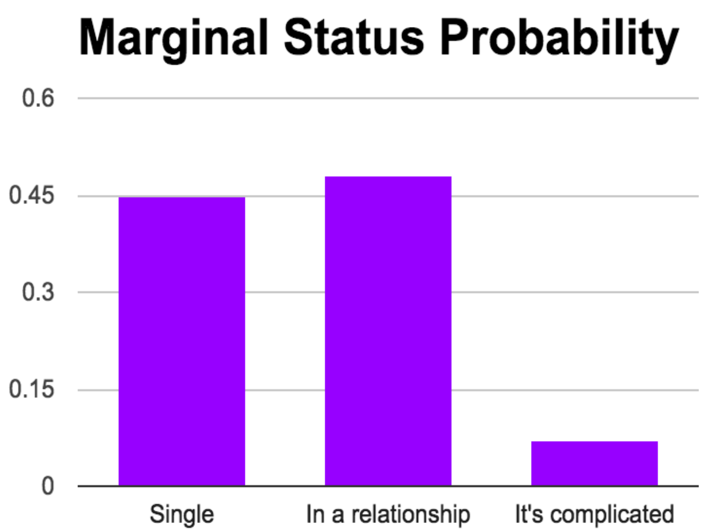
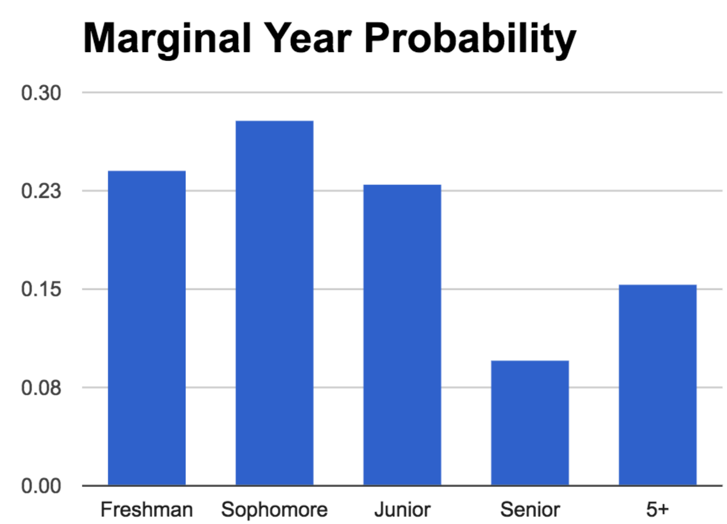
Properties of Joint Distributions

Chris Piech
CS109, Stanford University

Discrete Joint Random Variables

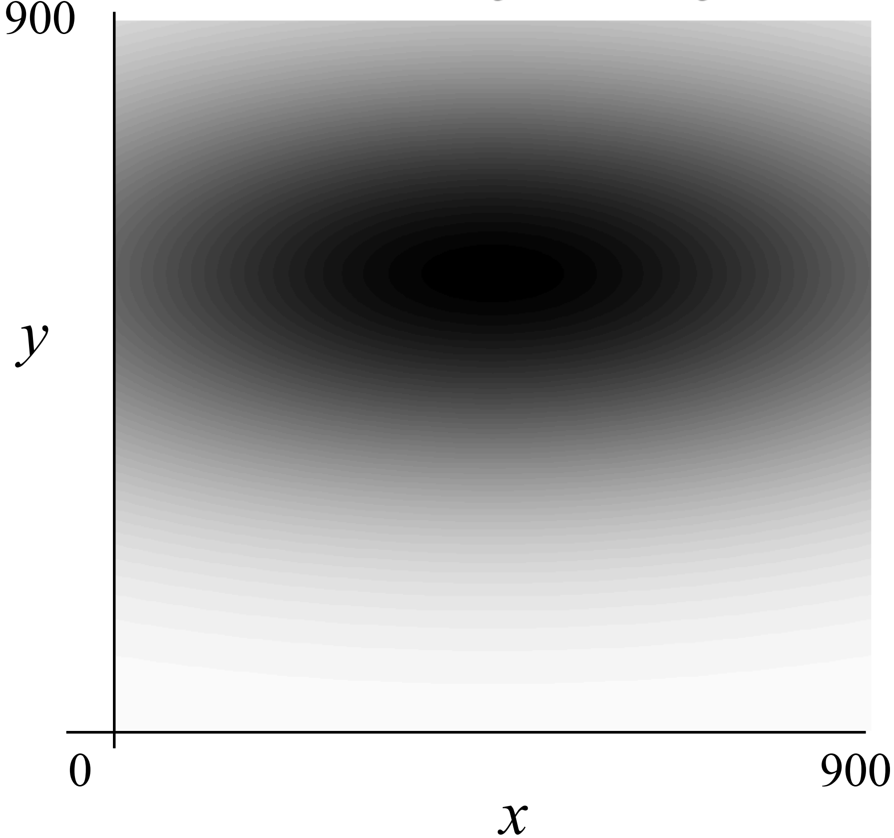
Joint Probability Table				
	Single	In a relationship	It's complicated	**Marginal Year**
Freshman	0.13	0.09	0.02	0.24
Sophomore	0.16	0.10	0.02	0.28
Junior	0.12	0.10	0.02	0.23
Senior	0.01	0.09	0.00	0.10
5+	0.03	0.12	0.01	0.15
Marginal Status	0.45	0.48	0.07	



