“The weak law of large numbers gives us a very valuable way of thinking about expectations.” ---Prof. Forsythe
Last time

- Random Variable
  - Expected value
  - Variance & covariance
Objectives

- Random Variable
  - Review
  - Covariance
  - *The weak law of large numbers*
  - *Simulation & example of airline overbooking*
The expected value (or expectation) of a random variable $X$ is

$$E[X] = \sum xP(x)$$

The expected value is a weighted sum of all the values $X$ can take.
Linearity of Expectation

\[ E[aX + bY] = a E[X] + b E[Y] \]

\[ E\left[ \sum_i c_i X_i \right] = \sum_i c_i E[X_i] \]
Expected value of a function of $X$

$$E[f(X)] = \sum_x f(x)P(x)$$

$$E[f(X,Y)] = \sum_x \sum_y f(x,y)P(x,y)$$

$P(X=x \cap Y=y)$
Motivation for covariance

- Study the relationship between random variables
- Note that it’s the un-normalized correlation
- Applications include the fire control of radar, communicating in the presence of noise.
Covariance

• The covariance of random variables $X$ and $Y$ is

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

• Note that

$$\text{cov}(X, X) = E[(X - E[X])^2] = \text{var}[X].$$

Here $f(x, y) = (X - E[X])(Y - E[Y])$.
A neater form for covariance

A neater expression for covariance (similar derivation as for variance)

\[
cov(X, Y) = E[XY] - E[X]E[Y]
\]

What is \( E[XY] \)?

\[
\sum_{xy} xy p(x, y)
\]

\[
\frac{1}{4} - \frac{1}{4}
\]
Correlation coefficient is normalized covariance

The correlation coefficient is

$$corr(X, Y) = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$$

When $X, Y$ takes on values with equal probability to generate data sets $\{(x,y)\}$, the correlation coefficient will be as seen in Chapter 2.
The correlation coefficient can also be written as:

$$corr(X, Y) = \frac{E[XY] - E[X]E[Y]}{\sigma_X \sigma_Y}$$
Correlation seen from scatter plots

Zero Correlation

No Correlation

Positive correlation

Positive Correlation

Negative correlation

Negative Correlation

Credit: Prof. Forsyth
Covariance seen from scatter plots

Zero Covariance → No Correlation

Positive Covariance → Positive Correlation

Negative Covariance → Negative Correlation

Credit: Prof. Forsyth
When correlation coefficient or covariance is zero

- The covariance is 0!
- That is: \( \text{Cov} = 0 \)

\[
E[XY] - E[X]E[Y] = 0
\]

\[
E[XY] = E[X]E[Y]
\]

- This is a necessary property of independence of random variables *(not equal to independence)*
Variance of the sum of two random variables

\[ \text{var}[X + Y] = \text{var}[X] + \text{var}[Y] + 2\text{cov}(X, Y) \]
If RVs $X$ & $Y$ are independent, then

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

$$E[XY] = \sum_{x} \sum_{y} x \cdot y \cdot p(x, y)$$

if $X$, $Y$ are indpt.

$$P(x, y) = p(x)p(y) \text{ for all } x, y$$

$$\text{RHS} = \sum_{x} \sum_{y} x \cdot p(x) \cdot y \cdot p(y)$$

$$= \sum_{y} y \cdot p(y) \sum_{x} x \cdot p(x)$$

$$= E[Y] \cdot E[X]$$

$$= E[X] \cdot E[Y]$$
These are equivalent! Uncorrelatedness

\[ E[XY] = E[X]E[Y] \]

if \( E[X] = E[X]E[Y] \)

then \( \text{cov}(X, Y) = E[XY] - E[X]E[Y] = 0 \)

\( \iff \left\langle \text{corr}(X, Y) = 0 \right\rangle \)

\[ \text{var}[X + Y] = \text{var}[X] + \text{var}[Y] \]
Q: What is this expectation?

We toss two identical coins A & B independently for three times and 4 times respectively, for each head we earn $1, we define $X$ is the earning from A and $Y$ is the earning from B. What is $E[XY]$?

