Moment Generating Functions

P. Sam Johnson

April 29, 2014

Motivation

One wishes to compute xy.

1. Obtain the values of $\log x$ and of $\log y$ (from suitable tables).
2. Evaluate $\log x + \log y$, which represents $\log xy$.
3. From the knowledge of $\log xy$ we are then able to obtain the value of xy (again with the aid of tables).

The above approach is useful for the following reasons.

1. To each positive number x there corresponds **exactly one number**, $\log x$, and this number is easily **obtained from tables**. Mathematically, there is a function f takes x to $\log x$.
2. To each value of $\log x$ there corresponds **exactly one value** of x, and this value is again **available from tables**. Mathematically, the function f is one to one.
3. Certain arithmetic operations involving the numbers x and y, such as **multiplication and division**, may be replaced by **simpler operations**, such as **addition and substraction**, by means of the "transformed" numbers $\log x$ and $\log y$.

Instead of performing the arithmetic directly with the numbers x and y, we first obtain the numbers $\log x$ and $\log y$, do our arithmetic with these numbers, and then transform back.

Suppose that X is a random variable.

In computing various characteristics of the random variable X, such as $E(X)$ or $V(X)$, we work directly with the probability distribution of X. Possibly, we can introduce **some other function (called moment generating function-mgf)** and make our required computation in terms of this new function. This is, in fact, precisely what we shall do.

Computing certain characteristics of X:

Start with probability distribution of X	$\Rightarrow$	Construct MGF
		$\Downarrow$
Come back to probability distribution of X	$\Leftarrow$	Solve the given problem with the help of MGF

Moments of a random variable

<u>**Those functions**</u> should characterize distribution functions of random variables.

The **mean and variance do not contain all the available information about the density function of a random variable**.

For instance, suppose X and Y are random variables, with distributions

$$p_X = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 1/4 & 1/2 & 0 & 0 & 1/4 \end{pmatrix},$$

$$p_Y = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/4 & 0 & 0 & 1/2 & 1/4 & 0 \end{pmatrix}.$$

Then with these choices, we have $E(X) = E(Y) = 7/2$ and $V(X) = V(Y) = 9/4$, and yet certainly p_X and p_Y are different distribution functions which have the same mean and the same variance.

This raises a question: Let X be a discrete random variable with range $\{x_1, x_2, \ldots\}$, and distribution function be $p = p_X$.

If we know its mean $\mu = E(X)$ and its variance $\sigma^2 = V(X)$, then what else do we need to know to determine p completely?

A nice answer to this question, at least in the case that X has finite range, can be given in terms of the **moments** of X, **which are numbers** defined as follows:

$$\begin{aligned} \mu_k &= k\text{th moment of } X \\ &= E(X^k) = \sum_{j=1}^{\infty} (x_j)^k p(x_j). \end{aligned}$$

provided the sum converges. Here $p(x_j) = P(X = x_j)$.

In terms of these moments, the mean μ and variance σ^2 of X are given simply by

$$\mu = \mu_1, \quad \sigma^2 = \mu_2 - \mu_1^2,$$

so that a 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 distribution function p completely. (The proof of the result will be given later.)

We need a function of a random variable which should generate moments of X.

We introduce a new variable t, and define a function $g(t)$ as follows:

$$\begin{aligned} g(t) &= E(e^{tX}) = \sum_{k=0}^{\infty} \frac{\mu_k t^k}{k!} \\ &= E\Big(\sum_{k=0}^{\infty} \frac{X^k t^k}{k!}\Big) = \sum_{j=1}^{\infty} e^{tx_j} p(x_j). \end{aligned}$$

We call $g(t)$ the **moment generating function** for X, and think of it **as a convenient device for describing the moments of** X. Indeed, if we differentiate $g(t)$, n times and then set $t = 0$, we get μ_n:

$$\begin{aligned} \frac{d^n}{dt^n} g(t)\Big|_{t=0} &= g^{(n)}(0) \\ &= \sum_{k=n}^{\infty} \frac{k!\mu_k t^{k-n}}{(k-n)!k!}\Big|_{t=0} = \mu_n. \end{aligned}$$

Existence of MGF

Mgf's may not always exist for all values of t. Hence it may happen that the mgf is not defined for all values of t. However, we shall not concern ourselves with this potential difficulty. Whenever we make use of the mgf, we shall always assume it exists.

At $t = 0$, the mgf always exists and equals 1.

Example

Suppose X has range $\{1, 2, \ldots\}$ and $p_X(j) = 1/n$ for $1 \leq j \leq n$ (uniform distribution). Find the first and second moments of X.