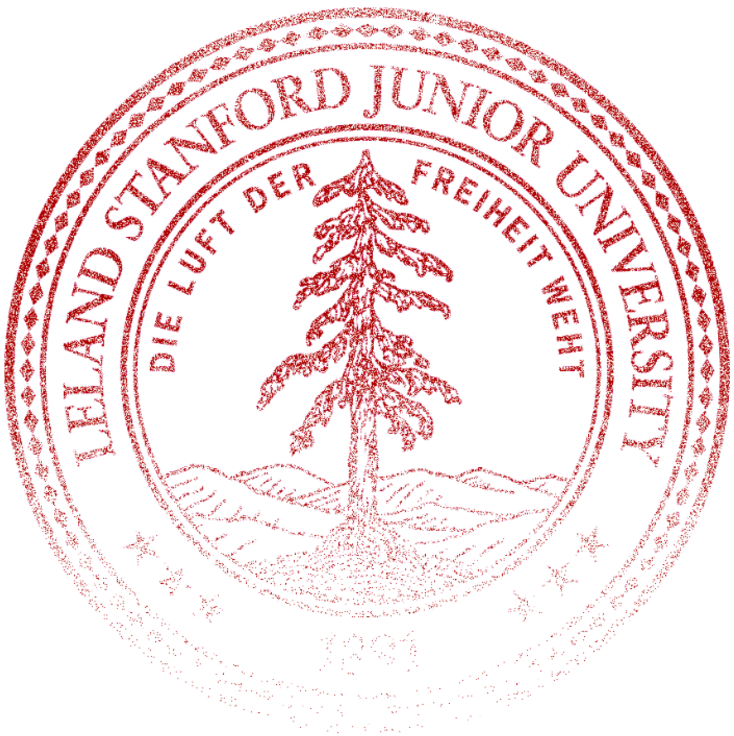


Continuous Joint Random Variables

Dart Probability Density



Dart Results



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 $\sum_{i=1}^{m} p_i = 1$
  - X_i = number of trials with outcome i

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

Joint distribution

Multinomial # ways of ordering the successes

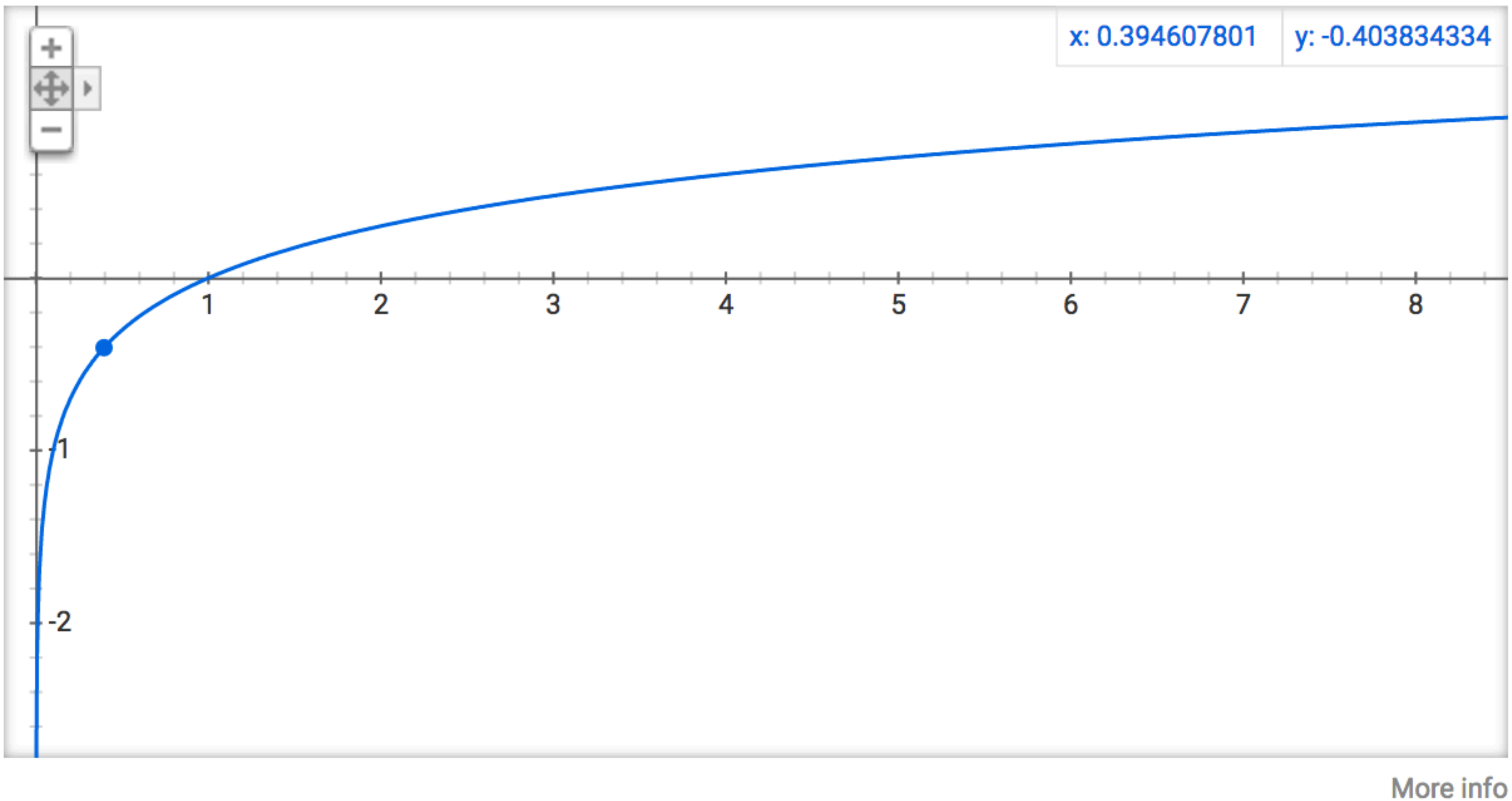
Probabilities of each ordering are equal and mutually exclusive

where $\sum_{i=1}^{m} c_i = n$ and $\binom{n}{c_1, c_2, \ldots, c_m} = \dfrac{n!}{c_1! c_2! \cdots c_m!}$

Log Review

$$e^y = x \qquad \log(x) = y$$

Graph for log(x)



Log Identities

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

$$\log(a/b) = \log(a) - \log(b)$$

$$\log(a^n) = n \cdot \log(a)$$

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.

Who wrote the federalist papers?

