Applying the Moment Generating Functions to the Study of Probability Distributions

Silvia SPĂTARU
Academy of Economic Studies, Bucharest

In this paper, we describe a tool to aid in proving theorems about random variables, called the moment generating function, which converts problems about probabilities and expectations into problems from calculus about function values and derivates. We show how the moment generating function determinates the moments and how the moments can be used to recover the moment generating function. Using of moment generating functions to find distributions of functions of random variables is presented. A standard form of the central limit theorem is also stated and proved.
Keywords: *probability distribution, probability density function, moment generating function, central limit theorem.*

Introduction

A generating function of a random variable (rv) is an expected value of a certain transformation of the variable. All generating functions have some very important properties. The most important property is that under mild conditions, the generating function completely determines the distribution. Often a random variable is shown to have a certain distribution by showing that the generating function has a certain known form. There is a process of recovering the distribution from a generating function and this is known as inversion. The second important property is that the moments of the random variable can be determined from the derivatives of the generating function. This property is useful because often obtaining moments from the generating function is easier than computing the moments directly from their definitions. Another important property is that the generating function of a sum of independent random variables is the product of the generating functions. This property is useful because the probability density function of a sum of independent variables is the convolution of the individual density functions, and this operation is much more complicated. The last important property is called the continuity theorem and asserts that ordinary convergence of a sequence of generating functions corresponds to convergence of the corresponding distributions. Often it is easier to demonstrate the convergence of the generating functions than to demonstrate convergence of the distributions directly.

Why is it necessary to study the probability distribution? The sample observations are frequently expressed as numerical events that corresponds to the values of the random variables. Certain types of random variables occurs frequently in practice, so it is useful to know the probability for each value of a random variable. The probability of an observed sample is needed to make inferences about a population.

Discrete and continuous distributions

A random variable is a function X, whose value is uncertain and depends on some random event. The space or range of X is the set S of possible values of X. A random variable X is said to be discrete if this set has a finite or countable infinite number of distinct values (i.e. can be listed as a sequence $x_1, x_2, \ldots$). The random variable X is said to have a continuous distribution if all values are possible in some real interval. Often, there are functions that assign probabilities to all events in a sample space. These functions are called *probability mass functions* if the events are discretely distributed, or *probability density functions* if the events are continuously distributed. All the possible value of a random variable and their associated probability values constitute the *probability distribution* of the random variable.

The discrete probability distributions are specified by the list of possible values and the probabilities attached to those values, and the continuous distributions are specified by probability density functions. The distribution of a random variable *X* can be also described by the cumulative distribution function $F_X(x) = P(X < x)$. In the case of discrete random variable, this is not particularly useful, although it does serve to unify discrete and continuous variables. There are other ways to characterize distributions. Thus, the probability distributions can be also specified by a variety of transforms, that is, by functions that somehow encode the properties of the distributions into a form more convenient for certain kinds of probability calculation.

For a discrete random variable X with a probability mass function $p(x) = P(X = x)$ we have $0 \leq p(x) \leq 1$ for all x and $\sum p(x) = 1$. The probability mass function or the probability density function of a random variable *X* contains all the information that one ever need about this variable.

The sequence of moments of a random variable

We know that the mean $\mu = E(X)$ and variance

$\sigma^2 = E(X - E(X))^2 = E(X^2) - (E(X))^2$ of a random variable enter into the fundamental limit theorems of probability, as well as into all sorts of practical calculations. These important attributes of a random variable contain important informations about the distribution function of that variable. But the mean and variance do not contain all the available information about density function of a random variable.

Besides the two numerical descriptive quantities μ and σ that locate the center and describe the spread of the values of a random variable, we define a set of numerical descriptive quantities, called moments, which uniquely determine the probability distribution of a random variable.

For a discrete or continuous random variable *X*, the k^{th} *moment of X* is a number defined as $\mu_k = E(X^k)$, $k = 1,2,...$, provided the defining sum or integral of the expectation converges.

