### Leftover: Independence, Covariance and Correlation

**18.05 Spring 2017**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
</tr>
<tr>
<td>2</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
</tr>
<tr>
<td>3</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
</tr>
<tr>
<td>4</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
</tr>
<tr>
<td>5</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
</tr>
<tr>
<td>6</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
</tr>
</tbody>
</table>
Independence

Events $A$ and $B$ are independent if

$$P(A \cap B) = P(A)P(B).$$

Random variables $X$ and $Y$ are independent if

$$F(x, y) = F_X(x)F_Y(y).$$

Discrete random variables $X$ and $Y$ are independent if

$$p(x_i, y_j) = p_X(x_i)p_Y(y_j).$$

Continuous random variables $X$ and $Y$ are independent if

$$f(x, y) = f_X(x)f_Y(y).$$
Concept question: independence

Roll two dice: \( X \) = value on first, \( Y \) = value on second

<table>
<thead>
<tr>
<th>( X \setminus Y )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>( p(x_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{6} )</td>
</tr>
<tr>
<td>2</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{6} )</td>
</tr>
<tr>
<td>3</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{6} )</td>
</tr>
<tr>
<td>4</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{6} )</td>
</tr>
<tr>
<td>5</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{6} )</td>
</tr>
<tr>
<td>6</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{6} )</td>
</tr>
</tbody>
</table>

| \( p(y_j) \) | \( \frac{1}{6} \) | \( \frac{1}{6} \) | \( \frac{1}{6} \) | \( \frac{1}{6} \) | \( \frac{1}{6} \) | \( \frac{1}{6} \) | 1 |

Are \( X \) and \( Y \) independent? 1. Yes 2. No
**Concept question: independence II**

Roll two dice: $X =$ value on first, $T =$ sum

<table>
<thead>
<tr>
<th>$X \backslash T$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>$p(x_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$1/6$</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$1/6$</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$1/6$</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$1/6$</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>0</td>
<td>0</td>
<td>$1/6$</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/36$</td>
<td>$1/6$</td>
</tr>
</tbody>
</table>

$\begin{array}{l}p(y_j) \\ \hline 1/36 & 2/36 & 3/36 & 4/36 & 5/36 & 6/36 & 5/36 & 4/36 & 3/36 & 2/36 & 1/36 & 1\end{array}$

Are $X$ and $Y$ independent? 1. Yes 2. No
Concept Question

Among the following pdf’s which are independent? (Each of the ranges is a rectangle chosen so that \( \int \int f(x, y) \, dx \, dy = 1 \).)

(i) \( f(x, y) = 4x^2y^3 \).
(ii) \( f(x, y) = \frac{1}{2}(x^3y + xy^3) \).
(iii) \( f(x, y) = 6e^{-3x-2y} \)

Put a 1 for independent and a 0 for not-independent.

(a) 111   (b) 110   (c) 101   (d) 100
(e) 011   (f) 010   (g) 001   (h) 000
Covariance

Measures the degree to which two random variables vary together, e.g. height and weight of people.

\(X, Y\) random variables with means \(\mu_X\) and \(\mu_Y\).

\[
\text{Cov}(X, Y) = \text{def} \ E((X - \mu_X)(Y - \mu_Y)).
\]

- **Vary together** might mean \(X\) is usually bigger than \(\mu_X\) when \(Y\) is bigger than \(\mu_Y\), and vice versa. In this case \((X - \mu_X)(Y - \mu_Y)\) is usually positive, so \(\text{Cov}(X, Y)\) is positive.

- **Vary together** might mean \(X\) is usually bigger than \(\mu_X\) when \(Y\) is smaller than \(\mu_Y\), and vice versa. In this case \((X - \mu_X)(Y - \mu_Y)\) is usually negative, so \(\text{Cov}(X, Y)\) is negative.

