Random Variables: Expectations and Variances

Charlie Gibbons
Political Science 236

September 22, 2008

Outline

Expectations of Random Variables

The *expectation*, $\mathbb{E}(X) \equiv \mu$ of a random variable X is simply the weighted average of the possible realizations, weighted by their probability.

Expectations of Random Variables

The *expectation*, $\mathbb{E}(X) \equiv \mu$ of a random variable X is simply the weighted average of the possible realizations, weighted by their probability.

For discrete random variables, this can be written as

$$\mathbb{E}(X) = \sum_{x \in X} \Pr(X = x)x = \sum_{x \in X} f(x)x.$$

Expectations of Random Variables

The *expectation*, $\mathbb{E}(X) \equiv \mu$ of a random variable X is simply the weighted average of the possible realizations, weighted by their probability.

For discrete random variables, this can be written as

$$\mathbb{E}(X) = \sum_{x \in X} \Pr(X = x)x = \sum_{x \in X} f(x)x.$$

For continuous random variables:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{-\infty}^{\infty} x\, dF$$

Expectations of Functions of Random Variables

The definition of expectation can be generalized for functions of random variables, $g(x)$.

Expectations of Functions of Random Variables

The definition of expectation can be generalized for functions of random variables, $g(x)$.

We have the equations

$$\begin{aligned}\mathbb{E}[g(x)] &= \sum_{x \in X} f(x)g(x) \\ &= \int_{-\infty}^{\infty} g(x)\, dF\end{aligned}$$

for discrete and continuous random variables respectively.

Properties of Expectations

Expectations are *linear operators*, *i.e.*,

$$\mathbb{E}(a \cdot g(X) + b \cdot h(X) + c) = a \cdot \mathbb{E}[g(X)] + b \cdot \mathbb{E}[h(X)] + c.$$

Properties of Expectations

Expectations are *linear operators*, *i.e.*,

$$\mathbb{E}(a \cdot g(X) + b \cdot h(X) + c) = a \cdot \mathbb{E}[g(X)] + b \cdot \mathbb{E}[h(X)] + c.$$

Note that, in general, $\mathbb{E}[g(x)] \neq g[\mathbb{E}(X)]$.

Jensen's inequality states that, for a convex function $g(x)$, $\mathbb{E}[g(x)] \geq g[\mathbb{E}(X)]$. For concave functions, the inequality is reversed.

Properties of Expectations

Expectations preserve monotonicity; for $g(x) \geq h(x)\ \forall x \in X$, $\mathbb{E}[g(X)] \geq \mathbb{E}[h(X)]$.

As a special case, let $h(x) = 0$. If $g(x) \geq 0$, then $\mathbb{E}[g(x)] \geq 0$.

Properties of Expectations

Expectations preserve monotonicity; for $g(x) \geq h(x) \; \forall x \in X$, $\mathbb{E}[g(X)] \geq \mathbb{E}[h(X)]$.

As a special case, let $h(x) = 0$. If $g(x) \geq 0$, then $\mathbb{E}[g(x)] \geq 0$.

Expectations also preserve equality; for $g(x) = h(x) \; \forall x \in X$, $\mathbb{E}[g(X)] = \mathbb{E}[h(X)]$.

Conditional Expectations

The expectation of a random variable Y *conditional on* or *given* X is defined analogously to the preceding formulations, but uses the conditional distribution $f_{Y|X}(y|X = x)$, rather than the unconditional $f_Y(y)$.

Conditional distributions are about changing the population that you are considering.

Conditional Expectations

Conditional distributions are about changing the population that you are considering.

For discrete random variables:

$$\mathbb{E}(Y|X = x) = \sum_{y \in Y} \Pr(Y = y|X = x)y = \sum_{y \in Y} f_{Y|X}(y)y.$$

Conditional Expectations

Conditional distributions are about changing the population that you are considering.