Probabilistic Text Analysis

- Ignoring order of words, what is probability of any given word you write in English?
  - P(word = “the”) > P(word = “transatlantic”)
  - P(word = “Stanford”) > P(word = “Cal”)
  - Probability of each word is just multinomial distribution
- What about probability of those same words in someone else’s writing?
  - P(word = “probability” | writer = you) >

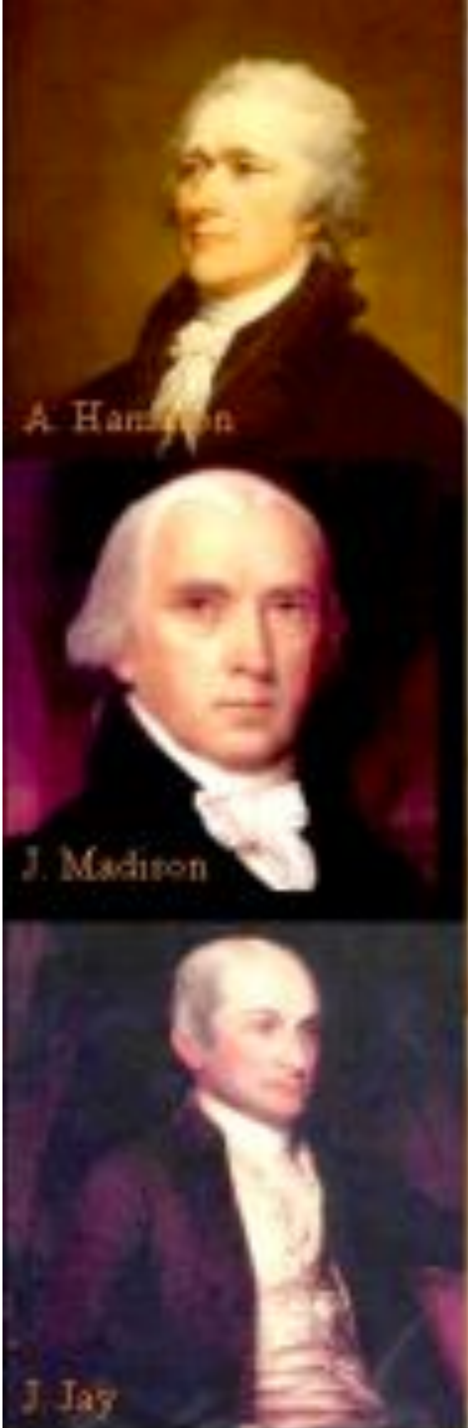
    P(word = “probability” | writer = non-CS109 student)
  - After estimating P(word | writer) from known writings, use Bayes’ Theorem to determine P(writer | word) for new writings!

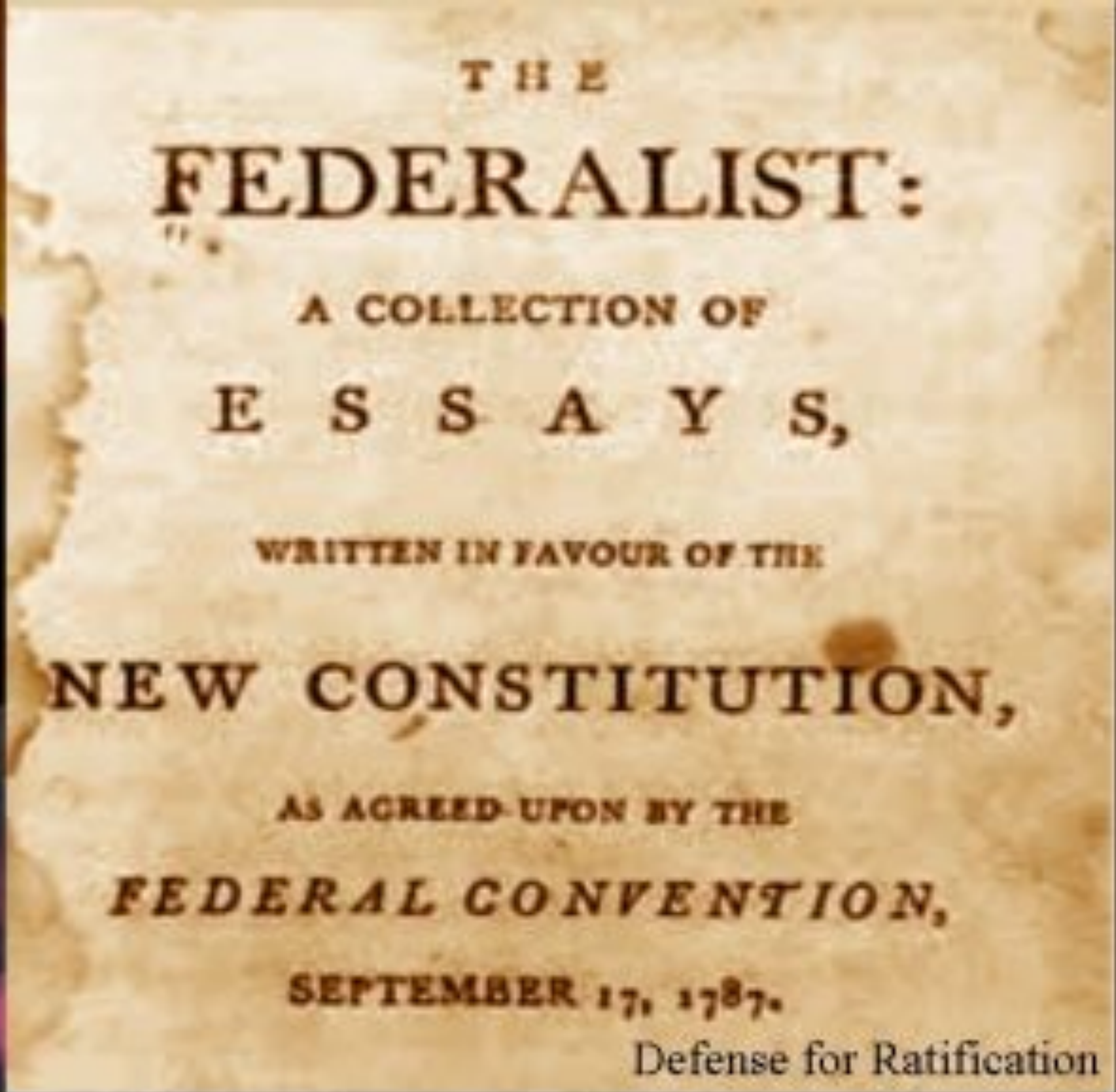
Text and the Multinomial

Example document:
"Pay for Viagra with a credit-card. Viagra is great. So are credit-cards. Risk free Viagra. Click for free."
n = 18