We have a sequence of moments associated to a random variable *X*. In many cases this sequence determines the probability distribution of *X*. However, the moments of *X* may not exist. In terms of these moments, the mean μ and variance σ^2 of *X* are given simply by $\mu = \mu_1$ and $\sigma^2 = \mu_2 - \mu_1^2$. The higher moments have more obscure meaning as *k* grows.

The moments give a lot of useful information about the distribution of *X*. The knowledge of the first two moments of *X* gives us its mean and variance, but a knowledge of all the moments of *X* determines its probability function completely. It turn out that different distributions can not have identical moments. That is what makes moments important.

Therefore, it seems that it should always be possible to calculate the expected value or mean of *X*, $E(X) = \mu$, the variance, $V(X) = \sigma^2$ or higher order moments of *X* from its probability density function, or to calculate the distribution of, say, sum of two independent random variables *X* and *Y*, whose distributions are known. In practice, it turn out that these calculations are often very difficult.

The generating functions

Roughly speaking, the generating functions transform problems about sequences into problems about functions. In this way we can use generating functions to solve all sorts of counting problems.

Suppose that $a_0, a_1, a_2,...$ is a finite or infinite sequence of real numbers. The ordinary generating function of the sequence is the power series $G(z) = a_0 + a_1 z + a_2 z^2 + \cdots = \sum_{k=0}^{\infty} a_k z^k$. In order to recover the original sequence from a given ordinary generating function, the following formula holds:

$$a_k = \frac{1}{k!}\frac{d^k G}{dz^k}(0),\ k = 0,1,2,\ldots.$$

Assume that $a_0, a_1, a_2, \ldots$ is a finite or infinite sequence of real numbers. The exponential generating function of this sequence is the power series

$$G(z) = a_0 + \frac{a_1}{1!}z + \frac{a_2}{2!}z^2 + \cdots = \sum_{k=0}^{\infty}\frac{a_k}{k!}z^k .$$

For recovering the original sequence of real numbers from the given exponential generating function, $G(z)$, the following formula holds:

$$a_k = \frac{d^k G}{dz^k}(0),\ k = 0,1,2,\ldots$$

For a random variable *X* taking only non-negative integer values *k*, with probabilities $p_k = P(X = k)$, the probability generating function is defined as: $G(z) = E\left(z^X\right) = \sum_{k=0}^{\infty} p_k z^k$ for $0 \le z \le 1$. The powers of the variable *z* serves as placeholders for the probabilities p_k that determine the distribution. We recover the probabilities p_k as coefficients in a power series expansion of the probability generating function. Expansion of a probability generating function in a power series is just one way of extracting information about the distribution. Repeated differentiation inside the expectation operator gives

$$G^{(k)}(z) = \frac{d^k}{dz^k}E\left(z^X\right) =$$
$$= E\left(X(X-1)\ldots(X-k+1)z^{X-k}\right),$$

whence $G^{(k)}(1) = E\left(X(X-1)\ldots(X-k+1)\right)$ for $k = 1,2,\ldots$. Thus we can recover the moment of *X*. An exact probability generating function uniquely determines a distribution and an approximation to the probability generating function approximately determines the distribution.

The moment generating functions

The beauty of moment generating functions is that they give many results with relative ease. Proofs using moment generating functions are often much easier than showing the same results using density functions (or some other ways).

There is a clever way of organizing all the moments of a random variable into one mathematical object. This is a function of a new variable *t*, called *the moment generating function* (mgf), which is defined by $g_X(t) = E(e^{tX})$, provided that the expectation exists for *t* in some neighborhood of 0. In the discrete case this is equal to $\sum e^{tx} p(x)$, and in the continuous case to $\int e^{tx} f(x)dx$. Hence, it is be important that the expectation be finite for all $t \in (-t_0, t_0)$ for some $t_0 > 0$. If the expectation does not exist in a neighbor of 0, we say that the moment generating function does not exist. Since the exponential function is positive, $E\left(e^{tX}\right)$ always exists, either as a real number or as a positive infinity.

