MOMENT GENERATING FUNCTIONS

Contents

1. Moment generating functions
2. Sum of a random number of random variables
3. Transforms associated with joint distributions

Moment generating functions, and their close relatives (probability generating functions and characteristic functions) provide an alternative way of representing a probability distribution by means of a certain function of a single variable.

These functions turn out to be useful in many different ways:

(a) They provide an easy way of **calculating the moments** of a distribution.

(b) They provide some powerful tools for addressing certain **counting and combinatorial** problems.

(c) They provide an easy way of characterizing the distribution of the **sum of independent random variables.**

(d) They provide tools for dealing with the distribution of the **sum of a random number of independent random variables.**

(e) They play a central role in the study of **branching processes**.

(f) They play a key role in **large deviations** theory, that is, in studying the asymptotics of tail probabilities of the form $\mathbb{P}(X \geq c)$, when c is a large number.

(g) They provide a bridge between **complex analysis** and probability, so that complex analysis methods can be brought to bear on probability problems.

(h) They provide powerful tools for proving **limit theorems**, such as laws of large numbers and the central limit theorem.

1 MOMENT GENERATING FUNCTIONS

1.1 Definition

Definition 1. *The* **moment generating function** *associated with a random variable X is a function $M_X : \mathbb{R} \to [0, \infty]$ defined by*

$$M_X(s) = \mathbb{E}[e^{sX}].$$

The **domain** D_X *of M_X is defined as the set $D_X = \{s \mid M_X(s) < \infty\}$.*

If X is a discrete random variable, with PMF p_X, then

$$M_X(s) = \sum_x e^{sx} p_X(x).$$

If X is a continuous random variable with PDF f_X, then

$$M_X(s) = \int e^{sx} f_X(x)\, dx.$$

Note that this is essentially the same as the definition of the Laplace transform of a function f_X, except that we are using s instead of $-s$ in the exponent.

1.2 The domain of the moment generating function

Note that $0 \in D_X$, because $M_X(0) = \mathbb{E}[e^{0X}] = 1$. For a discrete random variable that takes only a finite number of different values, we have $D_X = \mathbb{R}$. For example, if X takes the values 1, 2, and 3, with probabilities $1/2$, $1/3$, and $1/6$, respectively, then

$$M_X(s) = \frac{1}{2}e^s + \frac{1}{3}e^{2s} + \frac{1}{6}e^{3s}, \tag{1}$$

which is finite for every $s \in \mathbb{R}$. On the other hand, for the Cauchy distribution, $f_X(x) = 1/(\pi(1 + x^2))$, for all x, it is easily seen that $M_X(s) = \infty$, for all $s \neq 0$.

In general, D_X is an interval (possibly infinite or semi-infinite) that contains zero.

Exercise 1. Suppose that $M_X(s) < \infty$ for some $s > 0$. Show that $M_X(t) < \infty$ for all $t \in [0, s]$. Similarly, suppose that $M_X(s) < \infty$ for some $s < 0$. Show that $M_X(t) < \infty$ for all $t \in [s, 0]$.

Exercise 2. Suppose that

$$\limsup_{x \to \infty} \frac{\log \mathbb{P}(X > x)}{x} \triangleq -\nu < 0.$$

Establish that $M_X(s) < \infty$ for all $s \in [0, \nu)$.

1.3 Inversion of transforms

By inspection of the formula for $M_X(s)$ in Eq. (1), it is clear that the distribution of X is readily determined. The various powers e^{sx} indicate the possible values of the random variable X, and the associated coefficients provide the corresponding probabilities.

At the other extreme, if we are told that $M_X(s) = \infty$ for every $s \neq 0$, this is certainly not enough information to determine the distribution of X.

On this subject, there is the following fundamental result. It is intimately related to the inversion properties of Laplace transforms. Its proof requires sophisticated analytical machinery and is omitted.

Theorem 1. Inversion theorem

(a) *Suppose that $M_X(s)$ is finite for all s in an interval of the form $[-a, a]$, where a is a positive number. Then, M_X determines uniquely the CDF of the random variable X.*

(b) *If $M_X(s) = M_Y(s) < \infty$, for all $s \in [-a, a]$, where a is a positive number, then the random variables X and Y have the same CDF.*

There are explicit formulas that allow us to recover the PMF or PDF of a random variable starting from the associated transform, but they are quite difficult to use (e.g., involving "contour integrals"). In practice, transforms are usually inverted by "pattern matching," based on tables of known distribution-transform pairs.

