THE CALCULATION OF CUMULANTS VIA CONDITIONING

DAVID R. BRILLINGER

(Received Oct. 25, 1967; Revised Aug. 27, 1968)

The cumulants of random variables are important in deriving, for statistics of interest, exact sampling distributions, approximate sampling distributions (as via Cornish-Fisher expansions) and asymptotic sampling distributions (such as asymptotic normality). This note presents a means of calculating cumulants when two or more stages of sampling may be recognized.

Given the k-variate random variable $(x_1, \cdots, x_k)$, let A denote an event in the associated probability field. The following properties of first and second order cumulants are well known (see Hansen, Hurwitz and Madow ([4], pp. 61–66) or Feller ([3], p. 164)).

$$Ex_i = \underset{A}{E}\{E(x_i|A)\} \tag{1}$$

$$\operatorname{var} x_i = \underset{A}{E}\{\operatorname{var}(x_i|A)\} + \underset{A}{\operatorname{var}}\{E(x_i|A)\} \tag{2}$$

$$\operatorname{cov}(x_i, x_j) = \underset{A}{E}\{\operatorname{cov}(x_i, x_j|A)\} + \underset{A}{\operatorname{cov}}\{E(x_i|A), E(x_j|A)\} \tag{3}$$

for $i, j=1, \cdots, k$ where $|A)$ indicates that calculations are carried out conditionally on the event A, while the subscript A indicates that calculations are carried out over the various values of A. Let $\kappa(x_1, \cdots, x_k)$ denote the joint kth order cumulant of $(x_1, \cdots, x_k)$ and for integers $\beta_1, \cdots, \beta_k$ let $\kappa_{\beta_1\cdots\beta_k}(x_1, \cdots, x_k) = \kappa(x_1[\beta_1 \text{ times}], \cdots, x_k[\beta_k \text{ times}])$. In this note we generalize (1), (2), (3) to

$$\kappa(x_1, \cdots, x_k) = \sum_{\alpha} \underset{A}{\kappa}\{\kappa(x_{\alpha_1}|A), \cdots, \kappa(x_{\alpha_p}|A)\} \ . \tag{4}$$

The summation in (4) extends over all partitions $\alpha=(\alpha_1, \cdots, \alpha_p)$, $p=1, \cdots, k$ of the integers $(1, \cdots, k)$ and $x_{\alpha_j}=(x_{j_1}, \cdots, x_{j_J})$ if $\alpha_j=(j_1, \cdots, j_J)$. We may prove,

THEOREM. *Given the k-variate random variable* $(x_1, \cdots, x_k)$ *with* $E|x_i|^k < \infty$, $i=1, \cdots, k$, *the identity* (4) *is valid.*

PROOF. $\kappa(x_1, \cdots, x_k)$ is the coefficient of $t_1 \cdots t_k$ in the Taylor series expansion of $\log E(\exp \sum_i x_i t_i)$ about the origin. (This expansion may be carried out because of the assumed finiteness of moments.) However

$$(5)\qquad \log E(\exp \sum x_i t_i) = \log \underset{A}{E}\{E(\exp \sum x_i t_i | A)\}$$
$$= \log \underset{A}{E}\left\{\exp\left(\sum \kappa_{\beta_1\cdots\beta_k}(x_1, \cdots, x_k | A) \frac{t_1^{\beta_1}\cdots t_k^{\beta_k}}{\beta_1!\cdots\beta_k!} + o(\|t\|^k)\right)\right\}$$

where $\|t\|^2 = t_1^2 + \cdots + t_k^2$ and the summation extends over integers β_i, $0 \leqq \beta_i \leqq k$, $i = 1, \cdots, k$, with $0 < \sum_i \beta_i \leqq k$. We note that the expression on the right-hand side of (5) is essentially the cumulant generating function of the random variables $\kappa_{\beta_1\cdots\beta_k}(x_1, \cdots, x_k | A)$. The stated result now follows on identification of the coefficient of $t_1 \cdots t_k$.

COROLLARY. *The kth order cumulant $\kappa_k(x)$ of a univariate random variable x, with $E|x|^k < \infty$, is given by*

$$(6)\qquad \sum \frac{k!}{\mu_1!\,\mu_2!\cdots} \frac{1}{(p_1!)^{\mu_1}(p_2!)^{\mu_2}\cdots} \underset{A}{\kappa_{\mu_1\mu_2\cdots}}\{\kappa_{p_1}(x|A), \kappa_{p_2}(x|A), \cdots\}$$

where the summation extends over all partitions $(p_1^{\mu_1}, p_2^{\mu_2}, \cdots)$ of k with $p_1\mu_1 + p_2\mu_2 + \cdots = k$.

