Math 425
Introduction to Probability
Lecture 33

Kenneth Harris
kaharri@umich.edu

Department of Mathematics
University of Michigan

April 8, 2009

Expectation of Products

☞ It is NOT generally true that an expectation of a product is a product of expectations. Consider X where

$$\mathbf{P}\{X = 0\} = \mathbf{P}\{X = 1\} = \mathbf{P}\{X = -1\} = \frac{1}{3}$$

and define

$$Y = \begin{cases} 1 & \text{if } X = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Compute expectations.

$$E[X] = 0, \quad E[Y] = \frac{1}{3}$$
$$E[X]E[Y] = 0$$
$$E[XY] = \frac{1}{3}.$$

So, $E[XY] \neq E[X]E[Y]$.

Expectation of Products

Theorem

If X and Y are independent, then for any functions g and h

$$E[g(X) \cdot h(Y)] = E[g(X)] \cdot E[h(Y)].$$

Proof.

Let X and Y be independent. Recall, $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$.

$$\begin{aligned} E[g(X) \cdot h(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) \cdot h(y) \cdot f_{X,Y}(x, y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) \cdot h(y) \cdot f_X(x) \cdot f_Y(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\, dx \int_{-\infty}^{\infty} h(y) \cdot f_Y(y)\, dy \\ &= E[g(X)] \cdot E[h(Y)]. \end{aligned}$$

□

Independence and Variance

☞ Recall the definition of variance for sums

$$Var(X + Y) = E\left[(X + Y - E[X + Y])^2\right].$$

Theorem

If X and Y be independent, then

$$Var(X + Y) = Var(X) + Var(Y).$$

More generally, if $X_1, X_2, \ldots, X_n$ are mutually independent, then

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i).$$

Proof

☞ Let X and Y be independent. Apply the definition of variance and sums of expectations:

$$\begin{aligned} Var(X+Y) &= E\left[(X+Y-E[X+Y])^2\right] \\ &= E\left[\left((X-E[X]) + (Y-E[Y])\right)^2\right] \\ &= E\left[(X-E[X])^2\right] + E\left[(Y-E[Y])^2\right] \\ &\quad +2\cdot E\left[(X-E[X])(Y-E[Y])\right] \end{aligned}$$

Because X and Y are independent, so is $X-E[X]$ and $Y-E[Y]$, so

$$\begin{aligned} E\left[(X-E[X])(Y-E[Y])\right] &= E\left[X-E[X]\right]\cdot E\left[Y-E[Y]\right] \\ &= (E[X]-E[X])\cdot(E[Y]-E[Y]) = 0 \end{aligned}$$

Hence

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

Example: Variance for Binomial R.V.

Example. Let $X_1, \ldots, X_n$ be a collection of Bernoulli random variables for n independent trials with success probability p.
The sum $X = \sum_{i=1}^{n} X_i$ is a binomial random variable.

☞ The variance of each Bernoulli trial is

$$Var(X_i) = E[X_i^2] - E[X_i]^2 = p - p^2 = p(1-p)$$

Since the X_i are mutually indendent

$$Var(X) = \sum_{i=1}^{n} Var(X_i) = n\cdot p(1-p).$$

This agrees with our computation in Section 4.6.1.

Example: Variance for Negative Binomial R.V.

Example. Let Y be a negative binomial random variable for a Bernoulli trials process counting the number of trials for k successes.
Let $Y_1, \ldots, Y_k$ be a collection of geometric random variables, where Y_i counts the number of trials between the $i-1$st and ith success.
So, $Y = \sum_{i=1}^{n} Y_i$ is a binomial random variable.

☞ The variance of each geometric variable (Section 4.8.1) is

$$Var(Y_i) = \frac{1-p}{p^2}$$

Since the Y_i are mutually indendent

$$Var(Y) = \sum_{i=1}^{k} Var(Y_i) = n\cdot\frac{1-p}{p^2}.$$

This agrees with our computation in Section 4.8.2.

Definition: Covariance

☞ Covariance measures how much two variables change together.
It is especially useful when computing the variance of sums when the variables are not independent.