$$\underbrace{P\left(\begin{array}{l}\text{Viagra} = 2\\ \text{Free} = 2\\ \text{Risk} = 1\\ \text{Credit-card: } 2\\ \dots\\ \text{For} = 2\end{array}\middle|\,\text{spam}\right)}_{\text{Probability of seeing this document | spam}} = \overbrace{\frac{n!}{2!2!\dots 2!}\,p_{\text{viagra}}^2\,p_{\text{free}}^2 \cdots p_{\text{for}}^2}^{\text{It's a Multinomial!}}$$

The probability of a word in spam email being viagra





THE

FEDERALIST:

A COLLECTION OF

E S S A Y S,

WRITTEN IN FAVOUR OF THE

NEW CONSTITUTION,

AS AGREED UPON BY THE

FEDERAL CONVENTION,

SEPTEMBER 17, 1787.

Defense for Ratification

The Story so Far

Expectation

Joint Random Variables
(in discrete and in
continuous world)

Conditionals

Independence

Expectation with Multiple Variables?

Joint Expectation

$$E[X] = \sum_x x p(x)$$

- Expectation over a joint isn't nicely defined because it is not clear how to compose the multiple variables:
  - Add them? Multiply them?

- Lemma: For a function g(X,Y) we can calculate the expectation of that function:

$$E[g(X,Y)] = \sum_{x,y} g(x,y) p(x,y)$$

- By the way, this also holds for single random variables:

$$E[g(X)] = \sum_x g(x) p(x)$$

Expected Values of Sums

Big deal lemma: first stated without proof

E[X + Y] = E[X] + E[Y]

Generalized: $$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Holds regardless of dependency between X_i's

Skeptical Chris Wants a Proof!

Let g(X,Y) = [X + Y]

$$E[X+Y] = E[g(X,Y)] = \sum_{x,y} g(x,y)p(x,y)$$

What a useful lemma

$$= \sum_{x,y} [x+y]p(x,y)$$

By the definition of g(x,y)

Break that sum into parts!

$$= \sum_{x,y} xp(x,y) + \sum_{x,y} yp(x,y)$$

Change the sum of (x,y) into separate sums

$$= \sum_{x} x \sum_{y} p(x,y) + \sum_{y} y \sum_{x} p(x,y)$$

That is the definition of marginal probability

$$= \sum_{x} xp(x) + \sum_{y} yp(y)$$

That is the definition of expectation

$$= E[X] + E[Y]$$

Independence and Random Variables

Independent Discrete Variables

- Two discrete random variables X and Y are called **independent** if:

$$p(x, y) = p_X(x) p_Y(y) \quad \text{for all } x, y$$

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

- Intuitively: knowing the value of X tells us nothing about the distribution of Y (and vice versa)
  - If two variables are **not** independent, they are called ***dependent***
- Similar conceptually to independent *events*, but we are dealing with multiple ***variables***
  - Keep your events and variables distinct (and clear)!

Is Year Independent of Status?

Joint Probability Table				
	Single	In a relationship	It's complicated	**Marginal Year**
Freshman	$\boxed{0.06}$	0.04	0.03	$\boxed{0.13}$
Sophomore	0.21	0.16	0.02	0.39
Junior	0.13	0.06	0.02	0.21
Senior	0.04	0.07	0.01	0.12
5+	0.04	0.09	0.03	0.15
Marginal Status	$\boxed{0.47}$	0.43	0.10	1.00

For all values of Year, Status:

$$\underbrace{P(\text{Year} = y, \text{Status} = s)}_{0.06} = \underbrace{P(\text{Year} = y)}_{0.13}\underbrace{P(\text{Status} = s)}_{0.47}$$

Yes!

Is Year Independent of Status?

Joint Probability Table				
	Single	In a relationship	It's complicated	**Marginal Year**
Freshman	0.06	0.04	0.03	0.13
Sophomore	$\boxed{0.21}$	0.16	0.02	$\boxed{0.39}$
Junior	0.13	0.06	0.02	0.21
Senior	0.04	0.07	0.01	0.12
5+	0.04	0.09	0.03	0.15
Marginal Status	$\boxed{0.47}$	0.43	0.10	1.00

For all values of Year, Status:

$$\underbrace{P(\text{Year} = y, \text{Status} = s)}_{0.21} = \underbrace{\underbrace{P(\text{Year} = y)}_{0.39}\underbrace{P(\text{Status} = s)}_{0.47}}_{0.18}$$

No ☹

Aside: Butterfly Effect

Coin Flips

- Flip coin with probability *p* of "heads"
  - Flip coin a total of *n* + *m* times
  - Let X = number of heads in first *n* flips
  - Let Y = number of heads in next *m* flips

$$P(X = x, Y = y) = \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{y} p^y (1-p)^{m-y}$$

$$= P(X = x)P(Y = y)$$

  - X and Y are independent
  - Let Z = number of total heads in *n* + *m* flips
  - Are X and Z independent?
    - What if you are told Z = 0?

