

# STA257

Neil Montgomery | HTML is official | PDF versions good for in-class use only  
Last edited: 2016-11-29 20:05

a new way to completely  
characterize the distribution of a  
random variable

# what is important about a random variable?

One of the main points of this course: we essentially care about the *distribution* of a random variable  $X$ , which maps events  $X \in A$  to probabilities  $P(X \in A)$ .

So far we have the following ways (**only**) to characterize a random variable's distribution, depending on the circumstances:

1. Cumulative distribution function  $F(x) = P(X \leq x)$
2. Survival functions  $S(x) = P(X > x) = 1 - F(x)$
3. (Discrete only) Probability mass functions  $p(x)$
4. (Continuous only) Probability density functions  $f(x)$ .

We use whichever one is most convenient for a given situation.

# mean, variance, and "moments"

$E(X)$  gives *some* (but not all) information about a random variable.

$\text{Var}(X) = E(X^2) - E(X)^2$  gives some *more* (but still not all) information about a random variable. (e.g. a Binomial(20, 0.5) and a  $N(10, 5)$  distribution have the same means and variances, but are not (at all!) the same distributions.)

**Definition:** For integers  $k \geq 0$ ,  $E(X^k)$  (if it exists) is called the *kth moment* of a random variable.

Note: calculating all these moments is not the point at all. The concept itself is what is important.

But a few examples could be  $X \sim \text{Bernoulli}(p)$  and  $Z \sim N(0, 1)$ .

# moment sequence *characterizes*

If turns out (proof beyond this course, although I'll give the reason in a minute) that if *all* moments exist:

$$\{E(X), E(X^2), E(X^3), E(X^4), \dots\}$$

then this sequences *sometimes* gives a *characterization* of the distribution of  $X$ .

(I'll tell you how to know *when* this "sometimes" is, soon.)

But carrying around infinite sequences is not convenient. We need a nice, convenient package for them.

# "generating function"

It often happens in mathematics that some big concept can be characterized by an infinite sequence of numbers.

A common trick is to use those numbers as coefficients in some infinite series (usually constructed to be otherwise useful in some way as well.)

With the moment sequence we shall do exactly that:

$$\begin{aligned} &= 1 + E(X) \frac{t}{1!} + E(X^2) \frac{t^2}{2!} + E(X^3) \frac{t^3}{3!} + \dots \\ &= \sum_{k=0}^{\infty} E(X^k) \frac{t^k}{k!} \end{aligned}$$

Why this particular infinite series? Because I said so.

Let's call this function  $m(t)$ , which is a *generating function* using all moments.

# $m(t)$ is actually an expected value itself

A bit of manipulation could give:

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

Of particular importance is the *existence* of that radius of convergence  $|t| < R$  (although the actual radius  $R > 0$  doesn't matter.)

Summary: *if* all moments exist *and if*  $E(e^{tX})$  is defined with a positive radius of convergence, then we have a neatly packaged object containing *all* the moments.

Theorem: under these conditions,  $m(t)$  completely characterizes the distribution of  $X$ .

Proof: Too hard; depends on complex functions.

# a generating function using moments

Definition:  $m(t) = E(e^{tX})$  is called the *moment generating function* for  $X$ , if it exists in an interval containing 0.

Examples: Bernoilli, Binomial,  $N(0, 1)$

You can extract moments from the mgf using derivatives:

$$\frac{d^k}{dt^k} m(t) = \frac{d^k}{dt^k} E(e^{tX}) = E\left(\frac{d^k}{dt^k} e^{tX}\right) = E(X^k e^{tX})$$

and then setting  $t = 0$ .

Examples...



# the more important use of mgf

Suppose  $X$  and  $Y$  are independent random variables with mgfs  $m_X(t)$  and  $m_Y(t)$ . Then  $W = X + Y$  has mgf:

$$\begin{aligned} m_W(t) &= m_{X+Y}(t) = E(e^{t(X+Y)}) \\ &= E(e^{tX} e^{tY}) \\ &= E(e^{tX}) E(e^{tY}) = m_X(t) m_Y(t) \end{aligned}$$

Corollary: If  $X_1, \dots, X_n$  are independent and  $W = \sum X_i$ , then:

$$m_W(t) = \prod_{i=1}^n m_{X_i}(t)$$

## some sums of r.v.s

If  $X_1, \dots, X_n$  are i.i.d. Bernoulli( $p$ ), then  $\sum X_i \sim \text{Binomial}(n, p)$ ...

If  $X_1, \dots, X_n$  are i.i.d. Geometric( $p$ ), then  $\sum X_i \sim \text{NegBin}(n, p)$ ...

This is fundamentally a "lookup table" technique.

Others (exercises):

- sum of  $n$  independent  $\text{Exp}(\lambda)$  is  $\text{Gamma}(n, \lambda)$
- sum of  $n$  independent  $\text{Poisson}(\lambda)$  is  $\text{Poisson}(n\lambda)$
- sum of  $X_i \sim \text{Binomial}(n_i, p)$  is  $\text{Binomial}(\sum n_i, p)$

# the normal distributions

First, suppose  $X$  has mgf  $m_x(t)$  and  $Y = a + bX$ . What is  $m_y(t)$ ?

So what is the mfg of a general  $N(\mu, \sigma^2)$ ?

Finally, if  $X_1, \dots, X_n$  are independent with  $X_i \sim N(\mu_i, \sigma_i^2)$ ?, what distribution is  $X = \sum X_i$ ?