Definition

The covariance between X and Y is $Cov(X, Y)$, defined by

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

where $\mu_X = E[X]$ and $\mu_Y = E[Y]$ are the means of X and Y.

Covariance and variance

☞ Variance is a special case of covariance when the two variables are identical.

Variance and Covariance. The covariance of X and itself is the variance of X:

$$\begin{aligned} Cov(X, X) &= E[(X - \mu_X) \cdot (X - \mu_X)] \\ &= Var(X). \end{aligned}$$

Covariance reformulated

Key Reformulation. The following reformulation is often useful for computing covariance:

$$Cov(X, Y) = E[XY] - E[X]E[Y].$$

Reason.

$$\begin{aligned} Cov(X, Y) &= E[(X - \mu_X) \cdot (Y - \mu_Y)] \\ &= E[XY] - E[\mu_Y X] - E[\mu_X Y] + E[\mu_X \mu_Y] \\ &= E[XY] - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

Covariance and Independence

Independence. If X and Y are independent, then $Cov(X, Y) = 0$.

Reason.

$$Cov(X, Y) = E[XY] - E[X]E[Y];$$

so if X and Y are independent, then

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

Key Properties

☞ Covariance has many nice properties, which we usefully exploit.

Proposition

Let a and b be constants.

$$\begin{aligned} (i) &\quad Cov(X, Y) = Cov(Y, X) \\ (ii) &\quad Cov(X, a) = 0 \\ (iii) &\quad Cov(aX, bY) = abCov(X, Y) \\ (iv) &\quad Cov(X + a, Y + b) = Cov(X, Y) \\ (v) &\quad Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z). \end{aligned}$$

Property (v) generalizes to arbitrary finite sums

$$(vi) \quad Cov\Big(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\Big) = \sum_{i=1}^{n} \sum_{j=1}^{m} Cov(X_i, Y_j).$$

Variance and Covariance

☞ Variance and covariance are related by the following

Theorem

Let X and Y be random variables.

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

More generally, for any random variables $X_1, \ldots, X_n$

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j).$$

Proof of Theorem

☞ The connection between variance and covariance arose from our earlier argument:

Apply the definition of variance and sums of expectations:

$$\begin{aligned} Var(X+Y) &= E\left[(X+Y-E[X+Y])^2\right] \\ &= E\left[\left((X-E[X]) + (Y-E[Y])\right)^2\right] \\ &= E\left[(X-E[X])^2\right] + E\left[(Y-E[Y])^2\right] \\ &\quad +2 \cdot E\left[(X-E[X]) \cdot (Y-E[Y])\right] \\ &= Var(X) + Var(Y) + 2Cov(X,Y). \end{aligned}$$

☞ The general case is by induction.

Example

Example

A group of $N \geq 2$ people put their names in a hat to exchange gifts. The names are well-mixed so every person is equally likely to draw any other person's name.

What is the expectation and variance of the number of people who draw their own name?

Example – continued

☞ Let G count the number of people who draw their own name. In Lecture 31 we computed $E[G]$ using the indicator variables

$$G_i = \begin{cases} 1 & \text{if person } i \text{ draws their own name} \\ 0 & \text{otherwise.} \end{cases}$$

Since $E[G_i] = \mathbf{P}\{G_i = 1\} = \frac{1}{N}$ and $G = \sum_{i=1}^{N} G_i$, we computed

$$E[G] = \sum_{i=1}^{N} E[G_i] = 1.$$

☞ The random variables $G_1, \ldots, G_N$ are NOT mutually independent, so computing $Var(G)$ will require computing $Cov(G_i, G_j)$.

Example – continued

☞ For $i \neq j$,

$$G_i \cdot G_j = \begin{cases} 1 & \text{if persons } i \text{ and } j \text{ draw their own names,} \\ 0 & \text{otherwise.} \end{cases}$$

So, when $i \neq j$

$$\begin{aligned} E[G_i \cdot G_j] &= \mathbf{P}\{G_i \cdot G_j = 1\} \\ &= \mathbf{P}\{G_i = 1 \mid G_j = 1\} \cdot \mathbf{P}\{G_j = 1\} \\ &= \frac{1}{N(N-1)}. \end{aligned}$$

Compute covariances when $i \neq j$.

