HG

1 October 2017

The moment generating function (mgf) of the gamma (α, λ) distribution and its application

(Supplementary note to the lecture Monday 25 Sept. on mgf's)

I did not have time to go through the mgf of the gamma distribution on the lecture and supply the mgf here due to the importance of the gamma distribution.

Suppose $X \sim \Gamma(\alpha, \lambda)$ distributed with pdf

$$f(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

where $\alpha > 0, \quad \lambda > 0$ are parameters.

(1) The mgf of *X* is given by $M(t) = E\left(e^{tX}\right) = \left(\dfrac{\lambda}{\lambda - t}\right)^\alpha$, well defined for all $t < \lambda$.

Since $\lambda > 0$, the mgf is well defined in an open interval about 0, which implies that all moments of *X* exist (a fact we have proven before more directly).

Proof: If $t < \lambda$, we get

$$M(t) = E\left(e^{tX}\right) = \int_{-\infty}^{\infty} e^{tx} f(x) dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{(t-\lambda)x} dx =$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\lambda - t)^\alpha} \int_0^\infty \frac{(\lambda - t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda - t)x} dx \overset{\text{Integral = 1}}{=} \frac{\lambda^\alpha}{(\lambda - t)^\alpha} = \left(\frac{\lambda}{\lambda - t}\right)^\alpha$$

(Note that the integrand of the last integral is the pdf of the $\Gamma(\alpha, \ \lambda - t)$ distribution, which implies that the integral is 1.) (**End of proof**)

Having the mgf tool, we may derive a number of useful properties for the gamma distributions:

(2) If $X \sim \Gamma(\alpha, \lambda)$, then $Y = \lambda X \sim \Gamma(\alpha, 1)$

Proof: We have seen in general that if *Y* is a linear transformation, $Y = a + bX$ (*a,b* constants), the mgf of *Y* is $M_Y(t) = e^{at} M_X(bt)$. Using this on $Y = \lambda X$, we get

$M_Y(t) = e^0 M_X(\lambda t) = M_X(\lambda t) = \left(\frac{\lambda}{\lambda - \lambda t}\right)^\alpha = \left(\frac{1}{1-t}\right)^\alpha$, i.e., the mgf of $\Gamma(\alpha,1)$.Since the mgf of any distribution is unique, it follows that $Y \sim \Gamma(\alpha,1)$. **EOP**.

(**Note** that this result may be proven directly using cdf's.)

In general, sums of independent rv's ($Y = X_1 + X_2 + \cdots + X_n$) have very complicated distributions. Therefore, the following results sometimes turn out as useful.

(3)	If $X_1, X_2, \ldots, X_n$ are independent and gamma distributed with the same scale (λ), i.e., $X_i \sim \Gamma(\alpha_i, \lambda),\ i = 1,2,\ldots,n$, then $Y = X_1 + X_2 + \cdots + X_n \sim \Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_n, \lambda)$

Proof: A general mgf property is that, if X_1, X_2 are independent with mgf's $M_1(t), M_2(t)$ respectively, the mgf of $Y = X_1 + X_2$ is $M_Y(t) = M_1(t) M_2(t)$. Hence, if $X_i \sim \Gamma(\alpha_i, \lambda),\ i = 1,2$, then $M_Y(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1} \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_2} = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1+\alpha_2}$, i.e., the mgf of $\Gamma(\alpha_1 + \alpha_2, \lambda)$. The uniqueness of mgf's then implies that $Y = X_1 + X_2 \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$. Hence, the result is proven for $n = 2$. For $n = 3$ we have $Y = X_1 + X_2 + X_3 = U + X_3$, where U and X_3 are independent and gamma distributed, implying as above that $Y \sim \Gamma(\alpha_1 + \alpha_2 + \alpha_3, \lambda)$. Having thus proven the result for $n = 3$, it follows in the same way that it is valid for $n = 4$ and then for $n = 5$, and so on, step by step, for all natural numbers, *n*.

(By the way, it may be worth mentioning that this way of reasoning is called *induction proof* and is often used in mathematics.) **EOP**

Note. The assumption that the X_i's all have the same scale, is essential. If $X_i \sim \Gamma(\alpha_i, \lambda_i),\ i = 1,2,$ where $\lambda_1 \neq \lambda_2$, we get $M_Y(t) = \left(\frac{\lambda_1}{\lambda_1 - t}\right)^{\alpha_1} \left(\frac{\lambda_2}{\lambda_2 - t}\right)^{\alpha_2}$, that is a mgf for some distribution that *is not* a gamma distribution (since it cannot be written in the form of (1)). On the other hand, explain yourself why $Y = \lambda_1 X_1 + \lambda_2 X_2$ is gamma distributed (which one?). (Hint: use (2) and (3).)

