# 14.31/14.310 Lecture 9

# Probability---moments of a distribution

Ok, back to probability now.

Where were we? Ah, yes, talking about moments of distributions, expectation, in particular.

# Probability---moments of a distribution

What if, instead of wanting to know a certain feature of the distribution of $X$, say expectation, we are interested, instead in that feature of the distribution of $Y = g(X)$.

# Probability---moments of a distribution

What if, instead of wanting to know a certain feature of the distribution of $X$, say expectation, we are interested, instead in that feature of the distribution of $Y = g(X)$.

Well, we can obviously figure out how $Y$ is distributed---we know how to do that---and then use that distribution to compute, say, $E(Y)$.

# Probability---moments of a distribution

What if, instead of wanting to know a certain feature of the distribution of X, say expectation, we are interested, instead in that feature of the distribution of $Y = g(X)$.

Well, we can obviously figure out how Y is distributed---we know how to do that---and then use that distribution to compute, say, $E(Y)$.

There may be an easier way---it can be shown that

$$E(Y) = E(g(X)) = \int y f_Y(y)\,dy = \int g(x) f_X(x)\,dx$$

# Probability---St. Petersburg paradox

Classic example/paradox in probability theory, but one where economists come out looking particularly good.

This example was first discussed by 18th Century Swiss mathematician Nicolaus Bernoulli and published in the St. Petersburg Academy proceedings in 1738.

# Probability---St. Petersburg paradox

Classic example/paradox in probability theory, but one where economists come out looking particularly good.

This example was first discussed by 18th Century Swiss mathematician Nicolaus Bernoulli and published in the St. Petersburg Academy proceedings in 1738.

Here's the game: I flip a fair coin until it comes up heads. If the number of flips necessary is X, I pay you $2^X$ dollars. How much would you be willing to pay me to play this game?

# Probability---St. Petersburg paradox

Classic example/paradox in probability theory, but one where economists come out looking particularly good.

This example was first discussed by 18th Century Swiss mathematician Nicolaus Bernoulli and published in the St. Petersburg Academy proceedings in 1738.

What's this distribution?

Here's the game: I flip a fair coin until it comes up heads. If the number of flips necessary is X, I pay you $2^X$ dollars. How much would you be willing to pay me to play this game?

# Probability---St. Petersburg paradox

You should be willing to pay your expected winnings, right?

So let's calculate them:

Let $X$ = number of flips required.

# Probability---St. Petersburg paradox

You should be willing to pay your expected winnings, right?

So let's calculate them:

Let $X$ = number of flips required.

(Note that $X \sim G(.5)$ so can look up that $E(X) = 2$.)

# Probability---St. Petersburg paradox

You should be willing to pay your expected winnings, right?

So let's calculate them:

Let $X$ = number of flips required.

(Note that $X \sim G(.5)$ so can look up that $E(X) = 2$.)

Then $Y$ = winnings = $2^X$

# Probability---St. Petersburg paradox

You should be willing to pay your expected winnings, right?

So let's calculate them:

Let $X$ = number of flips required.

(Note that $X \sim G(.5)$ so can look up that $E(X) = 2$.)

Then $Y$ = winnings = $2^X$

$E(Y) = \sum_y y f_Y(y) = \sum_x r(x) f_X(x)$

# Probability---St. Petersburg paradox

You should be willing to pay your expected winnings, right?

So let's calculate them:

Let $X$ = number of flips required.

(Note that $X \sim G(.5)$ so can look up that $E(X) = 2$.)

Then $Y$ = winnings = $2^X$

$$E(Y) = \sum_y y f_Y(y) = \sum_x r(x) f_X(x)$$
$$= \sum_x 2^x (1/2)^x$$

# Probability---St. Petersburg paradox

You should be willing to pay your expected winnings, right?

So let's calculate them:

Let $X$ = number of flips required.

(Note that $X \sim G(.5)$ so can look up that $E(X) = 2$.)

