STA286 Lecture 13

Neil Montgomery

Last edited: 2017-02-08 22:16

a (strange) new way to completely characterize the distribution of a random variable (note: section 7.3 of book)

what is important about a random variable?

One of the main points of this course so far: 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 $F(x) = P(X \leqslant x)$ (you take on faith that this does the job)

what is important about a random variable?

One of the main points of this course so far: 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 $F(x) = P(X \leqslant x)$ (you take on faith that this does the job)
2. (Discrete only) Probability mass functions $p(x)$ (equivalent to cdf)

what is important about a random variable?

One of the main points of this course so far: 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 $F(x) = P(X \leqslant x)$ (you take on faith that this does the job)
2. (Discrete only) Probability mass functions $p(x)$ (equivalent to cdf)
3. (Continuous only) Probability density functions $f(x)$ (equivalent to cdf)

what is important about a random variable?

One of the main points of this course so far: 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 $F(x) = P(X \leqslant x)$ (you take on faith that this does the job)
2. (Discrete only) Probability mass functions $p(x)$ (equivalent to cdf)
3. (Continuous only) Probability density functions $f(x)$ (equivalent to cdf)

what is important about a random variable?

One of the main points of this course so far: 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 $F(x) = P(X \leqslant x)$ (you take on faith that this does the job)
2. (Discrete only) Probability mass functions $p(x)$ (equivalent to cdf)
3. (Continuous only) Probability density functions $f(x)$ (equivalent to cdf)

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

mean, variance, and “moments”

$E(X)$ gives a little bit of information about a random variable.

mean, variance, and “moments”

$E(X)$ gives a little bit of information about a random variable.

$\mathrm{Var}(X) = E(X^2) - E(X)^2$ gives a little bit *more* information about a random variable.

(c.f. Markov’s and Chebyshev’s inequalities.)

mean, variance, and “moments”

$E(X)$ gives a little bit of information about a random variable.

$\mathrm{Var}(X) = E(X^2) - E(X)^2$ gives a little bit *more* information about a random variable.

(c.f. Markov’s and Chebyshev’s inequalities.)

Definition: For integers $k \geqslant 0$, $E\left(X^k\right)$ (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.

the complete moment sequence *characterizes* a distribution

If turns out that if *all* moments exist:

$$\left\{E(X), E\left(X^2\right), E\left(X^3\right), E\left(X^4\right), \ldots\right\}$$

then this sequence *usually* gives a *characterization* of the distribution of X.

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

the complete moment sequence *characterizes* a distribution

If turns out that if *all* moments exist:

$$\left\{E(X), E\left(X^2\right), E\left(X^3\right), E\left(X^4\right), \ldots\right\}$$

then this sequence *usually* gives a *characterization* of the distribution of X.

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

But carrying an around infinite sequence is not convenient—they need to be neatly packaged. Here is the trick:

the complete moment sequence *characterizes* a distribution

If turns out that if *all* moments exist:

$$\left\{E(X), E\left(X^2\right), E\left(X^3\right), E\left(X^4\right), \ldots\right\}$$

then this sequence *usually* gives a *characterization* of the distribution of X.

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

But carrying an around infinite sequence is not convenient—they need to be neatly packaged. Here is the trick:

$$M_X(t) = 1 + E(X)\frac{t}{1!} + E\left(X^2\right)\frac{t^2}{2!} + E\left(X^3\right)\frac{t^3}{3!} + E\left(X^4\right)\frac{t^4}{4!}\cdots$$

the complete moment sequence *characterizes* a distribution

If turns out that if *all* moments exist:

$$\left\{E(X), E\left(X^2\right), E\left(X^3\right), E\left(X^4\right), \ldots\right\}$$

then this sequence *usually* gives a *characterization* of the distribution of X.

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

But carrying an around infinite sequence is not convenient—they need to be neatly packaged. Here is the trick:

$$\begin{aligned} M_X(t) &= 1 + E(X)\frac{t}{1!} + E\left(X^2\right)\frac{t^2}{2!} + E\left(X^3\right)\frac{t^3}{3!} + E\left(X^4\right)\frac{t^4}{4!}\cdots \\ &= E\left(1 + Xt + X^2\frac{t^2}{2!} + X^3\frac{t^3}{3!} + X^4\frac{t^4}{4!} + \cdots\right) \end{aligned}$$

the complete moment sequence *characterizes* a distribution

If turns out that if *all* moments exist:

$$\left\{E(X), E\left(X^2\right), E\left(X^3\right), E\left(X^4\right), \ldots\right\}$$

then this sequence *usually* gives a *characterization* of the distribution of X.

