11: Joint (Multivariate) Distributions

Lisa Yan
April 29, 2020
Quick slide reference

3  Normal Approximation  11a_normal_approx
13 Discrete Joint RVs       11b_discrete_joint
26 Multinomial RV           11c_multinomial
34 Exercises                  LIVE
43 Federalist Papers Demo    LIVE
Normal Approximation
Normal RVs

\[ X \sim \mathcal{N}(\mu, \sigma^2) \]

- Used to model many real-life situations because it maximizes entropy (i.e., randomness) for a given mean and variance
- Also useful for approximating the Binomial random variable!
Website testing

- 100 people are given a new website design.
- \( X = \# \) people whose time on site increases
- The design actually has no effect, so \( P(\text{time on site increases}) = 0.5 \) independently.
- CEO will endorse the new design if \( X \geq 65 \).

What is \( P(\text{CEO endorses change}) \)? *Give a numerical approximation.*

**Approach 1: Binomial**

Define

\[ X \sim \text{Bin}(n = 100, p = 0.5) \]

Want: \( P(X \geq 65) \)

Solve

\[
P(X \geq 65) = \sum_{i=65}^{100} \binom{100}{i} 0.5^i (1 - 0.5)^{100-i}
\]
Don’t worry, Normal approximates Binomial

Galton Board

(We’ll explain why in 2 weeks’ time)
Website testing

- 100 people are given a new website design.
- $X = \#$ people whose time on site increases.
- The design actually has no effect, so $P($time on site increases$) = 0.5$ independently.
- CEO will endorse the new design if $X \geq 65$.

What is $P($CEO endorses change$)$? *Give a numerical approximation.*

Approach 1: Binomial

Define

\[ X \sim \text{Bin}(n = 100, p = 0.5) \]

Want: $P(X \geq 65)$

Solve

\[ P(X \geq 65) \approx 0.0018 \]

Approach 2: approximate with Normal

Define

\[ Y \sim \mathcal{N}(\mu, \sigma^2) \]

\[ \mu = np = 50 \]
\[ \sigma^2 = np(1 - p) = 25 \]
\[ \sigma = \sqrt{25} = 5 \]

Solve

\[ P(X \geq 65) \approx P(Y \geq 65) = 1 - F_Y(65) = 1 - \Phi\left(\frac{65 - 50}{5}\right) = 1 - \Phi(3) \approx 0.0013 \]

⚠️⚠️🤔 (this approach is actually missing something)
Website testing (with continuity correction)

In our website testing, $Y \sim \mathcal{N}(50, 25)$ approximates $X \sim \text{Bin}(100, 0.5)$.

\[ P(X \geq 65) \approx P(Y \geq 64.5) \approx 0.0018 \]

You must perform a continuity correction when approximating a Binomial RV with a Normal RV.
Continuity correction

If $Y \sim \mathcal{N}(np, np(1 - p))$ approximates $X \sim \text{Bin}(n, p)$, how do we approximate the following probabilities?

<table>
<thead>
<tr>
<th>Discrete (e.g., Binomial) probability question</th>
<th>Continuous (Normal) probability question</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = 6)$</td>
<td></td>
</tr>
<tr>
<td>$P(X \geq 6)$</td>
<td></td>
</tr>
<tr>
<td>$P(X &gt; 6)$</td>
<td></td>
</tr>
<tr>
<td>$P(X &lt; 6)$</td>
<td></td>
</tr>
<tr>
<td>$P(X \leq 6)$</td>
<td></td>
</tr>
</tbody>
</table>

Discrete (e.g., Binomial) probability question

Continuous (Normal) probability question

- $P(X = 6)$
- $P(X \geq 6)$
- $P(X > 6)$
- $P(X < 6)$
- $P(X \leq 6)$
Continuity correction

If $Y \sim \mathcal{N}(np, np(1 - p))$ approximates $X \sim \text{Bin}(n, p)$, how do we approximate the following probabilities?

