COURSE OUTCOME 2

OPERATION ON ONE AND MULTIPLE RANDOM VARIABLE-EXPECTATIONS

Moments and the Moment Generating Function (MGF)

There are various reasons for studying moments and the moment generating functions. One of them that the moment generating function can be used to prove the central limit theorem.

Moments

The k^{th} *moment* of a random variable X is defined as $\mu_k = E(X^k)$. Thus, the mean is the first moment, $\mu = \mu_1$, and the variance can be found from the first and second moments, $\sigma^2 = \mu_2 - \mu_1^2$

The k^{th} *central moment* is defined as $E((X-\mu)^k)$. Thus, the variance is the second central moment.

The higher moments have more obscure meanings as k grows.

A third central moment of the standardized random variable $X^* = (X - \mu)/\sigma$,

$$\beta_3 = E((X^*)^3) = \frac{E((X-\mu)^3)}{\sigma^3}$$

is called the ***skewness*** of X. A distribution that is symmetric about its mean has 0 skewness. (In fact all the odd central moments are 0 for a symmetric distribution.) But if it has a long tail to the right and a short one to the left, then it has a positive skewness, and a negative skewness in the opposite situation.

A fourth central moment of X^*,

$$\beta_4 = E((X^*)^4) = \frac{E((X-\mu)^4)}{\sigma^4}$$

is called ***kurtosis*.** A fairly flat distribution with long tails has a high kurtosis, while a short-tailed distribution has a low kurtosis. A bimodal distribution has a very high kurtosis. A normal distribution has a kurtosis of 3. (The word kurtosis was made up in the early 19th century from the Greek word for curvature.)

It turns out that the whole distribution for X is determined by all the moments, that is different distributions cannot have identical moments. That is what makes moments important.

The ***k***th moment of a random variable X about origin is denoted by μ_k^1 and defined as

$$\mu_k^1 = \begin{cases} \sum_x x^r f(x) & for\ discrete\ random\ variable \\ \int x^r f(x)dx & for\ continuous\ random\ variable \end{cases}$$

Clearly,

$\mu_1^1 = E(X)$, it is also called mean of random variable X and denoted by μ or $\bar{X}$

The ***k^th^*** moment of a random variable X about mean μ or $\bar{X}$ is denoted by μ_k and defined as

$$\mu_k^1 = E[(x-\bar{X})^k] = \begin{cases} \sum_x (x-\overline{X})^r f(x) & \text{for discrete random variable} \\ \int (x-\overline{X})^r f(x)dx & \text{for continuous random variable} \end{cases}$$

Moment Generating Function (MGF)

There is a clever way of organizing all the moments into one mathematical object, and that object is called the *Moment Generating Function (MGF).* It is a function $m(t)$ of a new variable t defined by $m(t) = E(e^{tX})$.

Since the exponential function e^t has the power series

$$e^t = \sum_{k=0}^{\infty} \frac{t^k}{k!} = 1 + t + \frac{t^2}{2!} + \cdots + \frac{t^k}{k!} + \cdots ,$$

we can rewrite $m(t)$ as follows

$$\begin{aligned} & m(t) = E(e^{tX}) \\ = & \; E\left(1 + tX + \frac{(tX)^2}{2!} + \cdots + \frac{(tX)^k}{k!} + \cdots\right) \\ = & \; 1 + tE(X) + \frac{t^2E(X^2)}{2!} + \cdots + \frac{t^kE(X^k)}{k!} + \cdots \\ = & \; 1 + t\mu_1 + \frac{t^2\mu_2}{2!} + \cdots + \frac{t^k\mu_k}{k!} + \cdots \\ = & \; 1 + \mu_1 t + \frac{\mu_2}{2!}t^2 + \cdots + \frac{\mu_k}{k!}t^k + \cdots \end{aligned}$$

Theorem. The k^{th} derivative of $m(t)$ evaluated at $t = 0$ is the k^{th} moment μ_k of X.

In other words, the MGF generates the moments of X by differentiation.

The primary use of MGF is to develop the theory of probability. For instance, the easiest way to prove the central limit theorem is to use MGF.

For discrete distributions, we can also compute the MGF directly in terms of the probability mass function $f(x) = P(X=x)$ as

$$M_X(t) = E(e^{tx}) = \sum_x e^{tx} f(x)$$

For continuous distributions, the MGF can be expressed in terms of the probability density function *f(x)* as

$$M_X(t) = E(e^{tx}) = \int e^{tx} f(x)\, dx$$

Properties of Moment Generating Functions (MGF)

Translation

If $Y = X + a$, then $m_Y(t) = e^{at} m_X(t)$.