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

But carrying an around infinite sequence is not convenient—they need to be neatly packaged. Here is the trick:

$$\begin{aligned} M_X(t) &= 1 + E(X)\,\frac{t}{1!} + E\left(X^2\right)\frac{t^2}{2!} + E\left(X^3\right)\frac{t^3}{3!} + E\left(X^4\right)\frac{t^4}{4!}\cdots \\ &= E\left(1 + Xt + X^2\frac{t^2}{2!} + X^3\frac{t^3}{3!} + X^4\frac{t^4}{4!} + \cdots\right) \\ &= E\left(e^{tX}\right) \end{aligned}$$

moment "generating function"

Definition: $M_X(t)$ is called the *moment generating function*, or mgf, for X (small print: has to converge on an interval containing 0.)

moment “generating function”

Definition: $M_X(t)$ is called the *moment generating function*, or mgf, for X (small print: has to converge on an interval containing 0.)

Important fact: the mgf (if it exists) characterizes the distribution of X.

moment “generating function”

Definition: $M_X(t)$ is called the *moment generating function*, or mgf, for X (small print: has to converge on an interval containing 0.)

Important fact: the mgf (if it exists) characterizes the distribution of X.

Its less important use is that it can be used to calculate moments:

moment “generating function”

Definition: $M_X(t)$ is called the *moment generating function*, or mgf, for X (small print: has to converge on an interval containing 0.)

Important fact: the mgf (if it exists) characterizes the distribution of X.

Its less important use is that it can be used to calculate moments:

$$\frac{d^r}{dt^r} M_X(t)\bigg|_{t=0}$$

moment "generating function"

Definition: $M_X(t)$ is called the *moment generating function*, or mgf, for X (small print: has to converge on an interval containing 0.)

Important fact: the mgf (if it exists) characterizes the distribution of X.

Its less important use is that it can be used to calculate moments:

$$\frac{d^r}{dt^r} M_X(t)\Big|_{t=0} = \frac{d^r}{dt^r} \left[\int_{-\infty}^{\infty} e^{tx} f(x)\, dx \right]_{t=0}$$

moment "generating function"

Definition: $M_X(t)$ is called the *moment generating function*, or mgf, for X (small print: has to converge on an interval containing 0.)

Important fact: the mgf (if it exists) characterizes the distribution of X.

Its less important use is that it can be used to calculate moments:

$$
\begin{aligned}
\left.\frac{d^r}{dt^r} M_X(t)\right|_{t=0} &= \frac{d^r}{dt^r}\left[\int_{-\infty}^{\infty} e^{tx} f(x)\,dx\right]_{t=0} \\
&= \left.\int_{-\infty}^{\infty} \frac{d^r}{dt^r} e^{tx} f(x)\,dx\right|_{t=0}
\end{aligned}
$$

moment "generating function"

Definition: $M_X(t)$ is called the *moment generating function*, or mgf, for X (small print: has to converge on an interval containing 0.)

Important fact: the mgf (if it exists) characterizes the distribution of X.

Its less important use is that it can be used to calculate moments:

$$
\begin{aligned}
\left.\frac{d^r}{dt^r} M_X(t)\right|_{t=0} &= \frac{d^r}{dt^r}\left[\int_{-\infty}^{\infty} e^{tx} f(x)\,dx\right]_{t=0} \\
&= \left.\int_{-\infty}^{\infty} \frac{d^r}{dt^r} e^{tx} f(x)\,dx\right|_{t=0} \\
&= \left.\int_{-\infty}^{\infty} x^r e^{tx} f(x)\,dx\right|_{t=0} = \int_{-\infty}^{\infty} x^r f(x)\,dx = E(X^r)
\end{aligned}
$$

mgf example

Reconsider the generic "observe until the first defective item" example, in which a defective item is produced with probability p. The pmf of the number of items X is:

$$p(x) = (1-p)^{x-1}\,p \text{ for } x \in \{1, 2, 3, \ldots,\}$$

mgf example

Reconsider the generic "observe until the first defective item" example, in which a defective item is produced with probability p. The pmf of the number of items X is:

$$p(x) = (1-p)^{x-1}\,p \text{ for } x \in \{1, 2, 3, \ldots,\}$$

The mgf of X is:

