Introduction to Probability
Transforms and transformations

Moment generating function

So far we have two ways to describe a random variable X:

- its PMF p_X (discrete) / its PDF $f(x)$ (continuous)
- its CDF F.

We will now gain a third way: its **moment generating function**.

First, we define the ***n*****th moment** of a random variable X as $E(X^n)$ (which may or may not be finite). The **moment generating function** of X is given by $M_X(t) = E(e^{tX})$.

$$\textbf{discrete}: \; M_X(t) = E(e^{tX}) = \sum_{n=0}^{\infty} e^{tn} p_X(n)$$

$$\textbf{continuous}: \; M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

Examples

For $X \in Geom(p)$,

$$M_X(t) = \sum_{n=1}^{\infty} e^{tn}(1-p)^{n-1}p = \frac{p}{1-p}\sum_{n=1}^{\infty}(e^t(1-p))^n$$

$$= \frac{p}{1-p} \cdot \frac{e^t(1-p)}{1-e^t(1-p)} = \begin{cases} \infty & t \geq -\ln(1-p) \\ \frac{pe^t}{1-e^t(1-p)} & t < -\ln(1-p) \end{cases}$$

For $X \in Exp(\lambda)$,

$$M_X(t) = \int_0^{\infty} e^{tx}\lambda e^{-\lambda x}dx = \frac{\lambda e^{(t-\lambda)x}}{t-\lambda}\bigg|_0^{\infty} = \begin{cases} \infty & t \geq \lambda \\ \frac{\lambda}{\lambda - t} & t < \lambda \end{cases}$$

Properties of the MGF

What's the point of the MGF? As its name implies, it *generates the moments of* X.

Recall, the n**th moment** of X is $E(X^n)$.

To see how we can get $E(X^n)$ from $M_X(t)$, we need some calculus. The MGF $M_X(t)$ is a differentiable function of t (where it is defined).

We are able to switch integrals and derivatives on t and X to isolate the moments. For example, to recover the mean $E(X)$ from $M_X(t)$:

$$\begin{aligned} & M_X(t) = E(e^{tX}) \\ \implies & M_X'(t) = \frac{d}{dt}E(e^{tX}) = E\left(\frac{d}{dt}e^{tX}\right) = E\left(Xe^{tX}\right) \\ \implies & M_X'(0) = E\left(Xe^{0X}\right) = E(X). \end{aligned}$$

Properties of the MGF

Theorem

In general, if $M_X^{(n)}(t) = \frac{d^n}{dt^n} M_X(t)$*, then* $M_X^{(n)}(0) = E(X^n)$*.*

Proof This is easily seen: each successive derivative with respect to t brings down another copy of X in the MGF, since X is not a function of t, so it acts as a constant multiple upon successive differentiation with respect to t.

$$M_X^{(n)}(t) = E(X^n e^{tX}).$$

Thus,

$$M_X^{(n)}(0) = E(X^n e^{0X}) = E(X^n). \blacksquare$$

Calculating moments with the MGF

Using the MGF, what is the variance of $X \sim Exp(3)$?
From earlier, if $X \sim Exp(\lambda)$,

$$M_X(t) = \begin{cases} \infty & t \geq \lambda \\ \frac{\lambda}{\lambda - t} & t < \lambda \end{cases}$$

The first two moments and variance of X are:

$$E(X) = M_X'(0) = \left(\frac{3}{(3-t)^2} \right) \Big|_{t=0} = \frac{1}{3}$$

$$E(X^2) = M_X''(0) = \left(\frac{6}{(3-t)^3} \right) \Big|_{t=0} = \frac{2}{9}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}.$$

(In general, for $X \sim Exp(\lambda)$, $E(X) = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$.)

Independent random variables implies expectation factors

If two random variables, X and Y, are independent*, then the expected value of their product splits:

$$X \perp Y \implies E(XY) = E(X)E(Y),$$

and, in general, if g and h are functions,

$$X \perp Y \implies E(g(X)h(Y)) = E(g(X))E(h(Y)).$$

THE CONVERSE IS NOT TRUE IN GENERAL!

$$E(XY) = E(X)E(Y) \text{ does NOT imply that } X \perp Y.$$

---

*We will define this more rigorously later; for now, stick with the intuition of "independent experimental trials".

Independent random variables implies expectation factors

Example

Let X be a random variable, and $Y = X$. Clearly, X and Y are dependent on each other. In fact, the only way we get

$$E(XY) = E(X^2) = E(X)E(Y) = E(X)^2$$

is if X is a constant. (Why?)