Then $Y$ = winnings = $2^X$

$$E(Y) = \sum_y y f_Y(y) = \sum_x r(x) f_X(x)$$

$$= \sum_x 2^x (1/2)^x$$

$$= \sum_x 1 = \infty$$

# Probability---St. Petersburg paradox

You should be willing to pay your expected winnings, right?

So let's calculate them:

Let $X$ = number of flips required.

(Note that $X \sim G(.5)$ so can look up that $E(X) = 2$.)

Then $Y$ = winnings = $2^X$

$$E(Y) = \sum_y y f_Y(y) = \sum_x r(x) f_X(x)$$

$$= \sum_x 2^x (1/2)^x$$

$$= \sum_x 1 = \infty$$

# Probability---St. Petersburg paradox

No one would be willing to pay me an infinite amount to play this game.

I would guess that I wouldn't have any takers at $20, and that's a lot less than infinity.

That's the paradox, but is it really?

# Probability---St. Petersburg paradox

No one would be willing to pay me an infinite amount to play this game.

I would guess that I wouldn't have any takers at $20, and that's a lot less than infinity.

That's the paradox, but is it really?

Economists know that people have diminishing marginal utility of money. In other words, their valuation of additional money decreases as the amount of money they have increases.

So let $Z$ = valuation of winnings = $\log(Y) = \log(2^X)$

# Probability---St. Petersburg paradox

So let $Z$ = valuation of winnings = $\log(Y) = \log(2^X)$

Then, $E(Z) = \sum_x \log(2^x)(1/2)^x$

# Probability---St. Petersburg paradox

So let $Z$ = valuation of winnings = $\log(Y) = \log(2^X)$

Then, $E(Z) = \sum_x \log(2^x)(1/2)^x$

$= \log(2) \sum_x x(1/2)^x$

# Probability---St. Petersburg paradox

So let $Z$ = valuation of winnings = $\log(Y) = \log(2^X)$

Then, $E(Z) = \sum_x \log(2^x)(1/2)^x$

$= \log(2) \sum_x x(1/2)^x$

$= 2\log(2) < \infty$

# Probability---St. Petersburg paradox

So let $Z$ = valuation of winnings = $\log(Y) = \log(2^X)$

Then, $E(Z) = \sum_x \log(2^x)(1/2)^x$

$= \log(2) \sum_x x(1/2)^x$

$= 2\log(2) < \infty$

So this is only a paradox unless you know a little bit of economics.

# Probability---properties of expectation

1. $E(a) = a$, a constant

# Probability---properties of expectation

1. $E(a) = a$, $a$ constant
2. $E(Y) = aE(X) + b$, $Y = aX + b$

# Probability---properties of expectation

1. $E(a) = a$, $a$ constant
2. $E(Y) = aE(X) + b$, $Y = aX + b$
3. $E(Y) = E(X_1) + E(X_2) + \ldots + E(X_n)$,
   $Y = X_1 + X_2 + \ldots + X_n$

# Probability---properties of expectation

1. $E(a) = a$, a constant
2. $E(Y) = aE(X) + b$, $Y = aX + b$
3. $E(Y) = E(X_1) + E(X_2) + \ldots + E(X_n)$,
   $Y = X_1 + X_2 + \ldots + X_n$

Really, what if the X's aren't independent?

# Probability---properties of expectation

1. $E(a) = a$, $a$ constant
2. $E(Y) = aE(X) + b$, $Y = aX + b$
3. $E(Y) = E(X_1) + E(X_2) + \ldots + E(X_n)$,
   $Y = X_1 + X_2 + \ldots + X_n$

Really, what if the X's aren't independent? Yes, really, they don't have to be independent.

