1 Cumulants

1.1 Definition

The rth moment of a real-valued random variable X with density $f(x)$ is

$$\mu_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x) \, dx$$

for integer $r = 0, 1, \ldots$. The value is assumed to be finite. Provided that it has a Taylor expansion about the origin, the moment generating function

$$M(\xi) = E(e^{\xi X}) = E(1 + \xi X + \cdots + \xi^r X^r / r! + \cdots)$$
$$= \sum_{r=0}^{\infty} \mu_r \xi^r / r!$$

is an easy way to combine all of the moments into a single expression. The rth moment is the rth derivative of M at the origin.

The cumulants κ_r are the coefficients in the Taylor expansion of the cumulant generating function about the origin

$$K(\xi) = \log M(\xi) = \sum_r \kappa_r \xi^r / r!.$$

Evidently $\mu_0 = 1$ implies $\kappa_0 = 0$. The relationship between the first few moments and cumulants, obtained by extracting coefficients from the expansion, is as follows

$$\kappa_1 = \mu_1$$
$$\kappa_2 = \mu_2 - \mu_1^2$$
$$\kappa_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3$$
$$\kappa_4 = \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4.$$

In the reverse direction

$$\mu_2 = \kappa_2 + \kappa_1^2$$
$$\mu_3 = \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3$$
$$\mu_4 = \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4.$$

In particular, $\kappa_1 = \mu_1$ is the mean of X, κ_2 is the variance, and $\kappa_3 = E((X - \mu_1)^3)$. Higher-order cumulants are not the same as moments about the mean.

This definition of cumulants is nothing more than the formal relation between the coefficients in the Taylor expansion of one function $M(\xi)$ with $M(0) = 1$, and the coefficients in the Taylor expansion of $\log M(\xi)$. For example Student's t on five degrees of freedom has finite moments up to order four, with infinite moments of order five and higher. The moment generating function does not exist for real $\xi \neq 0$, but the characteristic function $M(i\xi)$ is $e^{-|\xi|}(1 + |\xi| + \xi^2/3)$. Both $M(i\xi)$ and $K(i\xi) = -|\xi| + \log(1 + |\xi| + \xi^2/3)$ have Taylor expansions about $\xi = 0$ up to order four only.

The normal distribution $N(\mu, \sigma^2)$ has cumulant generating function $\xi\mu + \xi^2\sigma^2/2$, a quadratic polynomial implying that all cumulants of order three and higher are zero. Marcinkiewicz (1935) showed that the normal distribution is the only distribution whose cumulant generating function is a polynomial, i.e. the only distribution having a finite number of non-zero cumulants. The Poisson distribution with mean μ has moment generating function $\exp(\mu(e^\xi - 1))$ and cumulant generating function $\mu(e^\xi - 1)$. Consequently all the cumulants are equal to the mean.

Two distinct distributions may have the same moments, and hence the same cumulants. This statement is fairly obvious for distributions whose moments are all infinite, or even for distributions having infinite higher-order moments. But it is much less obvious for distributions having finite moments of all orders. Heyde (1963) gave one such pair of distributions with densities

$$f_1(x) = \exp(-(\log x)^2/2)/(x\sqrt{2\pi})$$
$$f_2(x) = f_1(x)[1 + \sin(2\pi \log x)/2]$$

for $x > 0$. The first of these is called the log normal distribution. To show that these distributions have the same moments it suffices to show that

$$\int_0^\infty x^k f_1(x) \sin(2\pi \log x)\, dx = 0$$

for integer $k \geq 1$, which can be shown by making the substitution $\log x = y + k$.

Cumulants of order $r \geq 2$ are called semi-invariant on account of their behaviour under affine transformation of variables (Thiele 198?, Dressel 1942). If κ_r is the rth cumulant of X, the rth cumulant of the affine transformation $a + bX$ is $b^r \kappa_r$, independent of a. This behaviour is considerably simpler than that of moments. However, moments about the mean are also semi-invariant, so this property alone does not explain why cumulants are useful for statistical purposes.

The term cumulant was coined by Fisher (1929) on account of their behaviour under addition of random variables. Let $S = X + Y$ be the sum of two independent random variables. The moment generating function of the sum is the product

$$M_S(\xi) = M_X(\xi)M_Y(\xi),$$

and the cumulant generating function is the sum

$$K_S(\xi) = K_X(\xi) + K_Y(\xi).$$

Consequently, the rth cumulant of the sum is the sum of the rth cumulants. By extension, if $X_1, \ldots X_n$ are independent and identically distributed, the rth cumulant of the sum is $n\kappa_r$, and the rth cumulant of the standardized sum $n^{-1/2}(X_1 + \cdots + X_n)$ is $n^{1-r/2}\kappa_r$. Provided that the cumulants are finite, all cumulants of order $r \geq 3$ of the standardized sum tend to zero, which is a simple demonstration of the central limit theorem.