Web Server Requests

- Let N = # of requests to web server/day
  - Suppose N ~ Poi(λ)
  - Each request comes from a human (probability = *p*) or from a "bot" (probability = (1 – *p*)), independently
  - X = # requests from humans/day (X | N) ~ Bin(N, *p*)
  - Y = # requests from bots/day (Y | N) ~ Bin(N, 1 - *p*)

$$P(X=i,Y=j) = P(X=i,Y=j \mid X+Y=i+j)P(X+Y=i+j)$$
$$+\,P(X=i,Y=j \mid X+Y \neq i+j)P(X+Y \neq i+j)$$

Probability of *i* human requests and *j* bot requests

Probability of number of requests in a day was i + j

Probability of *i* human requests and *j* bot requests | we got *i + j requests*

Web Server Requests

- Let N = # of requests to web server/day
  - Suppose N ~ Poi(λ)
  - Each request comes from a human (probability = *p*) or from a "bot" (probability = (1 – *p*)), independently
  - X = # requests from humans/day (X | N) ~ Bin(N, *p*)
  - Y = # requests from bots/day (Y | N) ~ Bin(N, 1 - *p*)

$$
\begin{aligned}
P(X=i,Y=j) = {} & P(X=i,Y=j \mid X+Y=i+j)P(X+Y=i+j) \\
& + P(X=i,Y=j \mid X+Y \neq i+j)P(X+Y \neq i+j)
\end{aligned}
$$

  - Note: $P(X=i,Y=j \mid X+Y \neq i+j) = 0$

You got *i* human requests and *j* bot requests

You did not get *i* + *j* requests

Web Server Requests

- Let N = # of requests to web server/day
  - Suppose N ~ Poi(λ)
  - Each request comes from a human (probability = *p*) or from a "bot" (probability = (1 – *p*)), independently
  - X = # requests from humans/day (X | N) ~ Bin(N, *p*)
  - Y = # requests from bots/day (Y | N) ~ Bin(N, 1 - *p*)

$$P(X = i, Y = j) = P(X = i, Y = j \mid X + Y = i + j)P(X + Y = i + j)$$

Web Server Requests

- Let N = # of requests to web server/day
  - Suppose N ~ Poi(λ)
  - Each request comes from a human (probability = *p*) or from a "bot" (probability = (1 – *p*)), independently
  - X = # requests from humans/day   (X | N) ~ Bin(N, *p*)
  - Y = # requests from bots/day   (Y | N) ~ Bin(N, 1 - *p*)

$$P(X=i, Y=j) = P(X=i, Y=j \mid X+Y=i+j)P(X+Y=i+j)$$

$$P(X=i, Y=j \mid X+Y=i+j) = \binom{i+j}{i} p^i (1-p)^j$$

Binomial

$$P(X+Y=i+j) = e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!}$$

Poisson

$$P(X=i, Y=j) = \binom{i+j}{i} p^i (1-p)^j e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!}$$

Joint

Web Server Requests

- Let N = # of requests to web server/day
  - Suppose N ~ Poi(λ)
  - Each request comes from a human (probability = *p*) or from a "bot" (probability = (1 – *p*)), independently
  - X = # requests from humans/day (X | N) ~ Bin(N, *p*)
  - Y = # requests from bots/day (Y | N) ~ Bin(N, 1 - *p*)

$$P(X=i, Y=j) = \frac{(i+j)!}{i!\,j!} p^i (1-p)^j e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = e^{-\lambda} \frac{(\lambda p)^i}{i!} \cdot \frac{(\lambda(1-p))^j}{j!}$$

Reorder terms

$$= e^{-\lambda p} \frac{(\lambda p)^i}{i!} \cdot e^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!} = P(X=i)P(Y=j)$$

  - Where X ~ Poi(λp) and Y ~ Poi(λ(1 – *p*))
  - X and Y are independent!

Independent Continuous Variables

- Two continuous random variables X and Y are called **independent** if:

  P(X ≤ *a*, Y ≤ *b*) = P(X ≤ *a*) P(Y ≤ *b*) for any *a*, *b*

- Equivalently:

$$F_{X,Y}(a,b) = F_X(a)F_Y(b) \quad \text{for all } a,b$$

$$f_{X,Y}(a,b) = f_X(a)f_Y(b) \quad \text{for all } a,b$$

- More generally, joint density factors separately:

$$f_{X,Y}(x,y) = h(x)g(y) \quad \text{where } -\infty < x, y < \infty$$

Pop Quiz (just kidding)

- Consider joint density function of X and Y:

$$f_{X,Y}(x, y) = 6e^{-3x}e^{-2y} \quad \text{for } 0 < x, y < \infty$$

  - Are X and Y independent? Yes!

Let $h(x) = 3e^{-3x}$ and $g(y) = 2e^{-2y}$, so $f_{X,Y}(x, y) = h(x)g(y)$

- Consider joint density function of X and Y:

$$f_{X,Y}(x, y) = 4xy \quad \text{for } 0 < x, y < 1$$

  - Are X and Y independent? Yes!

Let $h(x) = 2x$ and $g(y) = 2y$, so $f_{X,Y}(x, y) = h(x)g(y)$

  - Now add constraint that: $0 < (x + y) < 1$
  - Are X and Y independent? No!
    - Cannot capture constraint on x + y in factorization!

Dating at Stanford

- Two people set up a meeting for 12pm
  - Each arrives independently at time uniformly distributed between 12pm and 12:30pm
  - X = # min. past 12pm person 1 arrives   X ~ Uni(0, 30)
  - Y = # min. past 12pm person 2 arrives   Y ~ Uni(0, 30)
  - What is P(first to arrive waits > 10 min. for other)?

$$P(X+10<Y)+P(Y+10<X)=2P(X+10<Y) \text{ by symmetry}$$

$$2P(X+10<Y)=2\iint_{x+10<y} f(x,y)\,dx\,dy=2\iint_{x+10<y} f_X(x)f_Y(y)\,dx\,dy$$

$$=2\int_{y=10}^{30}\int_{x=0}^{y-10}\left(\frac{1}{30}\right)^2 dx\,dy=\frac{2}{30^2}\int_{y=10}^{30}\left(\int_{x=0}^{y-10}dx\right)dy=\frac{2}{30^2}\int_{y=10}^{30}\left(x\Big|_0^{y-10}\right)dy=\frac{2}{30^2}\int_{y=10}^{30}(y-10)\,dy$$

$$=\frac{2}{30^2}\left(\frac{y^2}{2}-10y\right)\Bigg|_{10}^{30}=\frac{2}{30^2}\left[\left(\frac{30^2}{2}-300\right)-\left(\frac{10^2}{2}-100\right)\right]=\frac{4}{9}$$