1.4 Moment generating properties

There is a reason why M_X is called a moment generating function. Let us consider the derivatives of M_X at zero. Assuming for a moment we can interchange

3

the order of integration and differentiation, we obtain

$$\frac{dM_X(s)}{ds}\bigg|_{s=0} = \frac{d}{ds}\mathbb{E}[e^{sX}]\bigg|_{s=0} = \mathbb{E}[Xe^{sX}]\bigg|_{s=0} = \mathbb{E}[X],$$

$$\frac{d^m M_X(s)}{ds^m}\bigg|_{s=0} = \frac{d^m}{ds^m}\mathbb{E}[e^{sX}]\bigg|_{s=0} = \mathbb{E}[X^m e^{sX}]\bigg|_{s=0} = \mathbb{E}[X^m]$$

Thus, knowledge of the transform M_X allows for an easy calculation of the moments of a random variable X.

Justifying the interchange of the expectation and the differentiation does require some work. The steps are outlined in the following exercise. For simplicity, we restrict to the case of nonnegative random variables.

Exercise 3. Suppose that X is a nonnegative random variable and that $M_X(s) < \infty$ for all $s \in (-\infty, a]$, where a is a positive number.

(a) Show that $\mathbb{E}[X^k] < \infty$, for every k.

(b) Show that $\mathbb{E}[X^k e^{sX}] < \infty$, for every $s < a$.

(c) Show that $(e^{hX} - 1)/h \le Xe^{hX}$.

(d) Use the DCT to argue that

$$\mathbb{E}[X] = \mathbb{E}\left[\lim_{h \downarrow 0} \frac{e^{hX} - 1}{h}\right] = \lim_{h \downarrow 0} \frac{\mathbb{E}[e^{hX}] - 1}{h}.$$

1.5 The probability generating function

For discrete random variables, the following **probability generating function** is sometimes useful. It is defined by

$$g_X(s) = \mathbb{E}[s^X],$$

with s usually restricted to positive values. It is of course closely related to the moment generating function in that, for $s > 0$, we have $g_X(s) = M_X(\log s)$. One difference is that when X is a positive random variable, we can define $g_X(s)$, as well as its derivatives, for $s = 0$. So, suppose that X has a PMF $p_X(m)$, for $m = 1, \ldots.$. Then,

$$g_X(s) = \sum_{m=1}^{\infty} s^m p_X(m),$$

resulting in

$$\frac{d^m}{ds^m} g_X(s)\bigg|_{s=0} = m!\, p_X(m).$$

(The interchange of the summation and the differentiation needs justification, but is indeed legitimate for small s.) Thus, we can use g_X to easily recover the PMF p_X, when X is a positive integer random variable.

4

1.6   Examples

Example : $X \stackrel{d}{=} \mathrm{Exp}(\lambda)$. Then,

$$M_X(s) = \int_0^\infty e^{sx} \lambda e^{-\lambda x}\, dx = \begin{cases} \frac{\lambda}{\lambda - s}, & s < \lambda; \\ \infty, & \text{otherwise.} \end{cases}$$

Example : $X \stackrel{d}{=} \mathrm{Ge}(p)$

$$M_X(s) = \sum_{m=1}^\infty e^{sm} p(1-p)^{m-1} \begin{cases} \frac{e^s p}{1-(1-p)e^s}, & e^s < 1/(1-p); \\ \infty, & \text{otherwise.} \end{cases}$$

In this case, we also find $g_X(s) = ps/(1 - (1-p)s), s < 1/(1-p)$ and $g_X(s) = \infty$, otherwise.

Example : $X \stackrel{d}{=} N(0,1)$. Then,

$$\begin{aligned} M_X(s) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \exp(sx) \exp(-\frac{x^2}{2}) dx \\ &= \frac{\exp(s^2/2)}{\sqrt{2\pi}} \int_{-\infty}^\infty \exp(-\frac{x^2 + 2sx - s^2}{2}) dx \\ &= \exp(s^2/2). \end{aligned}$$

1.7   Properties of moment generating functions

We record some useful properties of moment generating functions.

Theorem 2.

(a) If $Y = aX + b$, then $M_Y(s) = e^{sb} M_X(as)$.

(b) If X and Y are independent, then $M_{X+Y}(s) = M_X(s) M_Y(s)$.

(c) Let X and Y be independent random variables. Let Z be equal to X, with probability p, and equal to Y, with probability $1 - p$. Then,

$$M_Z(s) = p M_X(s) + (1 - p) M_Y(s).$$

Proof: For part (a), we have

$$M_X(aX + b) = \mathbb{E}[\exp(saX + sb)] = \exp(sb)\mathbb{E}[\exp(saX)] = \exp(sb) M_X(as).$$

For part (b), we have

$$M_{X+Y}(s) = \mathbb{E}[\exp(sX + sY)] = \mathbb{E}[\exp(sX)]\mathbb{E}[\exp(sY)] = M_X(s)M_Y(s).$$

For part (c), by conditioning on the random choice between X and Y, we have

$$M_Z(s) = \mathbb{E}[e^{sZ}] = p\mathbb{E}[e^{sX}] + (1-p)\mathbb{E}[e^{sY}] = pM_X(s) + (1-p)M_Y(s).$$

$\square$

Example : (Normal random variables)

(a) Let X be a standard normal random variable, and let $Y = \sigma X + \mu$, which we know to have a $N(\mu, \sigma^2)$ distribution. We then find that $M_Y(s) = \exp(s\mu + \frac{1}{2}s^2\sigma^2)$.

