Week 10 notes

Expectation for two variables

01 Theory

Expectation for a function on two variables

Discrete case:

$$E[\,g(X,Y)\,] \quad = \quad \sum_{k,\ell} g(k,\ell)\,P_{X,Y}(k,\ell) \qquad \text{(sum over possible values)}$$

Continuous case:

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

These formulas are *not trivial to prove*, and we omit the proofs. (Recall the technical nature of the proof we gave for $E[\,g(X)\,]$ in the discrete case.)

Expectation sum rule

Suppose X and Y are *any* two random variables on the same probability model.

Then:

$$E[X+Y] = E[X] + E[Y]$$

We already know that expectation is linear in a single variable: $E[aX+b] = aE[X]+b$.

Therefore this two-variable formula implies:

$$E[aX+bY+c] = aE[X]+bE[Y]+c$$

Expectation product rule: independence

Suppose that X and Y are *independent*.

Then we have:

$$E[XY] = E[X]E[Y]$$

Extra - Proof: Expectation sum rule, continuous case

Suppose f_X and f_Y give marginal PDFs for X and Y, and $f_{X,Y}$ gives their joint PDF.

Then:

$$E[X+Y] \gg\gg \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)f_{X,Y}(x,y)\,dx\,dy$$

$$\gg\gg \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x f_{X,Y}\,dx\,dy+\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}y f_{X,Y}\,dx\,dy$$

$$\gg\gg \int_{-\infty}^{+\infty}x f_X(x)\,dx+\int_{-\infty}^{+\infty}y f_Y(y)\,dy$$

$$\gg\gg E[X]+E[Y]$$

Observe that this calculation relies on the formula for $E[\,g(X,Y)\,]$, specifically with $g(x,y)=x+y$.

Extra - Proof: Expectation product rule

$$E[XY] \gg\gg \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(xy)f_{X,Y}(x,y)\,dx\,dy$$

$$\gg\gg \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(xy)f_X(x)f_Y(y)\,dx\,dy$$

$$\gg\gg \int_{-\infty}^{+\infty}x f_X(x)\,dx\int_{-\infty}^{+\infty}y f_Y(y)\,dy$$

$$\gg\gg E[X]E[Y]$$

02 Illustration

$E[\,X^2+Y\,]$ from joint PMF chart

Suppose the joint PMF of X and Y is given by this chart:

$Y\downarrow\ X\rightarrow$	1	2
-1	0.2	0.2
0	0.35	0.1
1	0.05	0.1

Define $W=X^2+Y$. Find the expectation $E[W]$.

Solution

First compute the values of W for each pair (X,Y) in the chart:

$Y\downarrow\ X\rightarrow$	1	2
-1	0	3
0	1	4
1	2	5

Now take the sum, weighted by probabilities:

$$\begin{array}{c}0\cdot(0.2)+3\cdot(0.2)+1\cdot(0.35)\\+4\cdot(0.1)+2\cdot(0.05)+5\cdot(0.1)\end{array} \gg\gg 1.95 = E[W]$$

Exercise - Understanding expectation for two variables

Suppose you know *only* that $X \sim \text{Geo}(p)$ and $Y \sim \text{Bin}(n, q)$.

Which of the following can you calculate?

$$E[X+Y], \quad E[XY], \quad E[X^2+Y^2], \quad E[(X+Y)^2]$$

$E[Y]$ two ways, and $E[XY]$, from joint density

Suppose X and Y are random variables with the following joint density:

$$f_{X,Y}(x,y) = \begin{cases} \frac{3}{16}xy^2 & x,y \in [0,2] \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute $E[Y]$ using two methods.

(b) Compute $E[XY]$.