Independence of Multiple Variables

- *n* random variables X_1, X_2, …, X_n are called **independent** if:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i) \quad \text{for all subsets of } x_1, x_2, \ldots, x_n$$

- Analogously, for continuous random variables:

$$P(X_1 \le a_1, X_2 \le a_2, \ldots, X_n \le a_n) = \prod_{i=1}^{n} P(X_i \le a_i) \quad \text{for all subsets of } a_1, a_2, \ldots, a_n$$

Independence is Symmetric

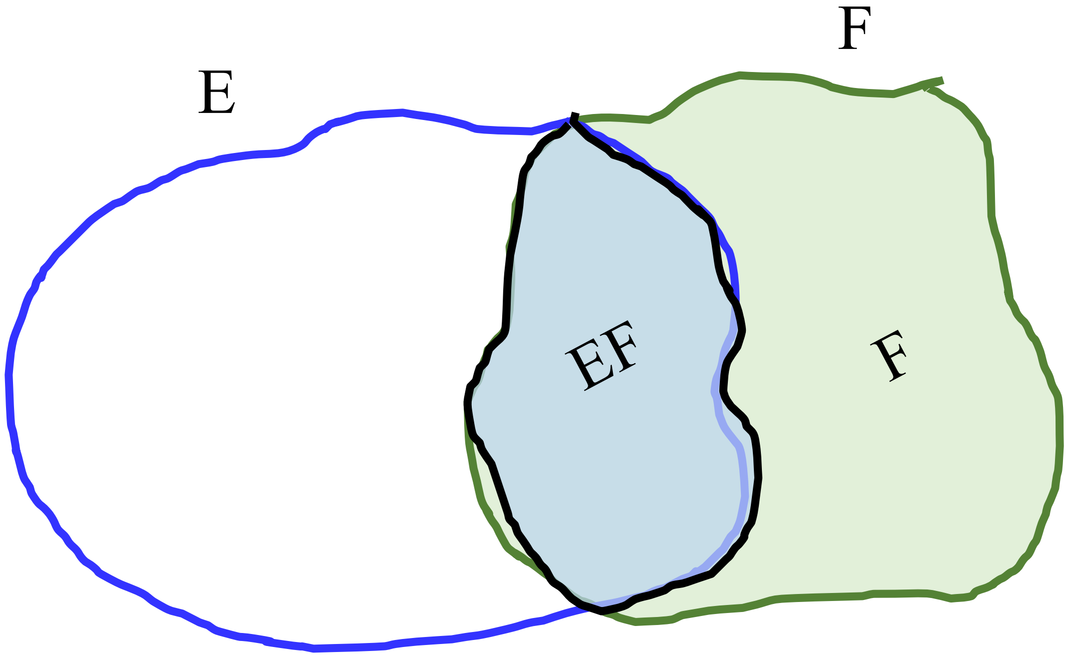
- If random variables X and Y independent, then
  - X independent of Y, and Y independent of X
- Duh!? Duh, indeed...
  - Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed (I.I.D.) continuous random vars
  - Say $X_n > X_i$ for all $i = 1, \ldots, n - 1$ (i.e. $X_n = \max(X_1, \ldots, X_n)$)
    - Call X_n a "record value"
  - Let event A_i indicate X_i is "record value"
    - Is A_{n+1} independent of A_n?
    - Is A_n independent of A_{n+1}?
    - Easier to answer: Yes!
    - By symmetry, $P(A_n) = 1/n$ and $P(A_{n+1}) = 1/(n+1)$
    - $P(A_n A_{n+1}) = (1/n)(1/(n+1)) = P(A_n)P(A_{n+1})$

Conditionals with multiple variables

Discrete Conditional Distribution

- Recall that for *events* E and F:

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



Discrete Conditional Distributions

- 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 (where $p_Y(y) > 0$):

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

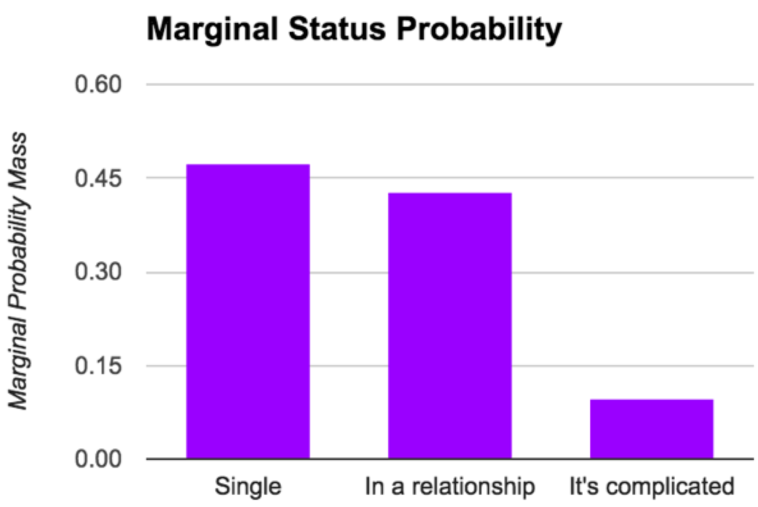
  - Conditional CDF of X given Y (where $p_Y(y) > 0$):