Good (195?) obtained an expression for the rth cumulant of X as the rth moment of the discrete Fourier transform of an independent and identically distributed sequence as follows. Let $X_1, X_2, \ldots$ be independent copies of X with rth cumulant κ_r, and let $\omega = e^{2\pi i/n}$ be a primitive nth root of unity. The discrete Fourier combination

$$Z = X_1 + \omega X_2 + \cdots + \omega^{n-1}X_n$$

is a complex-valued random variable whose distribution is invariant under rotation $Z \sim \omega Z$ through multiples of $2\pi/n$. The rth cumulant of the sum is $\kappa_r \sum_{j=1}^{n} \omega^{rj}$, which is equal to $n\kappa_r$ if r is a multiple of n, and zero otherwise. Consequently $E(Z^r) = 0$ for integer $r < n$ and $E(Z^n) = n\kappa_n$.

1.2 Multivariate cumulants

Somewhat surprisingly, the relation between moments and cumulants is simpler and more transparent in the multivariate case than in the univariate case. Let $X = (X^1, \ldots, X^k)$ be the components of a random vector. In a departure from the univariate notation, we write $\kappa^r = E(X^r)$ for the components of the mean vector, $\kappa^{rs} = E(X^r X^s)$ for the components of the second moment matrix, $\kappa^{rst} = E(X^r X^s X^t)$ for the third moments, and so on. It is convenient notationally to adopt Einstein's summation convention, so $\xi_r X^r$ denotes the linear combination $\xi_1 X^1 + \cdots + \xi_k X^k$, the square of the linear combination is $(\xi_r X^r)^2 = \xi_r \xi_s X^r X^s$ a sum of k^2 terms, and so on

3

for higher powers. The Taylor expansion of the moment generating function $M(\xi) = E(\exp(\xi_r X^r)$ is

$$M(\xi) = 1 + \xi_r \kappa^r + \tfrac{1}{2!}\xi_r\xi_s\kappa^{rs} + \tfrac{1}{3!}\xi_r\xi_s\xi_t\kappa^{rst} + \cdots.$$

The cumulants are defined as the coefficients $\kappa^{r,s}, \kappa^{r,s,t}, \ldots$ in the Taylor expansion

$$\log M(\xi) = \xi_r \kappa^r + \tfrac{1}{2!}\xi_r\xi_s\kappa^{r,s} + \tfrac{1}{3!}\xi_r\xi_s\xi_t\kappa^{r,s,t} + \cdots.$$

This notation does not distinguish first-order moments from first-order cumulants, but commas separating the superscripts serve to distinguish higher-order cumulants from moments.

Comparison of coefficients reveals that the each moment $\kappa^{rs}, \kappa^{rst}, \ldots$ is a sum over partitions of the superscripts, each term in the sum being a product of cumulants:

$$\kappa^{rs} = \kappa^{r,s} + \kappa^r \kappa^s$$
$$\kappa^{rst} = \kappa^{r,s,t} + \kappa^{r,s}\kappa^t + \kappa^{r,t}\kappa^s + \kappa^{s,t}\kappa^r + \kappa^r\kappa^s\kappa^t$$
$$= \kappa^{r,s,t} + \kappa^{r,s}\kappa^t[3] + \kappa^r\kappa^s\kappa^t$$
$$\kappa^{rstu} = \kappa^{r,s,t,u} + \kappa^{r,s,t}\kappa^u[4] + \kappa^{r,s}\kappa^{t,u}[3] + \kappa^{r,s}\kappa^t\kappa^u[6] + \kappa^r\kappa^s\kappa^t\kappa^u.$$

Each parenthetical number indicates a sum over distinct partitions having the same block sizes, so the fourth-order moment is a sum of 15 distinct cumulant products. In the reverse direction, each cumulant is also a sum over partitions of the indices. Each term in the sum is a product of moments, but with coefficient $(-1)^{\nu-1}(\nu-1)!$ where ν is the number of blocks:

$$\kappa^{r,s} = \kappa^{rs} - \kappa^r\kappa^s$$
$$\kappa^{r,s,t} = \kappa^{rst} - \kappa^{rs}\kappa^t[3] + 2\kappa^r\kappa^s\kappa^t$$
$$\kappa^{r,s,t,u} = \kappa^{rstu} - \kappa^{rst}\kappa^u[4] - \kappa^{rs}\kappa^{tu}[3] + 2\kappa^{rs}\kappa^t\kappa^u[6] - 6\kappa^r\kappa^s\kappa^t\kappa^u$$

Partition notation serves one additional purpose. It establishes moments and cumulants as special cases of generalized cumulants, which includes objects of the type $\kappa^{r,st} = \mathrm{cov}(X^r, X^s X^t)$, $\kappa^{rs,tu} = \mathrm{cov}(X^r X^s, X^t X^u)$, and $\kappa^{rs,t,u}$ with incompletely partitioned indices. These objects arise very naturally in statistical work involving asymptotic approximation of distributions. They are intermediate between moments and cumulants, and have characteristics of both.