Solution
(a) Method One: via marginal PDF $f_Y(y)$:

$$f_Y(y) \quad = \quad \int_0^2 \frac{3}{16}xy^2\,dx \quad \gg\gg \quad \begin{cases} \frac{3}{8}y^2 & y \in [0,2 \\ 0 & \text{otherwise} \end{cases}$$

Then expectation:

$$E[Y] \;=\; \int_0^2 y\,f_Y(y)\,dy \quad \gg\gg \quad \int_0^2 \frac{3}{8}y^3\,dy \quad \gg\gg \quad 3/2$$

(a) Method Two: directly, via two-variable formula:

$$E[Y] \;=\; \int_0^2\int_0^2 y\cdot\frac{3}{16}xy^2\,dy\,dx \gg\gg \quad \int_0^2 \frac{3}{4}x\,dx \quad \gg\gg \quad 3/2$$

(b) Directly, via two-variable formula:

$$E[XY] \quad = \quad \int_0^2\int_0^2 xy\cdot\frac{3}{16}xy^2\,dy\,dx$$

$$\gg\gg \quad \int_0^2 \frac{3}{4}x^2\,dx \quad \gg\gg \quad 2$$

Covariance and correlation

03 Theory

Write $\mu_X = E[X]$ and $\mu_Y = E[Y]$.

Observe that the random variables $X - \mu_X$ and $Y - \mu_Y$ are "centered at zero," meaning that $E[X - \mu_X] = 0 = E[Y - \mu_Y]$.

Covariance

Suppose X and Y are any two random variables on a probability model. The **covariance** of X and Y measures the *typical synchronous deviation* of X and Y from their respective means.

Then the *defining formula* for covariance of X and Y is:

$$\mathrm{Cov}[X,Y] \;=\; E[\,(X-\mu_X)(Y-\mu_Y)\,]$$

There is also a *shorter formula*:

$$\mathrm{Cov}[X,Y] \;=\; E[XY] - \mu_X\mu_Y$$

To derive the shorter formula, first expand the product $(X-\mu_X)(Y-\mu_Y)$ and then apply linearity.

Notice that covariance is always *symmetric*:

$$\mathrm{Cov}[X,Y] \quad = \quad \mathrm{Cov}[Y,X]$$

The *self* covariance equals the variance:

$$\mathrm{Cov}[X,X] \quad = \quad \mathrm{Var}[X]$$

The *sign* of $\mathrm{Cov}[X,Y]$ reveals the *correlation type* between X and Y:

Correlation	Sign
Positively correlated	$\mathrm{Cov}(X,Y) > 0$
Negatively correlated	$\mathrm{Cov}(X,Y) < 0$
Uncorrelated	$\mathrm{Cov}(X,Y) = 0$

Correlation coefficient

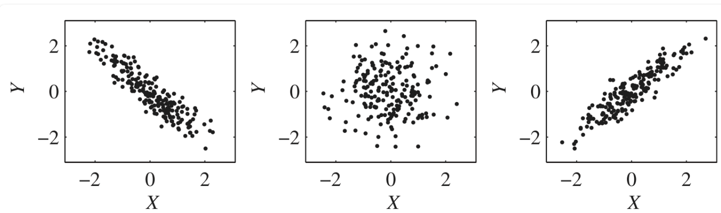
Suppose X and Y are any two random variables on a probability model.

Their **correlation coefficient** is a rescaled version of covariance that measures the *synchronicity of deviations*:

$$\rho[X,Y] \;=\; \frac{\mathrm{Cov}[X,Y]}{\sigma_X\sigma_Y}$$

The rescaling ensures:

$$-1 \;\leq\; \rho_{X,Y} \;\leq\; +1$$



(a) $\rho_{X,Y} = -0.9$ (b) $\rho_{X,Y} = 0$ (c) $\rho_{X,Y} = 0.9$

Covariance depends on the *separate variances* of X and Y as well as their relationship.

Correlation coefficient, because we have divided out $\sigma_X\sigma_Y$, depends only on their *relationship*.

04 Illustration

Covariance from PMF chart

Suppose the joint PMF of X and Y is given by this chart:

$Y \downarrow \quad X \rightarrow$	1	2
-1	0.2	0.2
0	0.35	0.1
1	0.05	0.1

Find $\mathrm{Cov}[X, Y]$.