Answer. The MGF is $g(t) = \frac{e^t(e^{nt}-1)}{n(e^t-1)}$. The moments are $\mu_1 = (n+1)/2$ and $\mu_2 = \frac{(n+1)(2n+1)}{6}$.

Formal definition of MGF

Definition

Let X be a **discrete random variable** *with probability distribution $p(x_i) = P(X = x_i), \quad i = 1, 2, \ldots$. The function M_X, called the* **moment generating function (mgf)** *of X is defined by*

$$M_X(t) = \sum_{j=1}^{\infty} e^{tx_j} p(x_j).$$

If X is a **continuous random variable** *with pdf f, we define the* **moment generating function** *by*

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx.$$

Examples

Example

Suppose that X is **uniformly distributed** *over the interval $[a, b]$. Therefore the mgf is given by*

$$\begin{aligned} M_X(t) &= \int_a^b \frac{e^{tx}}{b-a} dx \\ &= \frac{1}{b-a}[e^{bt} - e^{at}], \quad t \neq 0. \end{aligned}$$

Examples

Example

Suppose that X *is* **binomially** *distributed with parameters* n *and* p. *Then*

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{\infty} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{\infty} \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\ &= [pe^t + (1-p)]^n. \end{aligned}$$

Examples

Example

Suppose that X has a **Poisson** *distribution with parameter λ. Thus*

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\lambda}\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} \\ &= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}. \end{aligned}$$

Examples

Example

Suppose that X *has an* **exponential distribution** *with parameter* α*. Therfore*

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

(This integral converges only if $t < \alpha$*. Hence the mgf exists only for those values of* t*. Assuming that this condition is satisfied, we shall proceed.) Thus*

$$M_X(t) = \frac{\alpha}{t-\alpha}e^{x(t-\alpha)}\Big|_0^\infty = \frac{\alpha}{\alpha - t}, \quad t < \alpha.$$

Obtaining mgf without knowing pdf

Example

We can obtain the mgf of a function of a random variable without first obtaining its probability distribution.
For example, if X has distribution $N(0,1)$ and we want to find the mgf of $Y = X^2$, we can proceed without first obtaining the pdf of Y. We may simply write

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = E(e^{tX^2}) \\ &= \frac{1}{\sqrt{2}\pi} \int_{-\infty}^{+\infty} \exp(tx^2 - x^2/2)dx \\ &= (1-2t)^{-1/2} \end{aligned}$$

after a straightforward integration.

Obtain pdf of $Y = X^2$ and find $M_Y(t)$.

Examples

Example

Suppose that X has **(normal) distribution** $N(\mu, \sigma^2)$. *Hence*

$$M_X(t)) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{tx} e^{\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right)} dx.$$

$$\begin{aligned}
M_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{t(\sigma s+\mu)} e^{-s^2/2} ds \qquad (\textit{by letting } \frac{x-\mu}{\sigma} = s) \\
&= e^{t\mu+\sigma^2 t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\frac{-(s-\sigma t)^2}{2}} ds. \\
&= e^{t\mu+\sigma^2 t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-v^2/2} dv \qquad (\textit{by letting } s - \sigma t = v) \\
&= e^{(t\mu+\sigma^2 t^2/2)}
\end{aligned}$$

Example

Let X have a **Gramma distribution** *with parameters α and r. Then*

$$\begin{aligned} M_X(t) &= \frac{\alpha}{\Gamma(r)} \int_0^{\infty} e^{tx} (\alpha x)^{r-1} e^{-\alpha x} dx \\ &= \frac{\alpha^r}{\Gamma(r)} \int_0^{\infty} x^{r-1} e^{-x(\alpha - t)} dx \end{aligned}$$

(This integral converges provided $\alpha > 1$) Let $x(\alpha - t) = u$

$$\begin{aligned} M_X(t) &= \frac{\alpha^r}{(\alpha - t)\Gamma(r)} \int_0^{\infty} \left(\frac{u}{\alpha - t}\right)^{r-1} e^{-u} dx \\ &= \left(\frac{\alpha}{\alpha - t}\right)^r \frac{1}{\Gamma(r)} \int_0^{\infty} u^{r-1} e^{-u} du \\ &= \left(\frac{\alpha}{\alpha - t}\right)^r \{ \textit{ because the integral equals } \Gamma(r).\} \end{aligned}$$

MGF of chi-square distribution

Since the **chi-square** distribution is obtained as a special case of the Gamma distribution by letting $\alpha = 1/2$ and $r = n/2$ (n a positive integer), we have that if Z has distribution χ_n^2, then

$$M_Z(t) = (1 - 2t)^{-n/2}.$$

Reason for calling M_X the moment generating function.