Properties of MGF

Let $M_X(t) = E(e^{tX})$ be the moment generating function (MGF) of the random variable X. Then:

- If $X = c \in \mathbb{R}$ is a constant (not random), then

$$M_X(t) = E(e^{tX}) = e^{tc}.$$

- If $a, b \in \mathbb{R}$, and $Y = aX + b$, then

$$M_Y(t) = E(e^{tY}) = E(e^{taX+tb}) = e^{tb}E(e^{taX}) = e^{tb}M_X(ta).$$

Properties of MGF

- If $X_1, X_2, X_3, ..., X_n$ are IID, and $Y_n = \sum_{j=1}^{n} X_j$, then all the X_j have the same MGF, $m_{X_1}(t)$, and

$$\begin{aligned} M_{Y_n}(t) &= M_{\sum_{j=1}^{n} X_j}(t) = E\left(e^{t\sum_{j=1}^{n} X_j}\right) \\ &= E\left(\prod_{j=1}^{n} e^{tX_j}\right) = \prod_{j=1}^{n} E\left(e^{tX_j}\right) = \prod_{j=1}^{n} m_{X_j}(t) = M_{X_1}(t)^n. \end{aligned}$$

Equivalence of two RVs in distribution

Let X and Y be two random variables on $(\Omega, \mathcal{F}, P)$.
We say that X and Y are **equal in distribution**, denoted in any of the following ways:

$$X \overset{d}{=} Y,\ X \overset{(d)}{=} Y,\ X \overset{\mathcal{D}}{=} Y,$$

if for any[†] subset $B \subseteq \mathbb{R}$,

$$P(X \in B) = P(Y \in B).$$

Note that X and Y need not be equal on an outcome-by-outcome basis to be equal in distribution: consider the fair coin flip example

$$X(H) = Y(T) = 1,\ X(T) = Y(H) = -1$$

[†] Technically, only $B \in \mathcal{B}(\mathbb{R})$, the **Borel** subsets of $\mathbb{R}$, are needed, which then imply the statement for all **Lebesgue measurable** subsets of $\mathbb{R}$.

MGF uniquely identifies a distribution (just like CDF)

The moment generating function of a random variable X is a unique way to describe X, just as the CDF is.

We can calculate the MGF of various functions of random variables to see if their MGFs tell us anything interesting.

Example

Let $X_1, X_2, X_3, ... \sim Bern(p)$ be IID.
What is the distribution of $Y_n = \sum_{j=1}^{n} X_j$?

Answer: $Y_n \sim Bin(n, p)$... but how do we know?
We'll prove it via MGFs.

Sum of n IID Bernoulli(p) is Binomial(n,p): Proof via MGF

Proof First, we'll calculate the MGFs of $X_1 \sim Bern(p)$ and $B_n \sim Bin(n, p)$.

Recalling that the PMF of $X_1 \sim Bern(p)$ is

$$p_{X_1}(x) = p1_{\{1\}}(x) + (1-p)1_{\{0\}}(x),$$

we get the MGF of X_1:

$$M_{X_1}(t) = E(e^{X_1 t}) = pe^{1t} + (1-p)e^{0t} = pe^t + 1 - p.$$

Sum of n IID Bernoulli(p) is Binomial(n,p): Proof via MGF

Next, recalling that the PMF of $B_n \sim Bin(n,p)$ is

$$p_{B_n}(x) = \sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-j} 1_{\{j\}}(x), \;\; x = 0, 1, 2, ..., n,$$

we get the MGF of B_n:

$$\begin{aligned} M_{B_n}(t) = E(e^{B_n t}) &= \sum_{j=0}^{n} e^{jt} \binom{n}{j} p^j (1-p)^{n-j} \\ &= \sum_{j=0}^{n} \binom{n}{j} (pe^t)^j (1-p)^{n-j} \\ \text{(binomial theorem)} \;\; &= (pe^t + 1 - p)^n. \end{aligned}$$

