARDIA

(Received Aug. 24, 1964; revised Jan. 11, 1965)

1. Introduction and summary

Let X and Y be the minimum and maximum observations in a random sample of size n drawn from a continuous population having the probability density function (p.d.f.) $f(x)$ and distribution function $F(x)$. Let the sample range be $R(=Y-X)$. For convenience, we shall assume that x is limited by $a \leqq x \leqq b$ unless otherwise is mentioned.

We shall furnish alternative proofs of the Tippett's formulas (see, Tippett [7]) for $E(R)$ and $E[R-E(R)]^m$, where m is a positive integer. Tippett [7] have assumed in his proof of the latter formula given by his equation (9) that n is even. We shall show that the same equation holds for all n. Further, it is believed that our proofs are much simpler than all the known proofs. For various available proofs of the formula for $E(R)$, one can refer to Tippett [7], Gumbel [2], Kendall and Stuart [3]. The only known proof of the simplified formula for $E[R-E(R)]^m$ (Tippett [7], equation (9)), is given in Tippett [7] where n is restricted to the even values.

We shall obtain a formula for $\mathrm{Cov}(X^r, Y^s)$ similar to Tippett's formulas for $E(R)$ and $E[R-E(R)]^m$, and show that it is non-negative for the odd values of r and s. In particular $\mathrm{Cov}(X, Y)$ is non-negative.

The exact values of variance of sample range from the normal population for $n=2$ and $n=4$ are known (Ruben [6]). We shall fill up this gap by providing the exact value for $n=3$, with the help of the p.d.f. of the range obtained by Mackay and Pearson [4].

2. Alternative proof of $E(R)$ formula

We know that the p.d.f. of Y is given by

$$\frac{d[F(Y)]^n}{dY},$$

so that

$$E(Y)=\int_a^b Y\frac{d[F(Y)]^n}{dY}dY$$

(2.1)
$$=\Big[Y[F(Y)]^n\Big]_{Y=a}^{Y=b}-\int_a^b [F(Y)]^n dY$$

$$=b-\int_a^b [F(Y)]^n dY\,.$$

Similarly, we have

(2.2)
$$E(X)=a+\int_a^b [1-F(X)]^n dX\,.$$

Therefore from (2.1) and (2.2), we get

(2.3)
$$E(R)=\int_a^b [1-\{F(X)\}^n-\{1-F(X)\}^n]dX\,.$$

3. Alternative proof of $E[R-E(R)]^m$ formula

LEMMA 1. *If r is a number greater than or equal to 2 and $\phi(X, Y)$, $\frac{\partial\phi(X, Y)}{\partial X}$, $\frac{\partial\phi(X, Y)}{\partial Y}$ are continuous functions of X and Y for $a\leqq X$, $Y\leqq b$, and $\frac{\partial^2\phi(X, Y)}{\partial X\partial Y}$ exists for all such values of X and Y, then*

$$\int_a^b\int_a^Y (Y-X)^r\left\{\frac{\partial^2\phi(X, Y)}{\partial X\partial Y}\right\}dYdX$$

(3.1)
$$=r(r-1)\int_a^b\int_a^Y (Y-X)^{r-2}[\phi(a, Y)+\phi(X, b)-\phi(a, b)-\phi(X, Y)]dYdX\,.$$

PROOF. On carrying out integration of the left member of (3.1) with respect to X, this term becomes

$$-\int_a^b (Y-a)^r\frac{\partial\phi(a, Y)}{\partial Y}dY+r\int_a^b\int_a^Y (Y-X)^{r-1}\frac{\partial\phi(X, Y)}{\partial Y}dYdX\,.$$

We interchange the order of integration in the second term of the above and carry out integration with respect to Y in both terms by parts, it reduces to

(3.2)
$$r\int_a^b (Y-a)^{r-1}\phi(a, Y)dY+r\int_a^b (b-X)^{r-1}\phi(X, b)dX-(b-a)^r\phi(a, b)$$
$$-r(r-1)\int_a^b\int_a^Y (Y-X)^{r-1}\phi(X, Y)dYdX\,.$$

We can prove the following results:

$$r\int_a^b (Y-a)^{r-1}\phi(a,\ Y)dY=r(r-1)\int_a^b\int_a^Y (Y-X)^r\phi(a,\ Y)dYdX\,,$$

$$r\int_a^b (b-X)^{r-1}\phi(X,\ b)dX=r(r-1)\int_a^b\int_a^Y (Y-X)^r\phi(X,\ b)dYdX\,,$$

and

$$-(b-a)^r\phi(a,\ b)=r(r-1)\int_a^b\int_a^Y (Y-X)^r\phi(a,\ b)dYdX\,.$$

With the help of these results, we can rewrite (3.2) in the form of the right member of (3.1). Thus the lemma is established.