- If \(X\) and \(Y\) don’t vary together, then sign of \((X - \mu_X)\) tells nothing about sign of \((Y - \mu_Y)\). In this case \((X - \mu_X)(Y - \mu_Y)\) can be both positive and negative, so \(\text{Cov}(X, Y)\) might be zero or small.
Properties of covariance

Properties
1. $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$ for constants $a, b, c, d$.
2. $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$.
3. $\text{Cov}(X, X) = \text{Var}(X)$
4. $\text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y$.
5. If $X$ and $Y$ are independent then $\text{Cov}(X, Y) = 0$.
6. **Warning** The converse is not true: if covariance is 0, the variables might not be independent.
Concept question

Suppose we have the following joint probability table.

\[
\begin{array}{c|ccc|c}
Y \backslash X & -1 & 0 & 1 & p(y_j) \\
\hline
0 & 0 & 1/2 & 0 & 1/2 \\
1 & 1/4 & 0 & 1/4 & 1/2 \\
\hline
p(x_i) & 1/4 & 1/2 & 1/4 & 1
\end{array}
\]

At your table work out the covariance \( \text{Cov}(X, Y) \).
Concept question

Suppose we have the following joint probability table.

<table>
<thead>
<tr>
<th>Y \ X</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>$p(y_j)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1/2</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>1</td>
<td>1/4</td>
<td>0</td>
<td>1/4</td>
<td>1/2</td>
</tr>
</tbody>
</table>

At your table work out the covariance $\text{Cov}(X, Y)$.

Because the covariance is 0 we know that $X$ and $Y$ are independent

1. True 2. False
Concept question

Suppose we have the following joint probability table.

<table>
<thead>
<tr>
<th>Y \ X</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>p(y_j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1/2</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>1</td>
<td>1/4</td>
<td>0</td>
<td>1/4</td>
<td>1/2</td>
</tr>
</tbody>
</table>

p(x_i) 1/4 1/2 1/4 1

At your table work out the covariance Cov(X, Y).

Because the covariance is 0 we know that X and Y are independent

1. True 2. False

Key point: covariance measures the linear relationship between X and Y. It can completely miss a quadratic or higher order relationship.
Flip a fair coin 12 times.

Let \( X = \) number of heads in the first 7 flips

Let \( Y = \) number of heads on the last 7 flips.

Compute \( \text{Cov}(X, Y) \),
Correlation

Like covariance, but removes scale.
The *correlation coefficient* between $X$ and $Y$ is defined by

$$\text{Cor}(X, Y) = \rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$ 

1. $\rho$ = covariance of standardized versions of $X$ and $Y$.
2. $\rho$ is dimensionless (it’s a ratio).
3. $-1 \leq \rho \leq 1$.
4. $\rho = 1$ if and only if $Y = aX + b$ with $a > 0$.
5. $\rho = -1$ if and only if $Y = aX + b$ with $a < 0$. 
Real-life correlations

- Over time, amount of ice cream consumption is correlated with number of pool drownings.
- In 1685 (and today) being a student is the most dangerous profession.
- In 90% of bar fights ending in a death the person who started the fight died.
- Hormone replacement therapy (HRT) is correlated with a lower rate of coronary heart disease (CHD).
Correlation is not causation

Edward Tufte: "Empirically observed covariation is a necessary but not sufficient condition for causality."
Overlapping sums of uniform random variables

We made two random variables $X$ and $Y$ from overlapping sums of uniform random variables

For example:

$$X = X_1 + X_2 + X_3 + X_4 + X_5$$
$$Y = X_3 + X_4 + X_5 + X_6 + X_7$$

These are sums of 5 of the $X_i$ with 3 in common.

If we sum $r$ of the $X_i$ with $s$ in common we name it $(r, s)$.

Below are a series of scatterplots produced using R.
Scatter plots

(1, 0) cor=0.00, sample_cor=-0.07

(2, 1) cor=0.50, sample_cor=0.48

(5, 1) cor=0.20, sample_cor=0.21

(10, 8) cor=0.80, sample_cor=0.81
Concept question

Toss a fair coin $2n + 1$ times. Let $X$ be the number of heads on the first $n + 1$ tosses and $Y$ the number on the last $n + 1$ tosses.

If $n = 1000$ then $\text{Cov}(X, Y)$ is:

(a) 0  (b) $1/4$  (c) $1/2$  (d) 1

(e) More than 1  (f) tiny but not 0
Board question

Toss a fair coin $2n + 1$ times. Let $X$ be the number of heads on the first $n + 1$ tosses and $Y$ the number on the last $n + 1$ tosses.

Compute $\text{Cov}(X, Y)$ and $\text{Cor}(X, Y)$.
Review for Exam 1
18.05 Spring 2017
Extra office hours

- Tuesday:
  - David 3–5 in 2-355
  - Watch web site for more
- Friday, Saturday, Sunday March 10–12: no office hours
Exam 1

- Designed to be 1 hour long. You’ll have the entire 80 minutes.

