Moment Generating Function

Statistics 110

Summer 2006

Moments Revisited

So far I've really only talked about the first two moments. Lets define what is meant by moments more precisely.

Definition. *The r^{th} moment of a random variable X is $E[X^r]$, assuming that the expectation exists.*

So the mean of a distribution is its first moment.

Definition. *The r central moment of a random variable X is $E[(X - E[X])^r]$, assuming that the expectation exists.*

Thus the variance is the 2nd central moment of distribution.

The 1st central moment usually isn't discussed as its always 0.

The 3rd central moment is known as the **skewness** of a distribution and is used as a measure of asymmetry.

If a distribution is symmetric about its mean ($f(\mu - x) = f(\mu + x)$), the skewness will be 0. Similarly if the skewness is non-zero, the distribution is asymmetric. However it is possible to have asymmetric distribution with skewness $= 0$.

Examples of symmetric distribution are normals, $Beta(a, a)$, $Bin(n, p = 0.5)$. Example of asymmetric distributions are

Distribution	Skewness
$Bin(n, p)$	$np(1 - p)(1 - 2p)$
$Pois(\lambda)$	λ
$Exp(\lambda)$	$\frac{2}{\lambda}$
$Beta(a, b)$	Ugly formula

The 4th central moment is known as the kurtosis. It can be used as a measure of how heavy the tails are for a distribution. The kurtosis for a normal is $3\sigma^4$.

Note that these measures are often standardized as in their raw form they depend on the standard deviation.

Theorem. *If the rth moment of a RV is exists, then the sth moment exists for all $s < r$. Also the sth central moment exists for all $s \leq r$.*

So you can't have a distribution that has a finite mean, an infinite variance, and a finite skewness.

Proof. Postponed till later. $\square$

Why are moments useful? They can be involved in calculating means and variances of transformed RVs or other summaries of RVs.

Example: What are the mean and variance of $A = \pi R^2$

$$E[A] = \pi E[R^2]$$
$$\mathrm{Var}(A) = \pi^2 \mathrm{Var}(R^2) = \pi^2 \left(E[R^4] - (E[R^2])^2 \right)$$

So we need $E[R^4]$ in addition to $E[R]$ and $E[R^2]$.

Example: What is the skewness of X?

$$E[(X-\mu)^3] = E[X^3 - 3\mu X^2 + 3\mu^2 X - \mu^3] = E[X^3] - 3\mu E[X^2] + 2\mu^3$$

so $E[X], E[X^2]$, and $E[X^3]$ are needed to calculate the skewness.

Moment Generating Function

Definition. *The* **Moment Generating Function** *(MGF) of a random variable X, is $M_X(t) = E[e^{tX}]$ if the expectation is defined.*

$$M_X(t) = \sum_x e^{tx} p_X(x) \qquad \textit{(Discrete)}$$

$$M_X(t) = \int_{\mathcal{X}} e^{tx} f_X(x) dx \qquad \textit{(Continuous)}$$

Whether the MGF is defined depends on the distribution and the choice of t. For example, the $M_X(t)$ is defined for all t if X is normal, defined for no t if X is Cauchy, and for $t < \lambda$ if $X \sim Exp(\lambda)$.

For those that have done some analysis, for the continuous case, the moment generating function is related to the Laplace transform of the density function. Many of the results about it come from that theory.

Why should we care about the MGF?

- To calculate moments. It may be easier to work with the MGF than to directly calculate $E[X^r]$.
- To determine distributions of functions of random variables.
- Related to this, approximating distributions. For example can use it to show that as n increases, the $Bin(n, p)$ “approaches” a normal distribution.

The following theorems justify these uses of the MGF.

Theorem. *If $M_X(t)$ of a RV X is finite in an open iterval containing 0, then it has derivatives of all orders and*

$$M_X^{(r)}(t) = E[X^r e^{tX}]$$
$$M_X^{(r)}(0) = E[X^r]$$

Proof.

$$\begin{aligned} M_X^{(1)}(t) &= \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f_X(x)dx \\ &= \int_{-\infty}^{\infty} \left(\frac{d}{dt}e^{tx}\right) f_X(x)dx \\ &= \int_{-\infty}^{\infty} x e^{tx} f_X(x)dx \\ &= E[Xe^{tX}] \end{aligned}$$

$$
\begin{aligned}
M_X^{(2)}(t) &= \frac{d}{dt} M_X^{(1)} \\
&= \int_{-\infty}^{\infty} x \left(\frac{d}{dt} e^{tx} \right) f_X(x) dx \\
&= \int_{-\infty}^{\infty} x^2 e^{tx} f_X(x) dx = E[X^2 e^{tX}]
\end{aligned}
$$

