Conditional Joint Distributions

Chris Piech

CS109, Stanford University
Tracking in 2D Space?
Joint Random Variables

- Use a joint table, density function or CDF to solve probability question
- Think about \textit{conditional} probabilities with joint variables (which might be continuous)
- Use and find \textit{independence} of random variables
- Use and find \textit{expectation} of random variables
- What happens when you \textit{add} random variables?
Joint Random Variables

- Use a joint table, density function or CDF to solve probability question

- Think about **conditional** probabilities with joint variables (which might be continuous)

- Use and find **independence** of random variables

- Use and find **expectation** of random variables

- What happens when you **add** random variables?
## Joint Probability Table

<table>
<thead>
<tr>
<th></th>
<th>Roommates</th>
<th>2RoomDbl</th>
<th>Shared Partner</th>
<th>Single</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frosh</td>
<td>0.30</td>
<td>0.07</td>
<td>0.00</td>
<td>0.00</td>
<td>0.37</td>
</tr>
<tr>
<td>Soph</td>
<td>0.12</td>
<td>0.18</td>
<td>0.00</td>
<td>0.03</td>
<td>0.32</td>
</tr>
<tr>
<td>Junior</td>
<td>0.04</td>
<td>0.01</td>
<td>0.00</td>
<td>0.10</td>
<td>0.15</td>
</tr>
<tr>
<td>Senior</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td>5+</td>
<td>0.02</td>
<td>0.00</td>
<td>0.05</td>
<td>0.04</td>
<td>0.11</td>
</tr>
<tr>
<td>Total</td>
<td>0.49</td>
<td>0.27</td>
<td>0.07</td>
<td>0.18</td>
<td>1.00</td>
</tr>
</tbody>
</table>

### Marginal Bar Charts

- **Marginal Room type**
  - Roommates: 0.5
  - 2RoomDbl: 0.4
  - Shared Partner: 0.3
  - Single: 0.2

- **Marginal Year**
  - Frosh: 0.4
  - Soph: 0.3
  - Junior: 0.2
  - Senior: 0.1
  - 5+: 0.1
Continuous Joint Random Variables

Dart Probability Density

Dart Results

\[ y \]

\[ x \]

0 900
A joint probability density function gives the relative likelihood of more than one continuous random variable each taking on a specific value.

\[ P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x, y) \, dy \, dx \]
End Review
Jointly Continuous

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

- Cumulative Density Function (CDF):

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

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

\[ F_{X,Y}(x, y) = P(X \leq x, Y \leq y) \]

to 1 as 
\[ x \to +\infty, \]
\[ y \to +\infty \]

to 0 as 
\[ x \to -\infty, \]
\[ y \to -\infty \]

plot by Academo
Probabilities from Joint CDF

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

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

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

\[ P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) \]
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.
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**

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

**Joint CDF**

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

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

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

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?

\[
P(-0.5 < X < 0.5, -0.5 < Y < 0.5)
= P(X < 0.5, Y < 0.5) - P(X < 0.5, Y < -0.5)
- P(X < -0.5, Y < 0.5) + P(X < -0.5, Y < -0.5)
= \phi\left(\frac{0.5}{3}\right) \cdot \phi\left(\frac{0.5}{3}\right) - 2\phi\left(\frac{0.5}{3}\right) \cdot \phi\left(\frac{-0.5}{3}\right)
+ \phi\left(\frac{-0.5}{3}\right) \cdot \phi\left(\frac{-0.5}{3}\right)
= 0.5662^2 - 2 \cdot 0.5662 \cdot 0.4338 + 0.4338^2 = 0.206
\]
Pedagogic Pause
Properties of Joint Distributions
Recall: any boolean question about a random variable makes for an event. For example:

\[ P(X \leq 5) \]

\[ P(Y = 6) \]

\[ P(5 \leq Z \leq 10) \]
Four Prototypical Trajectories

Conditionals with multiple variables
Recall that for events $E$ and $F$:

$$P(E \mid F) = \frac{P(EF)}{P(F)}$$

where $P(F) > 0$
Recall that for events $E$ and $F$:

