Multinomial + Cont. Joint

Chris Piech
CS109, Stanford University
Discrete Joint
Distributions: General Case

Multinomial: A parametric
Discrete Joint

Cont. Joint
Distributions: General Case

Today
Learning Goals

1. Know how to use a multinomial
2. Be able to calculate large bayes problems using a computer
3. Use a Joint CDF
Motivating Examples

THE FEDERALIST
A COLLECTION OF ESSAYS
WRITTEN IN FAVOUR OF THE NEW CONSTITUTION
AS AGREED UPON BY THE FEDERAL CONVENT
SEPTEMBER 17, 1787
Defense for Ratification

Original
StDev = 3
StDev = 10
Recall logs
Log Review

\[ e^y = x \quad \log(x) = y \]
Log Identities

\[ \log(a \cdot b) = \log(a) + \log(b) \]

\[ \log\left(\frac{a}{b}\right) = \log(a) - \log(b) \]

\[ \log(a^n) = n \cdot \log(a) \]
Products become Sums!

$$\log(a \cdot b) = \log(a) + \log(b)$$

$$\log(\prod_{i} a_i) = \sum_{i} \log(a_i)$$

* Spoiler alert: This is important because the product of many small numbers gets hard for computers to represent.
Where we left off
### Joint Probability Table

<table>
<thead>
<tr>
<th>Roommates</th>
<th>2RoomDbl</th>
<th>Shared Partner</th>
<th>Single</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frosh</td>
<td>0.30</td>
<td>0.07</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Soph</td>
<td>0.12</td>
<td>0.18</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td>Junior</td>
<td>0.04</td>
<td>0.01</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td>Senior</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>5+</td>
<td>0.02</td>
<td>0.00</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>0.49</td>
<td>0.27</td>
<td>0.07</td>
<td>0.18</td>
</tr>
</tbody>
</table>

**Marginal Room type**

- Roommates: 0.49
- 2RoomDbl: 0.27
- Shared Partner: 0.07
- Single: 0.18

**Marginal Year**

- Frosh: 0.40
- Soph: 0.30
- Junior: 0.20
- Senior: 0.10
- 5+: 0.10
Change in Marginal Year

Fall quarter ‘18

Spr quarter ‘19
The Multinomial

• Multinomial distribution
  ▪ $n$ independent trials of experiment performed
  ▪ Each trial results in one of $m$ outcomes, with respective probabilities: $p_1, p_2, \ldots, p_m$ where
  ▪ $X_i =$ number of trials with outcome $i$

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

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

and
$$\binom{n}{c_1, c_2, \ldots, c_m} = \frac{n!}{c_1!c_2!\cdots c_m!}$$

Joint distribution
Multinomial # ways of ordering the successes
Probabilities of each ordering are equal and mutually exclusive
The Multinomial

- Multinomial distribution
  - \( n \) independent trials of experiment performed
  - Each trial results in one of \( m \) outcomes, with respective probabilities: \( p_1, p_2, \ldots, p_m \) where
  - \( X_i = \) number of trials with outcome \( i \)

\[
P(X_1 = c_1, X_2 = c_2, \ldots, X_m = c_m) = \binom{n}{c_1, c_2, \ldots, c_m} \prod_{i} p_i^{c_i}
\]

where

\[
\sum_{i=1}^{m} c_i = n
\]

and

\[
\binom{n}{c_1, c_2, \ldots, c_m} = \frac{n!}{c_1!c_2!\cdots c_m!}
\]

\[
\sum_{i=1}^{m} p_i = 1
\]
Hello Die Rolls, My Old Friends

• 6-sided die is rolled 7 times
  ▪ Roll results: 1 one, 1 two, 0 three, 2 four, 0 five, 3 six

\[
P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 2, X_5 = 0, X_6 = 3) = \frac{7!}{1!1!0!2!0!3!} \left( \frac{1}{6} \right)^1 \left( \frac{1}{6} \right)^1 \left( \frac{1}{6} \right)^0 \left( \frac{1}{6} \right)^2 \left( \frac{1}{6} \right)^0 \left( \frac{1}{6} \right)^3 = 420 \left( \frac{1}{6} \right)^7
\]

• This is generalization of Binomial distribution
  ▪ Binomial: each trial had 2 possible outcomes
  ▪ Multinomial: each trial has \( m \) possible outcomes
According to the Global Language Monitor there are 988,968 words in the English language used on the internet.
Example document:
“Pay for Viagra with a credit-card. Viagra is great. So are credit-cards. Risk free Viagra. Click for free.”

\[ n = 18 \]

\[
P(V, F, R, C, F | \text{spam}) = \frac{n!}{2!2! \ldots 2!} p_{\text{Viagra}}^2 p_{\text{Free}}^2 \ldots p_{\text{For}}^2
\]

Probability of seeing this document | spam

It’s a Multinomial!