# Probability---properties of expectation

1. $E(a) = a$, $a$ constant
2. $E(Y) = aE(X) + b$, $Y = aX + b$
3. $E(Y) = E(X_1) + E(X_2) + \ldots + E(X_n)$,
   $Y = X_1 + X_2 + \ldots + X_n$
4. $E(Y) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n) + b$,
   $Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n + b$

# Probability---properties of expectation

1. $E(a) = a$, a constant
2. $E(Y) = aE(X) + b$, $Y = aX + b$
3. $E(Y) = E(X_1) + E(X_2) + \ldots + E(X_n)$,
   $Y = X_1 + X_2 + \ldots + X_n$
4. $E(Y) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n) + b$,
   $Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n + b$
5. $E(XY) = E(X)E(Y)$ if $X, Y$ independent

# Probability---another moment: variance

In addition to describing the location, or center, of a distribution of a random variable, we often would like to describe how spread out it is.  There's a moment for that, variance.

$$\mathrm{Var}(X) = E[(X-\mu)^2]$$

# Probability---another moment: variance

In addition to describing the location, or center, of a distribution of a random variable, we often would like to describe how spread out it is. There's a moment for that, variance.

$$\mathrm{Var}(X) = E[(X-\mu)^2]$$

We often denote Var(X) with $\sigma^2$ (Greek "sigma" squared)

# Probability---another moment: variance

In addition to describing the location, or center, of a distribution of a random variable, we often would like to describe how spread out it is. There's a moment for that, variance.

$$Var(X) = E[(X-\mu)^2]$$

We often denote Var(X) with $\sigma^2$ (Greek "sigma" squared)

Note that variance is an expectation, so many of its properties will follow from that.

# Probability---properties of variance

1. Var(X) >= 0

# Probability---properties of variance

1. Var(X) >= 0
2. Var(a) = 0, a constant

# Probability---properties of variance

1. $Var(X) >= 0$
2. $Var(a) = 0$, a constant
3. $Var(Y) = a^2 Var(X)$, $Y = aX + b$

# Probability---properties of variance

1. $Var(X) >= 0$
2. $Var(a) = 0$, a constant
3. $Var(Y) = a^2 Var(X)$, $Y = aX + b$

In other words, shift a distribution and its variance doesn't change. Shrink or spread out a distribution and its variance changes by the square of the multiplicative factor.

# Probability---properties of variance

1. $Var(X) >= 0$
2. $Var(a) = 0$, a constant
3. $Var(Y) = a^2 Var(X)$, $Y = aX + b$
4. $Var(Y) = Var(X_1) + Var(X_2) + . . . + Var(X_n)$,
   $Y = X_1 + X_2 + . . . + X_n$, $X_1, . . . , X_n$ independent

# Probability---properties of variance

1. $Var(X) >= 0$
2. $Var(a) = 0$, a constant
3. $Var(Y) = a^2 Var(X)$, $Y = aX + b$
4. $Var(Y) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$,

   $Y = X_1 + X_2 + \ldots + X_n$, $X_1, \ldots, X_n$ independent

Ah, here we actually need independence.

# Probability---properties of variance

1. $Var(X) >= 0$
2. $Var(a) = 0$, a constant
3. $Var(Y) = a^2 Var(X)$, $Y = aX + b$
4. $Var(Y) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$,
   $Y = X_1 + X_2 + \ldots + X_n$, $X_1, \ldots, X_n$ independent
5. $Var(Y) = a_1^2 Var(X_1) + \ldots + a_n^2 Var(X_n)$,
   $Y = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n + b$, $X_1, \ldots, X_n$ independent

# Probability---properties of variance

1. $Var(X) >= 0$
2. $Var(a) = 0$, a constant
3. $Var(Y) = a^2 Var(X)$, $Y = aX + b$
4. $Var(Y) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$,
   $Y = X_1 + X_2 + \ldots + X_n$, $X_1, \ldots, X_n$ independent
5. $Var(Y) = a_1^2 Var(X_1) + \ldots + a_n^2 Var(X_n)$,
   $Y = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n + b$, $X_1, \ldots, X_n$ independent
6. $Var(X) = E(X^2) - [E(X)]^2$

# Probability---properties of variance

1. $Var(X) >= 0$
2. $Var(a) = 0$, a constant
3. $Var(Y) = a^2 Var(X)$, $Y = aX + b$
4. $Var(Y) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$,
   $Y = X_1 + X_2 + \ldots + X_n$, $X_1, \ldots, X_n$ independent
5. $Var(Y) = a_1^2 Var(X_1) + \ldots + a_n^2 Var(X_n)$,
   $Y = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n + b$, $X_1, \ldots, X_n$ independent
6. $Var(X) = E(X^2) - [E(X)]^2$

This last property can provide a handy way to compute variance.