The rest can be shown by induction. The second part of the theorem follows from $e^0 = 1$. $\square$

Another way to see this result is due to the Taylor series expansion of

$$
e^y = 1 + y + \frac{y^2}{2!} + \frac{y^3}{3!} + \ldots,
$$

which gives

$$
\begin{aligned}
M_X(t) &= E\left[1 + Xt + \frac{X^2t^2}{2!} + \frac{X^3t^3}{3!} + \ldots\right] \\
&= 1 + E[X]t + E[X^2]\frac{t^2}{2!} + E[X^3]\frac{t^3}{3!} + \ldots
\end{aligned}
$$

Example MGFs:

- $X \sim U(a, b)$

$$M_X(t) = \int_a^b \frac{e^{tx}}{b-a} dx = \frac{e^{bt} - e^{at}}{(b-a)t}$$

- $X \sim Exp(\lambda)$

$$M_X(t) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \int_0^\infty \lambda e^{-(\lambda - t)x} dx = \frac{\lambda}{\lambda - t}$$

Note that this integral is only defined when $t < \lambda$

- $X \sim Geo(p), (q = 1 - p)$

$$M_X(t) = \sum_{x=1}^\infty e^{tx} p q^{x-1} = pe^t \sum_{x=1}^\infty (e^t)^{x-1} q^{x-1} = \frac{pe^t}{1 - qe^t}$$

- $X \sim Pois(\lambda)$

$$M_X(t) = \sum_{x=0}^\infty e^{tx} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^\infty \frac{(e^t \lambda)^x}{x!} = e^{\lambda(e^t - 1)}$$

Examples of using the MGF to calculate moments

- $X \sim Exp(\lambda)$

$$M^{(1)}(t) = \frac{\lambda}{(\lambda - t)^2}; \qquad E[X] = \frac{1}{\lambda}$$

$$M^{(2)}(t) = \frac{2\lambda}{(\lambda - t)^3}; \qquad E[X^2] = \frac{2}{\lambda^2}$$

$$M^{(r)}(t) = \frac{\Gamma(r+1)\lambda}{(\lambda - t)^{r+1}}; \quad E[X^r] = \frac{\Gamma(r+1)}{\lambda^r}$$

- $X \sim Geo(p), (q = 1 - p)$

$$M^{(1)}(t) = \frac{pe^t}{1 - qe^t} + \frac{pqe^{2t}}{(1 - qe^t)^2}; \qquad E[X] = \frac{1}{p}$$

$$M^{(2)}(t) = \frac{pe^t}{1 - qe^t} + \frac{3pqe^{2t}}{(1 - qe^t)^2} + \frac{2pq^2e^{3t}}{(1 - qe^t)^3}; \quad E[X^2] = \frac{5 - 6p + 2p^2}{p}$$

Theorem. *If $Y = a + bX$ then*

$$M_Y(t) = e^{at} M_X(bt)$$

Proof.

$$M_Y(t) = E[e^{tY}] = E[e^{at+btX}] = e^{at}E[e^{(bt)X}] = e^{at}M_X(bt)$$

□

For example, this result can be used to verify the result that $E[a + bX] = a + bE[X]$ as

$$M_Y^{(1)}(t) = ae^{at}M_X(bt) + be^{at}M_X^{(1)}(bt)$$

$$M_Y^{(1)}(0) = aM_X(0) + bM_X^{(1)}(0) = a + bE[X]$$

Theorem. *If X and Y are independent RVs with MGFs M_X and M_Y and $Z = X + Y$, then $M_Z(t) = M_X(t)M_Y(t)$ on the common interval where both MGFs exist.*

Proof.

$$\begin{aligned} M_Z(t) &= E[e^{tZ}] = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] \\ &= E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t) \end{aligned}$$

□

By induction, this result can be extended to sums of many independent RVs.

One particular use of this result is that it can give an easy approach to showing what the distribution of a sum of RVs is without having the calculate the convolution of the densities. But first we need one more result.

Theorem. [Uniqueness theorem] *If the MGF of X exists for t in an open interval containing 0, then it uniquely determines the CDF.*

i.e no two different distributions can have the same values for the MGFs on an interval containing 0.