For discrete random variables:

$$\mathbb{E}(Y|X=x) = \sum_{y \in Y} \Pr(Y=y|X=x)y = \sum_{y \in Y} f_{Y|X}(y)y.$$

For continuous random variables:

$$\mathbb{E}(Y|X=x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|X=x)\,dy = \int_{-\infty}^{\infty} y\,dF_{Y|X}(y|X=X)$$

Conditional Expectations

Recall from the previous set of slides that, for independent random variables X and Y,

$$f_{Y|X}(y|X=x) = f_Y(y) \text{ and } f_{X|Y}(x|Y=y) = f_X(x)$$

Conditional Expectations

Recall from the previous set of slides that, for independent random variables X and Y,

$$f_{Y|X}(y|X=x) = f_Y(y) \text{ and } f_{X|Y}(x|Y=y) = f_X(x)$$

Hence,

$$\mathbb{E}(Y|X) = \mathbb{E}(Y) \text{ and } \mathbb{E}(X|Y) = \mathbb{E}(X)$$

This will be a heavily-used result later in the course.

Conditional Expectations

Let's pause and consider what is random here.

Conditional Expectations

Let's pause and consider what is random here.

First, there is one conditional distribution of Y given $X = x$, but a *family* of distributions of Y given X, since X takes many values and each value imparts its own conditional distribution for Y.

Conditional Expectations

Let's pause and consider what is random here.

First, there is one conditional distribution of Y given $X = x$, but a *family* of distributions of Y given X, since X takes many values and each value imparts its own conditional distribution for Y.

$Y|X = x$ derives its randomness solely from the fact that Y is random—X is fixed at x here.

Conditional Expectations

Let's pause and consider what is random here.

First, there is one conditional distribution of Y given $X = x$, but a *family* of distributions of Y given X, since X takes many values and each value imparts its own conditional distribution for Y.

$Y|X = x$ derives its randomness solely from the fact that Y is random—X is fixed at x here.

$\mathbb{E}(Y|X = x)$ is not random, since X is fixed and we are integrating over all possible values of Y.

Conditional Expectations

Let's pause and consider what is random here.

First, there is one conditional distribution of Y given $X = x$, but a *family* of distributions of Y given X, since X takes many values and each value imparts its own conditional distribution for Y.

$Y|X = x$ derives its randomness solely from the fact that Y is random—X is fixed at x here.

$\mathbb{E}(Y|X = x)$ is not random, since X is fixed and we are integrating over all possible values of Y.

$\mathbb{E}(Y|X)$ is random, since X is no longer fixed (*i.e.*, it is random), though Y is integrated over.

Conditional Expectations

The unconditional expected value of a function of X and Y, $g(x,y)$ can be written as:

$$\begin{aligned}\mathbb{E}[g(X,Y)] &= \sum_{x\in X}\sum_{y\in Y} g(x,y) f_{X,Y}(x,y) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\, dy\, dx\end{aligned}$$

Conditional Expectations

The following holds for the conditional expectation of Y given X:

$$\mathbb{E}[g(X)h(Y)|X] = g(X)\mathbb{E}[h(Y)|X]$$

Question: Why does this hold when X is a random variable?

Conditional Expectations

The *law of iterated expectations* states that

$$\mathbb{E}_Y(Y) = \mathbb{E}_X[\mathbb{E}(Y|X)],$$

which can be shown by

$$\begin{aligned}
\mathbb{E}_Y(Y) &= \int \int y f_{X,Y}(x,y)\,dx\,dy \\
&= \int \int y f_{Y|X}(y|x) f_X(x)\,dx\,dy \\
&= \int \left[y f_{Y|X}(y|x)\,dy \right] f_X(x)\,dx \\
&= \int \mathbb{E}(Y|X) f_X(x)\,dx \\
&= \mathbb{E}_X[\mathbb{E}(Y|X)]
\end{aligned}$$

Conditional Expectations

Note that the LIE only holds when the necessary marginal moments exist.