A. 2  B. 3  C. 4

\[
E[XY] = E(X)E(Y)
\]
\[ X = A_1 + A_2 + A_3 \]
\[ Y = B_1 + B_2 + B_3 + B_4 \]
\[ E[X] = E[A_1] + E[A_2] + E[A_3] = 0.5 + 0.5 + 0.5 = 1.5 \]
\[ E[Y] = E[\sum B_i] = \sum_{i=1}^{4} E[B_i] = 4 \times 0.5 = 2 \]
\[ \because X, Y \text{ are independent, } E[XY] = E[X]E[Y] \]
\[ E[XY] = E[X]E[Y] = 1.5 \times 2 = 3 \]
If two random variables are uncorrelated, does this mean they are independent? Investigate the case $X$ takes -1, 0, 1 with equal probability and $Y=X^2$.

$E[X] = \frac{2}{3}$

$E[Y] = 2$ (Corrected)

$E[XY] = 0$

$X, Y$ are dependent but uncorrelated.
Covariance example

It’s an underlying concept in principal component analysis in Chapter 10
Random Variable Example for WLLN

Money box

* shake and take one and put back

10¢ 25¢
dime quarter

$P_{all-p} >$

$E[X] = ?$

$X_1 = 10? 25?$

$X_n = 10? 25?$

$10P + (1-P) 25$
2 RVs have the same distribution

\[ P(X = x) = \begin{cases} \frac{1}{2} & \text{tail} \\ 1 & \text{head} \end{cases} \]

\[ P(Y = y) = \begin{cases} \frac{1}{2} & 4 \text{-die comes up even} \\ 1 & 3 \text{ or } 4 \end{cases} \]

\[ Y(w) = \begin{cases} 0 & 4 \text{-die comes up even} \\ 1 & \text{odd} \end{cases} \]

\[ Z(w) = \begin{cases} 0 & 4 \text{-die comes up } 1 \text{ or } 2 \\ 1 & 3 \text{ or } 4 \end{cases} \]
Three experiments of 2 students

Report the sum of random number each finds after rolling a fair 4-die.

① each roll once, then add them.
② one of them rolls twice, then add them.
③ one of rolls once, then times with 2.

\[
\begin{align*}
X & \quad Y \\
\mathbb{X} & \mathbb{Y} \\
\text{(1)} & X + Y \\
\text{(2)} & X_1 + X_2 \\
\text{(3)} & 2 \cdot X_1
\end{align*}
\]
Markov’s inequality

For any random variable $X$ that only takes $x \geq 0$ and constant $a > 0$

$$P(X \geq a) \leq \frac{E[X]}{a}$$

For example, if $a = 10E[X]$

$$P(X \geq 10E[X]) \leq \frac{E[X]}{10E[X]} = 0.1$$
Proof of Markov’s inequality

\[ \text{given: } x \geq 0, \quad a > 0 \]

\[ \mathbb{E}[x] = \sum_x x \cdot p(x) = \sum_{x \in (-\infty, 0]} x \cdot p(x) + \sum_{x \in [0, a)} x \cdot p(x) + \sum_{x \in [a, \infty)} x \cdot p(x) \]

\[ \text{RHS} \geq 0 + \sum_{x \in [a, \infty)} x \cdot p(x) \geq a \sum_{x \in [a, \infty)} p(x) \]

\[ \mathbb{E}[x] \geq a \cdot p(x \geq a) \]

\[ p(x \geq a) \leq \frac{\mathbb{E}[x]}{a} \]

\[ = a \sum_{a \leq x < \infty} p(x) \]

\[ = a \cdot p(X \geq a) \]
Chebyshev’s inequality

For any random variable $X$ and constant $a > 0$

$$P(|X - E[X]| \geq a) \leq \frac{\text{var}[X]}{a^2}$$

If we let $a = k\sigma$ where $\sigma = \text{std}[X] = \sqrt{\text{var}[X]}$

$$P(|X - E[X]| \geq k\sigma) \leq \frac{1}{k^2}$$

In words, the probability that $X$ is greater than $k$ standard deviation away from the mean is small.
Proof of Chebyshev’s inequality