Proof. Postponed $\square$

Example: Let $X_1, X_2, \ldots, X_n$ be iid $Exp(\lambda)$. What is the distribution of $S = \sum X_i$

$$M_S(t) = \prod_{i=1}^{n} \frac{\lambda}{\lambda - t} = \left(\frac{\lambda}{\lambda - t}\right)^n$$

Note that this isn't the form of the MGF for an exponential, so the sum isn't exponential. As shown in Example B on page 145, the MGF of a $Gamma(\alpha, \lambda)$ is

$$M(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}$$

so $S \sim Gamma(n, \lambda)$

This approach also leads to an easy proof that the sum of independent normals is also normal. The moment generating function for $N(\mu, \sigma^2)$ RV is $M(t) = e^{\mu t + \sigma^2 t^2/2}$. So if $X_i \overset{iid}{\sim} N(\mu_i, \sigma_i^2), i = 1, \ldots, n$, then

$$M_{\sum X_i} = \prod_{i=1}^{n} e^{\mu_i t + \sigma_i^2 t^2/2} = \exp(t \sum \mu_i + t^2/2 \sum \sigma_i^2)$$

which is the moment generating function of a $N(\sum \mu_i, \sum \sigma_i^2)$ RV.

There is one important thing with this approach. We must be able to identify what MGF goes with each density or PMF.

For example, let $X \sim Gamma(\alpha, \lambda)$ and $Y \sim Gamma(\beta, \mu)$ be independent. Then the MGF of $Z \sim X + Y$ is

$$M_Z(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha} \left(\frac{\mu}{\mu - t}\right)^{\beta}$$

This is not the MGF of a gamma distribution unless $\lambda = \mu$.

In fact I'm not quite sure what the density looks like beyond

$$f_Z(z) = \int_0^{\infty} \frac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \frac{\mu^{\beta}(z-x)^{\beta-1} e^{-\mu(z-x)}}{\Gamma(\beta)} dx$$

You can sometimes use tables of Laplace transforms or doing some complicated complex variable integration to invert the MGF to determine the density or PMF.

While we can't get the density easily in this case, we can still use the MGF to get the moments of this distribution.

It is also possible to work with more complicated situations described by hierarchical models. Suppose that the MGFs for $X (M_X(t))$ and $Y|X = x$ $(M_{Y|X}(t))$ are known. Then the marginal MGF of Y is

$$M_Y(t) = E[e^{tY}] = E[E[e^{tY}|X]] = E[M_{Y|X}(t)]$$

For example, this could be used to get the MGF of the Beta-Binomial model.

Another situation where this is useful is with a random sums model where

$$S = \sum_{i=1}^{N} X_i$$

and N is random. Then the MGF of S is given by

$$\begin{aligned} M_S(t) &= E[E[e^{tS}|N]] = E[(M_X(t))^N] \\ &= E[e^{N \log M_X(t)}] = M_N(\log M_X(t)) \end{aligned}$$

An example of this model is

$$S|N = n \sim Bin(n,p) \qquad \left(= \sum_{i=1}^{N} Bern(p)\right)$$

$$N \sim Pois(\lambda)$$

$$M_X(t) = 1 - p + pe^t$$

$$M_N(t) = e^{\lambda(e^t - 1)}$$

So the moment generating function for S is

$$\begin{aligned} M_S(t) &= M_N(log M_X(t)) \\ &= \exp(\lambda(e^{\log(1-p+pe^t)-1})) \\ &= \exp(\lambda(1-p+pe^t)e^{-1}) \end{aligned}$$

Another example is the compound Poisson model discussed in the text.

The MGF is also defined for joint distributions. It has the form

$$M_{X,Y}(s,t) = E[e^{sX+tY}]$$

It has similar properties as the univariate case. For example the mixed moments are given by

$$E[X^n Y^m] = \frac{\partial^{n+m}}{\partial x^n \partial y^m} M_{X,Y}(0,0)$$

The marginal MGFs can be determined directly from the joint MGF as

$$M_X(s) = M_{X,Y}(s,0); \qquad M_Y(t) = M_{X,Y}(0,t)$$

Also X and Y and independent if and only if $M_{X,Y}(s,t) = M_X(s)M_Y(t)$. This relates to the idea that if X and Y are independent so are $g(X) = e^{sX}$ and $h(Y) = e^{tY}$.