<table>
<thead>
<tr>
<th>Discrete (e.g., Binomial) probability question</th>
<th>Continuous (Normal) probability question</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = 6)$</td>
<td>$P(5.5 \leq Y \leq 6.5)$</td>
</tr>
<tr>
<td>$P(X \geq 6)$</td>
<td>$P(Y \geq 5.5)$</td>
</tr>
<tr>
<td>$P(X &gt; 6)$</td>
<td>$P(Y \geq 6.5)$</td>
</tr>
<tr>
<td>$P(X &lt; 6)$</td>
<td>$P(Y \leq 5.5)$</td>
</tr>
<tr>
<td>$P(X \leq 6)$</td>
<td>$P(Y \leq 6.5)$</td>
</tr>
</tbody>
</table>

... 5 6 7 ...
Who gets to approximate?

\[ X \sim \text{Bin}(n, p) \]
\[ E[X] = np \]
\[ \text{Var}(X) = np(1 - p) \]

\[ Y \sim \text{Poi}(\lambda) \]
\[ \lambda = np \]

\[ Y \sim \mathcal{N}(\mu, \sigma^2) \]
\[ \mu = np \]
\[ \sigma^2 = np(1 - p) \]
Who gets to approximate?

1. If there is a choice, use Normal to approximate.
2. When using Normal to approximate a discrete RV, use a continuity correction.

Poisson approximation
- $n$ large ($> 20$), $p$ small ($< 0.05$)
- slight dependence okay

Normal approximation
- $n$ large ($> 20$), $p$ mid-ranged ($np(1-p) > 10$)
- independence
Discrete Joint RVs
From last time

What is the probability that the Warriors win?
How do you model zero-sum games?

\[ P(A_W > A_B) \]

This is a probability of an event involving two random variables!
Joint probability mass functions

Roll two 6-sided dice, yielding values $X$ and $Y$.

$X$

random variable

$P(X = 1)$

probability of an event

$P(X = k)$

probability mass function
### Joint probability mass functions

Roll two 6-sided dice, yielding values $X$ and $Y$.

<table>
<thead>
<tr>
<th>$X$</th>
<th>$P(X = 1)$</th>
<th>$P(X = k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>random variable</td>
<td>probability of an event</td>
<td>probability mass function</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$X, Y$</th>
<th>$P(X = 1 \cap Y = 6)$</th>
<th>$P(X = a, Y = b)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>random variables</td>
<td>$P(X = 1, Y = 6)$</td>
<td>joint probability mass function</td>
</tr>
</tbody>
</table>

new notation: the comma

probability of the intersection of two events
Discrete joint distributions

For two discrete joint random variables $X$ and $Y$, the joint probability mass function is defined as:

$$p_{X,Y}(a, b) = P(X = a, Y = b)$$

The marginal distributions of the joint PMF are defined as:

$$p_X(a) = P(X = a) = \sum_y p_{X,Y}(a, y)$$

$$p_Y(b) = P(Y = b) = \sum_x p_{X,Y}(x, b)$$

Use marginal distributions to get a 1-D RV from a joint PMF.
Two dice

Roll two 6-sided dice, yielding values $X$ and $Y$.

1. What is the joint PMF of $X$ and $Y$?

The joint PMF of $X$ and $Y$ is given by:

$$p_{X,Y}(a, b) = \frac{1}{36} \quad (a, b) \in \{(1,1), \ldots, (6,6)\}$$

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

- Probability table
  - All possible outcomes for several discrete RVs
  - Not parametric (e.g., parameter $p$ in $\text{Ber}(p)$)
Two dice

Roll two 6-sided dice, yielding values $X$ and $Y$.

1. What is the joint PMF of $X$ and $Y$?

$$p_{X,Y}(a, b) = 1/36 \quad (a, b) \in \{(1,1), \ldots, (6,6)\}$$

2. What is the marginal PMF of $X$?

$$p_X(a) = P(X = a) = \sum_y p_{X,Y}(a, y) = \sum_{y=1}^{6} \frac{1}{36} = \frac{1}{6} \quad a \in \{1, \ldots, 6\}$$
A computer (or three) in every house.