The probability of a word in spam email being viagra
Who wrote the Federalist papers?
• Authorship of “Federalist Papers”
  
  ▪ 85 essays advocating ratification of US constitution
  
  ▪ Written under pseudonym “Publius”
    o Really, Alexander Hamilton, James Madison and John Jay
  
  ▪ Who wrote which essays?
    o Analyzed probability of words in each essay versus word distributions from known writings of three authors
Let’s write a program!
Example document:
“Pay for Viagra with a credit-card. Viagra is great. So are credit-cards. Risk free Viagra. Click for free.”

\( n = 18 \)

\[ P \left( \begin{array}{c}
\text{Viagra} = 2 \\
\text{Free} = 2 \\
\text{Risk} = 1 \\
\text{Credit-card: 2} \\
\text{...} \\
\text{For} = 2 \\
\end{array} \mid \text{spam} \right) = \frac{n!}{2!2! \ldots 2!} p_{\text{viagra}}^2 p_{\text{free}}^2 \cdots p_{\text{for}}^2 \]

It’s a Multinomial!

The probability of a word in spam email being viagra

Probability of seeing this document | spam
woot
Continuous Random Variables

Joint Distributions
Continuous Joint Distribution
You are running to the bus stop. 
You don’t know exactly when 
the bus arrives. You arrive at 
2:20pm.

What is $P(\text{wait} < 5 \text{ min})$?
Joint Dart Distribution

Dart Results  \[ P(\text{hit within } R \text{ pixels of center})? \]

What is the probability that a dart hits at \((456.234231234122355, 532.12344123456)\)?
Joint Dart Distribution

Dart Results

$P(\text{hit within } R \text{ pixels of center})$?
Joint Dart Distribution

Dart Results

\[ P(\text{hit within R pixels of center})? \]

\[ \text{Dart x location} \]

\[ \text{Dart y location} \]

0.12

0.005
Joint Dart Distribution

P(hit within R pixels of center)?
In the limit, as you break down continuous values into intestinally small buckets, you end up with multidimensional probability density.
A joint probability density function gives the relative likelihood of more than one continuous random variable each taking on a specific value.

\[ P(a_1 < X < a_2, b_1 < Y < b_2) = \int_{x=a_1}^{a_2} \int_{y=b_1}^{b_2} f(X = x, Y = y) \, dy \, dx \]
Joint Probability Density Function

\[ P(a_1 < X < a_2, b_1 < Y < b_2) = \int_{x=a_1}^{a_2} \int_{y=b_1}^{b_2} f(X = x, Y = y) \, \partial y \partial x \]
Joint Probability Density Function

\[ P(a_1 < X < a_2, b_1 < Y < b_2) = \int_{x=a_1}^{a_2} \int_{y=b_1}^{b_2} f(X = x, Y = y) \, dy \, dx \]
Let $X$ and $Y$ be two continuous random variables
- where $0 \leq X \leq 1$ and $0 \leq Y \leq 2$

We want to integrate $g(x,y) = xy$ w.r.t. $X$ and $Y$:
- First, do “innermost” integral (treat $y$ as a constant):

$$\int_0^2 \int_0^1 xy \, dx \, dy = \int_0^2 \left( \int_0^1 xy \, dx \right) \, dy = \int_0^2 y \left[ \frac{x^2}{2} \right]_0^1 \, dy = \int_0^2 y \frac{1}{2} \, dy$$

- Then, evaluate remaining (single) integral:

$$\int_0^2 y \frac{1}{2} \, dy = \left[ \frac{y^2}{4} \right]_0^2 = 1 - 0 = 1$$
Marginal probabilities give the distribution of a subset of the variables (often, just one) of a joint distribution.

Sum/integrate over the variables you don’t care about.

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

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

\[
f_Y(b) = \int_{-\infty}^{\infty} f_{X,Y}(x, b) \, dx
\]
Marginal probabilities give the distribution of a subset of the variables (often, just one) of a joint distribution.

Sum/integrate over the variables you don’t care about.

\[
p_X(a) = \sum_y P(X = a, Y = y) \\

f_X(a) = \int_{-\infty}^{\infty} f_{X,Y}(a, y) \, dy \\
f_Y(b) = \int_{-\infty}^{\infty} f_{X,Y}(x, b) \, dx
\]
Marginal probabilities give the distribution of a subset of the variables (often, just one) of a joint distribution.

Sum/integrate over the variables you don’t care about.

\[
p_X(a) = \sum_y P(X = a, Y = y)
\]

\[
f_X(a) = \int_{-\infty}^{\infty} f(X = a, Y = y)
\]

\[
f_Y(b) = \int_{-\infty}^{\infty} f_{X,Y}(x, b) \, dx
\]
Marginal probabilities give the distribution of a subset of the variables (often, just one) of a joint distribution.