$$\begin{aligned} Cov(G_i, G_j) &= E[G_i \cdot G_j] - E[G_i] \cdot E[G_j] \\ &= \frac{1}{N(N-1)} - \frac{1}{N^2} \end{aligned}$$

Example – continued

☞ The variance of the indicator variable G_i is

$$Var(G_i) = E[G_i^2] - E[G_i]^2 = \frac{1}{N} - \frac{1}{N^2} = \frac{N-1}{N^2}.$$

☞ The variance of G is

$$\begin{aligned} Var(G) &= \sum_{i=1}^{N} Var(G_i) + \sum_{i \neq j} Cov(G_i, G_j) \\ &= N \cdot Var(G_1) + N(N-1) \cdot Cov(G_1, G_2) \\ &= N \cdot \left(\frac{N-1}{N^2}\right) + N(N-1) \cdot \left(\frac{1}{N(N-1)} - \frac{1}{N^2}\right) \\ &= \frac{N-1}{N} + 1 - \frac{N-1}{N} \\ &= 1. \end{aligned}$$

Example

Example

An urn contains r red balls and b blue balls. A small number $n < r + b$ is to be drawn at random (so that all $\binom{r+b}{n}$ possible outcomes are equally likely).

What is the expectation and variance of the number of red balls drawn, when the drawing is done

- With replacement of the ball on each draw.
- Without replacement of the ball on each draw.

Example – with replacement

☞ When the drawing is done with replacement, the count R of red balls is a binomial random variable with probability $\frac{r}{r+b}$. Let $N = r + b$.

$$\begin{aligned} E[R] &= n \cdot \frac{r}{N} \\ Var(R) &= n \cdot \frac{r}{N}\left(1 - \frac{r}{N}\right) = \frac{nrb}{N^2} \end{aligned}$$

Example – without replacement

☞ Let S count the number of red balls out of n when the drawing is done without replacement. Consider the following indicator variables

$$S_i = \begin{cases} 1 & \text{if the } i\text{th ball drawn is red,} \\ 0 & \text{otherwise.} \end{cases}$$

Let $N = r + b$ be the total number of balls in the urn. For each i,

$$E[S_i] = \frac{r}{N}$$

Since $S = \sum_{i=1}^{n} S_i$,

$$E[S] = \sum_{i=1}^{n} E[S_i] = \frac{nr}{N}.$$

This is what we derived in Lecture 31.

Example – without replacement

☞ Since each S_i is an indicator variable, the variance is

$$Var(S_i) = \frac{r}{N}(1 - \frac{r}{N}) = \frac{rb}{N^2}.$$

☞ When $i \neq j$,

$$\begin{aligned} E[S_i \cdot S_j] &= \mathbf{P}\{S_i = 1, S_j = 1\} \\ &= \mathbf{P}\{S_i = 1 \mid S_j = 1\} \cdot \mathbf{P}\{S_j = 1\} \\ &= \frac{r-1}{N-1} \cdot \frac{r}{N} = \frac{r(r-1)}{N(N-1)} \end{aligned}$$

The covariance of S_i and S_j (for $i \neq j$) is

$$\begin{aligned} Cov(S_i, S_j) &= E[S_i \cdot S_j] - E[S_i] \cdot E[S_j] \\ &= \frac{r(r-1)}{N(N-1)} - \frac{r^2}{N^2} \\ &= -\frac{rb}{N^2(N-1)}. \end{aligned}$$

Example – without replacement

☞ The variance of S is

$$\begin{aligned} Var(S) &= \sum_{i=1}^{n} Var(S_i) + \sum_{i \neq j} Cov(S_i, S_j) \\ &= n \cdot Var(S_1) + n(n-1) \cdot Cov(S_1, S_2) \\ &= n\frac{rb}{N^2} - n(n-1)\frac{rb}{N^2(N-1)} \\ &= \frac{nrb}{N^2}(1 - \frac{n-1}{N-1}) \\ &= \frac{nrb}{N^2}(\frac{N-n}{N-1}) \\ &= \frac{N-n}{N-1} Var(R), \end{aligned}$$

where R was the variance when drawn with replacement.