Given Markov inequality, \( a > 0, \ x \geq 0 \)

\[
P(X \geq a) \leq \frac{E[X]}{a}
\]

We can rewrite it as

\[
\omega > 0 \quad P(|U| \geq w) \leq \frac{E[|U|]}{w}
\]
Proof of Chebyshev’s inequality

If \( U = (X - E[X])^2 \)

\[
P(|U| \geq w) \leq \frac{E[|U|]}{w} = \frac{E[U]}{w} = \frac{\text{var} \{ X \}}{w}
\]

\[
P(|X - E[X]| \geq w) \leq \frac{\text{var} \{ X \}}{w}
\]

\[
P((X - E[X])^2 \geq w) \leq \frac{\text{var} \{ X \}}{w}
\]

\[
\omega = \sigma^2
\]
Proof of Chebyshev’s inequality

- Apply Markov inequality to \( U = (X - E[X])^2 \)

\[
P(|U| \geq w) \leq \frac{E[|U|]}{w} = \frac{E[U]}{w} = \frac{\text{var}[X]}{w}
\]

- Substitute \( U = (X - E[X])^2 \) and \( w = a^2 \)

\[
P((X - E[X])^2 \geq a^2) \leq \frac{\text{var}[X]}{a^2} \quad \text{Assume } a > 0
\]

\[
\Rightarrow P(|X - E[X]| \geq a) \leq \frac{\text{var}[X]}{a^2}
\]
Now we are closer to the law of large numbers
We define the sample mean $\overline{X}$ to be the average of $N$ random variables $X_1, \ldots, X_N$.

If $X_1, \ldots, X_N$ are independent and have identical probability function $P(x)$ then the numbers randomly generated from them are called IID samples.

The sample mean is a random variable.
Assume we have a set of IID samples from $N$ random variables $X_1, \ldots, X_N$ that have probability function $P(x)$.

We use $\overline{X}$ to denote the sample mean of these IID samples:

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$
Expected value of sample mean of IID random variables

By linearity of expected value

\[ E[\bar{X}] = E\left[ \frac{\sum_{i=1}^{N} X_i}{N} \right] = \frac{1}{N} \sum_{i=1}^{N} E[X_i] \]
Expected value of sample mean of IID random variables

By linearity of expected value

$$E[\bar{X}] = E\left[\frac{\sum_{i=1}^{N} X_i}{N}\right] = \frac{1}{N} \sum_{i=1}^{N} E[X_i]$$

Given each $X_i$ has identical $P(x)$

$$E[\bar{X}] = \frac{1}{N} \sum_{i=1}^{N} E[X] = E[X]$$
By the scaling property of variance

\[ \text{var} \left[ \frac{1}{N} \sum_{i=1}^{N} X_i \right] = \frac{1}{N^2} \text{var} \left[ \sum_{i=1}^{N} X_i \right] \]

\[ \text{var} \left[ X + T \right] = \text{var} \left[ X \right] + \text{var} \left[ T \right] \]

if \( X, T \) are i.i.d.

\[ \text{var} \left[ kX \right] = k^2 \text{var} \left[ X \right] \]
Variance of sample mean of IID random variables

By the scaling property of variance

\[
\text{var}[\bar{X}] = \text{var}\left[\frac{1}{N} \sum_{i=1}^{N} X_i\right] = \left(\frac{1}{N^2}\right) \text{var}\left[\sum_{i=1}^{N} X_i\right]
\]

And by independence of these IID random variables

\[
\text{var}[\bar{X}] = \frac{1}{N^2} \sum_{i=1}^{N} \text{var}[X_i] = \frac{1}{N^2} \cdot N \cdot \text{var}[x] = \frac{1}{N} \cdot \text{var}[x]
\]
Variance of sample mean of IID random variables