Consider households in Silicon Valley.
- A household has $X$ Macs and $Y$ PCs.
- Each house has a maximum of 3 computers (Macs + PCs) in the house.

1. What is $P(X = 1, Y = 0)$, the missing entry in the probability table?

<table>
<thead>
<tr>
<th>$Y$ (# PCs)</th>
<th>$X$ (# Macs)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.16</td>
<td>?</td>
<td>.07</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.12</td>
<td>.14</td>
<td>.12</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.07</td>
<td>.12</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.04</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Consider households in Silicon Valley.

- A household has $X$ Macs and $Y$ PCs.
- Each house has a maximum of 3 computers (Macs + PCs) in the house.

1. What is $P(X = 1, Y = 0)$, the missing entry in the probability table?

<table>
<thead>
<tr>
<th>$X$ (# Macs)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.16</td>
<td>.12</td>
<td>.07</td>
<td>.04</td>
</tr>
<tr>
<td>1</td>
<td>.12</td>
<td>.14</td>
<td>.12</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>.07</td>
<td>.12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>.04</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

A joint PMF must sum to 1:

$$\sum_{x} \sum_{y} p_{X,Y}(x, y) = 1$$
A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has $X$ Macs and $Y$ PCs.
- Each house has a maximum of 3 computers (Macs + PCs) in the house.

2. How do you compute the marginal PMF of $X$?

<table>
<thead>
<tr>
<th>$X$ (# Macs)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Sum rows here</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.16</td>
<td>.12</td>
<td>.07</td>
<td>.04</td>
<td>.39</td>
</tr>
<tr>
<td>1</td>
<td>.12</td>
<td>.14</td>
<td>.12</td>
<td>.00</td>
<td>.38</td>
</tr>
<tr>
<td>2</td>
<td>.07</td>
<td>.12</td>
<td>.00</td>
<td>.00</td>
<td>.19</td>
</tr>
<tr>
<td>3</td>
<td>.04</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$Y$ (# PCs)</th>
<th>Sum cols here</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 A</td>
<td>.39</td>
</tr>
<tr>
<td>1 B</td>
<td>.38</td>
</tr>
<tr>
<td>2 C</td>
<td>.19</td>
</tr>
<tr>
<td>3</td>
<td>.04</td>
</tr>
</tbody>
</table>
A computer (or three) in every house.

Consider households in Silicon Valley.

- A household has $X$ Macs and $Y$ PCs.
- Each house has a maximum of 3 computers (Macs + PCs) in the house.

2. How do you compute the marginal PMF of $X$?

\[
p_{X,Y}(x, 0) = P(X = x, Y = 0)
\]

\[
p_X(x) = \sum_y p_{X,Y}(x, y)
\]

\[
p_Y(y) = \sum_x p_{X,Y}(x, y)
\]

To find a marginal distribution over one variable, sum over all other variables in the joint PMF.
A computer (or three) in every house.

Consider households in Silicon Valley.

• A household has $X$ Macs and $Y$ PCs.
• Each house has a maximum of 3 computers (Macs + PCs) in the house.

3. Let $C = X + Y$. What is $P(C = 3)$?

<table>
<thead>
<tr>
<th>$Y$ (# PCs)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.16</td>
<td>.12</td>
<td>.07</td>
<td>.04</td>
</tr>
<tr>
<td>1</td>
<td>.12</td>
<td>.14</td>
<td>.12</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>.07</td>
<td>.12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>.04</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Consider households in Silicon Valley.

- A household has $X$ Macs and $Y$ PCs.
- Each house has a maximum of 3 computers (Macs + PCs) in the house.