The infinite series converges

$$M_X(t) = 1 + tE(X) + \frac{t^2 E(X^2)}{2!} + \cdots + \frac{t^n E(X^n)}{n!} + \ldots$$

under fairly general condition. We shall assume that the required conditions are satisfied and proceed accordingly. Since M_X is a function of the real variable t, we may consider taking the derivative of $M_X(t)$ with respect to t.

The infinte series

$$M_X(t) = E(X) + tE(X^2) + \frac{t^2 E(X^3)}{2!} + \cdots + \frac{t^{n-1} E(X^n)}{(n-1)!} + \dots$$

converges under fairly general condition. We shall assume that the required conditions are satisfied and proceed accordingly.

Setting $t = 0$ we find that only the first term survives and we have $M'(0) = E(X)$. Continuing in this manner, we obatin [assuming that $M^{(n)}(0)$ exists] the following theorem.

Theorem

$$M^{(n)}(0) = E(X^n)$$

That is, the nth derivative of $M_X(t)$ evaluated at $t = 0$ yields $E(X^n)$.

Applying Maclaurin's series expansion to the function M_X, we may wirte

$$\begin{aligned} M_X(t) &= M_X(0) + M'_X(0)t + \cdots + \frac{M_X^{(n)}(0)t^n}{n!} + \cdots \\ &= 1 + \mu_1 t + \frac{\mu_2 t^2}{2!} + \cdots + \frac{\mu_n t^n}{n!} + \cdots \end{aligned}$$

where $\mu_i = E(X^i), \quad i = 1, 2, \ldots$ In particular,

$$V(X) = E(X^2) - (E(X)^2) = M''(0) - [M'(0)]^2$$

Why the above methods should be fruitful at all?

Would it not be simpler (and more straightforward) to compute the moments of X directly, rather than first obtain the mgf and then differentiate it?

The answer is that for many porblems the latter approach is more easily followed.

Moment Problem

Using the moment generating function, we can now show, at least in the case of a discrete random variable with finite range, that its distribution function is completely determined by **all the moments of X.**

Theorem

Let X be a discrete random variable with finite range $\{x_1, x_2, \ldots, x_n\}$, and moments $\mu_k = E(X^k)$. Then the moment series

$$g(t) = \sum_{k=1}^{\infty} \frac{\mu_k t^k}{k!}$$

converges for all t to an infinitely differentiable function $g(t)$.

Proof of the theorem

We know that $\mu_k = \sum_{j=1}^{n} (x_j)^k p(x_j)$. If we set $M = \max |x_j|$, then we have

$$|\mu_k| \leq \sum_{j=1}^{n} |x_j|^k p(x_j) \leq M^k . \sum_{j=1}^{n} p(x_j) = M^k .$$

Hence, for all N we have

$$\sum_{k=0}^{N} \left| \frac{\mu_k t^k}{k!} \right| \leq \sum_{k=0}^{N} \frac{(M|t|)^k}{k!} \leq e^{M|t|},$$

which shows that the moment series converges for all t. Since it is a power series, we know that its sum is infinitely differentiable.

This shows that the μ_k determines $g(t)$. Conversely, since $\mu_k = g^{(k)}(0)$, we see that $g(t)$ determines the μ_k.

Theorem

Let X be a discrete random variable with finite range $\{x_1, x_2, \ldots, x_n\}$, distribution function p, and moment generating function g. Then g is uniquely determined by p, and conversely.

If we delete the hypothesis that X have finite range in the above theorem, then the conclusion is no longer necessarily true.

Generating Functions for Continuous Densities

We introduced the concepts of moments and moment generating functions for discrete random variables. These concepts have natural analogues for continuous random variables, **provided some care is taken in arguments involving convergence**.

If X is a continuous random variable defined on the probability space Ω, with density function f_X, then we define the nth **moment** of X by the formula

$$\mu_n = E(X^n) = \int_{-\infty}^{\infty} x^n f_X(x)\ dx,$$

provided the integral

$$\mu_n = E(X^n) = \int_{-\infty}^{\infty} |x|^n f_X(x)\ dx,$$

is finite. Then, just as in the discrete case, we see that $\mu_0 = 1, \mu_1 = \mu$, and $\mu_2 - \mu_1^2 = \sigma^2$.

Example

Using the mgf, evaluate $E(X)$ and $V(X)$ when

1. X has binomial distribution
2. X has geometric distribution
3. X has Poisson distribution with mean λ

Solution.

