Announcements

Hooray midterm!
  ◦ Grades (hopefully) by Monday

Problem Set #3
  ◦ Should be graded by Monday as well (instead of Friday)

Quick note about Piazza
Goals for today

As engineers, why are we doing all of this?
Wrap-up of joint/independent RVs
Covariance
  ◦ Definition, intuition, and properties
  ◦ Properties of independent RVs
  ◦ Correlation
  ◦ Bringing back the Poisson Paradigm
Where are we going with all of this?

As engineers, we want to:

1. Model things that can be random.

2. Find average values of situations that let us make good decisions.

So far we have learned:

- If possibilities are finite, we can count equally likely outcomes
- If we are dealing with a 1-D domain, we can use 1-D RVs with parameters

But the real world is more complex:

- Multi-dimensional (e.g., 3-D world)
- We don’t know parameters, much less what RV to use

We perform multiple experiments to help us with engineering.
Where are we going with all of this?

As engineers, we want to:

1. Model things that can be random.
2. Find average values of situations that let us make good decisions.

We perform multiple experiments to help us with engineering.

i.i.d.: We often want to aggregate the results of repeated experiments:
- Throwing darts or flipping coins
- Counts of successes or arrivals
- Clinical trials

Or we just want to model more complex functions:
- Receiving a signal that had some added noise
- Receiving web server requests from different sources

\begin{align*}
\text{Averages/Counts} &= \text{sums} \\
\text{Sums of things distributed similarly} &\quad \text{(expectation)}
\end{align*}
Where are we going with all of this?

As engineers, we want to:

1. Model things that can be random.
2. Find average values of situations that let us make good decisions.

We perform multiple experiments to help us with engineering. Counts and averages are both sums of RVs.

This week + last week:
- Defining joint PMF/PDF
- How to model sums of 2 RVs
- How to deal with uncertainty of model (coin flips only)

Next week (week 6):
- Finding expectation of complex situations
- Defining sampling and using existing data
- Mathematical bounds w.r.t modeling the average

Week 7: bringing it all together
- Sampling + average = Central Limit Theorem
- Samples + modeling = finding the best model parameters given data
Summary of last time

\[ Z = X + Y , \text{ where } X \text{ and } Y \text{ are independent RVs:} \]

\[
p_Z(z) = \sum_{x} p_X(x)p_Y(z-x) \quad \text{(discrete } X \text{ and } Y) \]

\[
f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx \quad \text{(continuous } X \text{ and } Y) \]

Convenient sums of independent RVs:

- Binomial: if same \( p \) of trial success, sum up trial counts \( n_i \)
- Poisson: sum up rates \( \lambda_i \)
  - we proved that if \( X \sim \text{Poi}(\lambda p) \) and \( Y \sim \text{Poi}(\lambda(1-p)), X + Y \sim \text{Poi}(p). \) (Web Servers)
- Normal: sum up means, sum up variances
  - On homework, you will find distribution of \( aX + bY \) (linear comb. of normal)
Let $X$ and $Y$ be independent RVs, where $X \sim \text{Unif}(0,1)$ and $Y \sim \text{Unif}(0,1)$.

What is the distribution of $Z = X + Y$?

Solution:

\[
 f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a - y)f_Y(y) \, dy = \int_0^1 f_X(a - y)f_Y(y) \, dy = \int_0^1 f_X(a - y) \, dy
\]

Note: for different $a$, need bounds on $y$ and $a$ such that $f_X(a - y) = 1$
Defects on a hard drive

A single point defect is uniformly distributed over a disk of radius $R$.

Are $X$ and $Y$ independent?