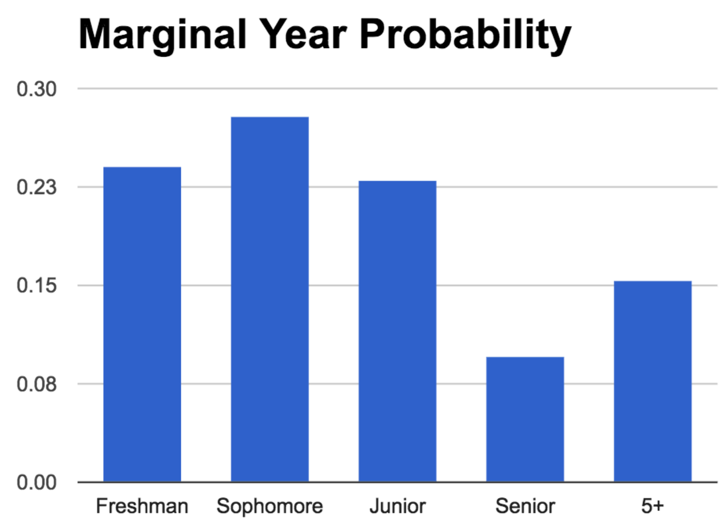
$$F_{X|Y}(a \mid y) = P(X \le a \mid Y = y) = \frac{P(X \le a, Y = y)}{P(Y = y)}$$

$$= \frac{\sum_{x \le a} p_{X,Y}(x, y)}{p_Y(y)} = \sum_{x \le a} p_{X|Y}(x \mid y)$$

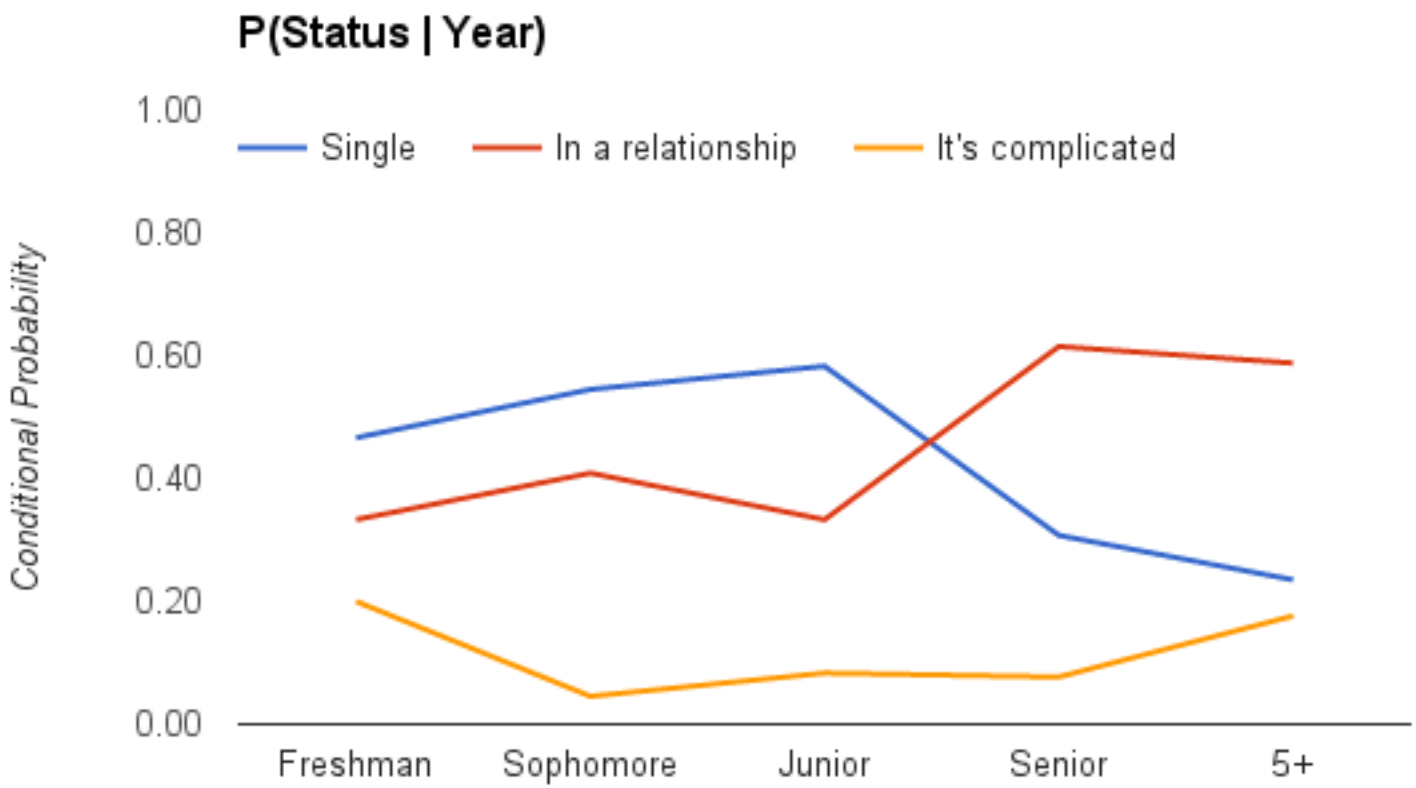
Probability Table

Joint Probability Table				
	Single	In a relationship	It's complicated	**Marginal Year**
Freshman	0.06	0.04	0.03	0.13
Sophomore	0.21	0.16	0.02	0.39
Junior	0.13	0.06	0.02	0.21
Senior	0.04	0.07	0.01	0.12
5+	0.04	0.09	0.03	0.15
Marginal Status	0.47	0.43	0.10	1.00





Relationship Status



And It Applies to Books Too

P(Buy Book Y | Bought Book X)

Continuous Conditional Distributions

- Let X and Y be continuous random variables
  - Conditional PDF of X given Y (where $f_Y(y) > 0$):