Solution
We need $E[X]$ and $E[Y]$ and $E[XY]$.

$$E[X] = 1 \cdot (0.2 + 0.35 + 0.05) + 2 \cdot (0.2 + 0.1 + 0.1) \quad \gg\gg \quad 1.4$$

$$E[Y] = -1 \cdot (0.2 + 0.2) + 0 \cdot (0.35 + 0.1) + 1 \cdot (0.05 + 0.1)$$

$$\gg\gg \quad -0.25$$

$$E[XY] = -1 \cdot (0.2) - 2 \cdot (0.2) + 0 + 1 \cdot (0.05) + 2 \cdot (0.1) \quad \gg\gg \quad -0.35$$

Therefore:

$$\mathrm{Cov}[X, Y] = E[XY] - E[X]E[Y]$$

$$\gg\gg \quad -0.35 - (1.4)(-0.25) \quad \gg\gg \quad 0$$

05 Theory

Covariance bilinearity

Given any three random variables X, Y, and Z, we have:

$$\mathrm{Cov}[\,X + Y,\ Z\,] = \mathrm{Cov}[X, Z] + \mathrm{Cov}[Y, Z]$$

$$\mathrm{Cov}[\,Z,\ X + Y\,] = \mathrm{Cov}[Z, X] + \mathrm{Cov}[Z, Y]$$

Covariance and correlation: shift and scale

Covariance *scales with each input, and ignores shifts*:

$$\mathrm{Cov}[\,aX + b,\ Y\,] = a\,\mathrm{Cov}[X, Y] = \mathrm{Cov}[\,X,\ aY + b\,]$$

Whereas shift or scale in correlation *only affects the sign*:

$$\rho[\,aX + b,\ Y\,] = \mathrm{sign}(a)\,\rho[X, Y] = \rho[\,X,\ aY + b\,]$$

Extra - Proof of covariance bilinearity

$$\mathrm{Cov}[X + Y,\ Z] \quad \gg\gg \quad E[(X + Y - (\mu_X + \mu_Y))(Z - \mu_Z)]$$

$$\gg\gg \quad E[(X - \mu_X + Y - \mu_Y)(Z - \mu_Z)]$$

$$\gg\gg \quad E[(X - \mu_X)(Z - \mu_Z)] + E[(Y - \mu_Y)(Z - \mu_Z)]$$

$$\gg\gg \quad \mathrm{Cov}[X, Z] + \mathrm{Cov}[Y, Z]$$

Extra - Proof of covariance shift and scale rule

$$\begin{aligned}\text{Cov}[aX+b,Y] \quad &\gg\gg \quad E[(aX+b)Y] - E[aX+b]E[Y] \\ &\gg\gg \quad E[aXY+bY] - aE[X]E[Y] - E[b]E[Y] \\ &\gg\gg \quad aE[XY] + bE[Y] - aE[X]E[Y] - bE[Y] \\ &\gg\gg \quad a\big(E[XY] - E[X]E[Y]\big)\end{aligned}$$

Independence implies zero covariance

Suppose that X and Y are any two random variables on a probability model.

If X and Y are independent, then:

$$\text{Cov}[X,Y] = 0$$

Sum rule for variance

Suppose that X and Y are any two random variables on a probability space.

Then:

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X,Y]$$

When X and Y are *independent*, the formula simplifies to:

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$$

Proof: Independence implies zero covariance

The product rule for expectation, since X and Y are independent:

$$E[XY] = E[X]E[Y]$$

The shorter formula for covariance:

$$\text{Cov}[X,Y] \;=\; E[XY] - \mu_X\mu_Y$$

But $E[XY] = E[X]E[Y] = \mu_X\mu_Y$, so those terms cancel and $\text{Cov}[X,Y] = 0$.