Solution:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy = \frac{1}{\pi R^2} \int_{y: x^2 + y^2 \leq R^2} dy$$

$$f_X(x) = \frac{1}{\pi R^2} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy$$

$$f_X(x) = \frac{2}{\pi R^2} \sqrt{R^2 - x^2}$$

if $0 \leq x \leq R$

$$f_Y(y) = \frac{2}{\pi R^2} \sqrt{R^2 - y^2}$$

if $0 \leq y \leq R$, by symmetry

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi R^2} & \text{if } x^2 + y^2 \leq R^2 \\ 0 & \text{otherwise} \end{cases}$$

$$f_X(x) f_Y(y) \neq f_{X,Y}(x,y)$$

No. $X$ and $Y$ are dependent.
Defects on a hard drive

A single point defect is uniformly distributed over a disk of radius $R$. Calculate the expected distance of the defect from the center, i.e., $E[D]$, where $D = \sqrt{X^2 + Y^2}$.

Solution:

$$F_D(a) = \frac{a^2}{R^2}$$

$$E[D] = \int_0^R P(D > a) da = \int_0^R (1 - F_D(a)) da$$

$$= \int_0^R \left(1 - \frac{a^2}{R^2}\right) da$$

$$= \left[a - \frac{a^3}{3R^2}\right]_0^R = \frac{2}{3}R$$
Expectation as a summary

For a single variable, we have a few summary values:

- Mean, $E[X]$
- Variance (a measure of spread), $\text{Var}(X)$

For two variables:

In both distributions: $E[X] = E[Y]$, $\text{Var}(X) = \text{Var}(Y)$

Difference: how the two variables vary with each other.
Covariance

The covariance of two variables X and Y is:

\[ \text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \]

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

Proof of second part:

\[ \text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \]


\[ = E[XY] - E[X]E[Y] \]
Covariance of a die roll

Roll our favorite 6-sided die.

Define two indicator variables:

\[ X = 1 \text{ if roll in } \{1, 2, 3, 4\} \]
\[ Y = 1 \text{ if roll in } \{3, 4, 5, 6\} \]

What is \( \text{Cov}(X, Y) \)?

Solution:

\[ E[X] = 2/3, \quad E[Y] = 2/3 \]

\[ E[XY] = \sum_x \sum_y xy p_{X,Y}(x, y) = (0 \cdot 0) + \left(0 \cdot \frac{1}{3}\right) + \left(0 \cdot \frac{1}{3}\right) + \left(1 \cdot \frac{1}{3}\right) = \frac{1}{3} \]

\[ \text{Cov}(X, Y) = E[XY] - E[X]E[Y] = \frac{1}{3} - \frac{4}{9} = -\frac{1}{9} \]

What does negative covariance mean?

Consider: \( P(X = 1) = 2/3 \), whereas \( P(X = 1|Y = 1) = 1/2 \)

Observing \( Y = 1 \) (positive) makes \( X = 1 \) (positive) less likely.
Covariance of humans

<table>
<thead>
<tr>
<th>Weight</th>
<th>Height</th>
<th>W · H</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>57</td>
<td>3648</td>
</tr>
<tr>
<td>71</td>
<td>59</td>
<td>4189</td>
</tr>
<tr>
<td>53</td>
<td>49</td>
<td>2597</td>
</tr>
<tr>
<td>67</td>
<td>62</td>
<td>4154</td>
</tr>
<tr>
<td>55</td>
<td>51</td>
<td>2805</td>
</tr>
<tr>
<td>58</td>
<td>50</td>
<td>2900</td>
</tr>
<tr>
<td>77</td>
<td>55</td>
<td>4235</td>
</tr>
<tr>
<td>57</td>
<td>48</td>
<td>2736</td>
</tr>
<tr>
<td>56</td>
<td>42</td>
<td>2352</td>
</tr>
<tr>
<td>51</td>
<td>42</td>
<td>2142</td>
</tr>
<tr>
<td>76</td>
<td>61</td>
<td>4636</td>
</tr>
<tr>
<td>68</td>
<td>57</td>
<td>3876</td>
</tr>
</tbody>
</table>

What is the covariance of weight (W) and height (H)?