- You may bring one 4 by 6 notecard. This will be turned in with your exam. (Be sure to write your name on the card.)

- Lots of practice problems posted on class web site.

- No calculators. (They won’t be necessary.)

- Be sure to get familiar with the table of normal probabilities (it’s easy).
### Normal Table

**Standard normal table of left tail probabilities.**

<table>
<thead>
<tr>
<th>$z$</th>
<th>$\Phi(z)$</th>
<th>$z$</th>
<th>$\Phi(z)$</th>
<th>$z$</th>
<th>$\Phi(z)$</th>
<th>$z$</th>
<th>$\Phi(z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.00</td>
<td>0.0000</td>
<td>-2.00</td>
<td>0.0228</td>
<td>0.00</td>
<td>0.5000</td>
<td>2.00</td>
<td>0.9772</td>
</tr>
<tr>
<td>-3.95</td>
<td>0.0000</td>
<td>-1.95</td>
<td>0.0256</td>
<td>0.05</td>
<td>0.5199</td>
<td>2.05</td>
<td>0.9798</td>
</tr>
<tr>
<td>-3.90</td>
<td>0.0000</td>
<td>-1.90</td>
<td>0.0287</td>
<td>0.10</td>
<td>0.5398</td>
<td>2.10</td>
<td>0.9821</td>
</tr>
<tr>
<td>-3.85</td>
<td>0.0001</td>
<td>-1.85</td>
<td>0.0322</td>
<td>0.15</td>
<td>0.5596</td>
<td>2.15</td>
<td>0.9842</td>
</tr>
<tr>
<td>-3.80</td>
<td>0.0001</td>
<td>-1.80</td>
<td>0.0359</td>
<td>0.20</td>
<td>0.5793</td>
<td>2.20</td>
<td>0.9861</td>
</tr>
<tr>
<td>-3.75</td>
<td>0.0001</td>
<td>-1.75</td>
<td>0.0401</td>
<td>0.25</td>
<td>0.5987</td>
<td>2.25</td>
<td>0.9878</td>
</tr>
<tr>
<td>-3.70</td>
<td>0.0001</td>
<td>-1.70</td>
<td>0.0446</td>
<td>0.30</td>
<td>0.6179</td>
<td>2.30</td>
<td>0.9893</td>
</tr>
<tr>
<td>-3.65</td>
<td>0.0001</td>
<td>-1.65</td>
<td>0.0495</td>
<td>0.35</td>
<td>0.6368</td>
<td>2.35</td>
<td>0.9906</td>
</tr>
<tr>
<td>-3.60</td>
<td>0.0002</td>
<td>-1.60</td>
<td>0.0548</td>
<td>0.40</td>
<td>0.6554</td>
<td>2.40</td>
<td>0.9918</td>
</tr>
<tr>
<td>-3.55</td>
<td>0.0002</td>
<td>-1.55</td>
<td>0.0606</td>
<td>0.45</td>
<td>0.6736</td>
<td>2.45</td>
<td>0.9929</td>
</tr>
<tr>
<td>-3.50</td>
<td>0.0002</td>
<td>-1.50</td>
<td>0.0668</td>
<td>0.50</td>
<td>0.6915</td>
<td>2.50</td>
<td>0.9938</td>
</tr>
<tr>
<td>-3.45</td>
<td>0.0003</td>
<td>-1.45</td>
<td>0.0735</td>
<td>0.55</td>
<td>0.7088</td>
<td>2.55</td>
<td>0.9946</td>
</tr>
<tr>
<td>-3.40</td>
<td>0.0003</td>
<td>-1.40</td>
<td>0.0808</td>
<td>0.60</td>
<td>0.7257</td>
<td>2.60</td>
<td>0.9953</td>
</tr>
<tr>
<td>-3.35</td>
<td>0.0004</td>
<td>-1.35</td>
<td>0.0885</td>
<td>0.65</td>
<td>0.7422</td>
<td>2.65</td>
<td>0.9960</td>
</tr>
<tr>
<td>-3.30</td>
<td>0.0005</td>
<td>-1.30</td>
<td>0.0968</td>
<td>0.70</td>
<td>0.7580</td>
<td>2.70</td>
<td>0.9965</td>
</tr>
<tr>
<td>-3.25</td>
<td>0.0006</td>
<td>-1.25</td>
<td>0.1056</td>
<td>0.75</td>
<td>0.7734</td>
<td>2.75</td>
<td>0.9970</td>
</tr>
</tbody>
</table>
Today