# Probability---standard deviation

Often it's convenient for the measure of dispersion to have the same units as the random variable. For this reason, we define <u>standard deviation</u>.

$$SD(X) = \sigma = \sqrt{Var(X)} = \sqrt{\sigma^2}$$

# Probability---variance of a function

Since variance is an expectation, we can apply the results of expectation of a function of a random variable to get variance of a function of a random variable.

So if $Y = r(X)$,

$$\mathrm{Var}(Y) = E(Y^2) - E(Y)^2 = E(r(X)^2) - E(r(X))^2$$

$$= \int r(x)^2 f_X(x)\,dx - \left[\int r(x) f_X(x)\,dx\right]^2$$

# Probability---conditional expectation

A <u>conditional expectation</u> is the expectation of a conditional distribution. In other words,

$E(Y|X) = \int y f_{Y|X}(y|x)dy$

Note that $E(Y|X)$ is a function of X, and, therefore, a random variable. $E(Y|X=x)$ is just a number.

# Probability---conditional expectation

A <u>conditional expectation</u> is the expectation of a conditional distribution. In other words,

$E(Y|X) = \int y f_{Y|X}(y|x)dy$

Note that $E(Y|X)$ is a function of X, and, therefore, a random variable. $E(Y|X=x)$ is just a number.

<u>Thm</u> $E(E(Y|X)) = E(Y)$ "Law of Iterated Expectations"

# Probability---conditional variance

The definition of <u>conditional variance</u> follows from that of variance and conditional expectation.

<u>Thm</u> $\mathrm{Var}(E(Y|X)) + E(\mathrm{Var}(Y|X)) = \mathrm{Var}(Y)$

"Law of Total Variance"

# Probability---two laws

"Law of Iterated Expectations"

$E(E(Y|X)) = E(Y)$

"Law of Total Variance"

$Var(E(Y|X)) + E(Var(Y|X)) = Var(Y)$

# Probability---two laws

"Law of Iterated Expectations"

$E(E(Y|X)) = E(Y)$

"Law of Total Variance"

$Var(E(Y|X)) + E(Var(Y|X)) = Var(Y)$

May seem a little mysterious, not clear how they're useful.

# Probability---example

A former student of mine started an innovation incubator in NYC. Suppose he's been doing this for a few years and has kept track of the number of patents produced every year in his incubator. He knows that $E(N) = 2$ and $Var(N) = 2$.

# Probability---example

A former student of mine started an innovation incubator in NYC. Suppose he's been doing this for a few years and has kept track of the number of patents produced every year in his incubator. He knows that $E(N) = 2$ and $Var(N) = 2$.

Let's also suppose that each patent is a commercial success with probability .2, and we can assume independence.

# Probability---example

A former student of mine started an innovation incubator in NYC. Suppose he's been doing this for a few years and has kept track of the number of patents produced every year in his incubator. He knows that $E(N) = 2$ and $Var(N) = 2$.

Let's also suppose that each patent is a commercial success with probability .2, and we can assume independence.

Suppose there are 5 patents this year. What is the probability that 3 are commercial successes?

# Probability---example

Suppose there are 5 patents this year. What is the probability that 3 are commercial successes?

$S|N=n \sim B(n,.2)$,

so $P(S=3|N=5) = 5!/(3!2!).2^3(1-.2)^2$

# Probability---example

Suppose there are 5 patents this year. What is the probability that 3 are commercial successes?