Example

Example

A pond contains 100 fish, of which 30 are carp. If 20 fish are caught, what are the mean and variance of the number of carp among these 20? Assume that any fish in the pond is equally likely to be caught.

☞ Let X count the number of carp in the 20 caught. So, X is hypergeometric, so the previous analysis applies with values

$$N = 100 \quad n = 20 \quad r = 30 \quad b = 70$$

Thus,

$$\begin{aligned} E[X] &= \frac{nr}{N} = \frac{20(30)}{100} = 6 \\ Var(X) &= \frac{nrb}{N^2}(\frac{N-n}{N-1}) = \frac{20(30)(70)}{100^2}(\frac{100-20}{100-1}) = \frac{112}{33}. \end{aligned}$$

Correlation and Covariance

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

Definition

Let X and Y be random variables.

- X and Y are positively correlated if $Cov(X, Y) > 0$,
  X and Y tend to vary above and below their means together.
- X and Y are negatively correlated if $Cov(X, Y) < 0$,
  X tends to vary above its mean when Y varies below its mean, and vice-versa.
- X and Y are uncorrelated if $Cov(X, Y) = 0$.
  There is no relation between the variance above and below their means for X and Y.

Example: positive correlation

Example. In the gift exchange problem with $N \geq 2$ people, the indicator variable G_i was that person i chose their own name. For $i \neq j$,

$$Cov(G_i, G_j) = \frac{1}{N^2(N-1)} > 0$$

☞ Why is it to be expected that G_i and G_j are positively correlated?

- Given person i selects their own name, (slightly) increases the likelihood of person j selecting their own name.
- Given person i selects their someone else's name, (slightly) decreases the likelihood of person j selecting their own name (since person i might choose person j).

Example: negative correlation

Example. In the urn example, the indicator variable S_i indicated that the ith ball selected was red. For $i \neq j$,

$$Cov(S_i, S_j) = -\frac{rb}{N^2(N-1)} < 0$$

☞ Why is it to be expected that S_i and S_j are negatively correlated?

- Given selection i is a red ball, (slightly) decreases the likelihood of selection j is NOT a red ball (by removing a red ball).
- Given selection i is NOT a red ball, (slightly) increases the likelihood selection j is a red ball (by removing a blue ball).

Example: no correlation

☞ If X and Y are independent, then they are uncorrelated.

Example. The indicator variables for different trials in a Bernoulli trials process are independent, so uncorrelated.

Example. Suppose (X, Y) is randomly in the unit square.

$$f_{X,Y}(x, y) = 1 \qquad 0 \leq x \leq 1,\ 0 \leq y \leq 1.$$

☞ X and Y are independent, so $Cov(X, Y) = 0$. More directly,

$$\begin{aligned} E[X] &= \int_0^1 x\,dx = \frac{1}{2} \\ E[Y] &= \int_0^1 y\,dy = \frac{1}{2} \\ E[X \cdot Y] &= \int_0^1 \int_0^1 xy\,dx\,dy = \frac{1}{4} \\ Cov(X, Y) &= E[X \cdot Y] - E[X] \cdot E[Y] = 0. \end{aligned}$$

Example: no correlation

☞ Dependent random variables can be uncorrelated.

Example. Let X be uniformly distributed in $[-1,1]$ and $Y = X^2$. So, X and Y are NOT independent

$$\mathbf{P}\{Y \le \frac{1}{2} \mid -\frac{1}{2} \le X \le \frac{1}{2}\} = 1 \qquad \mathbf{P}\{Y \le \frac{1}{2}\} = \frac{1}{2}.$$

☞ However, X and Y are uncorrelated:

$$\begin{aligned} E[X] &= \int_{-1}^{1} x\,dx = 0 \\ E[Y] &= \int_{-1}^{1} x^2\,dx = \frac{2}{3} \\ E[XY] &= \int_{-1}^{1} x^3\,dx = 0 \\ Cov(X,Y) &= E[X \cdot Y] - E[X]\cdot E[Y] = 0. \end{aligned}$$

Example: Dice

Example. Throw two dice. Let X and Y denote the value of each die. Let

$$U = X + Y \qquad V = X - Y.$$

☞ U and V are NOT independent.