1. Suppose that X has a binomial distribution with parameters n and p. Hence, $M_X(t) = [pe^t + q]^n$.

$$\begin{aligned} E(X) &= M'(0) = np \\ E(X^2) &= M''(0) = np[(n-1)p+1]. \\ V(X) &= M''(0) - [M'(0)]^2 = np(1-p). \end{aligned}$$

1. Suppose that X has distribution $N(\alpha, \beta^2)$. Therefore, $M_X(t) = \exp(\alpha + t + \frac{1}{2}\beta^2 t^2)$.

$$M'(0) = \alpha, M''(0) = \beta^2 + \alpha^2, E(X) = \alpha, V(X) = \beta^2.$$

2. Let X have a geometric probability distribution. That is, $P(X = k) = q^{k-1}p, k = 1, 2, \ldots (p + q = 1)$. Thus

$$M_X(t) = \sum_{k=1}^{\infty} e^{tk} q^{k-1} p = \frac{p}{q} \sum_{k=1}^{\infty} (qe^t)k.$$

$$E(X) = p/(1-q)^2 = 1/p, \quad E(X^2) = p(1+q)/(1-q)^3 = (1+q)/p^2$$

and

$$V(X) = (1+q)/p^2 - (1/p)^2 = q/p^2.$$

Theorem

Suppose that the random variable X has mgf M_X. Let $Y = \alpha X + \beta$. Then M_Y, the mgf of random variable Y, is given by

$$M_Y(t) = e^{\beta t} M_X(\alpha t).$$

In words: To find the mgf of $Y = \alpha X + \beta$, evaluate the mgf of X at αt (instead of t) and multiply by $e^{\beta t}$

Theorem

Let X and Y be two random variables with mgf's, $M_X(t)$ and $M_Y(t)$, respectively. If $M_X(t) = M_Y(t)$ for all values of X t, then X and Y have the same probability distribution.

Proof.

It says that if two random variables have the same mgf, then they have the same probability distribution. That is, the mgf uniquely determines the probability distribution of the random variable □

Example

Suppose that X has distribution $N(\mu, \sigma^2)$. Let $Y = \alpha X + \beta$. Then prove that Y is again normally distributed.

Theorem

Suppose that X and Y are independent random variables. Let $Z = X + Y$. Let $M_X(t), M_Y(t)$, and $M_Z(t)$ be the mgf's of the random variables X, Y, and Z respectively. Then $M_Z(t) = M_X(t)M_Y(t)$.

Proof.

$$\begin{aligned} M_Z(t) &= E(e^{Zt}) = E[e^{(X+Y)t}] = E(e^{Xt}e^{Yt}) \\ &= E(e^{Xt})E(e^{Yt}) = M_X(t)M_Y(t) \end{aligned}$$

Note: This theorem may be generalized as follows: If $X_1, X_2, \ldots, X_n$ are independent random variables with mgf's M_{X_i}, $i = 1, 2, \ldots, n$, then M_Z, the mgf of $Z = X_1 + X_2 + \cdots + X_n$, is given by $M_Z(t) = M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t)$.

Conclusion

A **generating function of a random variable** is an expected value of a certain transformation of the variable. Most generating functions share four important properties:

1. **Under mild conditions, the generating function completely determines the distribution of the random variable.**
   Often a random variable is shown to have a certain distribution by showing that the generating function has a certain form. The process of recovering the distribution from the generating function is known as **inversion**.
2. **The generating function of a sum of independent variables is the product of the generating functions.**
   This is frequently used to determine the distribution of a sum of independent variables. By contrast, recall that the probability density function of a sum of independent variables is the **convolution** of the individual density functions, a much more complicated operation.

3. The **moments of the random variable** can be obtained from the derivatives of the generating function.
   Computing moments from the generating function is easier than computing the moments directly from the definition.
4. **Ordinary (pointwise) convergence** of a sequence of generating functions corresponds to the **special convergence of the corresponding distributions**.
   The last property is known as the **continuity theorem**. Often it is easer to show the convergence of the generating functions than to prove convergence of the distributions directly.

Problems

1. The random variable X can assume the values 1 and -1 with probability $\frac{1}{2}$ each. Find
   (a) the moment generating function
   (b) the first four moments about the origin.

2. A random vairable X has density function given by

$$f(x) = \begin{cases} 2e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

   Find the moment generating function and the first four moments about the origin.