3. Let $C = X + Y$. What is $P(C = 3)$?

<table>
<thead>
<tr>
<th>$X$ (# Macs)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.16</td>
<td>.12</td>
<td>.07</td>
<td>.04</td>
</tr>
<tr>
<td>1</td>
<td>.12</td>
<td>.14</td>
<td>.12</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>.07</td>
<td>.12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>.04</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$P(C = 3) = P(X + Y = 3)$

$$= \sum_x \sum_y P(X + Y = 3 | X = x, Y = y)P(X = x, Y = y)$$

$$= P(X = 0, Y = 3) + P(X = 1, Y = 2) + P(X = 2, Y = 1) + P(X = 3, Y = 0)$$

We’ll come back to sums of RVs next lecture!
Multinomial RV
Recall the good times

Permutations $n!$

How many ways are there to order $n$ objects?
Counting unordered objects

**Binomial coefficient**

How many ways are there to group \( n \) objects into two groups of size \( k \) and \( n - k \), respectively?

\[
\binom{n}{k} = \frac{n!}{k!(n-k)!}
\]

Called the binomial coefficient because of something from Algebra

**Multinomial coefficient**

How many ways are there to group \( n \) objects into \( r \) groups of sizes \( n_1, n_2, ..., n_r \), respectively?

\[
\binom{n}{n_1, n_2, ..., n_r} = \frac{n!}{n_1!n_2!...n_r!}
\]

Multinomials generalize Binomials for counting.
Probability

**Binomial RV**

What is the probability of getting \( k \) successes and \( n - k \) failures in \( n \) trials?

\[
P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}
\]

- Binomial # of ways of ordering the successes
- Probability of each ordering of \( k \) successes is equal \& mutually exclusive

**Multinomial RV**

What is the probability of getting \( c_1 \) of outcome 1, \( c_2 \) of outcome 2, \ldots, and \( c_m \) of outcome \( m \) in \( n \) trials?

- Multinomial RVs also generalize Binomial RVs for probability!
Multinomial Random Variable

Consider an experiment of $n$ independent trials:

- Each trial results in one of $m$ outcomes. $P(\text{outcome } i) = p_i$, $\sum_{i=1}^{m} p_i = 1$
- Let $X_i =$ # trials with outcome $i$

Joint PMF

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

where $\sum_{i=1}^{m} c_i = n$ and $\sum_{i=1}^{m} p_i = 1$

**Multinomial** # of ways of ordering the outcomes

**Probability** of each ordering is equal + mutually exclusive
Hello dice rolls, my old friends

A 6-sided die is rolled 7 times.

What is the probability of getting:

• 1 one
• 0 threes
• 0 fives
• 1 two
• 2 fours
• 3 sixes
Hello dice rolls, my old friends

A 6-sided die is rolled 7 times. What is the probability of getting:

- 1 one
- 1 two
- 0 threes
- 2 fours
- 0 fives
- 3 sixes

\[ P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 2, X_5 = 0, X_6 = 3) \]

\[
\begin{align*}
&= \binom{7}{1,1,0,2,0,3} \left( \frac{1}{6} \right)^1 \left( \frac{1}{6} \right)^1 \left( \frac{1}{6} \right)^0 \left( \frac{1}{6} \right)^2 \left( \frac{1}{6} \right)^0 \left( \frac{1}{6} \right)^3 \\
&= 420 \left( \frac{1}{6} \right)^7
\end{align*}
\]
Hello dice rolls, my old friends

A 6-sided die is rolled 7 times.

What is the probability of getting:

- 1 one
- 0 threes
- 0 fives
- 1 two
- 2 fours
- 3 sixes

\[ P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 2, X_5 = 0, X_6 = 3) \]

\[ = \binom{7}{1,1,0,2,0,3} \left( \frac{1}{6} \right)^1 \left( \frac{1}{6} \right)^1 \left( \frac{1}{6} \right)^0 \left( \frac{1}{6} \right)^2 \left( \frac{1}{6} \right)^0 \left( \frac{1}{6} \right)^3 = 420 \left( \frac{1}{6} \right)^7 \]
11: Joint (Multivariate) Distributions

Slides by Lisa Yan
April 29, 2020
Normal RVs

\[ X \sim \mathcal{N}(\mu, \sigma^2) \]

- Used to model many real-life situations because it maximizes entropy (i.e., randomness) for a given mean and variance
- Also useful for approximating the Binomial random variable!
Who gets to approximate?

