Summary of basic probability theory, part 2

D. Joyce, Clark University

Math 218, Mathematical Statistics, Jan 2008

Expectation. The *expected value* $E(X)$, also called the *expectation* or *mean* μ_X, of a random variable X is defined differently for the discrete and continuous cases.

For a discrete random variable, it is a weighted average defined in terms of the probability mass function f as

$$E(X) = \mu_X = \sum_x x f(x).$$

For a continuous random variable, it is defined in terms of the probability density function f as

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

There is a physical interpretation where this mean is interpreted as a center of gravity.

Expectation is a linear operator. That means that the expectation of a sum or difference is the difference of the expectations

$$E(X + Y) = E(X) + E(Y),$$

and that's true whether or not X and Y are independent, and also

$$E(cX) = c\, E(X)$$

where c is any constant. From these two properties it follows that

$$E(X - Y) = E(X) - E(Y),$$

and, more generally, expectation preserves linear combinations

$$E(\sum_{i=1}^{n} c_i X_i) = \sum_{i=1}^{n} c_i E(X_i).$$

Furthermore, when X and Y are independent, then $E(XY) = E(X)\, E(Y)$, but that equation doesn't usually hold when X and Y are not independent.

Variance and standard deviation. The *variance* of a random variable X is defined as

$$\mathrm{Var}(X) = \sigma_X^2 = E((X - \mu_X)^2) = E(X^2) - \mu_X^2$$

where the last equality is provable. Standard deviation, σ, is defined as the square root of the variance.

Here are a couple of properties of variance. First, if you multiply a random variable X by a constant c to get cX, the variance changes by a factor of the square of c, that is

$$\mathrm{Var}(cX) = c^2 \mathrm{Var}(X).$$

That's the main reason why we take the square root of variance to normalize it—the standard deviation of cX is c times the standard deviation of X. Also, variance is translation invariant, that is, if you add a constant to a random variable, the variance doesn't change:

$$\mathrm{Var}(X + c) = \mathrm{Var}(X).$$

In general, the variance of the sum of two random variables is *not* the sum of the variances of the two random variables. But it is when the two random variables are independent.

Moments, central moments, skewness, and kurtosis. The k^{th} *moment* of a random variable X is defined as $\mu_k = E(X^k)$. Thus, the mean is the first moment, $\mu = \mu_1$, and the variance can be found from the first and second moments, $\sigma^2 = \mu_2 - \mu_1^2$.

The k^{th} *central moment* is defined as $E((X-\mu)^k$. Thus, the variance is the second central moment.

A third central moment of the standardized random variable $X^* = (X-\mu)/\sigma$,

$$\beta_3 = E((X^*)^3) = \frac{E((X-\mu)^3)}{\sigma^3}$$

is called the *skewness* of X. A distribution that's symmetric about its mean has 0 skewness. (In fact all the odd central moments are 0 for a symmetric distribution.) But if it has a long tail to the right and a short one to the left, then it has a positive skewness, and a negative skewness in the opposite situation.

A fourth central moment of X^*,

$$\beta_4 = E((X^*)^4) = \frac{E((X-\mu)^4)}{\sigma^4}$$

is callled *kurtosis*. A fairly flat distribution with long tails has a high kurtosis, while a short tailed distribution has a low kurtosis. A bimodal distribution has a very high kurtosis. A normal distribution has a kurtosis of 3. (The word kurtosis was made up in the early 19th century from the Greek word for curvature.)

The moment generating function. There is a clever way of organizing all the moments into one mathematical object, and that object is called the *moment generating function*. It's a function $m(t)$ of a new variable t defined by

$$m(t) = E(e^{tX}).$$

Since the exponential function e^t has the power series

$$e^t = \sum_{k=0}^{\infty} \frac{t^k}{k!} = 1 + t + \frac{t^2}{2!} + \cdots + \frac{t^k}{k!} + \cdots,$$

we can rewrite $m(t)$ as follows

$$m(t) = E(e^{tX})1 + \mu_1 t + \frac{\mu_2}{2!}\,t^2 + \cdots + \frac{\mu_k}{k!}\,t^k + \cdots.$$

That implies that $m^{(k)}(0)$, the k^{th} derivative of $m(t)$ evaluated at $t=0$, equals the k^{th} moment μ_k of X. In other words, the moment generating function generates the moments of X by differentiation.

For discrete distributions, we can also compute the moment generating function directly in terms of the probability mass function $f(x) = P(X{=}x)$ as

$$m(t) = E(e^{tX}) = \sum_x e^{tx} p(x).$$

For continuous distributions, the moment generating function can be expressed in terms of the probability density function f as

$$m(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx.$$

The moment generating function enjoys the following properties.

Translation. If $Y = X + a$, then

$$m_Y(t) = e^{ta} m_X(t).$$

Scaling. If $Y = bx$, then

$$m_Y(t) = m_X(bt).$$

Standardizing. From the last two properties, if

$$X^* = \frac{X-\mu}{\sigma}$$

is the standardized random variable for X, then

$$m_{X^*}(t) = e^{-\mu t/\sigma} m_X(t/\sigma).$$

Convolution. If X and Y are independent variables, and $Z = X + Y$, then

$$m_Z(t) = m_X(t)\, m_Y(t).$$

The primary use of moment generating functions is to develop the theory of probability. For instance, the easiest way to prove the central limit theorem is to use moment generating functions.

The median, quartiles, quantiles, and percentiles. The *median* of a distribution X, sometimes denoted $\tilde{\mu}$, is the value such that $P(X \leq \tilde{\mu}) =$

$\frac{1}{2}$. Whereas some distributions, like the Cauchy distribution, don't have means, all continuous distributions have medians.

If p is a number between 0 and 1, then the p^{th} *quantile* is defined to be the number θ_p such that

$$P(X \leq \theta_p) = F(\theta_p) = p.$$

Quantiles are often expressed as percentiles where the p^{th} quantile is also called the $100p^{\text{th}}$ *percentile.* Thus, the median is the 0.5 quantile, also called the 50^{th} percentile.

The *first quartile* is another name for $\theta_{0.25}$, the 25^{th} percentile, while the *third quartile* is another name for $\theta_{0.75}$, the 75^{th} percentile