- David will work examples on one side of the room.
- Gus and Sam and Lucas will hold office hours on the other side of the room.
- You should feel free to go back and forth between the sides.
Topics

1. Sets.
2. Counting.
3. Sample space, outcome, event, probability function.
4. Probability: conditional probability, independence, Bayes’ theorem.
5. Discrete random variables: events, pmf, cdf.
6. Bernoulli($p$), binomial($n$, $p$), geometric($p$), uniform($n$)
7. $E(X)$, $Var(X)$, $\sigma$
9. uniform($a,b$), exponential($\lambda$), normal($\mu,\sigma^2$)
10. Transforming random variables.
11. Quantiles.
12. Central limit theorem, law of large numbers, histograms.
Sets and counting

- Sets:
  - $\emptyset$, union, intersection, complement Venn diagrams, products

- Counting:
  - inclusion-exclusion, rule of product, permutations $nP_k$, combinations $nC_k = \binom{n}{k}$
Sample space, outcome, event, probability function.

Rule: \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \).

Special case: \( P(A^c) = 1 - P(A) \)

\( (A \text{ and } B \text{ disjoint } \Rightarrow P(A \cup B) = P(A) + P(B).) \)

Conditional probability, multiplication rule, trees, law of total probability, independence

Bayes’ theorem, base rate fallacy
Random variables, expectation and variance

- Discrete random variables: events, pmf, cdf
- Bernoulli($p$), binomial($n$, $p$), geometric($p$), uniform($n$)
- $E(X)$, meaning, algebraic properties, $E(h(X))$
- Var($X$), meaning, algebraic properties
- Continuous random variables: pdf, cdf
  - uniform($a$, $b$), exponential($\lambda$), normal($\mu$, $\sigma$)
- Transforming random variables
- Quantiles
Central limit theorem

- Law of large numbers averages and histograms
- Central limit theorem
Joint distributions

- Joint pmf, pdf, cdf.
- Marginal pmf, pdf, cdf
- Covariance and correlation.
Hospitals (binomial, CLT, etc)

- A certain town is served by two hospitals.
- Larger hospital: about 45 babies born each day.
- Smaller hospital about 15 babies born each day.
- For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys.

(a) Which hospital do you think recorded more such days?
   (i) The larger hospital.  (ii) The smaller hospital.
   (iii) About the same (that is, within 5% of each other).

(b) Assume exactly 45 and 15 babies are born at the hospitals each day. Let $L_i$ (resp., $S_i$) be the Bernoulli random variable which takes the value 1 if more than 60% of the babies born in the larger (resp., smaller) hospital on the $i^{th}$ day were boys. Determine the distribution of $L_i$ and of $S_i$.

Continued on next slide
Let $L$ (resp., $S$) be the number of days on which more than 60% of the babies born in the larger (resp., smaller) hospital were boys. What type of distribution do $L$ and $S$ have? Compute the expected value and variance in each case.

Via the CLT, approximate the 0.84 quantile of $L$ (resp., $S$). Would you like to revise your answer to part (a)?

What is the correlation of $L$ and $S$? What is the joint pmf of $L$ and $S$? Visualize the region corresponding to the event $L > S$. Express $P(L > S)$ as a double sum.
Counties with high kidney cancer death rates
Counties with low kidney cancer death rates

Discussion and reference on next slide
Discussion

The maps were taken from

*Teaching Statistics: A Bag of Tricks* by Andrew Gelman, Deborah Nolan

- The first map shows with the lowest 10% age-standardized death rates for cancer of kidney/ureter for U.S. white males 1980-1989.
- The second map shows the highest 10%
- We see that both maps are dominated by low population counties. This reflects the higher variability around the national mean rate among low population counties and conversely the low variability about the mean rate among high population counties. As in the hospital example this follows from the central limit theorem.
Problem correlation

1. Flip a coin 3 times. Use a joint pmf table to compute the covariance and correlation between the number of heads on the first 2 and the number of heads on the last 2 flips.

2. Flip a coin 5 times. Use properties of covariance to compute the covariance and correlation between the number of heads on the first 3 and last 3 flips.