\( X \sim \text{Bin}(n, p) \)

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

\( Y \sim \text{Poi}(\lambda) \)

- \( \lambda = np \)

\( n \) large (> 20), \( p \) small (< 0.05)

- slight dependence okay

\( Y \sim \mathcal{N}(\mu, \sigma^2) \)

- \( \mu = np \)
- \( \sigma^2 = np(1 - p) \)

\( n \) large (> 20), \( p \) mid-ranged (\( np(1 - p) > 10 \))

- independence

- need continuity correction

Review

- Computing probabilities on Binomial RVs is often computationally expensive.
- Two reasonable approximations, but when to use which?
Think

Check out the question on the next slide. Post any clarifications here!

https://us.edstem.org/courses/667/discussion/90049
Stanford Admissions (a while back)

Stanford accepts 2480 students.

• Each accepted student has 68% chance of attending (independent trials)
• Let \( X = \# \) of students who will attend

What is \( P(X > 1745) \)? Give a numerical approximation.

Strategy:
A. Just Binomial
B. Poisson
C. Normal
D. None/other
Stanford Admissions (a while back)

Stanford accepts 2480 students.
- Each accepted student has 68% chance of attending (independent trials)
- Let $X = \#$ of students who will attend

What is $P(X > 1745)$? *Give a numerical approximation.*

Strategy:
- A. Just Binomial
- B. Poisson
- C. Normal
- D. None/other

Define an approximation

Solve

Let $Y \sim \mathcal{N}(E[X], \text{Var}(X))$

$E[X] = np = 1686$
$\text{Var}(X) = np(1 - p) \approx 540 \rightarrow \sigma = 23.3$

$P(X > 1745) \approx P(Y \geq 1745.5)$

Continuous correction

$P(Y \geq 1745.5) = 1 - F(1745.5) = 1 - \Phi\left(\frac{1745.5 - 1686}{23.3}\right)$

$= 1 - \Phi(2.54) \approx 0.0055$
Changes in Stanford Admissions

Stanford accepts 2480 students.
- Each accepted student has 68% chance of attending (independent trials)
- Let $X = \#$ of students who will attend

What is $P(X > 1745)$? Give a numerical approximation.

Overview for the Class of 2022
- Total Applicants: 47,451
- Total Admits: 2,071
- Total Enrolled: 1,706

Admit rate: 4.3%
Yield rate: 81.9%

People love coming to Stanford!
Multinomial Random Variable

Consider an experiment of \( n \) independent trials:
- Each trial results in one of \( m \) outcomes. \( P(\text{outcome } i) = p_i, \sum_{i=1}^{m} p_i = 1 \)
- Let \( X_i = \# \) trials with outcome \( i \)

**Joint PMF**

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

where \( \sum_{i=1}^{m} c_i = n \) and \( \sum_{i=1}^{m} p_i = 1 \)

**Example:**
- Rolling 2 twos, 3 threes, and 5 fives on 10 rolls of a fair-sided die
- Generating a random 5-word phrase with 1 “the”, 2 “bacon”, 1 “put”, 1 “on”
Hello dice rolls, my old friends

A 6-sided die is rolled 7 times. What is the probability of getting:

• 1 one  
• 1 two  
• 0 threes  
• 2 fours  
• 0 fives  
• 3 sixes

\[ P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 2, X_5 = 0, X_6 = 3) \]

\[ = \binom{7}{1,1,0,2,0,3} (\frac{1}{6})^1 (\frac{1}{6})^1 (\frac{1}{6})^0 (\frac{1}{6})^2 (\frac{1}{6})^0 (\frac{1}{6})^3 = 420 \left( \frac{1}{6} \right)^7 \]

Review

# of times a six appears

# of times

choose where the sixes appear

probability of rolling a six

this many times
Parameters of a Multinomial RV?

$X \sim \text{Bin}(n, p)$ has parameters $n, p$...