Scaling

If $Y = bX$, then $m_Y(t) = m_X(bt)$.

Standardizing. From the last two properties, if

$$X^* = \frac{X - \mu}{\sigma}$$

is the standardized random variable for X, then

$$m_{X*}(t) = e^{-\mu t/\sigma} m_X({}^{t}/_{\sigma})$$

Convolution. If X and Y are independent variables, and $Z = X + Y$, then

$$m_Z(t) = m_X(t). m_Y(t)$$

Note that this property of convolution on MGF implies that for a sample sum $S_n = X_1 + X_2 + \cdots + X_n$, the MGF is $m_{Sn}(t) = (m_X(t))^n$.

We can couple that with the standardizing property to determine the moment generating function for the standardized sum

$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

The moment generating function for a uniform distribution on [0,1].

Let X be uniform on [0,1] so that the probability density function f_X has the value 1 on [0,1] and 0 outside this interval. Let's first compute the moments.

$$\begin{aligned}\mu_n = E(X^n) &= \int_{-\infty}^{\infty} x^n f_X(x)\,dx \\ &= \int_0^1 x^n\,dx \\ &= \left.\frac{x^{n+1}}{n+1}\right|_0^1 = \frac{1}{n+1}\end{aligned}$$

Next, let's compute the moment generating function.

$$\begin{aligned}m(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx \\ &= \int_0^1 e^{tx}\,dx \\ &= \left.\frac{1}{t}e^{tx}\right|_0^1 = \frac{e^t - 1}{t}\end{aligned}$$

Note that the expression for $g(t)$ does not allow $t = 0$ since there is a t in the denominator. Still $g(0)$ can be evaluated by using power series or L'Hôpital's rule.

The moment generating function for an exponential distribution with parameter λ.

Recall that when events occur in a Poisson process uniformly at random over time at a rate of λ events per unit time, then the random variable X giving the time to the first event has an exponential distribution. The density function for X is $f_X(x) = \lambda e^{-\lambda x}$, for $x \in [0,\infty)$.

Let us compute its moment generating function.

$$\begin{aligned}m(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx \\ &= \int_0^{\infty} e^{tx}\lambda e^{-\lambda x}\,dx \\ &= \lambda\int_0^{\infty} e^{(t-\lambda)x}\,dx \\ &= \left.\lambda\frac{e^{(t-\lambda)x}}{t-\lambda}\right|_0^{\infty} \\ &= \left(\lim_{x\to\infty}\lambda\frac{e^{(t-\lambda)x}}{t-\lambda}\right) - \lambda\frac{e^0}{t-\lambda}\end{aligned}$$

Now if $t < \lambda$, then the limit in the last line is 0, so in that case

$$m(t) = \frac{\lambda}{\lambda - t}.$$

This is a minor, yet important point. The moment generating function does not have to be defined for all t. We only need it to be defined for t near 0 because we are only interested in its derivatives evaluated at 0.

The moment generating function for the standard normal distribution

Let Z be a random variable with a standard normal distribution. Its probability density function is

$$f_Z(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}.$$

Its moments can be computed from the definition, but it takes repeated applications of integration by parts to compute

$$\mu_n = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, x^n e^{-x^2/2}\, dx.$$

We will not do that computation here, but it turns out that when n is odd, the integral is 0, so μ_n is 0 if n is odd. On the other hand, when n is even, say $n = 2m$, then it turns out that

$$\mu_{2m} = \frac{(2m)!}{2^m m!}.$$

From these values of all the moments, we can compute the moment generating function.

$$\begin{aligned} m(t) &= \sum_{n=0}^{\infty} \frac{\mu_n}{n!}\, t^n \\ &= \sum_{m=0}^{\infty} \frac{\mu_{2m}}{(2m)!}\, t^{2m} \\ &= \sum_{m=0}^{\infty} \frac{(2m)!}{2^m m!}\, \frac{1}{(2m)!}\, t^{2m} \\ &= \sum_{m=0}^{\infty} \frac{1}{2^m m!}\, t^{2m} \\ &= e^{t^2/2} \end{aligned}$$

Thus, the moment generating function for the standard normal distribution Z is

$$m_Z(t) = e^{t^2/2}$$

More generally, if $X = \sigma Z + \mu$ is a normal distribution with mean μ and variance σ^2, then the moment generating function is

$$g_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

Transformations of Random Variables

We begin with a random variable X and we want to start looking at the random variable $Y = g(X) = g \circ X$ where the function

$$g : \mathbb{R} \to \mathbb{R}.$$

The **inverse image** of a set A,

$$g^{-1}(A) = \{x \in \mathbb{R}; g(x) \in A\}.$$

In other words,

$$x \in g^{-1}(A) \text{ if and only if } g(x) \in A.$$

For example, if $g(x) = x^3$, then $g^{-1}([1, 8]) = [1, 2]$

For the singleton set $A = \{y\}$, we sometimes write $g^{-1}(\{y\}) = g^{-1}(y)$. For $y = 0$ and $g(x) = \sin x$, $g^{-1}(0) = \{k\pi; k \in \mathbb{Z}\}$.