Every generalized cumulant can be expressed as a sum of certain products of ordinary cumulants. Some examples are as follows:

$$\kappa^{rs,t} = \kappa^{r,s,t} + \kappa^r \kappa^{s,t} + \kappa^s \kappa^{r,t}$$
$$= \kappa^{r,s,t} + \kappa^r \kappa^{s,t}[2]$$
$$\kappa^{rs,tu} = \kappa^{r,s,t,u} + \kappa^{r,s,t} \kappa^u[4] + \kappa^{r,t} \kappa^{s,u}[2] + \kappa^{r,t} \kappa^s \kappa^u[4]$$
$$\kappa^{rs,t,u} = \kappa^{r,s,t,u} + \kappa^{r,t,u} \kappa^s[2] + \kappa^{r,t} \kappa^{s,u}[2]$$

Each generalized cumulant is associated with a partition τ of the given set of indices. For example, $\kappa^{rs,t,u}$ is associated with the partition $\tau = rs|t|u$ of four indices into three blocks. Each term on the right is a cumulant product associated with a partition σ of the same indices. The coefficient is one if the least upper bound $\sigma \vee \tau$ has a single block, otherwise zero. Thus, with $\tau = rs|t|u$, the product $\kappa^{r,s} \kappa^{t,u}$ does not appear on the right because $\sigma \vee \tau = rs|tu$ has two blocks.

As an example of the way these formulae may be used, let X be a scalar random variable with cumulants $\kappa_1, \kappa_2, \kappa_3, \ldots$. By translating the second formula in the preceding list, we find that the variance of the squared variable is

$$\mathrm{var}(X^2) = \kappa_4 + 4\kappa_3\kappa_1 + 2\kappa_2^2 + 4\kappa_2\kappa_1^2,$$

reducing to $\kappa_4 + 2\kappa_2^2$ if the mean is zero.

1.3 Exponential families

2 Approximation of distributions

2.1 Edgeworth approximation

2.2 Saddlepoint approximation

3 Samples and sub-samples

A function $f\colon \mathcal{R}^n \to \mathcal{R}$ is symmetric if $f(x_1, \ldots, x_n) = f(x_{\pi(1)}, \ldots, x_{\pi(n)})$ for each permutation π of the arguments. For example, the total $T_n = x_1 + \cdots + x_n$, the average T_n/n, the min, max and median are symmetric functions, as are the sum of squares $S_n = \sum x_i^2$, the sample variance $s_n^2 = (S_n - T_n^2/n)/(n-1)$ and the mean absolute deviation $\sum |x_i - x_j|/(n(n-1))$.

A vector x in $\mathcal{R}^n$ is an ordered list of n real numbers $(x_1, \ldots, x_n)$ or a function $x\colon [n] \to \mathcal{R}$ where $[n] = \{1, \ldots, n\}$. For $m \le n$, a 1–1 function $\varphi\colon [m] \to [n]$ is a sample of size m, the sampled values being

$x\varphi = (x_{\varphi(1)}, \ldots, x_{\varphi(m)})$. All told, there are $n(n-1)\cdots(n-m+1)$ distinct samples of size m that can be taken from a list of length n. A *sequence* of functions $f_n \colon \mathcal{R}^n \to \mathcal{R}$ is consistent under subsampling if, for each f_m, f_n,

$$f_n(x) = \text{ave}_\varphi\, f_m(x\varphi),$$

where ave_φ denotes the average over samples of size m. For $m = n$, this condition implies only that f_n is a symmetric function.

Although the total and the median are both symmetric functions, neither is consistent under subsampling. For example, the median of the numbers $(0, 1, 3)$ is one, but the average of the medians of samples of size two is $4/3$. However, the average $\bar{x}_n = T_n/n$ is sampling consistent. Likewise the sample variance $s_n^2 = \sum(x_i - \bar{x})^2/(n-1)$ with divisor $n-1$ is sampling consistent, but the mean squared deviation $\sum(x_i - \bar{x}_n)^2/n$ with divisor n is not. Other sampling consistent functions include Fisher's k-statistics, the first few of which are $k_{1,n} = \bar{x}_n$, $k_{2,n} = s_n^2$ for $n \geq 2$,

$$k_{3,n} = n \sum (x_i - \bar{x}_n)^3/((n-1)(n-2))$$
$$k_{4,n} =$$

defined for $n \geq 3$ and $n \geq 4$ respectively.

For a sequence of independent and identically distributed random variables, the k-statistic of order $r \leq n$ is the unique symmetric function such that $E(k_{r,n}) = \kappa_r$. Fisher (1929) derived the variances and covariances. The connection with finite-population sub-sampling was developed by Tukey (1954).