$S|N=n \sim B(n,.2)$,

so $P(S=3|N=5) = 5!/(3!2!).2^3(1-.2)^2$

$= .05$

# Probability---example

Suppose there are 5 patents this year. What is the probability that 3 are commercial successes?

$S|N=n \sim B(n,.2)$,

so $P(S=3|N=5) = 5!/(3!2!).2^3(1-.2)^2$

$= .05$

Suppose there are 5 patents this year. What is the expected number of commercial successes?

# Probability---example

Suppose there are 5 patents this year. What is the probability that 3 are commercial successes?

$S|N=n \sim B(n,.2)$,

so $P(S=3|N=5) = 5!/(3!2!).2^3(1-.2)^2$

$= .05$

Suppose there are 5 patents this year. What is the expected number of commercial successes?

$E(S|N=5) = np = 5x.2 = 1$

# Probability---example

Suppose there are 5 patents this year. What is the probability that 3 are commercial successes?

$S|N=n \sim B(n,.2)$,

so $P(S=3|N=5) = 5!/(3!2!).2^3(1-.2)^2$

$= .05$

Suppose there are 5 patents this year. What is the expected number of commercial successes?

$E(S|N=5) = np = 5x.2 = 1$

How do we get this?

# Probability---example

Suppose there are 5 patents this year. What is the probability that 3 are commercial successes?

$S|N=n \sim B(n, .2)$,

so $P(S=3|N=5) = 5!/(3!2!).2^3(1-.2)^2$

$= .05$

Suppose there are 5 patents this year. What is the expected number of commercial successes?

$E(S|N=5) = np = 5x.2 = 1$

How do we get this? Compute the expectation of a Bernoulli random variable and add it up n times

# Probability---example

What is the (unconditional) expected number of commercial successes?

# Probability---example

What is the (unconditional) expected number of commercial successes? Can use the Law of Iterated Expectations.

# Probability---example

What is the (unconditional) expected number of commercial successes?

$$E(S) = E(E(S|N)) = E(Np) = .2E(N) = .4$$

# Probability---example

What is the (unconditional) variance of number of commercial successes?

# Probability---example

What is the (unconditional) variance of number of commercial successes? Can use the Law of Total Variance.

# Probability---example

What is the (unconditional) variance of number of commercial successes?

$$Var(S) = Var(E(S|N)) + E(Var(S|N))$$

# Probability---example

What is the (unconditional) variance of number of commercial successes?

$$\mathrm{Var}(S) = \mathrm{Var}(E(S|N)) + E(\mathrm{Var}(S|N))$$

$$= \mathrm{Var}(Np) + E(Np(1-p))$$

# Probability---example

What is the (unconditional) variance of number of commercial successes?

$$\begin{aligned}
Var(S) &= Var(E(S|N)) + E(Var(S|N)) \\
&= Var(Np) + E(Np(1-p)) \\
&= .2^2 Var(N) + .2(1-.2)E(N) = .4
\end{aligned}$$

# Probability---covariance and correlation

We now have moments to describe the location, or center, of a distribution of a random variable and how spread out that distribution is. We are often interested in the relationship between random variables, and we have a moment of joint distributions to describe one aspect of that relationship, <u>covariance</u>.

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

# Probability---covariance and correlation

We now have moments to describe the location, or center, of a distribution of a random variable and how spread out that distribution is. We are often interested in the relationship between random variables, and we have a moment of joint distributions to describe one aspect of that relationship, <u>covariance</u>.

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

We often denote $Cov(X,Y)$ with $\sigma_{XY}$

# Probability---covariance and correlation

We now have moments to describe the location, or center, of a distribution of a random variable and how spread out that distribution is. We are often interested in the relationship between random variables, and we have a moment of joint distributions to describe one aspect of that relationship, covariance.

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

And we have a standardized version, correlation.

$$\rho(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]/\sqrt{Var(X)}\sqrt{Var(Y)}$$

# Probability---covariance and correlation

$$\rho(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]/\sqrt{Var(X)}\sqrt{Var(Y)}$$

We say that X&Y are "positively correlated" if $\rho > 0$.