3. Find the first four moments about the origin and about the mean, for a random variable X having density function

$$f(x) = \begin{cases} 4x(9 - x^2)/81 & 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

Problems

4. If $M(t)$ is the moment generating function for a random variable X, prove that the mean is $\mu = M'(0)$ and the variance is $\sigma^2 = M''(0) - [M'(0)]^2$.
5. Find the moment generating function of a random vairable X that is binomially distributed.
6. Find the moment generating function for the general normal distribution.
7. Show that the moment generating function of a random variable X, which is chi square distributed with v degrees of freedom, is $M(t) = (1 - 2t)^{-\nu/2}$.

Problems

8. Let X_1 and X_2 be independent random variables that are chi square distributed with ν_1 and ν_2 degrees of freedom, respectively.
   (a) Show that the moment generating function of $Z = X_1 + X_2$ is $(1-2t)^{-(\nu_1+\nu_2)/2}$, thereby
   (b) show that Z is chi square distributed with $\nu_1 + \nu_2$ degrees of freedom.
9. Suppose that X has pdf given by

$$f(x) = 2x, \quad 0 \leq x \leq 1.$$

   (a) Determine the mgf of X.
   (b) Using the mgf, evaluate $E(X)$ and $V(X)$ and check your answer.

Problems

10. Suppose that S, a random voltage, varies between 0 and 1 volt and is uniformly distributed over that interval. Suppose that the signal S is perturbed by an additive, independent random noise N which is uniformly distributed between 0 and 2 volts.
    (a) Find the mgf of the voltage (including noice).
    (b) Using the mgf, obtain the expected value and variance of this voltage.

11. Suppose that X has the following pdf:

$$f(x) = \lambda e{-}\lambda(x-a), \quad x \geq a.$$

This is known as a **two-parameter exponential distribution.**
    (a) Find the mgf of X.
    (b) Using the mgf, find $E(X)$ and $V(X)$.

Problems

12. Let X be the outcome when a fair die is tossed.
    (a) Find the mgf of X.
    (b) Using the mgf, find $E(X)$ and $V(X)$.
13. Suppose that the continuous random variables X had pdf

$$f(x) = e^{-|x|}/2, \quad -\infty < x < \infty.$$

    (a) Obtain the mgf of X.
    (b) Using the mgf, find $E(X)$ and $V(X)$.

Problems

14. Use the mgf to show that if X and Y are independent random variables with distribution $N(\mu_x, \sigma_x^2)$ and $N(\mu_y, \sigma_y^2)$, respectively, then $Z = aX + bY$ is again normally distributed, where a and b are constants.

15. Suppose that the mgf of a random variable X is of the form $M_X(t) = (0.4e^t + 0.6)^8$.
    (a) What is the mgf of the random variable $Y = 3X + 2$?
    (b) Evaluate $E(X)$.
    (c) Can you check your answer to (b) by some other method?

Problems

16. A number of resistances, R_i, $i = 1, 2, \ldots, n$, are put into a series arrangement in a circuit. Suppose that each resistance is normally distributed with $E(R_i) = 10$ ohms and $V(R_i) = 0.16$.
    (a) If $n = 5$, what is the probability that the resistance of the circuit exceeds 49 ohms?
    (b) How large should n be so that the probability that the total resistance exceeds 100 ohms is approximately 0.05?
17. In a circuit n resistances are hooked up into a series arrangement. Suppose that each resistance is uniformly distributed over $[0, 1]$ and suppose, furthermore, that all resistances are independent. Let R be the total resistance.
    (a) Find the mgf of R.
    (b) Using the mgf, obtain $E(R)$ and $V(R)$. Check your answers by direct computation.
18. If X has distribution χ^2_n, using the mgf, show that $E(X) = n$ and $V(X) = 2n$.

Problems

19. Suppose that V, the velocity (cm/sec) of an object, has distribution $N(0,4)$. If $K = mV^2/2$ ergs is the kinetic energy of the object (where $m =$ mass), find the pdf of K. If $m = 10$ grams, evaluate $P(K \leq 3)$.
20. Suppose that the life length of an item is exponentially distributed with parameter 0.5. Assume that 10 such items are installed successively, so that the ith item is installed "immediately" after the $(i-1)$-item has failed. Let T_i be the time to failure of the ith item, $i = 1, 2, \ldots, 10$, always measured from the time of installation. Hence $S = T_1 + \cdots + T_{10}$ represents the total time of functioning of the 10 items. Assuming that the T_i's are independent, evaluate $P(S \geq 15.5)$.
21. Suppose that $X_1, \ldots, X_{80}$ are independent random variables, each having distribution $N(0,1)$. Evaluate $P[X_1^2 + \cdots + X_{80}^2 > 77]$.