The moment generating functions may not be defined for all values of *t*, and some well-known distributions do not have moment generating function (e.g. the Cauchy distribution).

Observe that $g_X(t)$ is a function of *t*, not of *X*. The moment generating function of a random variable packages all the moments for a random variable into one simple expression. Formally, the moment generating function is obtained by substituting $z = e^t$ in the probability generating function.

Note that there is a substitute for mgf which is defined for every distribution, the complex numbers version of the mgf, namely the characteristic function.

Fundamental properties of the moments generating functions

The moment generating function has many useful properties in the study of random variables, but we consider only a few here. Suppose that *X* is a random variable with the moment generating function $g_X(t)$. Henceforth we assume that $g_X(t)$ exists in some neighbourhood of the origin. In this case some useful properties can be proved.

1. If $g_X(t)$ is the moment generating function of a random variable *X* then $g_X(0)=1$.
Actually, we have $g_X(0)=E\left(e^{0\times X}\right)=E(1)=1$.

2. The moments of the random variable *X* may be found by power series expansion. The moment generating function of a random variable *X* is the exponential generating function of its sequence of moments $g(t)=\sum_{k=0}^{\infty}\frac{\mu_k t^k}{k!}$.

Since the exponential function e^t has the power series $e^t=\sum_{k=0}^{\infty}\frac{t^k}{k!}$, by the series expansion of the function e^{tX} we have the equality of random variables $e^{tX}=\sum_{k=0}^{\infty}\frac{X^k t^k}{k!}$. Then we take the expectation of both sides and use the fact that the operator *E* commutes with sum to get

$$E\left(e^{tX}\right)=E\left(\sum_{k=0}^{\infty}\frac{X^k t^k}{k!}\right)=$$

$$=\sum_{k=0}^{\infty}E\left(\frac{X^k t^k}{k!}\right)=\sum_{k=0}^{\infty}\frac{t^k}{k!}E\left(X^k\right).$$

3. Calculating moments. We call $g_X(t)$ the moment generating function because all of the moments of *X* can be obtained by successively differentiating $g_X(t)$ and then evaluating the result at $t=0$.

The k^{th} derivative of $g_X(t)$ evaluated at the point $t=0$ is the k^{th} moment μ_k of *X*, i.e. $\mu_k=g^{(k)}(0)$, where $g^{(k)}(0)=\frac{d^k}{dt^k}g(t)\big|_{t=0}$.

In this way, the moments of *X* may also be found by differentiation. We can easily see that $\frac{d^k}{dt^k}g(t)=\frac{d^k}{dt^k}E\left(e^{tX}\right)=E\left(\frac{d^k}{dt^k}e^{tX}\right)=$ $=E\left(X^k e^{tX}\right)$ (interchange of *E* and differentiation is valid). Therefore, we obtain $\frac{d^k}{dt^k}g(t)\big|_{t=0}=E\left(X^k\right)=\mu_k$.

In other words, the moment generating function generates all the moments of *X* by differentiation. We can find the moments of *X* by calculating the moment generating function and then differentiating. Sometimes it is easier to get moments this way than directly. Together, all the moments of a distribution pretty much determine the distribution.

In addition to producing the moments of *X*, the mgf is useful in identifying the distribution of *X*.

4. If $g_X(t)$ exists in an interval around 0, then knowledge of the mgf of a rv is equivalent to knowledge of its probability density function. This means that the mgf uniquely determines the probability density function.

In general, the series defining $g_X(t)$ will not converge for all *t*. But in the important special case where *X* is bounded (i.e. where the range of *X* is contained in a finite interval) we can show that the series does converges for all *t*. The distribution function is completely determined by its moments.

Theorem. Suppose *X* is a continuous random variable with range contained in the real interval $[-M,M]$. Then the series $g_X(t)=\sum_{k=0}^{\infty}\frac{\mu_k t^k}{k!}$ converges for all *t* to an infinitely differentiable function $g_X(t)$ and $g_X^{(k)}(0)=\mu_k$.

Proof: We know that $\mu_k=\int_{-M}^{M}x^k f_X(x)dx$.