$$P(E \mid F) = \frac{P(EF)}{P(F)} \quad \text{where} \quad P(F) > 0$$

- Now, have $X$ and $Y$ as discrete random variables
  - Conditional PMF of $X$ given $Y$:

$$P_{X \mid Y}(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

Different notations, same idea.
### Joint Probability Table

<table>
<thead>
<tr>
<th></th>
<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>
<td>0.37</td>
</tr>
<tr>
<td>Soph</td>
<td>0.12</td>
<td>0.18</td>
<td>0.00</td>
<td>0.03</td>
<td>0.32</td>
</tr>
<tr>
<td>Junior</td>
<td>0.04</td>
<td>0.01</td>
<td>0.00</td>
<td>0.10</td>
<td>0.15</td>
</tr>
<tr>
<td>Senior</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td>5+</td>
<td>0.02</td>
<td>0.00</td>
<td>0.05</td>
<td>0.04</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>0.49</td>
<td>0.27</td>
<td>0.07</td>
<td>0.18</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Marginal Room type**

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

**Marginal Year**

- Frosh: 0.49
- Soph: 0.38
- Junior: 0.15
- Senior: 0.11
- 5+: 0.19
A line graph titled "P(Room | Year)", showing the probability distribution of different room categories (Roommates, 2RoomDbi, Shared Partner, Single) across different years (Frosh, Soph, Junior, 5+).
Relationship Status | Year

P(Status | Year)

- Single
- In a relationship
- It's complicated

Conditional Probability

Freshman | Sophomore | Junior | Senior | 5+

The graph shows the conditional probability of different relationship statuses across different years of study.
Number or Function?

$$P(X = 2|Y = 5)$$

Number
Function
(or 1D table)

\[ P(X = 2 | Y = y) \]
Number or Function?

\[ P(X = x \mid Y = y) \]

2D Function (or 2D table)
And It Applies to Books Too

<table>
<thead>
<tr>
<th>Title</th>
<th>Author</th>
<th>Price</th>
<th>Rating</th>
<th>Customer Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harry Potter and the Sorcerer's Stone (Book 1) (Hardcover)</td>
<td>J.K. Rowling</td>
<td>$15.92</td>
<td>4.5</td>
<td>5.0</td>
</tr>
<tr>
<td>Harry Potter and the Prisoner of Azkaban (Book 3) by J.K. Rowling</td>
<td>(2.599) $16.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harry Potter and the Goblet of Fire (Book 4)</td>
<td>J.K. Rowling</td>
<td>$19.79</td>
<td>4.5</td>
<td>5.0</td>
</tr>
<tr>
<td>Harry Potter and the Order of the Phoenix (Book 5) by J.K. Rowling</td>
<td>(5.876) $10.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harry Potter and the Half-Blood Prince (Book 6)</td>
<td>J.K. Rowling</td>
<td>$10.18</td>
<td>4.5</td>
<td>5.0</td>
</tr>
<tr>
<td>The Tales of Beedle the Bard, Collector's Ed...</td>
<td>J.K. Rowling</td>
<td>(176)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P(Buy Book Y | Bought Book X)

![Amazon screenshot](image)
Continuous Conditional Distributions

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

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

$$f_{X|Y}(x|y) \cdot \epsilon_x = \frac{f_{X,Y}(x, y) \cdot \epsilon_x \cdot \epsilon_y}{f_Y(y) \cdot \epsilon_y}$$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$
Warmup: Bayes Revisited

Posterior belief

\[ P(B|E) = \frac{P(E|B) \cdot P(B)}{P(E)} \]