We say that X&Y are "negatively correlated" if $\rho < 0$.

We say that X&Y are "uncorrelated" if $\rho = 0$.

# Probability---properties of covariance

1. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$

# Probability---properties of covariance

1. $Cov(X,X) = Var(X)$
2. $Cov(X,Y) = Cov(Y,X)$

# Probability---properties of covariance

1. $Cov(X,X) = Var(X)$
2. $Cov(X,Y) = Cov(Y,X)$
3. $Cov(X,Y) = E(XY) - E(X)E(Y)$

# Probability---properties of covariance

1. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$
2. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
3. $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$
4. $X, Y$ indep $\longrightarrow \mathrm{Cov}(X,Y) = 0$

# Probability---properties of covariance

1. $Cov(X,X) = Var(X)$
2. $Cov(X,Y) = Cov(Y,X)$
3. $Cov(X,Y) = E(XY) - E(X)E(Y)$
4. $X,Y$ indep $\longrightarrow Cov(X,Y) = 0$
5. $Cov(aX+b, cY+d) = acCov(X,Y)$

# Probability---properties of covariance

1. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$
2. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
3. $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$
4. $X,Y$ indep $\longrightarrow \mathrm{Cov}(X,Y) = 0$
5. $\mathrm{Cov}(aX+b, cY+d) = ac\,\mathrm{Cov}(X,Y)$
6. $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$

# Probability---properties of covariance

1. $Cov(X,X) = Var(X)$
2. $Cov(X,Y) = Cov(Y,X)$
3. $Cov(X,Y) = E(XY) - E(X)E(Y)$
4. $X,Y$ indep $\longrightarrow Cov(X,Y) = 0$
5. $Cov(aX+b, cY+d) = acCov(X,Y)$
6. $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
7. $|\rho(X,Y)| <= 1$

# Probability---properties of covariance

1. $Cov(X,X) = Var(X)$
2. $Cov(X,Y) = Cov(Y,X)$
3. $Cov(X,Y) = E(XY) - E(X)E(Y)$
4. $X,Y$ indep $\longrightarrow Cov(X,Y) = 0$
5. $Cov(aX+b, cY+d) = acCov(X,Y)$
6. $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
7. $|\rho(X,Y)| <= 1$
8. $|\rho(X,Y)| = 1$ iff $Y = aX + b,\ a \neq 0$

# Probability---a preview of regression

We have two random variables, X&Y.

$EX = \mu_X$, $\mathrm{Var}X = \sigma_X^2$

$EY = \mu_Y$, $\mathrm{Var}Y = \sigma_Y^2$

$\rho_{XY} = \mathrm{Cov}(X,Y)/(\sigma_X \sigma_Y)$

# Probability---a preview of regression

We have two random variables, X&Y.

$EX = \mu_X$, $\text{Var}X = \sigma_X^2$

$EY = \mu_Y$, $\text{Var}Y = \sigma_Y^2$

$\rho_{XY} = \text{Cov}(X,Y)/(\sigma_X \sigma_Y)$

We know that, if $\rho_{XY} = 1$ then $Y = a + bX$, $b > 0$, and if
$\rho_{XY} = -1$ then $Y = a + bX$, $b < 0$.

# Probability---a preview of regression

We have two random variables, X&Y.

$EX = \mu_X$, $VarX = \sigma_X^2$

$EY = \mu_Y$, $VarY = \sigma_Y^2$

$\rho_{XY} = Cov(X,Y)/(\sigma_X\sigma_Y)$

We know that, if $\rho_{XY} = 1$ then $Y = a + bX$, $b > 0$, and if $\rho_{XY} = -1$ then $Y = a + bX$, $b < 0$.

If $|\rho_{XY}| < 1$, then we can write $Y = \alpha + \beta X + V$.