Then we have $|\mu_k|\le\int_{-M}^{M}\left|x^k\right|f_X(x)dx\le M^k$.

Hence, for all *n* we have $\sum_{k=0}^{n}\left|\frac{\mu_k t^k}{k!}\right|\le\sum_{k=0}^{n}\frac{\left(M|t|\right)^k}{k!}\le e^{M|t|}$.

This inequality shows that the moment series converges for all *t* and that its sum is infinitely differentiable, because it is a power series. In this way we have shown that the moments μ_k determine the function $g_X(t)$.

Conversely, since $\mu_k=g_X^{(k)}(0)$, we can see that the function $g_X(t)$ determines the moments μ_k.

If *X* is a bounded rv, then we can show that the mgf $g_X(t)$ of *X* determinates the probability density function $f_X(x)$ of X uniquely. This is important since, on occasion, manipulating generating functions is simpler than manipulating probability density functions.

5. Uniqueness theorem asserts that two random variables with the same mgf have the same distribution. Let X and Y be two random variables with moment generating functions $g_X(t)$ and $g_Y(t)$ and with corresponding distribution functions $F_X(x)$ and $F_Y(y)$. If $g_X(t) = g_Y(t)$, then $F_X(x) = F_Y(x)$. This ensures that the distribution of a random variable can be identified by its moment generating function.

A consequence of the above theorem is that if all moments of a rv X exist, they characterize completely the mgf (since the moments are derivatives of the mgf in its Taylor series expansion) and the moments also completely characterize the distribution, as well as the cumulative distribution function, probability density function and probability mass function.

When a moment generating function exists, there is a unique distribution corresponding to that moment generating function. Hence, there is an injective mapping between moment generating functions and probability distributions. This allows us to use moment generating functions to find distriutions of transformed random variables in some cases. This technique is most commonly used for linear combinations of independent random variables.

6. When the mgf exists, it characterizes an infinite set of moments. The obvious question that then arises is if can two different distributions have the same infinite set of moment. The answer is that, when the mgf exists in a neighborhood around 0, the infinite sequence of moments does uniquely determine the distribution. This knowledge allows us to determine the limiting distribution of a sequence of random variable by examining the associated moment generating functions.

Theorem. Suppose $X_1, X_2, \ldots$ is a sequence of random variables, each having mgf $g_{X_n}(t)$ and $\lim_{n\to\infty} g_{X_n}(t) = g_X(t)$ is finite for all t in a neighborhood of 0. Then there is a unique cumulative distribution function $F_X(x)$ whose moments are determined by $g_X(t)$, and we have $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ whenever x is a continuity point of F_X. Thus, convergence of moment generating functions to a moment generating function in a neighborhood around 0 implies convergence of the associated cumulative distribution functions.

7. Sums of independent random variables. Moment generating functions are useful in establishing distributions of sums of independent random variables.

i) If $Y = aX + b$, where a and b are two real constants, then we have $g_Y(t) = e^{bt} g_X(at)$.

We have $g_Y(t) = E\left(e^{(aX+b)t}\right) = E\left(e^{bt} e^{atX}\right)$ $= e^{bt} E\left(e^{atX}\right) = e^{bt} g_X(at)$.

ii) The function which generates central moments of a random variable with mean μ is given by $g_{X-\mu}(t) = e^{-\mu t} g_X(t)$.

This result is understood by considering the following identity: $g_{X-\mu}(t) = E\left(e^{(X-\mu)t}\right) =$ $e^{-\mu t} E\left(e^{Xt}\right) = e^{-\mu t} g_X(t)$.

iii) If X and Y are two independent random variables with probability density functions $f_X(x)$ and $f_Y(y)$ and with corresponding moment generating functions $g_X(t)$ and $g_Y(t)$, then their sum $X + Y$ has the mgf $g_{X+Y}(t) = g_X(t) g_Y(t)$.