This corollary follows from the theorem on taking $x_i = x$, $i = 1, \cdots, k$, and counting the identical terms.

We now turn to several examples of the theorem and corollary.

Example 1. Mixtures. Suppose that the probability measure of $(x_1, \cdots, x_k)$ is in fact a mixture, that is its c.d.f. $F(x_1, \cdots, x_k)$ is of the form

$$(7)\qquad F(x_1, \cdots, x_k) = \int G(x_1, \cdots, x_k; \theta)\, dU(\theta)$$

where, for fixed θ, $G(x_1, \cdots, x_k; \theta)$ is a c.d.f. and $U(\theta)$ is a probability measure in θ. The theorem allows us to express the kth order joint cumulant of $(x_1, \cdots, x_k)$ in terms of the cumulants calculated from $G(x_1, \cdots, x_k; \theta)$ for fixed θ. The required expression is given by (4) taking A to refer to θ.

This result is given for the first and second order cases in Feller ([3], p. 164).

The next example refers to the sum of a random number of random variables.

Example 2. Cumulants of random sums. Let $x_1, x_2, \cdots$ be a sequence of independent, identically distributed random variables with $\kappa_j(x) = \kappa_j$, $j = 1, \cdots, k$ existing, and n an integer valued random variable distributed independently of the sequence, whose moments exist up to order k. Let $S_n=x_1+\cdots+x_n$. From (1) and (2) above, letting A refer to n, we see $ES_n=(Ex)(En)$ and $\operatorname{var} S_n=(En)\operatorname{var} x+\operatorname{var} n(Ex)^2$. In general we have from the corollary, taking A to refer to n,

$$(8)\qquad \begin{aligned}\kappa_k(S_n)&=\sum\frac{k!}{\mu_1!\,\mu_2!\cdots}\,\frac{1}{(p_1!)^{\mu_1}(p_2!)^{\mu_2}\cdots}\,\underset{n}{\kappa}_{\mu_1\mu_2\cdots}\{n\kappa_{p_1},\,n\kappa_{p_2},\,\cdots\}\\ &=\sum\frac{k!}{\mu_1!\,\mu_2!\cdots}\,\frac{1}{(p_1!)^{\mu_1}(p_2!)^{\mu_2}\cdots}\,\kappa_{p_1}^{\mu_1}\kappa_{p_2}^{\mu_2}\cdots\kappa_{\mu_1+\mu_2+\cdots}(n)\,,\end{aligned}$$

the summation extending over all partitions $(p_1^{\mu_1}, p_2^{\mu_2},\cdots)$ of k with $p_1\mu_1+p_2\mu_2+\cdots=k$.

The expression (8) may be used to derive a central limit theorem for a random number of random variables. Suppose all moments of x and n exist with $\kappa_1(x)=0$. Suppose the distribution of n depends on a parameter N with $\lim_{N\to\infty}\kappa_1(n)=\infty$. Consider the standardized variate $Z_n=S_n/(\operatorname{var} S_n)^{1/2}$. We see that $EZ_n=0$, $\operatorname{var} Z_n=1$ and $\kappa_k(Z_n)=\kappa_k(S_n)/(En\cdot\operatorname{var} x)^{k/2}$. By inspection we see that if $\kappa_k(n)/(E(n))^{k/2}\to 0$ as $N\to\infty$ for $k=3, 4, \cdots$, then $\kappa_k(Z_n)\to 0$ as $N\to\infty$ for $k=3, 4, \cdots$. We see that Z_n is asymptotically standardized normal. Central limit theorems for random sums are considered in Robbins [5] and Wittenberg [7].

