The Annals of Statistics
1973, Vol. 1, No. 2, 392-394

MOMENTS OF ORDER STATISTICS

BY KOSABURO HIRAKAWA

Science University of Tokyo

Let $X_{(i|n)}$ be the ith smallest order statistic from a population with pdf $f(x)$ and cdf $F(x)$. When $\tilde{x}$ is the population median, n is the sample size and $G(x) = F(x)(1 - F(x))$, the following are proved.

$$E(X_{(n-i+1|n)}) + E(X_{(i|n)}) = n(n-2i+1)\binom{n-1}{i-1} \sum_{\nu=0}^{[(n-2i+1)/2]} \frac{(-1)^\nu}{n-2i+1-\nu} \binom{n-2i+1-\nu}{\nu} \times \Psi_1(i-1+\nu),$$

$$E(X_{(n-i+1|n)}) - E(X_{(i|n)}) = n\binom{n-1}{i-1} \sum_{\nu=0}^{[(n-2i)/2]} (-1)^\nu \binom{n-2i-\nu}{\nu} \Psi_2(i-1+\nu)$$

for $i = 1, 2, \cdots, [n/2]$ and $E(X_{(i|n)}) = n\binom{n-1}{i-1}\Psi_1(i-1)$ for $i = (n+1)/2$ when n is odd, where $\Psi_1(l) = \int_{-\infty}^{\infty} x f(x)\{G(x)\}^l \, dx$ and

$$\Psi_2(l) = -\int_{-\infty}^{\tilde{x}} x f(x)\{G(x)\}^l (1-4G(x))^{\frac{1}{2}} \, dx + \int_{\tilde{x}}^{\infty} x f(x)\{G(x)\}^l (1-4G(x))^{\frac{1}{2}} \, dx .$$

Parallel formulas are obtained for the kth order moment.

The expected value of the ith order statistic $X_{(i|n)}$ for a sample of size n from a population with probability density function $f(x)$ and cumulative distribution function $F(x)$ is given by the equation

$$(1) \qquad E(X_{(i|n)}) = n\binom{n-1}{i-1} \int_{-\infty}^{\infty} x f(x)\{F(x)\}^{i-1}\{1 - F(x)\}^{n-i} \, dx \qquad (i = 1, 2, \cdots, n),$$

which we can evaluate by numerical integration. But straightforward integration has shortcomings in view of the quantity of computation and the accuracy. Here some expressions are proposed for computing the moments of order statistics.

If we denote the population median by $\tilde{x}$ and define

$$\Psi_1(k, l) = \int_{-\infty}^{\infty} x^k f(x)\{F(x)(1 - F(x))\}^l \, dx ,$$

$$\Psi_2(k, l) = -\int_{-\infty}^{\tilde{x}} x^k f(x)\{F(x)(1 - F(x))\}^l [1 - 4F(x)(1 - F(x))]^{\frac{1}{2}} \, dx + \int_{\tilde{x}}^{\infty} x^k f(x)\{F(x)(1 - F(x))\}^l [1 - 4F(x)(1 - F(x))]^{\frac{1}{2}} \, dx$$

$$(k = 1, 2, \cdots; l = 0, 1, 2, \cdots),$$

then we have

$$(2) \qquad E(X_{(n-i+1|n)}) + E(X_{(i|n)}) = n(n-2i+1)\binom{n-1}{i-1} \sum_{\nu=0}^{[(n-2i+1)/2]} \frac{(-1)^\nu}{n-2i+1-\nu} \binom{n-2i+1-\nu}{\nu} \times \Psi_1(1, i-1+\nu),$$

$$(3) \qquad E(X_{(n-i+1|n)}) - E(X_{(i|n)}) = n\binom{n-1}{i-1} \sum_{\nu=0}^{[(n-2i)/2]} (-1)^\nu \binom{n-2i-\nu}{\nu} \Psi_2(1, i-1+\nu)$$

Received October 13, 1971; revised June 26, 1972.

for $i = 1, 2, \cdots, [n/2]$, and

$$E(X_{(i|n)}) = n\binom{n-1}{i-1}\Psi_1(1, i-1) \tag{4}$$

for $i = (n+1)/2$ when n is odd.