By the scaling property of variance

\[ \text{var}[\bar{X}] = \text{var}\left[ \frac{1}{N} \sum_{i=1}^{N} X_i \right] = \frac{1}{N^2} \text{var}\left[ \sum_{i=1}^{N} X_i \right] \]

And by independence of these IID random variables

\[ \text{var}[\bar{X}] = \frac{1}{N^2} \sum_{i=1}^{N} \text{var}[X_i] \]

Given each \( X_i \) has identical \( P(x) \), \( \text{var}[X_i] = \text{var}[X] \)

\[ \text{var}[\bar{X}] = \frac{1}{N^2} \sum_{i=1}^{N} \text{var}[X] = \frac{\text{var}[X]}{N} \]
Expected value and variance of sample mean of IID random variables

- The expected value of sample mean is the same as the expected value of the distribution:
  \[ E[\overline{X}] = E[X] \]

- The variance of sample mean is the distribution’s variance divided by the sample size \( N \):
  \[ var[\overline{X}] = \frac{var[X]}{N} \]
Weak law of large numbers

Given a random variable $X$ with finite variance, probability distribution function $P(x)$ and the sample mean $\overline{X}$ of size $N$.

For any positive number $\epsilon > 0$

$$\lim_{N \to \infty} P(\overline{X} - E[X] \geq \epsilon) = 0$$

That is: the value of the mean of IID samples is very close with high probability to the expected value of the population when sample size is very large.
Proof of Weak law of large numbers

Apply Chebyshev’s inequality

\[ P\left( \left| \overline{X} - E[\overline{X}] \right| \geq \epsilon \right) \leq \frac{\text{var}[\overline{X}]}{\epsilon^2} \]

\[ E[\overline{X}] = E[X] \]

\[ \text{var}[\overline{X}] = \frac{\text{var}[X]}{N} \]
Proof of Weak law of large numbers

- Apply Chebyshev’s inequality

\[ P\left(\left|\overline{X} - E[\overline{X}]\right| \geq \epsilon\right) \leq \frac{\text{var}[\overline{X}]}{\epsilon^2} \]

- Substitute \( E[\overline{X}] = E[X] \) and \( \text{var}[\overline{X}] = \frac{\text{var}[X]}{N} \)
Proof of Weak law of large numbers

Apply Chebyshev’s inequality

$$P(|X - E[X]| \geq \epsilon) \leq \frac{\text{var}[X]}{\epsilon^2}$$

Substitute $E\bar{X} = E[X]$ and $\text{var}[\bar{X}] = \frac{\text{var}[X]}{N}$

$$P(|\bar{X} - E[X]| \geq \epsilon) \leq \frac{\text{var}[X]}{N \epsilon^2} \quad \sim \quad 0$$

If $N \to \infty$

RHS $\to 0$
Proof of Weak law of large numbers

- Apply Chebyshev’s inequality

\[ P\left( |\bar{X} - E[\bar{X}]| \geq \epsilon \right) \leq \frac{\text{var}[\bar{X}]}{\epsilon^2} \]

- Substitute \( E[\bar{X}] = E[X] \) and \( \text{var}[\bar{X}] = \frac{\text{var}[X]}{N} \)

\[ P\left( |\bar{X} - E[X]| \geq \epsilon \right) \leq \frac{\text{var}[X]}{N\epsilon^2} \]

\( N \to \infty \)

0
Proof of Weak law of large numbers

- Apply Chebyshev’s inequality

\[ P(|\bar{X} - E[\bar{X}]| \geq \epsilon) \leq \frac{var[\bar{X}]}{\epsilon^2} \]

- Substitute \( E[\bar{X}] = E[X] \) and \( var[\bar{X}] = \frac{var[X]}{N} \)

\[ P(|\bar{X} - E[X]| \geq \epsilon) \leq \frac{var[X]}{N\epsilon^2} \]

\[ N \to \infty \]

\[ \lim_{N \to \infty} P(|\bar{X} - E[X]| \geq \epsilon) = 0 \]
Applications of the Weak law of large numbers