Solution:
\[ \text{Cov}(W, H) = 3355.83 - (62.75)(52.75) = 45.77 \]

What does positive covariance mean?

Observing high W makes high H more likely.
Covariance

Q: Is the covariance positive, negative, or zero?

A: positive!
Properties of Covariance

Let $X$ and $Y$ be arbitrary random variables


Symmetry:

$$\text{Cov}(X, Y) = \text{Cov}(Y, X)$$

Variance, redefined:

$$\text{Cov}(X, X) = E[XX] - E[X]E[X] = \text{Var}(X)$$

Non-linear:

$$\text{Cov}(aX + b, Y) = a\text{Cov}(X, Y)$$

Covariance and sums of $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$:

$$\text{Cov} \left( \sum_i X_i, \sum_j Y_j \right) = \sum_i \sum_j \text{Cov} \left( X_i, Y_j \right)$$

(covariance of all pairs)
Variance of Sums of Variables

\[
\text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \text{Cov}(X_i, X_j)
\]

For 2 variables:
\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)
\]

Proof:
\[
\text{Var}\left(\sum_{i=1}^{n} X_i \right) = \text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i \right)
\]
\[
= \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov}(X_i, X_j)
\]
\[
= \sum_{i=1}^{n} \text{Var}(X_i) + \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \text{Cov}(X_i, X_j)
\]
\[
= \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \text{Cov}(X_i, X_j)
\]

(Cov(X, X) = Var(X))

(covariance of all pairs)

(symmetry)
Break

Attendance: tinyurl.com/cs109summer2018
Independent RVs

Definition: Two random variables $X$ and $Y$ are independent if:

- **Discrete case:**
  \[ p_{X,Y}(x, y) = p_X(x)p_Y(y) \]

- **Continuous case:**
  \[ f_{X,Y}(x, y) = f_X(x)f_Y(y) \]

- **General case:**
  \[ F_{X,Y}(x, y) = F_X(x)F_Y(y) \]

Properties of independence (next 2 slides)

Not bidirectional!

If $X$ and $Y$ are independent, then these properties hold.

If these properties hold, $X$ and $Y$ are not necessarily independent.
Product of independent RVs

If $X$ and $Y$ are independent, then:

$$E[XY] = E[X]E[Y]$$
$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Proof:

$$E[g(X)h(Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_{X,Y}(x,y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty} h(y)f_Y(y)\,dy \int_{-\infty}^{\infty} g(x)f_X(x)\,dx$$
$$= (\int_{-\infty}^{\infty} g(x)f_X(x)\,dx)(\int_{-\infty}^{\infty} h(y)f_Y(y)\,dy)$$
$$= E[g(X)]E[h(Y)]$$
Independent RVs

Definition: Two random variables $X$ and $Y$ are independent if:

- $(\text{discrete})$
  \[ p_{X,Y}(x, y) = p_X(x)p_Y(y) \]

- $(\text{continuous})$
  \[ f_{X,Y}(x, y) = f_X(x)f_Y(y) \]

- $(\text{any } X \text{ and } Y)$
  \[ F_{X,Y}(x, y) = F_X(x)F_Y(y) \]

Properties of independence:

\[ E[XY] = E[X]E[Y] \]
\[ \text{Cov}(X, Y) = 0 \]
\[ \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \]
\[ = \text{Var}(X) + \text{Var}(Y) \]
Covariance does not imply independence

Let $X$ take on values $\{-1, 0, 1\}$ with equal probability $(1/3)$.