Variance

The *variance* of a random variable is a measure of its dispersion around its mean. It is defined as the second central moment of X:

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

Variance

The *variance* of a random variable is a measure of its dispersion around its mean. It is defined as the second central moment of X:

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

Multiplying this out yields:

$$\begin{aligned} &= \mathbb{E}\left(X^2 - 2\mu X + \mu^2\right) \\ &= \mathbb{E}\left(X^2\right) - 2\mu\mathbb{E}(x) + \mu^2 \\ &= \mathbb{E}\left(X^2\right) - [\mathbb{E}(X)]^2 \end{aligned}$$

Variance

The *variance* of a random variable is a measure of its dispersion around its mean. It is defined as the second central moment of X:

$$\text{Var}(X) = \mathbb{E}\left[(X - \mu)^2\right]$$

Multiplying this out yields:

$$\begin{aligned} &= \mathbb{E}\left(X^2 - 2\mu X + \mu^2\right) \\ &= \mathbb{E}\left(X^2\right) - 2\mu\mathbb{E}(x) + \mu^2 \\ &= \mathbb{E}\left(X^2\right) - [\mathbb{E}(X)]^2 \end{aligned}$$

The *standard deviation*, σ, of a random variable is the square root of its variance; *i.e.*, $\sigma = \sqrt{\text{Var}(X)}$.

Variance

The *variance* of a random variable is a measure of its dispersion around its mean. It is defined as the second central moment of X:

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

Multiplying this out yields:

$$\begin{aligned} &= \mathbb{E}\left(X^2 - 2\mu X + \mu^2\right) \\ &= \mathbb{E}\left(X^2\right) - 2\mu\mathbb{E}(x) + \mu^2 \\ &= \mathbb{E}\left(X^2\right) - [\mathbb{E}(X)]^2 \end{aligned}$$

The *standard deviation*, σ, of a random variable is the square root of its variance; *i.e.*, $\sigma = \sqrt{\mathrm{Var}(X)}$.

See that $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$.

Conditional Variance

The *conditional variance of* Y *given* X is

$$\mathbb{E}\left[(Y - \mathbb{E}(Y|X))^2 \middle| X \right] = \mathbb{E}\left(Y^2 \middle| X \right) - [\mathbb{E}(Y|X)]^2$$

Conditional Variance

The *conditional variance of* Y *given* X is

$$\mathbb{E}\left[(Y - \mathbb{E}(Y|X))^2 \middle| X \right] = \mathbb{E}\left(Y^2 \middle| X \right) - [\mathbb{E}(Y|X)]^2$$

Note that

$$\mathrm{Var}[g(X)h(Y)|X] = [g(X)]^2 \mathrm{Var}[h(Y)|X]$$

Conditional Variance

$$
\begin{aligned}
\mathrm{Var}(Y) &= \mathbb{E}_Y\left[(Y - \mathbb{E}_Y(Y))^2\right] \\
&= \mathbb{E}_Y\left[\left(Y - \mathbb{E}_{Y|X}(Y|X) + \mathbb{E}_{Y|X}(Y|X) - \mathbb{E}_Y(Y)\right)^2\right] \\
&= \mathbb{E}_Y\left[\left(Y - \mathbb{E}_{Y|X}(Y|X)\right)^2\right] + \mathbb{E}_Y\left[\left(\mathbb{E}_{Y|X}(Y|X) - \mathbb{E}_Y(Y)\right)^2\right] \\
&= \mathbb{E}_X\left[\mathbb{E}\left[\left(Y - \mathbb{E}_{Y|X}(Y|X)\right)^2 \middle| X\right]\right] \\
&\quad + \mathbb{E}_X\left[\mathbb{E}\left[\left(\mathbb{E}_{Y|X}(Y|X) - \mathbb{E}_Y(Y)\right)^2 \middle| X\right]\right] \\
&= \mathbb{E}_X\left[\mathrm{Var}(Y|X)\right] + \mathbb{E}_X\left[\left(\mathbb{E}_{Y|X}(Y|X) - \mathbb{E}_Y(Y)\right)^2\right] \\
&= \mathbb{E}_X\left[\mathrm{Var}(Y|X)\right] + \mathrm{Var}\left[\mathbb{E}_{Y|X}(Y|X)\right]
\end{aligned}
$$