Actually, we have $g_{X+Y}(t) = E\left(e^{t(X+Y)}\right) =$ $E\left(e^{tX}\right)E\left(e^{tY}\right) = g_X(t) g_Y(t)$, because X and Y being independent, so are also e^{tX} and e^{tY}.

Note that this is a very useful property of mgf's, but the above formula would not be of any use if we did not know that the mgf determines the probability distribution.

Using this result it is also possible to obtain, in a very simple way, the mgf of a finite sequence of independent identically distributed random variables. At this point, the mgf's may seem to be a panacea when it comes to calculating the distribution of sums of independent identically distributed random variables. Sometimes we cannot write down the distribution in closed form but, because there are many numerical methods for inverting transforms, we can calculate probabilities from a mgf.

8. There are various reasons for studying moment generating functions, and one of them is that they can be used to prove the central limit theorem.

Central limit theorem. Let $X_1, X_2, ...$ be a sequence of independent identically distributed random variables, each having mean μ and variance σ^2. If $S_n = X_1 + X_2 + \cdots + X_n$ and $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$, then Z_n has a limiting distribution $N(0,1)$ as $n \to \infty$.

Proof: Let $Y_i = X_i - \mu$, for $i = 1,2,...$. In this case the variables $Y_1, Y_2, ...$ are independent identically distributed and we have $S_n - n\mu = X_1 + \cdots + X_n - n\mu$ $= Y_1 + Y_2 + \cdots + Y_n$. We know that $g_{S_n - n\mu}(t) = g_{Y_1}(t) \cdots g_{Y_n}(t) = \left(g_{Y_1}(t)\right)^n$. Then we can write

$$g_{Z_n}(t) = E\left(e^{tZ_n}\right) = E\left(e^{(S_n - n\mu)(t/\sigma\sqrt{n})}\right) =$$

$$= g_{S_n - n\mu}\left(\frac{t}{\sigma\sqrt{n}}\right) = \left(g_{Y_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n.$$

Now we use the power series expansion

$$g_{Y_1}(t) = 1 + E(Y_1)\frac{t}{1!} + E(Y_1^2)\frac{t^2}{2!} + E(Y_1^3)\frac{t^3}{3!} + \ldots$$

$= 1 + \frac{1}{2}\sigma^2 t^2 + o(t^2)$. We used the fact that $E(Y_1) = 0$ and $E(Y_1^2) = \sigma^2$, and we denoted $o(t^2)$ a function $h(t)$ such that $\frac{h(t)}{t^2} \to 0$ as $t \to 0$. Then, for fixed *t*, we obtain

$$g_{Z_n}(t) = \left(1 + \frac{1}{2}\sigma^2\left(\frac{t^2}{\sigma^2 n}\right)t^2 + o\left(\frac{1}{n}\right)\right)^n$$

$$= \left(1 + \frac{t^2}{2}\frac{1}{n} + o\left(\frac{1}{n}\right)\right)^n.$$

Here $o\left(\frac{1}{n}\right)$ denotes a function $h(n)$ such that $\frac{h(n)}{1/n} \to 0$ as $n \to 0$. We deduce that $g_{Z_n}(t) \to e^{\frac{t^2}{2}}$ as $n \to \infty$ and this is precisely the mgf of a variable $N(0,1)$. Thus, we have proved that the distribution of Z_n converges to the standard normal distribution $N(0,1)$ as $n \to \infty$.

Corollary: If we consider the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, then $\overline{X}_n$ has a limiting distribution $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ as $n \to \infty$.

The theorem can be generalised to independent random variables having different means and variances.

References

[1] Cuculescu, I., (1998), *Teoria Probabilitatilor*, Ed. All.

[2] Grinstead, C., Snell, L., (1998), *Introduction to Probability*.

[3] Kallenberg, O., (2001) *Foundations of Modern Probability*, Springer.

[4] Petrov, V.V., (1975), *Sums of Independent Random Variables*, Springer.

[5] Petrov, V.V., (1995), *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Oxford University Press.

[6] Roussas, G., (2005), *An Introduction to Measure-Theoretic Probability*, Elsevier Academic Press.