If g is a one-to-one function, then the inverse image of a singleton set is itself a singleton set. In this case, the inverse image naturally defines an inverse function. For $g(x) = x^3$, this inverse function is the cube root. For $g(x) = \sin x$ or $g(x) = x^2$ we must limit the domain to obtain an inverse function.

Exercise 1. *The inverse image has the following properties:*

- $g^{-1}(\mathbb{R}) = \mathbb{R}$
- *For any set A, $g^{-1}(A^c) = g^{-1}(A)^c$*
- *For any collection of sets $\{A_\lambda; \lambda \in \Lambda\}$,*

$$g^{-1}\left(\bigcup_\lambda A_\lambda\right) = \bigcup_\lambda g^{-1}(A).$$

As a consequence the mapping

$$A \mapsto P\{g(X) \in A\} = P\{X \in g^{-1}(A)\}$$

satisfies the axioms of a probability. The associated probability $\mu_{g(X)}$ is called the **distribution** of $g(X)$.

1 Discrete Random Variables

For X a discrete random variable with probabiliity mass function f_X, then the probability mass function f_Y for $Y = g(X)$ is easy to write.

$$f_Y(y) = \sum_{x \in g^{-1}(y)} f_X(x).$$

Example 2. *Let X be a uniform random variable on $\{1, 2, \ldots n\}$, i. e., $f_X(x) = 1/n$ for each x in the state space. Then $Y = X + a$ is a uniform random variable on $\{a + 1, 2, \ldots a + n\}$*

Example 3. *Let X be a uniform random variable on $\{-n, -n + 1, \ldots, n - 1, n\}$. Then $Y = |X|$ has mass function*

$$f_Y(y) = \begin{cases} \frac{1}{2n+1} & \text{if } x = 0, \\ \frac{2}{2n+1} & \text{if } x \neq 0. \end{cases}$$

2 Continuous Random Variable

The easiest case for transformations of continuous random variables is the case of g one-to-one. We first consider the case of g increasing on the range of the random variable X. In this case, g^{-1} is also an increasing function.

To compute the cumulative distribution of $Y = g(X)$ in terms of the cumulative distribution of X, note that

$$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\} = P\{X \leq g^{-1}(y)\} = F_X(g^{-1}(y)).$$

Now use the chain rule to compute the density of Y

$$f_Y(y) = F_Y'(y) = \frac{d}{dy} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y).$$

For g decreasing on the range of X,

$$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\} = P\{X \geq g^{-1}(y)\} = 1 - F_X(g^{-1}(y)),$$

and the density

$$f_Y(y) = F_Y'(y) = -\frac{d}{dy} F_X(g^{-1}(y)) = -f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y).$$

For g decreasing, we also have g^{-1} decreasing and consequently the density of Y is indeed positive,

We can combine these two cases to obtain

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|.$$

Example 4. *Let U be a uniform random variable on $[0, 1]$ and let $g(u) = 1 - u$. Then $g^{-1}(v) = 1 - v$, and $V = 1 - U$ has density*

$$f_V(v) = f_U(1 - v)|-1| = 1$$

on the interval $[0, 1]$ and 0 otherwise.

Example 5. *Let X be a random variable that has a uniform density on $[0,1]$. Its density*

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } 0 \le x \le 1, \\ 0 & \text{if } x > 1. \end{cases}$$

Let $g(x) = x^p,\ p \neq 0$. Then, the range of g is $[0,1]$ and $g^{-1}(y) = y^{1/p}$. If $p > 0$, then g is increasing and

$$\frac{d}{dy} g^{-1}(y) = \begin{cases} 0 & \text{if } y < 0, \\ \frac{1}{p} y^{1/p-1} & \text{if } 0 \le y \le 1, \\ 0 & \text{if } y > 1. \end{cases}$$

This density is unbounded near zero whenever $p > 1$.