Sum of n IID Bernoulli(p) is Binomial(n,p): Proof via MGF

Therefore, since

$$M_{X_1}(t) = E(e^{X_1 t}) = pe^{1t} + (1-p)e^{0t} = pe^t + 1 - p.$$

and

$$M_{B_n}(t) = E(e^{B_n t}) = (pe^t + 1 - p)^n,$$

we can easily see that

$$M_{B_n}(t) = M_{X_1}(t)^n.$$

However, we know that, since $Y_n = \sum_{j=1}^n X_j$, we also have

$$M_{Y_n}(t) = M_{\sum_{j=1}^n X_j}(t) = m_{X_1}(t)^n.$$

Therefore,

$$M_{Y_n}(t) = M_{B_n}(t) \implies Y_n \sim B_n \sim Bin(n,p).$$

MGF uniquely identifies a distribution (just like CDF)

This last statement is true because of the following theorem:

Theorem

If X and Y are two RVs on $(\Omega, \mathcal{F}, P)$, and $\exists \delta > 0$ such that

$$M_X(t) = M_Y(t) < \infty$$

for all $t \in (-\delta, \delta)$, then $X \overset{d}{=} Y$.

Functions of Continuous RV: CDF, PDF

If X is an RV, and $g : \mathbb{R} \to \mathbb{R}$ is a function, then $Y = g(X)$ is also a RV.

The CDF of $Y = g(X)$ can be calculated relatively easily if g is an *invertible* function. For example, if g is increasing, g^{-1} is as well:

$$\begin{aligned} F_Y(a) = P(Y \leq a) &= P(g(X) \leq a) \\ &= P(X \leq g^{-1}(a)) = F_X(g^{-1}(a)), \end{aligned}$$

or if g is decreasing, note the flip in the inequality:

$$\begin{aligned} F_Y(a) = P(Y \leq a) &= P(g(X) \leq a) \\ &= P(X \geq g^{-1}(a)) = 1 - F_X(g^{-1}(a)). \end{aligned}$$

Functions of Continuous RV: CDF, PDF

To get the PDF of $Y = g(X)$ from X, differentiate the CDF: in either case, the chain rule states

$$f_Y(a) = f_x(g^{-1}(a))|(g^{-1})'(a)|.$$

For non-invertible functions, this is still sometimes possible but must be calculated on a case-by-case basis.

Transforming $Unif(0, 1)$ into other RVs

Recall, a **standard uniform random variable** is a continuous RV with distribution $U \sim Unif(0, 1)$.

This kind of RV is particularly easy to transform into other kinds of RVs; in fact, this is the basis for **Monte Carlo simulation**.

Transforming $\mathit{Unif}(0,1)$ into $\mathit{Unif}(a,b)$

Example

If $U \sim \mathit{Unif}(0,1)$, then, for any $a, b \in \mathbb{R}$ such that $a < b$,

$$V = a + (b-a)U \sim \mathit{Unif}(a,b).$$

Why? Look at F_V:

$$F_V(x) = P(V \leq x) = P(a + (b-a)U \leq x) = P\left(U \leq \frac{x-a}{b-a}\right)$$
$$= \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b, \end{cases}$$

which is precisely the CDF of $\mathit{Unif}(a,b)$.

Transforming $Unif(a, b)$ into something else

For $X \sim Unif(4, 10)$, what are the CDF and PDF of $Y = X^3 - 50$?

Our function is $g(x) = x^3 - 50$, so that $Y = g(X)$.

g is an increasing function, with inverse function

$$x = g^{-1}(y) = (y + 50)^{1/3}.$$

The CDF and PDF of X are

$$F_X(a) = \frac{a - 4}{6} 1_{(4,10)}(a) + 1_{[10,\infty)}(a),$$
$$f_X(a) = \frac{1}{6} 1_{(4,10)}(a).$$

Transforming $Unif(a, b)$ into something else

Hence, the CDF and PDF of Y are

$$\begin{aligned} F_Y(a) &= F_X((a+50)^{1/3}) \\ &= \frac{(a+50)^{1/3}-4}{6} 1_{(4,10)}((a+50)^{1/3}) \\ &= \frac{(a+50)^{1/3}-4}{6} 1_{(14,950)}(a) + 1_{[950,\infty)}(a) \end{aligned}$$

$$\begin{aligned} f_Y(a) &= f_X((a+50)^{1/3}) |(a+50)^{1/3})'| \\ &= \frac{1}{6} 1_{(14,950)}(a) \left(\frac{1}{3}(a+50)^{-2/3} \right). \end{aligned}$$

Inverse Transform Method

Let us go in the other direction. If you have a target distribution V, what function $y = g(x)$ transforms $U \sim \mathit{Unif}(0, 1)$ to $V = g(U)$?

Assume V is a continuous RV. We will discover an invertible function g via the **inverse transform method**.