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

1. What are $E[X]$ and $E[Y]$?
   
   $E[X] = -1 \left(\frac{1}{3}\right) + 0 \left(\frac{1}{3}\right) + 1 \left(\frac{1}{3}\right) = 0$
   
   $E[Y] = -0 \left(\frac{2}{3}\right) + 1 \left(\frac{1}{3}\right) = \frac{1}{3}$

2. What is the covariance of $X$ and $Y$?
   
   $XY = 0$ for all $X$ and $Y$
   
   $\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0 - 0 \left(\frac{1}{3}\right) = 0$

3. Are $X$ and $Y$ independent?
   
   $P(Y = 0|X = 1) = 0 \neq 2/3 = P(Y = 0)$

   Even though $X$ and $Y$ are clearly dependent!
Statistics of functions of 2 RVs

Sum: always can be separated, regardless of independence
\[ E[X + Y] = E[X] + E[Y] \]

Product: can only be separated if \( X, Y \) independent
\[ E[XY] = E[X]E[Y] \]

Variance for any \( X \) and \( Y \):
\[ \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y) \]

Variance for \( X, Y \) independent:
\[ \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \]

If all \( X_1, X_2, ..., X_n \) are independent:
\[ \text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i) \]
Variance of Binomial

For a binomial RV \( X \sim \text{Bin}(n,p) \),

\[
E[X] = np \quad \text{(expectation of sum = sum of expectation)}
\]

\[
\text{Var}(X) = np(1 - p)
\]

Proof of variance:

Define: \( X_k \) as an indicator variable for success on trial \( k \). \( P(X_k = 1) = p \)

\[
\text{Var}(X) = \text{Var} \left( \sum_{k=1}^{n} X_k \right) = \sum_{k=1}^{n} \text{Var}(X_k) = \sum_{k=1}^{n} p(1 - p) = np(1 - p)
\]

Variance of sum of independent RVs = sum of variance

Variance of Bernoulli
Correlation

The correlation of two random variables $X$ and $Y$ is

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Note: $-1 \leq \rho(X, Y) \leq 1$

Measures the linearity of the relationship between $X$ and $Y$:

If $Y = aX + b$, then

$$\text{Cov}(X, Y) = \text{Cov}(X, aX + b) = a\text{Cov}(X, X) = a\text{Var}(X)$$

$$= \text{sgn}(a) \sqrt{a^2(\text{Var}(X))^2} = \text{sgn}(a) \sqrt{(a^2\text{Var}(X))\text{Var}(X)}$$

$$= \text{sgn}(a) \sqrt{\text{Var}(Y)\text{Var}(X)}$$

$$\rho(X, Y) = \text{sgn}(a)$$

($\text{sgn}(a)$: sign of $a$)
Correlation quantifies linearity

\[ \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} \]

- \( \rho(X, Y) = -1 \)
  - \( Y = aX + b \)
  - \( a = -\frac{\sigma_Y}{\sigma_X} \)

- \( \rho(X, Y) = 0 \)
  - "uncorrelated"

- \( \rho(X, Y) = 1 \)
  - \( Y = aX + b \)
  - \( a = \frac{\sigma_Y}{\sigma_X} \)

\[ \text{Var}(Y) = \text{Var}(aX + b) = \left(\frac{\sigma_Y}{\sigma_X}\right)^2 \text{Var}(X) = \sigma_Y^2 \]

\( Y = X^2 \)

No linear relationship
\[ \text{== Cov}(X,Y) = 0, \]
but \( X \) and \( Y \) could be nonlinearly related!
Spurious Correlations

$\rho(X, Y)$ is used a lot in statistics to quantify the relationship between $X$ and $Y$.

“Correlation does not imply causation”
Do indicator variables correlate?

Define indicator variables $I_A$ and $I_B$ for events A and B.

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

What is $\text{Cov}(I_A, I_B)$? (in terms of $P(A)$, $P(B)$, $P(A|B)$)

**Solution:**

$$E[I_A] = P(A), \quad E[I_B] = P(B), \quad E[I_AI_B] = P(AB) = P(A|B)P(B)$$

$$\text{Cov}(I_A, I_B) = E[I_AI_B] - E[I_A]E[I_B] = P(A|B)P(B) - P(A)P(B)$$

$$= P(B)[P(A|B) - P(A)]$$

$\rightarrow$ Sign of $\text{Cov}(I_A, I_B)$ determined by

$$P(A|B) - P(A)$$

$$P(A|B) > P(A) \implies \rho(I_A, I_B) > 0$$

$$P(A|B) = P(A) \implies \rho(I_A, I_B) = 0 \quad \text{(independence, Cov}(I_A, I_B) = 0$$

$$P(A|B) < P(A) \implies \rho(I_A, I_B) < 0$$
Multinomial Random Variable

Consider \( n \) independent trials of an experiment.

