18.443, revised Aug. 28, 2014

Gamma and beta probabilities

The *gamma function* is defined for any $a > 0$ by

$$\Gamma(a) \ := \ \int_0^\infty x^{a-1}e^{-x}dx. \tag{1}$$

The integral is finite if (and only if) $a > 0$, because $\int_0^1 x^{a-1}dx = 1/a < \infty$, and $x^{a-1} < e^{x/2}$ for x large enough.

Integration by parts shows that $\Gamma(a+1) = a\Gamma(a)$ for any $a > 0$. We have $\Gamma(1) = 1$. It follows by induction that $\Gamma(k+1) = k!$ for any nonnegative integer k.

For any $a > 0$ the function defined by

$$\gamma_a(x) \ := \ x^{a-1}e^{-x}/\Gamma(a) \tag{2}$$

for $x > 0$, and 0 for $x \leq 0$, is a probability density. The corresponding distribution is called a *gamma distribution with parameter* a.

If the random variable X has a gamma distribution with parameter a then $EX = a$ since $EX = \Gamma(a+1)/\Gamma(a)$. Likewise $EX^2 = \Gamma(a+2)/\Gamma(a) = (a+1)a$ so $\text{Var}(X) = a$ and $\sigma_X = a^{1/2}$.

Recall that for any random variable X with density f and any $c > 0$, cX has a density $c^{-1}f(x/c)$. Applying that to $c = 1/\lambda$ for any $\lambda > 0$, if X has density γ_a then X/λ has the density $\gamma_{a,\lambda}$ defined by

$$\gamma_{a,\lambda}(x) \ = \ \lambda^a x^{a-1}e^{-\lambda x}/\Gamma(a)$$

for $0 < x < +\infty$ and 0 otherwise. A random variable Y with this density will be said to have a gamma(a, λ) distribution. It is easily seen and known to have $EY = a/\lambda$ and $\text{Var}(Y) = a/\lambda^2$.

The *Beta function* is defined for any $a > 0$ and $b > 0$ by

$$B(a,b) \ := \ \int_0^1 x^{a-1}(1-x)^{b-1}dx. \tag{3}$$

Clearly, $0 < B(a,b) < \infty$ for any $a > 0$ and $b > 0$. Letting $y \ := \ 1-x$ shows that $B(b,a) \equiv B(a,b)$. Let $\beta_{a,b}(x) := x^{a-1}(1-x)^{b-1}/B(a,b)$ for $0 < x < 1$ and 0 for $x \leq 0$ or $x \geq 1$. Then $\beta_{a,b}$ is a probability density. The probability distribution with this density is called a *beta distribution with parameters* a, b, or beta(a, b). Its distribution function is then defined as

$$I_x(a,b) \ := \ \int_0^x \beta_{a,b}(t)dt, \quad 0 \le x \le 1. \tag{4}$$

The following fact relates gamma distributions with different parameters with each other and relates gamma and beta functions.

Theorem 1 *For any $a > 0$ and $b > 0$,*
(a) $B(a,b) \equiv B(b,a) \equiv \Gamma(a)\Gamma(b)/\Gamma(a+b)$.
(b) If X and Y are independent random variables having gamma(a,λ) and $\gamma(b,\lambda)$ distributions respectively, for the same $\lambda > 0$, then $U := X + Y$ has a gamma$(a+b,\lambda)$ distribution.

Proof. First consider (b) and suppose $\lambda = 1$. U has a density u given by a convolution of those of X and Y, namely, for any $x > 0$,

$$u(x) \ = \ \int_0^x \gamma_a(x-y)\gamma_b(y)dy$$

$$= \ \int_0^x (x-y)^{a-1}e^{-(x-y)}y^{b-1}e^{-y}dy/(\Gamma(a)\Gamma(b))$$

$$= \ e^{-x}\int_0^x (x-y)^{a-1}y^{b-1}dy/(\Gamma(a)\Gamma(b)).$$

The substitution $y = tx, \quad 0 \le t \le 1$ gives

$$= e^{-x}x^{a+b-1}B(b,a)/(\Gamma(a)\Gamma(b)).$$

Since u must be a probability density, it must be the gamma$(a+b,1)$ density as desired, and the normalizing constants must agree, so (a) follows. To get (b) for a general $\lambda > 0$, just consider X/λ and Y/λ. $\square$

Iterating Theorem 1, it follows that if X_i are independent identically distributed variables, each having the standard exponential distribution with density e^{-x} for $x \ge 0$ and 0 for $x < 0$, so that the X_i have gamma distributions with parameter 1, then for each $n = 1, 2, ...$, $S_n = X_1 + \cdots + X_n$ has a γ_n density. If each X_i has a $\gamma_{a,\lambda}$ density then S_n has a $\gamma_{na,\lambda}$ density.