Proof: Sum rule for variance

$$\begin{aligned}\text{Var}[X+Y] \quad &\gg\gg \quad E\Big[\big(X+Y-(\mu_X+\mu_Y)\big)^2\Big] \\ &\gg\gg \quad E\Big[\big((X-\mu_X)+(Y-\mu_Y)\big)^2\Big] \\ &\gg\gg \quad E\Big[(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y)\Big] \\ &\gg\gg \quad \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X,Y]\end{aligned}$$

Proof that $-1 \leq \rho \leq +1$

1. Create standardizations:

$$\tilde{X} \;=\; \frac{X-\mu_X}{\sigma_X}, \qquad \tilde{Y} \;=\; \frac{Y-\mu_Y}{\sigma_Y}$$

2. Now $\tilde{X}$ and $\tilde{Y}$ satisfy $E[\tilde{X}] = 0 = E[\tilde{Y}]$ and $\mathrm{Var}[\tilde{X}] = 1 = \mathrm{Var}[\tilde{Y}]$.
3. Observe that $\mathrm{Var}[W] \geq 0$ for any W. *Variance can't be negative.*
4. Apply the variance sum rule.
   - Apply to $\tilde{X}$ and $\tilde{Y}$:

   $$0 \leq \mathrm{Var}[\tilde{X}+\tilde{Y}] \;=\; \mathrm{Var}[\tilde{X}] + \mathrm{Var}[\tilde{Y}] + 2\mathrm{Cov}[\tilde{X},\tilde{Y}]$$

   - Simplify:

   $$\gg\gg \quad 1+1+2\mathrm{Cov}[\tilde{X},\tilde{Y}] \geq 0$$

   $$\gg\gg \quad 1+\mathrm{Cov}[\tilde{X},\tilde{Y}] \geq 0$$

   - Notice effect of standardization:

   $$\mathrm{Cov}[\tilde{X},\tilde{Y}] \quad = \quad \rho[X,Y]$$

   - Therefore $\rho[X,Y] \geq -1$.
5. Modify and reapply variance sum rule.
   - Change to $\tilde{X}-\tilde{Y}$:

   $$0 \leq \mathrm{Var}[\tilde{X}-\tilde{Y}] \;=\; \mathrm{Var}[\tilde{X}] + \mathrm{Var}[-\tilde{Y}] + 2\mathrm{Cov}[\tilde{X},\,-\tilde{Y}]$$

   - Simplify:

   $$\gg\gg \quad 1+1-2\mathrm{Cov}[\tilde{X},\tilde{Y}] \geq 0$$

   $$\gg\gg \quad 1-\mathrm{Cov}[\tilde{X},\tilde{Y}] \geq 0$$

06 Illustration

Exercise - Covariance rules

Simplify:

$$\mathrm{Cov}[\,2X+5Y+1,\; Z+8W+X+9\,]$$

Exercise - Independent variables are uncorrelated

Let X be given with possible values $\{-1,0,+1\}$ and PMF given uniformly by $P_X(k) = 1/3$ for all three possible k. Let $Y = X^2$.

Show that X and Y are dependent but uncorrelated.

Hint: To speed the calculation, notice that $X^3 = X$.

Variance of sum of indicators

An urn contains 3 red balls and 2 yellow balls.

Suppose 2 balls are drawn without replacement, and X counts the number of red balls drawn.

Find $\mathrm{Var}(X)$.

Solution

Let X_1 indicate (one or zero) whether the first ball is red, and X_2 indicate whether the second ball is red, so $X = X_1 + X_2$.

Then X_1X_2 indicates whether both drawn balls are red; so it is Bernoulli with success probability $\frac{3}{5}\frac{2}{4} = \frac{3}{10}$. Therefore $E[X_1X_2] = \frac{3}{10}$.

We also have $E[X_1] = E[X_2] = \frac{3}{5}$.

The variance sum rule gives:

$$\mathrm{Var}[X] \quad = \quad \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + 2\mathrm{Cov}[X_1, X_2]$$

$$\gg\gg \quad E[X_1^2] - E[X_1]^2 + E[X_2^2] - E[X_2]^2 + 2(E[X_1X_2] - E[X_1]E[X_2])$$

$$\gg\gg \quad \frac{3}{5} - \left(\frac{3}{5}\right)^2 + \frac{3}{5} - \left(\frac{3}{5}\right)^2 + 2\left(\frac{3}{10} - \frac{3}{5}\cdot\frac{3}{5}\right) \quad \gg\gg \quad \frac{9}{25}$$