Special case I. If *X* is *exponentially distributed* with parameter $\lambda > 0$, $(X \sim \exp(\lambda))$, we know that $X \sim \Gamma(1, \lambda)$ distributed. The mgf of *X* then follows from (1):

$$M_X(t) = \frac{\lambda}{\lambda - t}$$, well defined for $t < \lambda$.

We also get from (3) that if $X_1, X_2, \ldots, X_n$ are *iid* and exponential (λ), then
$Y = X_1 + X_2 + \cdots + X_n \sim \Gamma(n, \lambda)$.

Special case II. ***Chi-square distributions*.**

(See also supplementary exercise 4 on the net.) The chi-square distributions turn out to be important as approximate or exact inference distributions for many test-problems involving several parameters tested jointly (e.g., Pearson's chi-square test and likelihood ratio testing - that both will be discussed later - as well as various regression problems).

DEF. If $Z \sim \Gamma\left(\frac{d}{2}, \frac{1}{2}\right)$, where d is a natural number (1,2,3,…), we say that Z is *chi square distributed with d degrees of freedom* (written shortly: $Z \sim \chi^2_d$)

Some properties:

(4) If $Z \sim \chi^2_d$, $E(Z) = d$, $\text{var}(Z) = 2d$

Proof. If $X \sim \Gamma(\alpha, \lambda)$, we have from before that $E(X) = \frac{\alpha}{\lambda}$, $\text{var}(X) = \frac{\alpha}{\lambda^2}$.

If $Z \sim \chi^2_d = \Gamma\left(\frac{d}{2}, \frac{1}{2}\right)$, we get $E(Z) = \frac{d/2}{1/2} = d$ and $\text{var}(Z) = \frac{d/2}{1/4} = 2d$

EOP

From (1) we get immediately

(5) If $Z \sim \chi^2_d$, the mgf is $M_Z(t) = \left(\frac{1/2}{1/2 - t}\right)^{d/2} = \frac{1}{(1-2t)^{d/2}} = (1-2t)^{-d/2}$, for $t < 1/2$

From (3) we get immediately

(6) If $Z_1, Z_2, \ldots, Z_n$ are independent and chi-square distributed, i.e., $Z_i \sim \chi^2_{d_i}$, $i = 1, 2, \ldots, n$, then $Y = Z_1 + Z_2 + \cdots + Z_n$ is chi square distributed with $d = d_1 + d_2 + \cdots + d_n$ degrees of freedom (i.e., $\Gamma\left(\frac{d}{2}, \frac{1}{2}\right)$ distributed).

Note. The term "degrees of freedom" comes from applications when one wants to estimate a linear model with *r* unknown parameters based on *n* observations. The *r* parameters often imply *r* restrictions on the observations to estimate, leaving $d = n - r$ observations (degrees of freedom) for the estimation of variances.

(7) If *X* is standard normally distributed, i.e., $X \sim N(0,1)$, then $Y = X^2$ is chi-square distributed with 1 degree of freedom, i.e., $Y \sim \chi_1^2$.

(This is a classical result that Rice proves in the basic way using cdf's. As an illustration we may also prove it using mgf's instead:)

Proof. Let $X \sim N(0,1)$ and $Y = X^2$. The pdf of *X* is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty$$

The mgf of $Y = X^2$ then becomes

$$M_Y(t) = E\left(e^{tX^2}\right) = \int_{-\infty}^{\infty} e^{tx^2} f(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{(t-\frac{1}{2})x^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-2t)^{-1}}x^2} dx$$

Multiplying and dividing by $(1-2t)^{-1/2}$, we obtain the total integral of a normal pdf, of $N\left[0,\ (1-2t)^{-1}\right]$, which is equal to 1. So

$$M_Y(t) = (1-2t)^{-1/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,(1-2t)^{-1/2}} e^{-\frac{1}{2(1-2t)^{-1}}x^2} dx = (1-2t)^{-1/2} \quad \text{for } t < 1/2,$$

which is the mgf (see (5)) of the χ_1^2 distribution. The uniqueness of the mgf then implies that $Y \sim \chi_1^2$. **EOP**

Using (6) and (7) we obtain immediately an important technical result that often underlies various distribution theorems relevant for handling statistically linear models with unknown parameters to be estimated.

(**As an example**, Rice uses it as part of the proof (see the first sentence) of theorem B of section 6.3 (optional reading), where he proves that if $X_1, X_2, \ldots, X_n$ are iid and normal, $X_i \sim N(\mu, \sigma^2)$, then

$$\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{\sigma^2} \sim \chi_{n-1}^2$$)

(8) If $X_1, X_2, \ldots, X_n$ are independent and standard normally distributed, i.e., $X_i \sim N(0,1),\ i = 1,2,\ldots,n$, then $Y = X_1^2 + X_2^2 + \cdots + X_n^2 \sim$ chi squared distributed with *n* degrees of freedom.