We used the following relations (5) and (6) for the proof of (2) and (3).

$$F^{\alpha}(x) + (1 - F(x))^{\alpha} = \alpha \sum_{\nu=0}^{[\alpha/2]} \frac{(-1)^{\nu}}{\alpha - \nu}\binom{\alpha - \nu}{\nu}\{F(x)(1 - F(x))\}^{\nu}, \tag{5}$$

$$\begin{aligned} F^{\alpha}(x) &- (1 - F(x))^{\alpha} \\ &= \pm \sum_{\nu=0}^{[(\alpha-1)/2]} (-1)^{\nu}\binom{\alpha-\nu-1}{\nu}\{F(x)(1 - F(x))\}^{\nu}[1 - 4F(x)(1 - F(x))]^{\frac{1}{2}}, \end{aligned} \tag{6}$$

where the right-hand side is preceded by the plus sign for $x \geqq \tilde{x}$ and by the minus sign for $x \leqq \tilde{x}$. (5) is an equivalent form of the expression derived by Godwin [1] and (6) may be proved by his similar argument.

Using Ψ_1 and Ψ_2 after the application of (5) and (6) with $\alpha = n - 2i + 1$ we have (2) and (3) directly. It is obvious that (4) is true. We can also obtain the expressions for the kth order moment corresponding to the 1st order moment, if we replace $\Psi_1(1, l)$ and $\Psi_2(1, l)$ in (2), (3) and (4) by $\Psi_1(k, l)$ and $\Psi_2(k, l)$ respectively.

When the integrand includes two parameters such as the formula (1), a great quantity of numerical integration is needed for the tabulation. Since the integrands in our formulas (2), (3) and (4) are constructed by $F(x)(1 - F(x))$ and include only one parameter, the truncation error may be relatively small and the integration may be reduced. In the symmetric case where $F(-x) = 1 - F(x)$ we have

$$\begin{aligned} E(X^{2r}_{(n-i+1|n)}) &= E(X^{2r}_{(i|n)}) \\ &= \tfrac{1}{2}n(n - 2i + 1)\binom{n-1}{i-1} \sum_{\nu=0}^{[(n-2i+1)/2]} \frac{(-1)^{\nu}}{n - 2i + 1 - \nu}\binom{n - 2i + 1 - \nu}{\nu} \\ &\quad \times \Psi_1(2r, i - 1 + \nu), \end{aligned} \tag{7}$$

$$\begin{aligned} E(X^{2r-1}_{(n-i+1|n)}) &= -E(X^{2r-1}_{(i|n)}) \\ &= \tfrac{1}{2}n\binom{n-1}{i-1} \sum_{\nu=0}^{[(n-2i)/2]} (-1)^{\nu}\binom{n-2i-\nu}{\nu}\Psi_2(2r - 1, i - 1 + \nu) \end{aligned} \tag{8}$$

for $r = 1, 2, \cdots; i = 1, 2, \cdots, [n/2]$, and

$$E(X^{2r}_{(i|n)}) = n\binom{n-1}{i-1}\Psi_1(2r, i - 1), \tag{9}$$

$$E(X^{2r-1}_{(i|n)}) = 0 \tag{10}$$

for $i = (n+1)/2$ when n is odd.

We computed $E(X^{k}_{(n-i+1|n)})$ for samples from $N(0, 1)$ by (7), (8) and (9) for $k = 1(1)4$, $n = 2(1)100$, $i = 1(1)[(n+1)/2]$. The variance, skewness and kurtosis were also computed for those cases. In some cases, instead of the cdf $U(x)$ of $X_{(n-i+1|n)}$, an approximation $U^*(x)$ which is the integral of the sum of the first four terms of the Edgeworth series may be used, since it is seen by investigation that

$\max_x |U(x) - U^*(x)|$ is less than 0.004 for $X_{(100|100)}$ which has the largest skewness and kurtosis.

REFERENCE

[1] GODWIN, H. J. (1949). On the estimation of dispersion by linear systematic statistics. *Biometrika* **36** 92–100.

2-5-103 NISHIYAMATO-DANCHI
WAKOSHI SAITAMA
JAPAN