If $p < 0$, then g is decreasing. Its range is $[1, \infty)$, and

$$\frac{d}{dy} g^{-1}(y) = \begin{cases} 0 & \text{if } y < 1, \\ -\frac{1}{p} y^{1/p-1} & \text{if } y \ge 1, \end{cases}$$

In this case, Y is a Pareto distribution with $\alpha = 1$ and $\beta = -1/p$. We can obtain a Pareto distribution with arbitrary α and β by taking

$$g(x) = \left(\frac{x}{\alpha}\right)^{1/\beta}.$$

If the transform g is not one-to-one then special care is necessary to find the density of $Y = g(X)$. For example if we take $g(x) = x^2$, then $g^{-1}(y) = \sqrt{y}$.

$$F_y(y) = P\{Y \le y\} = P\{X^2 \le y\} = P\{-\sqrt{y} \le X \le \sqrt{y}\} = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Thus,

$$\begin{aligned} f_Y(y) &= f_X(\sqrt{y}) \frac{d}{dy}(\sqrt{y}) - f_X(-\sqrt{y}) \frac{d}{dy}(-\sqrt{y}) \\ &= \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) \end{aligned}$$

If the density f_X is symmetric about the origin, then

$$f_y(y) = \frac{1}{\sqrt{y}} f_X(\sqrt{y}).$$

Example 6. *A random variable Z is called a* **standard normal** *if its density is*

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}).$$

A calculus exercise yields

$$\phi'(z) = -\frac{1}{\sqrt{2\pi}} z \exp(-\frac{z^2}{2}) = -z\phi(z), \qquad \phi''(z) = \frac{1}{\sqrt{2\pi}} (z^2 - 1) \exp(-\frac{z^2}{2}) = (z^2 - 1)\phi(z).$$

Thus, ϕ has a global maximum at $z = 0$, it is concave down if $|z| < 1$ and concave up for $|z| > 1$. This show that the graph of ϕ has a bell shape.

*$Y = Z^2$ is called a χ^2 (***chi-square***) random variable with one degree of freedom. Its density is*

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} \exp(-\frac{y}{2}).$$

3 The Probability Transform

Let X a continuous random variable whose distribution function F_X is strictly increasing on the possible values of X. Then F_X has an inverse function.

Let $U = F_X(X)$, then for $u \in [0, 1]$,

$$P\{U \le u\} = P\{F_X(X) \le u\} = P\{U \le F_X^{-1}(u)\} = F_X(F_X^{-1}(u)) = u.$$

In other words, U is a uniform random variable on $[0, 1]$. Most random number generators simulate independent copies of this random variable. Consequently, we can simulate independent random variables having distribution function F_X by simulating U, a uniform random variable on $[0, 1]$, and then taking

$$X = F_X^{-1}(U).$$

Example 7. *Let X be uniform on the interval $[a, b]$, then*

$$F_X(x) = \begin{cases} 0 & \text{if } x < a, \\ \frac{x-a}{b-a} & \text{if } a \le x \le b, \\ 1 & \text{if } x > b. \end{cases}$$

Then

$$u = \frac{x-a}{b-a}, \quad (b-a)u + a = x = F_X^{-1}(u).$$

Example 8. *Let T be an exponential random variable. Thus,*

$$F_T(t) = \begin{cases} 0 & \text{if } t < 0, \\ 1 - \exp(-t/\beta) & \text{if } t \ge 0. \end{cases}$$

Then,

$$u = 1 - \exp(-t/\beta), \quad \exp(-t/\beta) = 1 - u, \quad t = -\frac{1}{\beta}\log(1-u).$$

Recall that if U is a uniform random variable on $[0, 1]$, then so is $V = 1 - U$. Thus if V is a uniform random variable on $[0, 1]$, then

$$T = -\frac{1}{\beta}\log V$$

is a random variable with distribution function F_T.

Example 9. *Because*

$$\int_\alpha^x \frac{\beta\alpha^\beta}{t^{\beta+1}}\,dt = -\alpha^\beta t^{-\beta}\Big|_\alpha^x = 1 - \left(\frac{\alpha}{x}\right)^\beta.$$

A Pareto random variable X has distribution function

$$F_X(x) = \begin{cases} 0 & \text{if } x < \alpha, \\ 1 - \left(\frac{\alpha}{x}\right)^\beta & \text{if } x \ge \alpha. \end{cases}$$

Now,

$$u = 1 - \left(\frac{\alpha}{x}\right)^\beta \quad 1 - u = \left(\frac{\alpha}{x}\right)^\beta, \quad x = \frac{\alpha}{(1-u)^{1/\beta}}.$$

As before if $V = 1 - U$ is a uniform random variable on [0,1], then

$$X = \frac{\alpha}{V^{1/\beta}}$$

is a Pareto random variable with distribution function F.