$$\begin{aligned}
M_X(t) = E\left(e^{tX}\right) &= \sum_{x=1}^{\infty} e^{tx}(1-p)^{x-1}p \\
&= pe^t \sum_{x=1}^{\infty} [e^t(1-p)]^{x-1} \\
&= \frac{pe^t}{1-(1-p)e^t}
\end{aligned}$$

mgf example

Reconsider the generic "observe until the first defective item" example, in which a defective item is produced with probability p. The pmf of the number of items X is:

$$p(x) = (1-p)^{x-1}\, p \text{ for } x \in \{1, 2, 3, \ldots, \}$$

The mgf of X is:

$$\begin{aligned} M_X(t) = E\left(e^{tX}\right) &= \sum_{x=1}^{\infty} e^{tx}(1-p)^{x-1}p \\ &= pe^t \sum_{x=1}^{\infty}[e^t(1-p)]^{x-1} \\ &= \frac{pe^t}{1-(1-p)e^t} \end{aligned}$$

Tedious calculus shows $M_X'(t) = \frac{pe^t}{(1-(1-p)e^t)^2}$, which when evaluated at 0 is $\frac{1}{p}$. (I hope!)

the more important use of of mgfs

Recall that if X and Y are independent, $E(XY) = E(X)E(Y)$. This can extend easily to:

$$E(g(X)h(Y)) = E(g(X))E(h(Y))$$

the more important use of of mgfs

Recall that if X and Y are independent, $E(XY) = E(X)E(Y)$. This can extend easily to:

$$E(g(X)h(Y)) = E(g(X))E(h(Y))$$

Now, consider the random variable $X + Y$ with X and Y independent. What could be said about the *distribution* of $X + Y$? This is a difficult problem!

the more important use of of mgfs

Recall that if X and Y are independent, $E(XY) = E(X)E(Y)$. This can extend easily to:

$$E(g(X)h(Y)) = E(g(X))E(h(Y))$$

Now, consider the random variable $X + Y$ with X and Y independent. What could be said about the *distribution* of $X + Y$? This is a difficult problem!

But the following result will be useful over the next few weeks:

$$M_{X+Y}(t)$$

the more important use of of mgfs

Recall that if X and Y are independent, $E(XY) = E(X)E(Y)$. This can extend easily to:

$$E(g(X)h(Y)) = E(g(X))E(h(Y))$$

Now, consider the random variable $X + Y$ with X and Y independent. What could be said about the *distribution* of $X + Y$? This is a difficult problem!

But the following result will be useful over the next few weeks:

$$M_{X+Y}(t) = E\left(e^{t(X+Y)}\right)$$

the more important use of of mgfs

Recall that if X and Y are independent, $E(XY) = E(X)E(Y)$. This can extend easily to:

$$E(g(X)h(Y)) = E(g(X))E(h(Y))$$

Now, consider the random variable $X + Y$ with X and Y independent. What could be said about the *distribution* of $X + Y$? This is a difficult problem!

But the following result will be useful over the next few weeks:

$$M_{X+Y}(t) = E\left(e^{t(X+Y)}\right) = E\left(e^{tX}e^{tY}\right)$$

the more important use of of mgfs

Recall that if X and Y are independent, $E(XY) = E(X)E(Y)$. This can extend easily to:

$$E(g(X)h(Y)) = E(g(X))E(h(Y))$$

Now, consider the random variable $X + Y$ with X and Y independent. What could be said about the *distribution* of $X + Y$? This is a difficult problem!

But the following result will be useful over the next few weeks:

$$M_{X+Y}(t) = E\left(e^{t(X+Y)}\right) = E\left(e^{tX}e^{tY}\right) = E\left(e^{tX}\right)E\left(e^{tY}\right)$$

the more important use of of mgfs

Recall that if X and Y are independent, $E(XY) = E(X)E(Y)$. This can extend easily to:

$$E(g(X)h(Y)) = E(g(X))E(h(Y))$$

Now, consider the random variable $X + Y$ with X and Y independent. What could be said about the *distribution* of $X + Y$? This is a difficult problem!

But the following result will be useful over the next few weeks:

$$M_{X+Y}(t) = E\left(e^{t(X+Y)}\right) = E\left(e^{tX}e^{tY}\right) = E\left(e^{tX}\right)E\left(e^{tY}\right) = M_X(t)M_Y(t)$$

distributions with names, because they are used all the time to model actual things

a few conventions

Some families of random variables are common enough that they get their own names.

a few conventions

Some families of random variables are common enough that they get their own names.

"The Foo Distribution"

a few conventions

Some families of random variables are common enough that they get their own names.

"The Foo Distribution"

Typically the Foo Distribution will only be specified up to some constants that are called parameters. The parameters often have some meaning or another.

a few conventions

Some families of random variables are common enough that they get their own names.

"The Foo Distribution"

Typically the Foo Distribution will only be specified up to some constants that are called parameters. The parameters often have some meaning or another.

Notation: $X \sim \text{Foo}(\delta, \nu)$

a few conventions

Some families of random variables are common enough that they get their own names.

"The Foo Distribution"

Typically the Foo Distribution will only be specified up to some constants that are called parameters. The parameters often have some meaning or another.

Notation: $X \sim \text{Foo}(\delta, \nu)$

This is pronounced: "X has a Foo distribution with parameters δ and ν."

a few conventions

Some families of random variables are common enough that they get their own names.

"The Foo Distribution"

Typically the Foo Distribution will only be specified up to some constants that are called parameters. The parameters often have some meaning or another.

Notation: $X \sim \text{Foo}(\delta, \nu)$

This is pronounced: "X has a Foo distribution with parameters δ and ν."

Finally, a collection of random variables is often called a "process". We will examine two processes that are central to probability modeling.

the Bernoulli distributions

We've seen this one a few times. X has two outcomes: 0 and 1. We say $X \sim$ Bernoulli(p). The parameter is $p = P(X = 1)$.

the Bernoulli distributions

We've seen this one a few times. X has two outcomes: 0 and 1. We say $X \sim$ Bernoulli(p). The parameter is $p = P(X = 1)$.

pmf: $p(x) = p^x(1-p)^{1-x}, \quad x \in 0, 1$

the Bernoulli distributions

We've seen this one a few times. X has two outcomes: 0 and 1. We say $X \sim$ Bernoulli(p). The parameter is $p = P(X = 1)$.

pmf: $p(x) = p^x(1-p)^{1-x}, \quad x \in 0, 1$

Sometimes $1 - p = q$ is used when convenient.

the Bernoulli distributions

We've seen this one a few times. X has two outcomes: 0 and 1. We say $X \sim$ Bernoulli(p). The parameter is $p = P(X = 1)$.

pmf: $p(x) = p^x(1-p)^{1-x}, \quad x \in 0, 1$

Sometimes $1 - p = q$ is used when convenient.

$E(X) = (0)(1-p) + (1)(p) = p$

the Bernoulli distributions

We've seen this one a few times. X has two outcomes: 0 and 1. We say $X \sim$ Bernoulli(p). The parameter is $p = P(X = 1)$.

pmf: $p(x) = p^x(1-p)^{1-x}, \quad x \in 0, 1$

Sometimes $1 - p = q$ is used when convenient.

$E(X) = (0)(1-p) + (1)(p) = p$

For variance, use $E(X^2) = 0^2(1-p) + 1^2 p = p$, so that
$\mathrm{Var}(X) = p - p^2 = p(1-p) = pq$

the Bernoulli distributions

We've seen this one a few times. X has two outcomes: 0 and 1. We say $X \sim$ Bernoulli(p). The parameter is $p = P(X = 1)$.

pmf: $p(x) = p^x(1-p)^{1-x}, \quad x \in 0, 1$

Sometimes $1 - p = q$ is used when convenient.

$E(X) = (0)(1-p) + (1)(p) = p$

For variance, use $E(X^2) = 0^2(1-p) + 1^2 p = p$, so that
$\mathrm{Var}(X) = p - p^2 = p(1-p) = pq$

$SD(X) = \sqrt{p(1-p)}$

the Bernoulli distributions

We've seen this one a few times. X has two outcomes: 0 and 1. We say $X \sim$ Bernoulli(p). The parameter is $p = P(X = 1)$.

pmf: $p(x) = p^x(1-p)^{1-x}, \quad x \in 0, 1$

Sometimes $1 - p = q$ is used when convenient.

$E(X) = (0)(1-p) + (1)(p) = p$

For variance, use $E(X^2) = 0^2(1-p) + 1^2 p = p$, so that
$\text{Var}(X) = p - p^2 = p(1-p) = pq$

$SD(X) = \sqrt{p(1-p)}$

$M_X(t) = E(e^{tX}) = e^{t \cdot 0}q + e^{t \cdot 1}p = q + e^t p$

Bernoulli process

The Bernoulli(p) distributions are used as models for anything with two possible outcomes.

Bernoulli process

The Bernoulli(p) distributions are used as models for anything with two possible outcomes.

There is usually more than one at a time. A *Bernoulli process* is a sequence of independent Bernoulli(p) random variables with the same p:

$$X_1, X_2, X_3, X_4 \dots$$

Often each X_i is called a (Bernoulli) *trial*, or an *experiment*.

after n trials, how many 1's?

Fix the number of Bernoulli trials at n, and let X be the count of the 1's that occurred. What is the *distribution* of X?

after n trials, how many 1's?

Fix the number of Bernoulli trials at n, and let X be the count of the 1's that occurred. What is the *distribution* of X?

First, X could be any integer between 0 and n.

after n trials, how many 1's?

Fix the number of Bernoulli trials at n, and let X be the count of the 1's that occurred. What is the *distribution* of X?

First, X could be any integer between 0 and n.

To illustrate a probability calculation, fix $n = 4$; we'll consider $P(X = 2)$.

Trial outcomes	Probability

after n trials, how many 1's?

Fix the number of Bernoulli trials at n, and let X be the count of the 1's that occurred. What is the *distribution* of X?

First, X could be any integer between 0 and n.

To illustrate a probability calculation, fix $n = 4$; we'll consider $P(X = 2)$.

Trial outcomes	Probability
1 1 0 0	$p^2(1-p)^2$

after n trials, how many 1's?

Fix the number of Bernoulli trials at n, and let X be the count of the 1's that occurred. What is the *distribution* of X?

First, X could be any integer between 0 and n.

To illustrate a probability calculation, fix $n = 4$; we'll consider $P(X = 2)$.

Trial outcomes	Probability
1 1 0 0	$p^2(1-p)^2$
1 0 1 0	$p(1-p)p(1-p)$

after n trials, how many 1's?

Fix the number of Bernoulli trials at n, and let X be the count of the 1's that occurred. What is the *distribution* of X?

First, X could be any integer between 0 and n.

To illustrate a probability calculation, fix $n = 4$; we'll consider $P(X = 2)$.

Trial outcomes	Probability
1 1 0 0	$p^2(1-p)^2$
1 0 1 0	$p(1-p)p(1-p) = p^2(1-p)^2$

after n trials, how many 1's?

Fix the number of Bernoulli trials at n, and let X be the count of the 1's that occurred. What is the *distribution* of X?

First, X could be any integer between 0 and n.

To illustrate a probability calculation, fix $n = 4$; we'll consider $P(X = 2)$.

Trial outcomes	Probability
1 1 0 0	$p^2(1-p)^2$
1 0 1 0	$p(1-p)p(1-p) = p^2(1-p)^2$
1 0 0 1	$p^2(1-p)^2$
0 1 1 0	$p^2(1-p)^2$
0 1 0 1	$p^2(1-p)^2$
0 0 1 1	$p^2(1-p)^2$

the Binomial(n, p) distributions

There are 6 ways to get 2 1s in 4 trials, so $P(X = 2) = 6p^2(1 - p)^2$.

the Binomial(n, p) distributions

There are 6 ways to get 2 1s in 4 trials, so $P(X = 2) = 6p^2(1 - p)^2$.

In general, the number of ways to get k 1s in n trials is:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

the Binomial(n, p) distributions

There are 6 ways to get 2 1s in 4 trials, so $P(X = 2) = 6p^2(1 - p)^2$.

In general, the number of ways to get k 1s in n trials is:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

If X is the number of 1s in n Bernoulli(p) trials, we say:

$$X \sim \text{Binomial}(n, p)$$

the Binomial(n, p) distributions

There are 6 ways to get 2 1s in 4 trials, so $P(X = 2) = 6p^2(1-p)^2$.

In general, the number of ways to get k 1s in n trials is:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

If X is the number of 1s in n Bernoulli(p) trials, we say:

$$X \sim \text{Binomial}(n, p)$$

pmf: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, 1, \ldots, n\}$

the Binomial(n, p) distributions

There are 6 ways to get 2 1s in 4 trials, so $P(X = 2) = 6p^2(1 - p)^2$.

In general, the number of ways to get k 1s in n trials is:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

If X is the number of 1s in n Bernoulli(p) trials, we say:

$$X \sim \text{Binomial}(n, p)$$

pmf: $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ for $k \in \{0, 1, \dots, n\}$

Is this actually a valid pmf? Yes, if you recall the *Binomial theorem*:

$$(a + b)^n = \sum_{k=0} \binom{n}{k} a^k b^{n-k}$$

Let $a = p$ and $b = 1 - p$.