(b) Let $X \stackrel{d}{=} N(\mu_1, \sigma_1^2)$ and $Y = N(\mu_2, \sigma_2^2)$. Then,

$$M_{X+Y}(s) = \exp\left(\left\{s(\mu_1 + \mu_2) + \frac{1}{2}s^2(\sigma_1^2 + \sigma_2^2)\right\}\right).$$

Using the inversion property of transforms, we conclude that $X + Y \stackrel{d}{=} N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, thus corroborating a result we first obtained using convolutions.

2 SUM OF A RANDOM NUMBER OF INDEPENDENT RANDOM VARIABLES

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables, with mean μ and variance σ^2. Let N be another independent random variable that takes nonnegative integer values. Let $Y = \sum_{i=1}^{N} X_i$. Let us derive the mean, variance, and moment generating function of Y.

We have

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid N]] = \mathbb{E}[N\mu] = \mathbb{E}[N]\mathbb{E}[X].$$

Furthermore, using the law of total variance,

$$\begin{aligned}
\mathrm{var}(Y) &= \mathbb{E}\big[\mathrm{var}(Y \mid N)\big] + \mathrm{var}\big(\mathbb{E}[Y \mid N]\big) \\
&= \mathbb{E}[N\sigma^2] + \mathrm{var}(N\mu) \\
&= \mathbb{E}[N]\sigma^2 + \mu^2\mathrm{var}(N).
\end{aligned}$$

Finally, note that

$$\mathbb{E}[\exp(sY) \mid N = n] = M_X^n(s) = \exp(n \log M_X(s)),$$

6

implying that

$$M_Y(s) = \sum_{n=1}^{\infty} \exp(n \log M_X(s)) \mathbb{P}(N = n) = M_N(\log M_X(s)).$$

The reader is encouraged to take the derivative of the above expression, and evaluate it at $s = 0$, to recover the formula $\mathbb{E}[Y] = \mathbb{E}[N]\mathbb{E}[X]$.

Example : Suppose that each X_i is exponentially distributed, with parameter λ, and that N is geometrically distributed, with parameter $p \in (0, 1)$. We find that

$$M_Y(s) = \frac{e^{\log M_X(s)}p}{1 - e^{\log M_X(s)}(1 - p)} = \frac{p\lambda/(\lambda - s)}{1 - \lambda(1 - p)/(\lambda - s)} = \frac{\lambda p}{\lambda p - s}$$

which we recognize as a moment generating function of an exponential random variable with parameter λp. Using the inversion theorem, we conclude that Y is exponentially distributed. In view of the fact that the sum of a fixed number of exponential random variables is far from exponential, this result is rather surprising. An intuitive explanation will be provided later in terms of the Poisson process.

3   TRANSFORMS ASSOCIATED WITH JOINT DISTRIBUTIONS

If two random variables X and Y are described by some joint distribution (e.g., a joint PDF), then each one is associated with a transform $M_X(s)$ or $M_Y(s)$. These are the transforms of the marginal distributions and do not convey information on the dependence between the two random variables. Such information is contained in a multivariate transform, which we now define.

Consider n random variables $X_1, \ldots, X_n$ related to the same experiment. Let $s_1, \ldots, s_n$ be real parameters. The associated **multivariate transform** is a function of these n parameters and is defined by

$$M_{X_1,\ldots,X_n}(s_1, \ldots, s_n) = \mathbb{E}\big[e^{s_1 X_1 + \cdots + s_n X_n}\big].$$

The inversion property of transforms discussed earlier extends to the multivariate case. That is, if $Y_1, \ldots, Y_n$ is another set of random variables and $M_{X_1,\ldots,X_n}(s_1, \ldots, s_n)$, $M_{Y_1,\ldots,Y_n}(s_1, \ldots, s_n)$ are the same functions of $s_1, \ldots, s_n$, in a neighborhood of the origin, then the joint distribution of $X_1, \ldots, X_n$ is the same as the joint distribution of $Y_1, \ldots, Y_n$.

Example :

(a) Consider two random variables X and Y. Their joint transform is

$$M_{X,Y}(s,t) = \mathbb{E}[e^{sX}e^{tY}] = \mathbb{E}[e^{sX+tY}] = M_Z(1),$$

where $Z = sX+tY$. Thus, calculating a multivariate transform essentially amounts to calculating the univariate transform associated with a single random variable that is a linear combination of the original random variables.

(b) If X and Y are independent, then $M_{X,Y}(s,t) = M_X(s)M_Y(t)$.

6.436J / 15.085J Fundamentals of Probability
Fall 2008