- The law of large numbers justifies using simulations (instead of calculation) to estimate the expected values of random variables

\[ \lim_{N \to \infty} P\left( \left| \overline{X} - E[X] \right| \geq \epsilon \right) = 0 \]

- The law of large numbers also justifies using histogram of large random samples to approximate the probability distribution function \( P(x) \)

see proof on Pg. 353 of the textbook by DeGroot, et al.
\( Ru \times = \{ \} \quad \text{if any } E \text{ of interest happens} \)

\[
E[x] = 1 \times P(\text{any Event of interest}) \\
+ 0 \times P(\text{ow}) \\
= P(\text{any Event of interest})
\]
The law of large numbers justifies using histograms to approximate the probability distribution. Given \( N \) IID random variables \( X_1, ..., X_N \),

According to the law of large numbers

\[
\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} \quad \stackrel{N \to \infty}{\longrightarrow} \quad E[Y_i]
\]

As we know for indicator function

\[
E[Y_i] = P(c_1 \leq X_i < c_2) = P(c_1 \leq X < c_2)
\]
Simulation of the sum of two-dice

http://www.randomservices.org/random/apps/DiceExperiment.html
An airline has a flight with \( s \) seats. They always sell \( t \) (\( t > s \)) tickets for this flight. If ticket holders show up independently with probability \( p \), what is the probability that the flight is overbooked?

\[
P(\text{overbooked}) = \sum_{u=s+1}^{t} C(t, u) p^u (1 - p)^{t-u}
\]
An airline has a flight with 7 seats. They always sell 12 tickets for this flight. If ticket holders show up independently with probability $p$, estimate the following values:

- Expected value of the number of ticket holders who show up
- Probability that the flight being overbooked
- Expected value of the number of ticket holders who can’t fly due to the flight is overbooked.
Conditional expectation

(Expected value of $X$ conditioned on event $A$):

$$E[X|A] = \sum_{x \in D(X)} xP(X = x|A)$$

(Expected value of the number of ticketholders not flying)

$$E[\text{NF|overbooked}] = \sum_{u=s+1}^{t} (u - s) \frac{\binom{t}{u} p^u (1 - p)^{t-u}}{\sum_{v=s+1}^{t} \binom{t}{v} p^v (1 - p)^{t-v}}$$
### Simulate the arrival

- **Expected value of the number of ticket holders who show up**

\[ nt = 100000, \quad t = 12, \quad s = 7, \quad p = 0.1, 0.2, \ldots 1.0 \]

<table>
<thead>
<tr>
<th>Num of tickets (t)</th>
<th>Num of trials (nt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

We generate a matrix of random numbers from uniform distribution in [0,1],

*Any number < p is considered an arrival*
Simulate the arrival

Expected value of the number of ticket holders who show up

$n_t=100000$, $t=12$, $s=7$, $p=0.1$, 0.2, ... 1.0
Simulate the expected probability of overbooking

Expected probability of the flight being overbooked

\[ t = 12, \ s = 7, \ p = 0.1, 0.2, \ldots, 1.0 \]

Expected probability is equal to the expected value of indicator function. Whenever we have Num of arrival > Num of seats, we mark it with an indicator function. Then estimate with the sample mean of indicator functions.
Simulate the expected probability of overbooking

Expected probability of the flight being overbooked

\[ n_t = 100000, \quad t = 12, \quad s = 7, \quad p = 0.1, 0.2, \ldots 1.0 \]
Simulate the expected value of the number of grounded ticket holders given overbooked

Expected value of the number of ticket holders who can’t fly due to the flight being overbooked

\[ N_t = 200000, \]
\[ t = 12, s = 7, \]
\[ p = 0.1, 0.2, \ldots 1.0 \]
Assignments

- Continue to work on HW4
- Module Week 5
- Next time: Continuous random variable, classic known probability distributions
Additional References

Charles M. Grinstead and J. Laurie Snell
"Introduction to Probability”

Morris H. Degroot and Mark J. Schervish
"Probability and Statistics”
See you next time

See You!