Prior belief

Likelihood of evidence

Normalization constant
Mixing Discrete and Continuous

Let $X$ be a continuous random variable
Let $N$ be a discrete random variable

$$P(X = x | N = n) = \frac{P(N = n | X = x)P(X = x)}{P(N = n)}$$

$$P_{X|N}(x|n) = \frac{P_{N|X}(n|x)P_X(x)}{P_N(n)}$$

$$f_{X|N}(x|n) \cdot \epsilon_x = \frac{P_{N|X}(n|x)f_X(x) \cdot \epsilon_x}{P_N(n)}$$

$$f_{X|N}(x|n) = \frac{P_{N|X}(n|x)f_X(x)}{P_N(n)}$$
All the Bayes Belong to Us

M, N are discrete. X, Y are continuous

\[ p_{M|N}(m|n) = \frac{P_{N|M}(n|m)p_M(m)}{p_N(n)} \]

\[ f_{X|N}(x|n) = \frac{P_{N|X}(n|x)f_X(x)}{P_N(n)} \]

\[ p_{N|X}(n|x) = \frac{f_{X|N}(x|n)p_N(n)}{f_X(x)} \]

\[ f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} \]
Warmup: Bayes Revisited

\[ P(B|E) = \frac{P(E|B) \cdot P(B)}{P(E)} \]

- **Posterior belief**
- **Likelihood of evidence**
- **Prior belief**

**Normalization constant**
Warmup: Bivariate Normal

- X, Y follow a symmetric bivariate normal distribution if they have joint PDF:

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

Here is an example where:

- $\mu_x = 3$
- $\mu_y = 3$
- $\sigma = 2$
Tracking in 2D Space?
You have a **prior** belief about the 2D location of an object.

What is your **updated belief** about the 2D location of the object after observing a **noisy distance** measurement?
Tracking in 2D Space: Prior

Prior belief:

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

Relative to Satellite at (0, 0)

\[ \mu_x = 3 \]
\[ \mu_y = 3 \]
\[ \sigma = 2 \]

Prior belief with K:

\[ f_{X,Y}(x, y) = K \cdot e^{-\frac{[(x-3)^2 + (y-3)^2]}{8}} \]
You will observe a noisy distance reading. It will say that your object is distance D away.

We can say how likely that reading is if we know the actual location of the object…

$P(D \mid X, Y)$ is knowable!
Observe a ping of the object that is distance $D$ away from satellite!

$$D|X, Y \sim N(\mu = \sqrt{x^2 + y^2}, \sigma^2 = 1)$$

Know that the distance of a ping is normal with respect to the true distance.
Observe a ping of the object that is distance $D = 4$ away!

Know that the distance of a ping is normal with respect to the true distance
Last known possible position of MH370 based on satellite data (somewhere on red lines)

Satellite 37.800 km above sea level.
Observe a ping of the object that is distance $D = 4$ away!

$\sqrt{x^2 + y^2} = 4$

Know that the distance of a ping is normal with respect to the true distance

$\mu = \text{actual distance}$

$\sigma = 1$
Observe a ping of the object that is distance $D = 4$ away from satellite!

$$D|X,Y \sim N(\mu = \sqrt{x^2 + y^2}, \sigma^2 = 1)$$

$$f(D = d|X = x, Y = y) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(d-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{(d-\mu)^2}{2}}$$

$$= K_2 \cdot e^{-\frac{(d-\mu)^2}{2}}$$

$$= K_2 \cdot e^{-\frac{(d-\sqrt{x^2+y^2})^2}{2}}$$
What is your new belief for the location of the object being tracked? Your joint probability density function can be expressed with a constant
Tracking in 2D Space: New Belief

\[ f(X = x, Y = y | D = 4) = \frac{f(D = 4 | X = x, Y = y) \cdot f(X = x, Y = y)}{f(D = 4)} \]

\[ = K_1 \cdot e^{-\frac{(4-\sqrt{x^2+y^2})^2}{2}} \cdot K_2 \cdot e^{-\frac{(x-3)^2+(y-3)^2}{8}} \]

\[ = K_3 \cdot e^{-\left[\frac{(4-\sqrt{x^2+y^2})^2}{2} + \frac{(x-3)^2+(y-3)^2}{8}\right]} \]

\[ = K_4 \cdot e^{-\left[\frac{(4-\sqrt{x^2+y^2})^2}{2} + \frac{(x-3)^2+(y-3)^2}{8}\right]} \]

For your notes…
Tracking in 2D Space: Posterior

Prior

Posterior

\[ f_{X,Y} \]

\[ f_{X,Y|D} \]

(top view)
Tracking in 2D Space: CS221