# Probability---a preview of regression

We have two random variables, X&Y.

$EX = \mu_X$, $\mathrm{Var}X = \sigma_X^2$

$EY = \mu_Y$, $\mathrm{Var}Y = \sigma_Y^2$

$\rho_{XY} = \mathrm{Cov}(X,Y)/(\sigma_X \sigma_Y)$

We know that, if $\rho_{XY} = 1$ then $Y = a + bX$, $b > 0$, and if $\rho_{XY} = -1$ then $Y = a + bX$, $b < 0$.

If $|\rho_{XY}| < 1$, then we can write $Y = \alpha + \beta X + V$.

V is another random variable, but what can we say about it?

# Probability---a preview of regression

What we can say about $V$ depends on how we define $\alpha$ & $\beta$.

Let $\beta = \rho_{XY}\sigma_Y/\sigma_X$

Let $\alpha = \mu_Y - \beta\mu_X$

# Probability---a preview of regression

What we can say about $V$ depends on how we define $\alpha$ & $\beta$.

Let $\beta = \rho_{XY}\sigma_Y/\sigma_X$

Let $\alpha = \mu_Y - \beta\mu_X$

Then, $V = Y - \alpha - \beta X$ has the following properties:

$E(V) = 0$ and $Cov(X,V) = 0$. (You can show this easily using properties of expectation, variance, and covariance that we've seen.)

# Probability---a preview of regression

What we can say about $V$ depends on how we define $\alpha$ & $\beta$.

Let $\beta = \rho_{XY}\sigma_Y/\sigma_X$

Let $\alpha = \mu_Y - \beta\mu_X$

Then, $V = Y - \alpha - \beta X$ has the following properties:

$E(V) = 0$ and $Cov(X,V) = 0$. (You can show this easily using properties of expectation, variance, and covariance that we've seen.)

We then call $\alpha$ & $\beta$ "regression coefficients," and think of $\alpha + \beta X$ as the part of $Y$ "explained by" $X$ and $V$ as the "unexplained" part.

# Probability---inequalities

Two inequalities involving moments of distributions and tail probabilities often come in handy:

## Markov Inequality

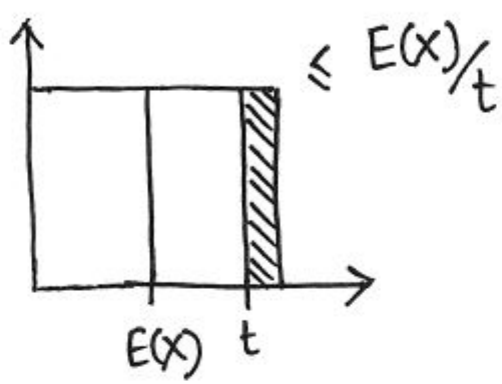
X is a random variable that is always non-negative.
Then for any t > 0, P(X>=t) <= E(X)/t.

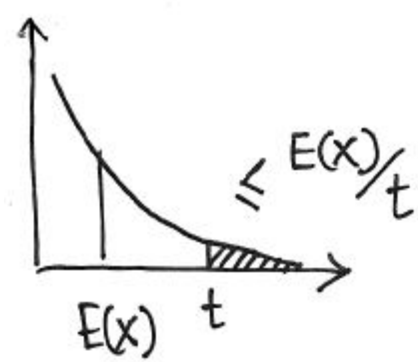
# Probability---inequalities

Two inequalities involving moments of distributions and tail probabilities often come in handy:

Markov Inequality

X is a random variable that is always non-negative.
Then for any $t > 0$, $P(X>=t) <= E(X)/t$.



# Probability---inequalities

## Chebyshev Inequality

X is a random variable for which Var(X) exists. Then for any t>0, P(|X-E(X)| >= t) <= Var(X)/$t^2$.

# Probability---inequalities

## Chebyshev Inequality

X is a random variable for which Var(X) exists. Then for any t>0, P(|X-E(X)| >= t) <= Var(X)/t².