- Each trial results in one of \( m \) outcomes, with respective probabilities \( p_1, p_2, \ldots, p_m \) where \( \sum_{k=1}^{m} p_k = 1 \).
- \( X_k \) is number of trials with outcome \( k \) for \( k = 1, \ldots, m \).

\[
X_1, \ldots, X_m \sim \text{Multi}(n, p_1, p_2, \ldots, p_m)
\]

\[
P(X_1 = c_1, X_2 = c_2, \ldots, X_m = c_m) = \binom{n}{c_1, c_2, \ldots, c_m} p_1^{c_1} p_2^{c_2} \ldots p_m^{c_m}
\]

where \( \sum_{k=1}^{m} c_k = n \) and \( \binom{n}{c_1, c_2, \ldots, c_m} \) is a multinomial coefficient.

We expect that the \( X_k \) are negatively correlated.
Covariance of the multinomial

Consider $n$ independent trials of an experiment.
- Each trial results in one of $m$ outcomes, w.p. $p_1, p_2, \ldots, p_m$ where $\sum_{k=1}^{m} p_k = 1$
- $X_k$ is number of trials with outcome $k$ for $k = 1, \ldots, m$.

What is $\text{Cov}(X_j, X_k)$, where $j \neq k$? Say, rolls of 3 and rolls of 5

Solution:

Define: Indicator $I_k(r) = 1$ if trial $r$ has outcome $k$, 0 otherwise
\[
E[I_k(r)] = p_k, \quad X_k = \sum_{r=1}^{n} I_k(r) \quad E[I_j(r)] = p_j, \quad X_j = \sum_{r=1}^{n} I_j(r)
\]

\[
\text{Cov}(X_j, X_k) = \sum_{a=1}^{n} \sum_{b=1}^{n} \text{cov}(I_j(a), I_k(b)) = \sum_{a=b=1}^{n} \text{cov}(I_j(a), I_k(b))
\]

- $a \neq b$: trial a and b independent, $\text{Cov}(I_j(a), I_k(b)) = 0$
- $a = b$: trial a cannot have outcome $i$ and $j$, $E[I_j(a)I_k(a)] = 0$

\[
= \sum_{a=1}^{n} E[I_j(a)I_k(a)] - E[I_j(a)]E[I_k(a)]
= \sum_{a=1}^{n} -E[I_j(a)]E[I_k(a)]
= \sum_{a=1}^{n} (-p_j p_k) = -np_jp_k
\]

$X_j, X_k$ are negatively correlated.
The Poisson Paradigm

Multinomial distributions:
- Count of strings hashed to buckets in hash table
- Number of server requests across machines in cluster
- Distribution of words/tokens in an email
- Etc.

When \( m \) (# outcomes) is large, \( p_k \) is small, and for equally likely outcomes, where \( p_k = \frac{1}{m} \):

\[
\text{Cov}(X_j, X_k) = -np_j p_k = -\frac{n}{m^2} \rightarrow 0 \text{ as } m \text{ gets large!}
\]

For large \( m \), \( X_j, X_k \) are very mildly negatively correlated, and the Poisson paradigm is applicable.
Summary of this time


Covariance is a linear relationship.

Independence:
- Implies uncorrelated (but uncorrelated does NOT imply independence)
- For \( n \) independent variables: \( \text{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{Var}(X_i) \)

Correlation is a normalized version of Covariance:

\[ \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \]