STA286 Lecture 11

Neil Montgomery

Last edited: 2017-02-03 10:52

reminder "theorem"

$$E(g(X)) = \begin{cases} \sum\limits_{x} g(x)p(x) & : \text{discrete} \\ \\ \int\limits_{-\infty}^{\infty} g(x)f(x)\,dx & : \text{continuous} \end{cases}$$

“constant” random variables

From last time: $E(a + bX) = a + bE(X)$. The left had side is a bit of an odd expression.

“constant” random variables

From last time: $E(a + bX) = a + bE(X)$. The left had side is a bit of an odd expression.

For convenience we can treat constants as random variables. Formally, we can consider a random variable X that is always, say, some real a no matter what. The pmf of X is:

$$P(X = a) = 1$$

and 0 otherwise.

“constant” random variables

From last time: $E(a + bX) = a + bE(X)$. The left had side is a bit of an odd expression.

For convenience we can treat constants as random variables. Formally, we can consider a random variable X that is always, say, some real a no matter what. The pmf of X is:

$$P(X = a) = 1$$

and 0 otherwise.

Informally, for convenience, we dispense with the X notation and just treat the constant a as a random variable, allowing for statements like:

$$E(a) = a$$

expected values and joint distributions

Given X and Y with joint density $f(x, y)$, and given a function $g : \mathbb{R}^2 \to \mathbb{R}$, a sophisticated application of the theorem from last time is:

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy$$

(discrete version is the same, with sums)

first key example

Consider $g(x, y) = x + y$.

$$E(g(X, Y)) = E(X + Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f(x, y) \, dx \, dy$$

first key example

Consider $g(x, y) = x + y$.

$$E(g(X,Y)) = E(X+Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x f(x,y)\,dy\right] dx + \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} y f(x,y)\,dx\right] dy$$

first key example

Consider $g(x, y) = x + y$.

$$
\begin{aligned}
E(g(X, Y)) = E(X + Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f(x, y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x f(x, y)\, dy \right] dx + \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} y f(x, y)\, dx \right] dy \\
&= \int_{-\infty}^{\infty} x \left[\int_{-\infty}^{\infty} f(x, y)\, dy \right] dx + \int_{-\infty}^{\infty} y \left[\int_{-\infty}^{\infty} f(x, y)\, dx \right] dy
\end{aligned}
$$

first key example

Consider $g(x,y) = x + y$.

$$
\begin{aligned}
E(g(X,Y)) = E(X+Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x f(x,y)\,dy\right] dx + \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} y f(x,y)\,dx\right] dy \\
&= \int_{-\infty}^{\infty} x\left[\int_{-\infty}^{\infty} f(x,y)\,dy\right] dx + \int_{-\infty}^{\infty} y\left[\int_{-\infty}^{\infty} f(x,y)\,dx\right] dy \\
&= \int_{-\infty}^{\infty} x f_X(x)\,dx + \int_{-\infty}^{\infty} y f_Y(y)\,dy
\end{aligned}
$$

first key example

Consider $g(x, y) = x + y$.

$$
\begin{aligned}
E(g(X, Y)) = E(X + Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f(x, y) \, dx \, dy \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x f(x, y) \, dy \right] dx + \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} y f(x, y) \, dx \right] dy \\
&= \int_{-\infty}^{\infty} x \left[\int_{-\infty}^{\infty} f(x, y) \, dy \right] dx + \int_{-\infty}^{\infty} y \left[\int_{-\infty}^{\infty} f(x, y) \, dx \right] dy \\
&= \int_{-\infty}^{\infty} x f_X(x) \, dx + \int_{-\infty}^{\infty} y f_Y(y) \, dy \\
&= E(X) + E(Y)
\end{aligned}
$$

second key example

Suppose X and Y are independent. Consider $g(x, y) = xy$.

$$E(g(X, Y)) = E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f(x, y)\, dx\, dy$$

second key example

Suppose X and Y are independent. Consider $g(x, y) = xy$.

$$
\begin{aligned}
E(g(X,Y)) = E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f(x,y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f_X(x) f_Y(y)\, dx\, dy
\end{aligned}
$$

second key example

Suppose X and Y are independent. Consider $g(x, y) = xy$.

$$
\begin{aligned}
E(g(X, Y)) = E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f(x, y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f_X(x) f_Y(y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} x f_X(x)\, dx \int_{-\infty}^{\infty} y f_Y(y)\, dy
\end{aligned}
$$

second key example

Suppose X and Y are independent. Consider $g(x, y) = xy$.

$$
\begin{aligned}
E(g(X, Y)) = E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f(x, y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f_X(x) f_Y(y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} x f_X(x)\, dx \int_{-\infty}^{\infty} y f_Y(y)\, dy \\
&= E(X)E(Y)
\end{aligned}
$$

notation

It is common to use μ as a shorter stand-in for $E(X)$.

I'm not so convinced this is a good idea.

measuring variation

BIG MONEY versus BIG MONEY SCHNAPPS VERSION

$E(X) = E(Y) = 0$

Roll	1, 2	3	4, 5	6
BIG MONEY Outcome X \$	-2	0	1	2
BIG MONEY SCHNAPPS Outcome Y \$	-200	0	100	200
Probability	$2/6$	$1/6$	$2/6$	$1/6$

BIG MONEY versus BIG MONEY SCHNAPPS VERSION

$E(X) = E(Y) = 0$

Roll	1, 2	3	4, 5	6
BIG MONEY Outcome X \$	-2	0	1	2
BIG MONEY SCHNAPPS Outcome Y \$	-200	0	100	200
Probability	2/6	1/6	2/6	1/6

But Y is clearly a riskier game. The distribution is more spread out, by a factor of 100.

BIG MONEY versus BIG MONEY SCHNAPPS VERSION

$E(X) = E(Y) = 0$

Roll	1, 2	3	4, 5	6
BIG MONEY Outcome X \$	-2	0	1	2
BIG MONEY SCHNAPPS Outcome Y \$	-200	0	100	200
Probability	$2/6$	$1/6$	$2/6$	$1/6$

But Y is clearly a riskier game. The distribution is more spread out, by a factor of 100.

The question is—how to measure this difference in variation?

variance

Recall the sample variance, expressed a little differently:

$$s^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 \frac{1}{n-1}$$

This is a weighted sum of squared deviations with weights $w_i = 1/(n-1)$

variance

Recall the sample variance, expressed a little differently:

$$s^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 \frac{1}{n-1}$$

This is a weighted sum of squared deviations with weights $w_i = 1/(n-1)$

The theoretical analogue is called the *variance*:

$$\mathrm{Var}(X) = E\left((X - E(X))^2\right) = E\left((X - \mu)^2\right)$$

provided the expectations all exist.

variance

Recall the sample variance, expressed a little differently:

$$s^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 \frac{1}{n-1}$$

This is a weighted sum of squared deviations with weights $w_i = 1/(n-1)$

The theoretical analogue is called the *variance*:

$$\mathrm{Var}(X) = E\left((X - E(X))^2\right) = E\left((X - \mu)^2\right)$$

provided the expectations all exist.

It is also possible to view this as an application of the $E(g(X))$ theorem with $g(x) = (x - \mu)^2$.

examples - BIG MONEY

$$\mathrm{Var}(X) = (-2-0)^2\frac{2}{6} + (0-0)^2\frac{1}{6} + (1-0)^2\frac{2}{6} + (2-0)^2\frac{1}{6} = \frac{14}{6}$$

examples - BIG MONEY

$$\mathrm{Var}(X) = (-2-0)^2\frac{2}{6} + (0-0)^2\frac{1}{6} + (1-0)^2\frac{2}{6} + (2-0)^2\frac{1}{6} = \frac{14}{6}$$

Schnapps version:

$$\mathrm{Var}(Y) = (-200-0)^2\frac{2}{6} + (0-0)^2\frac{1}{6} + (100-0)^2\frac{2}{6} + (200-0)^2\frac{1}{6} = \frac{140000}{6}$$

how to actually calculate variance

Using the two rules $E(a + bX) = a + bE(X)$ and $E(X + Y) = E(X) + E(Y)$ we can derive the better way to calculate variance:

$$\begin{aligned}
\mathrm{Var}(X) &= E\Big((X - \mu)^2\Big) \\
&= E\Big(X^2 - 2X\mu + \mu^2\Big)
\end{aligned}$$

how to actually calculate variance

Using the two rules $E(a+bX) = a + bE(X)$ and $E(X+Y) = E(X) + E(Y)$ we can derive the better way to calculate variance:

$$\begin{aligned}
\text{Var}(X) &= E\Big((X-\mu)^2\Big) \\
&= E\Big(X^2 - 2X\mu + \mu^2\Big) \\
&= E\Big(X^2\Big) - 2E(X)\mu + \mu^2
\end{aligned}$$

how to actually calculate variance

Using the two rules $E(a + bX) = a + bE(X)$ and $E(X + Y) = E(X) + E(Y)$ we can derive the better way to calculate variance:

$$
\begin{aligned}
\text{Var}(X) &= E\Big((X - \mu)^2\Big) \\
&= E\Big(X^2 - 2X\mu + \mu^2\Big) \\
&= E\Big(X^2\Big) - 2E(X)\mu + \mu^2 \\
&= E\Big(X^2\Big) - \mu^2
\end{aligned}
$$

how to actually calculate variance

Using the two rules $E(a + bX) = a + bE(X)$ and $E(X + Y) = E(X) + E(Y)$ we can derive the better way to calculate variance:

$$
\begin{aligned}
\mathrm{Var}(X) &= E\Big((X - \mu)^2\Big) \\
&= E\Big(X^2 - 2X\mu + \mu^2\Big) \\
&= E\Big(X^2\Big) - 2E(X)\mu + \mu^2 \\
&= E\Big(X^2\Big) - \mu^2
\end{aligned}
$$

Or as I prefer: $\mathrm{Var}(X) = E(X^2) - E(X)^2$

another variance example

Gas pipes revisited:

	X			
Y	1	1.5	1.75	Marginal
0.5	0.075	0.100	0.150	0.325
1	0.110	0.080	0.090	0.280
2	0.160	0.140	0.095	0.395
Marginal	0.345	0.320	0.335	1.000

another variance example

Gas pipes revisited:

Y	X 1	1.5	1.75	Marginal
0.5	0.075	0.100	0.150	0.325
1	0.110	0.080	0.090	0.280
2	0.160	0.140	0.095	0.395
Marginal	0.345	0.320	0.335	1.000

$$E(X) = 1 \cdot 0.345 + 1.5 \cdot 0.320 + 1.75 \cdot 0.335 = 1.41125$$

another variance example

Gas pipes revisited:

	X			
Y	1	1.5	1.75	Marginal
0.5	0.075	0.100	0.150	0.325
1	0.110	0.080	0.090	0.280
2	0.160	0.140	0.095	0.395
Marginal	0.345	0.320	0.335	1.000

$$E(X) = 1 \cdot 0.345 + 1.5 \cdot 0.320 + 1.75 \cdot 0.335 = 1.41125$$

$$E(X^2) = 1^2 \cdot 0.345 + 1.5^2 \cdot 0.320 + 1.75^2 \cdot 0.335 = 2.0909375$$

another variance example

Gas pipes revisited:

	X			
Y	1	1.5	1.75	Marginal
0.5	0.075	0.100	0.150	0.325
1	0.110	0.080	0.090	0.280
2	0.160	0.140	0.095	0.395
Marginal	0.345	0.320	0.335	1.000

$$E(X) = 1 \cdot 0.345 + 1.5 \cdot 0.320 + 1.75 \cdot 0.335 = 1.41125$$

$$E(X^2) = 1^2 \cdot 0.345 + 1.5^2 \cdot 0.320 + 1.75^2 \cdot 0.335 = 2.0909375$$

$$\mathrm{Var}(X) = 0.0993109$$

variance non-example

Reconsider X with density $f(x) = 2x^{-3}$ for $x > 1$, and 0 otherwise. Last time we found $E(X) = 2$.

But X does not have a variance, because $E(X^2)$ does not exist.

notation; standard deviation

It is common to use σ^2 as a shorter stand-in for $\mathrm{Var}(X)$.

notation; standard deviation

It is common to use σ^2 as a shorter stand-in for $\mathrm{Var}(X)$.

I'm not so convinced this is a good idea.

notation; standard deviation

It is common to use σ^2 as a shorter stand-in for $\text{Var}(X)$.

I'm not so convinced this is a good idea.

Recall the "unit" problem with the sample variance s^2; hence the use of the sample standard deviation $s = \sqrt{s^2}$.

notation; standard deviation

It is common to use σ^2 as a shorter stand-in for $\mathrm{Var}(X)$.

I'm not so convinced this is a good idea.

Recall the "unit" problem with the sample variance s^2; hence the use of the sample standard deviation $s = \sqrt{s^2}$.

The theoretical analogue is the *standard deviation*, defined as:

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$$

notation; standard deviation

It is common to use σ^2 as a shorter stand-in for $\text{Var}(X)$.

I'm not so convinced this is a good idea.

Recall the "unit" problem with the sample variance s^2; hence the use of the sample standard deviation $s = \sqrt{s^2}$.

The theoretical analogue is the *standard deviation*, defined as:

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

It is common to use σ as a shorter stand-in for $\text{SD}(X)$.

notation; standard deviation

It is common to use σ^2 as a shorter stand-in for $\text{Var}(X)$.

I'm not so convinced this is a good idea.

Recall the "unit" problem with the sample variance s^2; hence the use of the sample standard deviation $s = \sqrt{s^2}$.

The theoretical analogue is the *standard deviation*, defined as:

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

It is common to use σ as a shorter stand-in for $\text{SD}(X)$.

I will leave to you to guess as to whether or not I am convinced this is a good idea.

variance rules

$$\mathrm{Var}(a + bX) = b^2\mathrm{Var}(X)$$

variance rules

$$\mathrm{Var}(a + bX) = b^2\mathrm{Var}(X)$$

What about $\mathrm{Var}(X + Y)$?