Noting that we require, $\forall x \in \mathbb{R}$, $0 \leq g^{-1}(x) \leq 1$, we have the CDF of V in the form

$$
\begin{aligned}
F_V(x) &= P(V \leq x) = P(g(U) \leq x) \\
&= P(U \leq g^{-1}(x)) = \begin{cases} 1 & \text{if } x > \max(V), \\ g^{-1}(x) & \min(V) \leq x \leq \max(V), \\ 0 & x < \min(V). \end{cases}
\end{aligned}
$$

Inverse Transform Method

Thus, we have the **Inverse Transform Method**:

The increasing, invertible function g that transforms

$$U \sim Unif(0, 1)$$

into the continuous random variable

$$V = g(U)$$

is the inverse of the CDF of V; that is, g is V's quantile function.

$$V = g(U) \iff g(p) = F_V^{-1}(p) = Q_V(p),\ 0 \leq p \leq 1.$$

Transforming $Unif(0,1)$ into $Exp(\lambda)$

Example

What function g turns $U \sim Unif(0,1)$ into $V \sim Exp(\lambda)$?

$$F_V(x) = (1 - e^{-\lambda x})1_{(0,\infty)}(x) \implies g(p) = Q_V(p) = -\frac{\ln(1-p)}{\lambda}.$$

Thus, $U \sim Unif(0,1) \implies V = -\frac{\ln(1-U)}{\lambda} \sim Exp(\lambda)$.

Check: If $0 < x < \infty$,

$$\begin{aligned} F_V(x) = P(V \leq x) = P\left(-\frac{\ln(1-U)}{\lambda} \leq x\right) &= P\left(\ln(1-U) \geq -\lambda x\right) \\ = P\left(1 - U \geq e^{-\lambda x}\right) &= P\left(U \leq 1 - e^{-\lambda x}\right) \\ = 1 - e^{-\lambda x}. \checkmark \end{aligned}$$

Transforming $Unif(0,1)$ into a Discrete RV

We can also turn $U \sim Unif(0,1)$ into a discrete RV.‡

If $X(\Omega) = \{a_1, a_2, ..., a_n\}$ for a discrete RV X, we can use U to model X by creating the PMF

$$X = \sum_{i=1}^{n} a_i 1_{A_i}(\omega),$$

where $\{A_1, A_2, ..., A_n\}$ is a partition of $[0,1]$ into subintervals such that the length of A_i is $P(X = a_i)$.

We will order the a_i so that $a_i < a_{i+1}$.

‡This method can be extended to $X(\Omega)$ with a countable number of values, but we will only show a finite example here.

Transforming $Unif(0,1)$ into a Discrete RV

Example

Let X be the discrete RV with PMF

$$p_X(a) = \begin{cases} 0.4 & \text{if } a = 4 \\ 0.25 & a = 12 \\ 0.15 & a = 25 \\ 0.2 & a = 60. \end{cases}$$

U can be used to model X via the function

$$X = 4 \cdot 1_{[0,0.4)}(U) + 12 \cdot 1_{[0.4,0.65)}(U) + 25 \cdot 1_{[0.65,0.8)}(U) + 60 \cdot 1_{[0.8,1)}(U).$$

This is the quantile of X! We do have $X = g(U) = Q_X(U)$.

Properties of Normal Random Variables

- **change of variable**: If $X \sim N(\mu, \sigma^2)$, and $a, b \in \mathbb{R}$, then $aX + b \sim N(a\mu + b, a^2\sigma^2)$.
- The MGF of $X \sim N(\mu, \sigma^2)$ is $m_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$.
- The MGF of $Z \sim N(0, 1)$ is $m_Z(t) = e^{\frac{t^2}{2}}$.

Properties of Normal Random Variables: Chi Square

If $g(x) = x^2$, and $Z \sim N(0,1)$ is a standard normal, then $g(Z) = Z^2$ is called a **chi square** random variable with one degree of freedom.

It is denoted $\chi^2(1)$, and its PDF is

$$f_{\chi^2(1)}(x) = \frac{1}{\sqrt{2\pi x}} e^{-x/2} 1_{(0,\infty)}(x),$$

and, in general, a **chi square** random variable with n degrees of freedom, denoted $\chi^2(n)$, has mean n, variance $2n$, and PDF

$$f_{\chi^2(n)}(x) = \frac{x^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} e^{-x/2} 1_{(0,\infty)}(x).$$