Sum/integrate over the variables you don’t care about.

\[ P(X = a) = \sum_y P(X = a, Y = y) \]

\[ f(X = a) = \int_{-\infty}^{\infty} f(X = a, Y = y) \, dy \]

\[ f_Y(b) = \int_{-\infty}^{\infty} f_{X,Y}(x, b) \, dx \]
Darts!

Dart PDF

\[ x \sim N \left( \frac{900}{2}, \frac{900}{2} \right) \]

\[ y \sim N \left( \frac{900}{3}, \frac{900}{5} \right) \]
Joint Cumulative Density Function

Cumulative Density Function (CDF):

\[ F_{X,Y}(a, b) = P(X < a, Y < b) \]

\[
F_{X,Y}(a, b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x, y) \, dy \, dx
\]

\[
f_{X,Y}(a, b) = \frac{\partial^2}{\partial a \partial b} F_{X,Y}(a, b)
\]
Joint CDF

\[ F_{X,Y}(a, b) = P(X < a, Y < b) \]

to 0 as
\( x \to -\infty, \quad y \to -\infty \)

to 1 as
\( x \to +\infty, \quad y \to +\infty \)
Jointly Continuous

\[ P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x, y) \, dy \, dx \]
Probabilities from Joint CDF

\[ P(a_1 < X \leq a_2, b_1 < Y \leq b_2) \]
Probabilities from Joint CDF

\[ P\left(a_1 < X \leq a_2, b_1 < Y \leq b_2\right) = F_{X,Y}(a_2, b_2) \]
Probabilities from Joint CDF

\[ P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2) \]
Proabilities from Joint CDF

\[ P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2,b_2) - F_{X,Y}(a_1,b_2) \]
Probabilities from Joint CDF

\[ P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) \]
Probabilities from Joint CDF

\[ P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(a_2, b_1) \]
Probabilities from Joint CDF

\[ P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(a_2, b_1) \]
Probabilities from Joint CDF

\[ P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(a_2, b_1) + F_{X,Y}(a_1, b_1) \]
Probabilities from Joint CDF

\[ P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(a_2, b_1) + F_{X,Y}(a_1, b_1) \]
Probability for Instagram!
In image processing, a Gaussian blur is the result of blurring an image by a Gaussian function. It is a widely used effect in graphics software, typically to reduce image noise.

Gaussian blurring with StDev = 3, is based on a joint probability distribution:

**Joint PDF**

\[
\begin{align*}
    f_{X,Y}(x, y) &= \frac{1}{2\pi \cdot 3^2} e^{-\frac{x^2 + y^2}{2 \cdot 3^2}} \\
    \end{align*}
\]

**Joint CDF**

\[
\begin{align*}
    F_{X,Y}(x, y) &= \Phi \left( \frac{x}{3} \right) \cdot \Phi \left( \frac{y}{3} \right) \\
    \end{align*}
\]

Used to generate this weight matrix
Gaussian Blur

Joint PDF

\[ f_{X,Y}(x, y) = \frac{1}{2\pi \cdot 3^2} e^{-\frac{x^2 + y^2}{2 \cdot 3^2}} \]

Joint CDF

\[ F_{X,Y}(x, y) = \Phi \left( \frac{x}{3} \right) \cdot \Phi \left( \frac{y}{3} \right) \]

Weight Matrix

Each pixel is given a weight equal to the probability that \( X \) and \( Y \) are both within the pixel bounds. The center pixel covers the area where

\[ -0.5 \leq x \leq 0.5 \text{ and } -0.5 \leq y \leq 0.5 \]

What is the weight of the center pixel?

\[ P(-0.5 < X < 0.5, -0.5 < Y < 0.5) \]
\[ = P(X < 0.5, Y < 0.5) - P(X < 0.5, Y < -0.5) \]
\[ - P(X < -0.5, Y < 0.5) + P(X < -0.5, Y < -0.5) \]
\[ = \Phi \left( \frac{0.5}{3} \right) \cdot \Phi \left( \frac{0.5}{3} \right) - 2 \Phi \left( \frac{0.5}{3} \right) \cdot \Phi \left( \frac{-0.5}{3} \right) \]
\[ + \Phi \left( \frac{-0.5}{3} \right) \cdot \Phi \left( \frac{-0.5}{3} \right) \]
\[ = 0.5662^2 - 2 \cdot 0.5662 \cdot 0.4338 + 0.4338^2 = 0.206 \]
How do you integrate under a circle?

\[ f(X = x, Y = y) \]
Have a great weekend!