Robbins also considers an alternate form of standardization of S_n, namely $Y_n=(S_n-ES_n)/n^{1/2}$. Here we see from (1) and (2) that $EY_n=0$ and $\operatorname{var} Y_n=\operatorname{var} x$. From the corollary we have for $k>2$,

$$(9)\qquad \begin{aligned}\kappa_k(Y_n)=\sum\frac{k!}{\mu_1!\,\mu_2!\cdots}\,&\frac{1}{(p_1!)^{\mu_1}(p_2!)^{\mu_2}\cdots}\,\kappa_{p_1}^{\mu_1}\kappa_{p_2}^{\mu_2}\cdots\\ &\times\kappa_{\mu_1\mu_2\cdots}(n^{-(p_1-2)/2},\,n^{-(p_2-2)/2},\,\cdots)\end{aligned}$$

the summation extending over all partitions $(p_1^{\mu_1}, p_2^{\mu_2},\cdots)$ of k with $p_1\mu_1+p_2\mu_2+\cdots=k$ and $p_1, p_2, \cdots>1$.

Example 3. Two-stage sampling. Consider a sampling plan involving the selection of n first-stage units with or without replacement and with possibly unequal probabilities, followed by a second stage of sampling, carried out independently within the selected first-stage units, followed by the measurement of the k-variate random variable $x(j)=(x_1(j),\cdots,x_k(j))$ in the jth unit. Define indicator variables as follows; $a_j=1$ if the jth unit is in the sample and $a_j=0$ otherwise. Consider sample totals. We see that these have the form $X_i=\sum_j a_jx_i(j)$, $i=1,\cdots,k$ where $(a_1, a_2,\cdots)$ is independent of the $(x_1(j),\cdots,x_k(j))$, $j=1, 2,\cdots$, which are independent of each other.

Letting A refer to the variate $(a_1, a_2, \cdots)$ and $X=(X_1,\cdots,X_k)$, we have from the theorem

$$\kappa(X)=\sum_{\alpha}\kappa_{A}\{\kappa(X_{\alpha_1}|A),\cdots,\kappa(X_{\alpha_p}|A)\}, \tag{10}$$

the summation extending over all partitions $\alpha=(\alpha_1,\cdots,\alpha_p)$, $p=1,\cdots,k$ of the integers $(1,\cdots,k)$.

Since the $(x_1(j),\cdots,x_k(j))$, $j=1,2,\cdots$ are independent and $a_j^m=a_j$, $m=1,2,\cdots$

$$\kappa(X_\beta|A)=\sum_j a_j\kappa(x_\beta;\,j) \tag{11}$$

where $\kappa(x_\beta;\,j)=\kappa(x_{i_1}(j),\cdots,x_{i_I}(j))$ if $\beta=(i_1,\cdots,i_I)$. We have therefore

$$\kappa(X)=\sum_{\alpha}\sum_{j_1}\cdots\sum_{j_p}\kappa(x_{\alpha_1};\,j_1)\cdots\kappa(x_{\alpha_p};\,j_p)\kappa(a_{j_1},\cdots,a_{j_p})\,. \tag{12}$$

We note that the cumulants of the variate $(a_1,a_2,\cdots)$ are needed and that these depend solely on the form of sampling employed in the selection of the first-stage units. We see that in order to obtain an unbiased estimate of $\kappa(X)$, we require unbiased estimates of the products of the cumulants of the $x(j)$. If the first-stage units are infinite in size and one employs simple random sampling within them, these estimates have been provided in Dressel [1] and Tukey [6].

After this note had been prepared, the author learned that D. S. Robson of Cornell University had previously obtained the result contained in the corollary. Ebner [2] employed it in an investigation of the balanced one-way nested classification and work has continued at Cornell on its use in sampling from finite populations.

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

REFERENCES

[1] P. L. Dressel, "Statistical seminvariants and their estimates with particular emphasis on their relation to algebraic invariants," *Ann. Math. Statist.* 9 (1940), 33-57.
[2] A. Ebner, "Cumulant component estimation in the balanced one-way nested classification," M. Sc. Thesis at Cornell University, 1959.
[3] W. Feller, *An Introduction to Probability Theory and its Applications*, II, Wiley, New York, 1966.
[4] M. H. Hansen, W. N. Hurwitz and W. G. Madow, *Sample Survey Methods and Theory*, II, Wiley, New York, 1953.
[5] H. E. Robbins, "The asymptotic distribution of the sum of a random number of random variables," *Bull. Amer. Math. Soc.*, 54 (1948), 1151-1161.
[6] J. W. Tukey, "Some sampling simplified," *J. Amer. Statist. Ass.*, 45 (1950), 501-519.
[7] H. Wittenberg, "The limiting distributions of random sums of independent random variables," *Zeit. für Wahrschein.*, 3 (1964), 7-18.