Conditional Variance

$$\mathrm{Var}(Y) = \mathbb{E}_X\left[\mathrm{Var}(Y|X)\right] + \mathrm{Var}\left[\mathbb{E}_{Y|X}(Y|X)\right]$$

So what?

Conditional Variance

$$\mathrm{Var}(Y) = \mathbb{E}_X \left[\mathrm{Var}(Y|X)\right] + \mathrm{Var}\left[\mathbb{E}_{Y|X}(Y|X)\right]$$

So what?

$$\mathrm{Var}(Y) \geq \mathrm{Var}\left[\mathbb{E}_{Y|X}(Y|X)\right]$$

Conditioning on more X variables yields a smaller variance.

Conditional Variance

$$\mathrm{Var}(Y) = \mathbb{E}_X \left[\mathrm{Var}(Y|X)\right] + \mathrm{Var}\left[\mathbb{E}_{Y|X}(Y|X)\right]$$

So what?

$$\mathrm{Var}(Y) \geq \mathrm{Var}\left[\mathbb{E}_{Y|X}(Y|X)\right]$$

Conditioning on more X variables yields a smaller variance.

Intuition: Adding more predictors can only make your guess of Y better. If you add something that isn't relevant or too noisy, you can just ignore it (in a regression, *e.g.*, you give that predictor a 0 coefficient).

See Casella and Berger, pp. 167–168 for a similar, detailed exposition.

Covariance and Correlation

The *covariance* of random variables X and Y is defined as

$$\begin{aligned}\mathrm{Cov}(X,Y) \equiv \sigma_{XY} &= \mathbb{E}_{X,Y}\left[\left(X - \mathbb{E}_X(X)\right)\left(Y - \mathbb{E}_Y(Y)\right)\right] \\ &= \mathbb{E}(XY) - \mu_X\mu_Y\end{aligned}$$

Covariance and Correlation

The *covariance* of random variables X and Y is defined as

$$\begin{aligned}\mathrm{Cov}(X,Y) \equiv \sigma_{XY} &= \mathbb{E}_{X,Y}\left[(X - \mathbb{E}_X(X))\,(Y - \mathbb{E}_Y(Y))\right] \\ &= \mathbb{E}(XY) - \mu_X\mu_Y\end{aligned}$$

We have

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X,Y)$$

Covariance and Correlation

The *covariance* of random variables X and Y is defined as

$$\begin{aligned}\mathrm{Cov}(X,Y) \equiv \sigma_{XY} &= \mathbb{E}_{X,Y}\left[\left(X - \mathbb{E}_X(X)\right)\left(Y - \mathbb{E}_Y(Y)\right)\right] \\ &= \mathbb{E}(XY) - \mu_X \mu_Y\end{aligned}$$

We have

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X,Y)$$

Note that covariance only measures the *linear* relationship between two random variables.

Covariance and Correlation

The *correlation* of random variables X and Y is defined as

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Covariance and Correlation

The *correlation* of random variables X and Y is defined as

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Correlation is a normalized version of covariance—how big is the correlation relative to the variation in X and Y? Both will have the same sign.

Covariance and Correlation

If X and Y are independent, then

$$\mathbb{E}\left[g(X)h(Y)\right] = \mathbb{E}\left[g(X)\right]\mathbb{E}\left[h(Y)\right]$$

Covariance and Correlation

If X and Y are independent, then

$$\mathbb{E}\left[g(X)h(Y)\right] = \mathbb{E}\left[g(X)\right]\mathbb{E}\left[h(Y)\right]$$

Hence, for independent random variables, the covariance is 0.