It is now easy to find the means and variances of beta distributions. If X has a beta distribution with parameters a, b, in other words has distribution function (4), then $EX = B(a+1,b)/B(a,b)$. Similarly $EX^2 =$

$B(a+2,b)/B(a,b) = a(a+1)/[(a+b)(a+b+1)]$. Thus

$$EX = a/(a+b), \quad \mathrm{Var}(X) \;=\; \frac{ab}{(a+b)^2(a+b+1)}. \tag{5}$$

Note that $1-X$ has a beta distribution with parameters b, a. Thus $E(1-X) = b/(a+b)$ which equals $1 - a/(a+b)$ as it should. Also, $1-X$ has the same variance as X, and so the expression for $\mathrm{Var}(X)$ is preserved by interchanging a and b as it should be.

Let $0 < \lambda < \infty$ and let Y be a Poisson random variable with parameter λ. Then some notations are, for any integer $k \geq 0$,

$$P(k,\lambda) \;=\; \Pr(Y \leq k) \;=\; e^{-\lambda}\sum_{j=0}^{k} \lambda^j/j!,$$

$$Q(k,\lambda) \;=\; \Pr(Y \geq k) \;=\; e^{-\lambda}\sum_{j=k}^{\infty} \lambda^j/j!.$$

There are identities relating the Poisson and gamma distributions:

Theorem 2 *For any positive integer k, if X has a γ_k density, we have for any $x \geq 0$,*

$$Q(k,x) = P(X \leq x) \tag{6}$$

and

$$P(k-1,x) = P(X > x). \tag{7}$$

For $0 < \lambda < \infty$, if Y has a $\gamma_{k,\lambda}$ density and $0 < t < \infty$, then

$$P(Y \leq t) = Q(k, \lambda t) \tag{8}$$

and

$$P(Y > t) = P(k-1, \lambda t). \tag{9}$$

Proof. To prove equation (7), differentiate with respect to x and note that the derivative of $P(k-1,x)$ is

$$-e^{-x} + e^{-x} - xe^{-x} + \frac{2}{2!}xe^{-x} - \cdots - \frac{x^{k-1}}{(k-1)!} = -\frac{x^{k-1}}{(k-1)!} = -\gamma_k(x),$$

a telescoping sum. Both sides of (7) equal 1 when $x = 0$, so (7) follows. Equation (6) follows by taking complements.

Then letting $Y = X/\lambda$, Y has the given density, (9) follows from (7), and (8) follows by taking complements or from (6). $\square$

A similar identity relates beta and binomial probabilities. Let $0 < p < 1$, $q = 1 - p$, let X be a binomial (n, p) random variable and

$$B(k, n, p) = \Pr(X \le k) = \sum_{j=0}^{k} b(j, n, p),$$

$$E(k, n, p) = \Pr(X \ge k) = \sum_{j=k}^{n} b(j, n, p).$$

Theorem 3 *If $0 < p < 1$, and $0 \le k \le n$ are integers, then*

$$E(k, n, p) = I_p(k, n - k + 1), \quad \text{if } k \ge 1;$$

$$B(k, n, p) = I_{1-p}(n - k, k + 1), \quad \text{if } k < n.$$

Proof. The first equality again follows from differentiating a finite sum with respect to p which gives a telescoping sum. The second then follows from $B(k, n, p) \equiv E(n - k, n, 1 - p)$.

$\square$

A $\chi^2(d)$ distribution, or χ^2 distribution with d degrees of freedom, is defined as the distribution of $Z_1^2 + \cdots + Z_d^2$ where $Z_1, Z_2, ..., Z_d$ are i.i.d. $N(0, 1)$. The following known fact will be proved:

Theorem 4 *For any positive integer d, $\chi^2(d)$ has a $\gamma(d/2, 1/2)$ distribution.*

Proof. First let $d = 1$. Let Z have $N(0, 1)$ distribution. Then for any $t \ge 0$,

$$\Pr(Z^2 \le t) = \Pr(|Z| \le \sqrt{t}) = \Phi(\sqrt{t}) - \Phi(-\sqrt{t})$$

where Φ is the standard normal distribution function. Thus by the chain rule the density of $\chi^2(1) = Z^2$ is

$$2\phi(\sqrt{t}) \cdot (1/(2t^{1/2}) = (2\pi t)^{-1/2} e^{-t/2}$$

which is the $\gamma(1/2, 1/2)$ density, since $\Gamma(1/2) = \sqrt{\pi}$ (if one did not know that, it would follow by unique normalization of probability densities), proving the statement for $d = 1$. The statement for a general positive integer d then follows by Theorem 1(b) for $\lambda = 1/2$ and induction on d. $\square$