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$p$: probability of success outcome on a single trial

A Multinomial RV has parameters $n, p_1, p_2, \ldots, p_m$ (Note $p_m = 1 - \sum_{i=1}^{m-1} p_i$)

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

$p_i$: probability of outcome $i$ on a single trial

Where do we get $p_i$ from?
Interlude for fun/announcements
Announcements

More OH!
Interesting probability news

Estimating Coronavirus Prevalence by Cross-Checking Countries

We’ll make the modeling assumption that $N_{ij}$ is a Poisson distribution with rate parameter $A_{ij} \times \lambda_i \times \alpha_j$. What this means is that the expected number of cases should be equal to the total amount of travel, times some source-dependent multiplier $\alpha_j$, times some country-dependent multiplier $\lambda_i$ (the infection prevalence in country i).”

https://medium.com/@jsteinhardt/estimating-coronavirus-prevalence-by-cross-checking-countries-c7e4211f0e18
Amazon scraps secret AI recruiting tool that showed bias against women

“In effect, Amazon’s system *taught itself that male candidates were preferable*. It penalized resumes that included the word ‘women’s,’ as in ‘women’s chess club captain.’”

**Basic Bayes Algorithm:** Pick highest \( P(H|\text{resume}) \)

Let \( \text{resumeF} \) be a resume associated with a female applicant.

\[
P(H|\text{resumeF}) = \frac{P(\text{resumeF}|H)P(H)}{P(\text{resumeF})}
\]

Because of biased historical data, \( P(\text{resumeF}|H) \) is small \( \xrightarrow{\frown} \) Therefore \( P(H|\text{resumeF}) \) may be higher than \( P(H|\text{resumeM}) \) *simply because of biased historical data, rather than comparative candidate skillsets.*

---

What if we ignore gender traits?

Amazon edited the programs to make them neutral to these particular terms. But that was no guarantee that the machines would not devise other ways of sorting candidates that could prove discriminatory.

[After re-training...] the technology favored candidates who described themselves using verbs more commonly found on male engineers’ resumes, such as “executed” and “captured,” one person said.

This is an open question in a field called Algorithmic Fairness.

The Federalist Papers
Probabilistic text analysis

Ignoring the order of words...

What is the probability of any given word that you write in English?

• $P(\text{word = “the”}) > P(\text{word = “pokemon”})$
• $P(\text{word = “Stanford”}) > P(\text{word = “Cal”})$

Probabilities of counts of words = Multinomial distribution

A document is a large multinomial.

(according to the Global Language Monitor, there are 988,968 words in the English language used on the internet.)
Probabilistic text analysis

Probabilities of counts of words = Multinomial distribution

Example document:

“When my late husband was alive he deposited some amount of Money with china Bank in which the amount will be declared to you once you respond to this message indicating your interest in helping to receive the fund and use it for Gods work as my wish.”

\[
P\left(\begin{array}{c}
\text{bank} = 1 \\
\text{fund} = 1 \\
\text{money} = 1 \\
\text{wish} = 1 \\
\text{to} = 3 \\
\end{array} \mid \text{spam} \right) = \frac{n!}{1! \ 1! \ 1! \ \ldots \ 3!} p_1^{\text{bank}} p_1^{\text{fund}} \ldots p_3^{\text{to}}
\]

Note: \( P(\text{bank} \mid \text{spam writer}) \gg P(\text{bank} \mid \text{writer=} \text{you}) \)
Probabilistic text analysis

Probabilities of counts of words = Multinomial distribution

What about probability of those same words in someone else’s writing?

- \[ P\left(\text{word} = \text{“probability”} \mid \text{writer} = \text{you}\right) > P\left(\text{word} = \text{“probability”} \mid \text{writer} = \text{non-CS109 student}\right) \]

To determine authorship:

1. Estimate \( P(\text{word} | \text{writer}) \) from known writings
2. Use Bayes’ Theorem to determine \( P(\text{writer} | \text{document}) \) for a new writing!

Who wrote the Federalist Papers?
See recordings 10e_all....
Up next: Independent RVs and Sums!