THEOREM 1. *For any positive integral value of r greater than one and for any n,*

$$(3.3)\qquad E(R^r)=r(r-1)\int_a^b\int_a^Y (Y-X)^{r-2}[1-\{F(Y)\}^n-\{1-F(X)\}^n+\{F(Y)-F(X)\}^n]dYdX\,.$$

PROOF. We can write the joint p.d.f. of X and Y as

$$-\frac{\partial^2[F(Y)-F(X)]^n}{\partial X\partial Y}\,,$$

so that

$$(3.4)\qquad E(R^r)=-\int_a^b\int_a^Y (Y-X)^r\frac{\partial^2[F(Y)-F(X)]^n}{\partial X\partial Y}dYdX\,.$$

By applying lemma 1 to (3.4), we obtain the desired result. Now

$$(3.5)\qquad E[R-E(R)]^m=-(m-1)[-E(R)]^m+\sum_{r=2}^{m}\binom{m}{r}[-E(R)]^{m-r}E(R^r)\,.$$

We substitute the value of $E(R^r)$ from (3.3) in the right member of (3.5), and interchange the signs of summation and integration, and then sum the series by the binomial theorem, so that (3.5) reduces to

$$(3.6)\qquad E[R-E(R)]^m=\int_a^b\int_a^Y[1-\{F(Y)\}^n-\{1-F(X)\}^n+\{F(Y)-F(X)\}^n][Y-X-E(R)]^{m-2}dYdX-(m-1)[-E(R)]^m\,,$$

which is the Tippett's formula.

It is worth noting that Tippett [7] in his equation (10), has deduced the variance of range from equation (3.6) but his equation (10) has a misprint of negative sign in place of positive sign in its fourth term. This misprint is reproduced in some standard texts such as Gumbel [2] and Kendall and Stuart [3].

4. Formula for Cov (X^r, Y^s)

THEOREM 2. *If r and s are any integers, then*

$$\text{Cov}(X^r, Y^s)=\int_a^b\int_a^b X^{r-1}Y^{s-1}[F(Y)]^n[1-F(X)]^n dYdX - \int_a^b\int_a^Y X^{r-1}Y^{s-1}[F(Y)-F(X)]^n dYdX. \tag{4.1}$$

PROOF. We have

$$E(X^rY^s)=-\int_a^b\int_a^Y X^rY^s\frac{\partial^2[F(Y)-F(X)]^n}{\partial X\partial Y}dYdX.$$

On first integrating out by parts with respect to X and then with respect to Y as done in lemma 1, we get

$$E(X^rY^s)=a^rb^s-sa^r\int_a^b Y^{s-1}[F(Y)]^n dY+rb^s\int_a^b X^{r-1}[1-F(X)]^n dX - \int_a^b\int_a^Y X^{r-1}Y^{s-1}[F(Y)-F(X)]^n dYdX. \tag{4.2}$$

Also we can show as done in (2.1) and (2.2) that

$$E(X^r)E(Y^s)=a^rb^s-sa^r\int_a^b Y^{s-1}[F(Y)]^n dY+rb^s\int_a^b X^{r-1}[1-F(X)]^n dX - \int_a^b\int_a^b X^{r-1}Y^{s-1}[F(Y)]^n[1-F(X)]^n dYdX. \tag{4.3}$$

The proof follows on substituting (4.2) and (4.3) in the expression for Cov (X^r, Y^s).

Now, since

$$F(Y)[1-F(X)]\geqq[F(Y)-F(X)], \tag{4.4}$$

we conclude from (4.1) that Cov (X^r, Y^s) is non-negative if r and s are odd integers.

5. Effect of $a\to-\infty$ and $b\to+\infty$

It can be easily seen that the theorems in sections 2, 3 and 4 hold even when $a\to-\infty$ and $b\to+\infty$ provided the integrals appeared therein are convergent. The following theorems give sufficient conditions for their convergence which are based only on the population moments. We shall denote the population rth moment by μ'_r.

THEOREM 3. *For any integers r and s, $E(X^r Y^s)$ is finite if μ'_k is finite where $k=\max(r, s)$.*

PROOF. Suppose $X_n(=X)$ and $Y_n(=Y)$ are the minimum and maximum observations in a random sample of size n, then with the help of (4.4) we can show that

$$E[\,|X_n^r Y_n^s|\,] \leqq \frac{n}{n-1}[E\,|X_{n-1}|^r][E\,|Y_{n-1}|^s]\,. \tag{5.1}$$

Now on using $[1-F(X_n)] \geqq 0$, we also have

$$E[\,|X_n|^r] < nE[\,|x|^r]\,. \tag{5.2}$$

Similarly as $F(Y_n) \geqq 0$, we get

$$E[\,|Y_n|^s] \leqq nE[\,|x|^s]\,. \tag{5.3}$$

From (5.2) and (5.3) we observe that $E[\,|X_n|^r]$ and $E[\,|Y_n|^s]$ exist if μ'_k exists where $k=\max(r, s)$. On using this information in (5.1), we obtain the desired result.

COROLLARY 1. $\mathrm{Cov}\,(X^r, Y^s)$ *exists if μ'_k exists where $k=\max(r, s)$.*

THEOREM 4. *$E(R^r)$ is finite if μ'_r exists.*

PROOF. We have

$$E(R^r) = \sum_{k=1}^{r} (-1)^k \binom{r}{k} E[Y^{r-k} X^k]\,,$$

and on utilizing theorem 3, we deduce that $E(R^r)$ exists if μ'_k exists for $k=1, \cdots, r$, i.e. by Cramér [1] if μ'_r is finite.

6. Variance of normal range for $n=3$

It is interesting to note from Pearson and Hartley [5] that

$$V_n(R) \leqq V_3(R) \tag{6.1}$$

for all n, where $V_n(R)$ stands for the variance of range of a random sample of size n from the normal population. Thus $V_3(R)$ has distinction of becoming the maximum of $V_n(R)$.

Without any loss of generality, we assume that the population standard deviation is unity. We shall show that

(6.2) $$V_3(R)=2-\frac{3\sqrt{3}}{\pi}(\sqrt{3}-1)\,.$$

On utilizing the p.d.f. of the normal range for $n=3$, given by Mackay and Pearson [4], we have in this case

(6.3) $$E(R^2)=\frac{6}{\pi\sqrt{2}}\left[\int_0^\infty \left\{w^2e^{-w^2/4}\left(\int_0^{w/\sqrt{6}} e^{-u^2/2}du\right)\right\}dw\right].$$

On changing the order of integration, (6.3) becomes

$$E(R^2)=\frac{6}{\pi\sqrt{2}}\left[\int_0^\infty e^{-u^2/2}\left(\int_{\sqrt{6}u}^\infty w^2e^{-w^2/4}dw\right)du\right],$$

and now integrating with respect to w by parts (taking parts as w and $we^{-w^2/4}$), it reduces to

(6.4) $$E(R^2)=\frac{6\sqrt{2}}{\pi}\left[\sqrt{6}\int_0^\infty ue^{-2u^2}du+\int_0^\infty\int_{\sqrt{6}u}^\infty e^{-u^2/2-w^2/4}\,dudw\right]$$

$$=\frac{6\sqrt{2}}{\pi}\left[\sqrt{3/8}+\int_0^\infty\int_0^\infty \exp\left[-\frac{1}{4}(8u^2+2\sqrt{6}uv+v^2)\right]dudv\right].$$

Now, since the value of the integral in (6.4) is known by the Sheppard's formula, we get

(6.5) $$E(R^2)=\frac{3\sqrt{3}}{\pi}+2\,.$$

Also

(6.6) $$E(R)=3/\sqrt{\pi}\,.$$

The result follows on using (6.5) and (6.6).

It may be noted that the computed value of $V_3(R)$ given in Pearson and Hartley [5] is .78922 while its value from the above formula is .78920. So, the tabulated value differs only in the last decimal place.

7. Acknowledgments

The author is grateful to Dr. G. C. Patni for his encouragement and interest, to Professor R. L. Plackett for the suggestions and to the referee for his helpful comments.

UNIVERSITY OF RAJASTHAN & UNIVERSITY OF NEWCASTLE

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[2] E. G. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, 1958.

[3] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. 1, Charles Griffin and Co. Ltd., London, 1958.

[4] A. T. Mackay and E. S. Pearson, "A note on the distribution of range in sample of n," *Biometrika*, 25 (1933), 415-420.

[5] E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, Part 1, Cambridge University Press, London, 1958.

[6] H. Ruben, "On the moments of range and product of extreme order statistics in normal samples," *Biometrika*, 43 (1956), 458-460.

[7] L. H. C. Tippett, "On the extreme individuals and the range of the samples taken from a normal population," *Biometrika*, 17 (1925), 364-387.