$$\mathbf{P}\{V = 0 \mid U = 2\} = 1 \qquad \mathbf{P}\{V = 0\} = \frac{1}{6}.$$

However, U and V are uncorrelated:

$$\begin{aligned} Cov(U,V) &= E[U\cdot V] - E[U]\cdot E[V] \\ &= E\big[(X+Y)\cdot(X-Y)\big] - E[X+Y]\cdot E[X-Y] \\ &= E[X^2] - E[Y^2] - E[X]^2 + E[Y]^2 \\ &= \big(E[X^2] - E[X]^2\big) - \big(E[Y^2] - E[Y]^2\big) \\ &= \sigma_X^2 - \sigma_Y^2 \\ &= 0. \end{aligned}$$

Correlation and Indicator Variables

Example. Let A and B be events, and I_A and I_B be their indicator variables:

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases} \qquad I_B = \begin{cases} 1 & \text{if } B \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

☞ Then

$$E[I_A] = \mathbf{P}\{A\} \qquad E[I_B] = \mathbf{P}\{B\} \qquad E[I_A I_B] = \mathbf{P}\{A \,\&\, B\}$$

so

$$\begin{aligned} Cov(I_A, I_B) &= \mathbf{P}\{A \,\&\, B\} - \mathbf{P}\{A\}\cdot\mathbf{P}\{B\} \\ &= \mathbf{P}\{B\}\,\big[\mathbf{P}\{A \mid B\} - \mathbf{P}\{A\}\big] \end{aligned}$$

- I_A and I_B are positively correlated if $\mathbf{P}\{A \mid B\} > \mathbf{P}\{A\}$.
- I_A and I_B are negatively correlated if $\mathbf{P}\{A \mid B\} < \mathbf{P}\{A\}$.
- I_A and I_B are uncorrelated if $\mathbf{P}\{A \mid B\} = \mathbf{P}\{A\}$ (i.e. independent).

Example

Example

Consider the following dice game played at some casinos. Players throw a pair of dice in turn. The bank then throws a pair of dice to determine the outcome as follows

- Player i wins if his roll is strictly greater than the banks.

Consider two players 1 and 2 who are in the game. For $i = 1, 2$, let

$$X_i = \begin{cases} 1 & \text{if player } i \text{ wins} \\ 0 & \text{otherwise.} \end{cases}$$

Are X_1 and X_2 correlated?

Example – continued

☞ We can determine if X_1 and X_2 are correlated by computing $\mathbf{P}\{X_2 = 1 \mid X_1 = 1\}$ and $\mathbf{P}\{X_2 = 1\}$.

☞ Let Z_b be the bank's score and Z_i ($i = 1, 2$) the score of player i.

$$\begin{aligned}
\mathbf{P}\{X_i = 1\} &= \mathbf{P}\{Z_b < Z_i\} \\
&= \sum_{k=2}^{12} \mathbf{P}\{Z_b < Z_i \mid Z_b = k\} \cdot \mathbf{P}\{Z_b = k\} \\
&= \frac{575}{1296} \approx 0.44 \\
\mathbf{P}\{X_1 = 1, X_2 = 1\} &= \mathbf{P}\{Z_b < Z_1, Z_b < Z_2\} \\
&= \sum_{k=2}^{12} \mathbf{P}\{Z_b < Z_1, Z_b < Z_2 \mid Z_b = k\} \cdot \mathbf{P}\{Z_b = k\} \\
&= \frac{13,037}{46,656} \\
\mathbf{P}\{X_2 = 1 \mid X_1 = 1\} &= \frac{13,037}{20,700} \approx 0.63
\end{aligned}$$

Correlation

Example – continued

☞ X_1 and X_2 are positively correlated, since $\mathbf{P}\{X_2 = 1 \mid X_1 = 1\} > \mathbf{P}\{X_2 = 1\}$:

$$\begin{aligned}
\mathbf{P}\{X_i = 1\} &\approx 0.44 \\
\mathbf{P}\{X_2 = 1 \mid X_1 = 1\} &\approx 0.63 \\
\mathbf{P}\{X_2 = 1 \mid X_1 = 1\} &> \mathbf{P}\{X_2 = 1\}
\end{aligned}$$