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

$$f_{X|Y}(x \mid y)\, dx = \frac{f_{X,Y}(x,y)\, dx\, dy}{f_Y(y)\, dy}$$

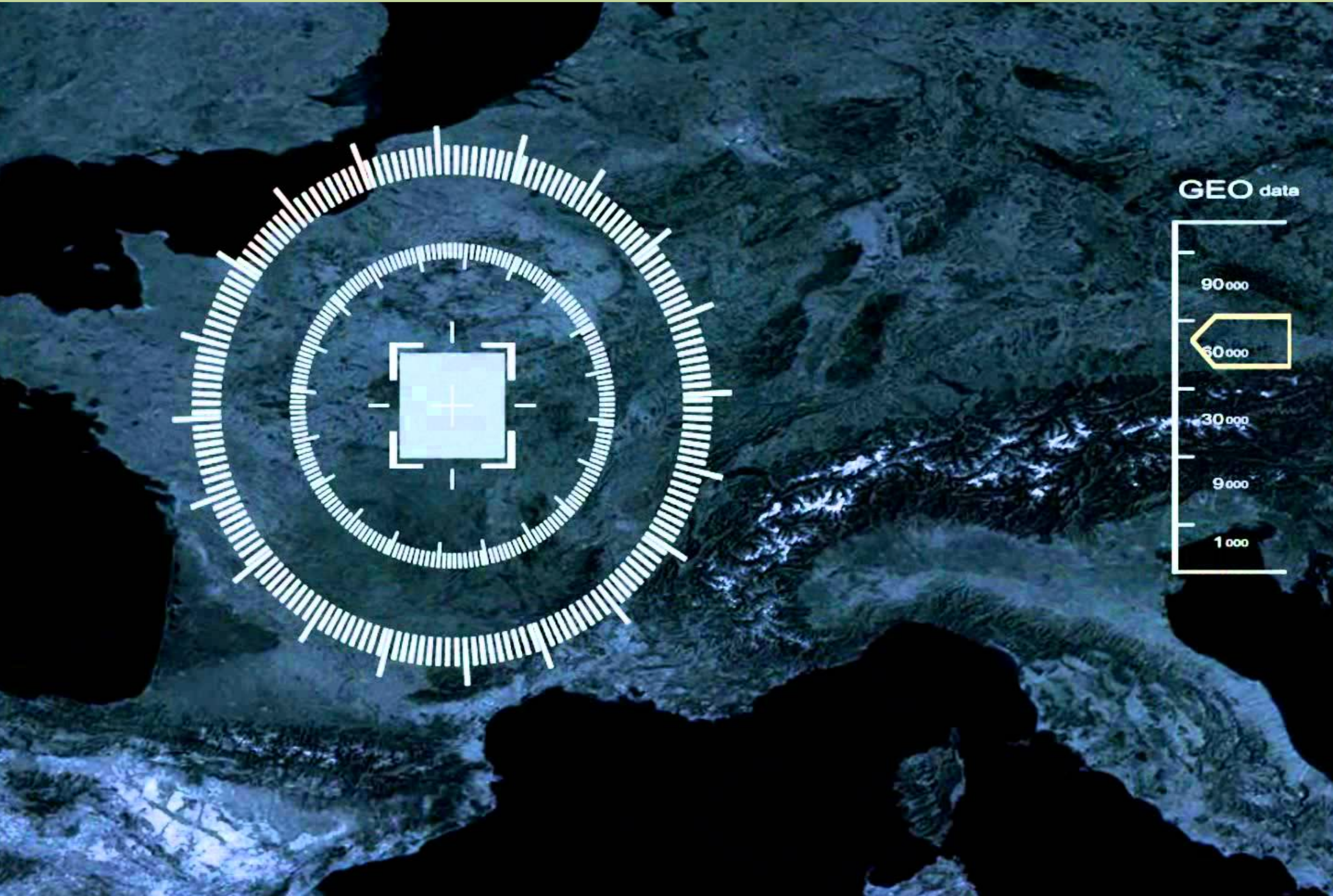
$$\approx \frac{P(x \le X \le x + dx, y \le Y \le y + dy)}{P(y \le Y \le y + dy)} = P(x \le X \le x + dx \mid y \le Y \le y + dy)$$

  - Conditional CDF of X given Y (where $f_Y(y) > 0$):

$$F_{X|Y}(a \mid y) = P(X \le a \mid Y = y) = \int_{-\infty}^{a} f_{X|Y}(x \mid y)\, dx$$

  - Note: Even though P(Y = *a*) = 0, can condition on Y = *a*
    - Really considering: $P(a - \frac{\varepsilon}{2} \le Y \le a + \frac{\varepsilon}{2}) = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f_Y(y) dy \approx \varepsilon f(a)$

Tracking without a grid?



Thanks!

Let's Do an Example

- X and Y are continuous RVs with PDF:

$$f(x,y)=\begin{cases}\frac{12}{5}x(2-x-y) & \text{where } 0<x,y<1\\ 0 & \text{otherwise}\end{cases}$$

  - Compute conditional density: $f_{X|Y}(x\mid y)$

$$f_{X|Y}(x\mid y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}=\frac{f_{X,Y}(x,y)}{\int_0^1 f_{X,Y}(x,y)\,dx}$$

$$=\frac{\frac{12}{5}x(2-x-y)}{\int_0^1 \frac{12}{5}x(2-x-y)\,dx}=\frac{x(2-x-y)}{\int_0^1 x(2-x-y)\,dx}=\frac{x(2-x-y)}{\left[x^2-\frac{x^3}{3}-\frac{x^2y}{2}\right]_0^1}$$

$$=\frac{x(2-x-y)}{\frac{2}{3}-\frac{y}{2}}=\frac{6x(2-x-y)}{4-3y}$$

Web Server Requests Redux

- Requests received at web server in a day
  - X = # requests from humans/day $X \sim \text{Poi}(\lambda_1)$
  - Y = # requests from bots/day $Y \sim \text{Poi}(\lambda_2)$
  - X and Y are independent → $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$
  - What is P(X = *k* | X + Y = *n*)?

$$P(X = k \mid X + Y = n) = \frac{P(X = k, Y = n-k)}{P(X+Y=n)} = \frac{P(X=k)P(Y=n-k)}{P(X+Y=n)}$$

$$= \frac{e^{-\lambda_1}\lambda_1^k}{k!} \cdot \frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!} \cdot \frac{n!}{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n} = \frac{n!}{k!(n-k)!} \cdot \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n}$$

$$= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$$

$$(X \mid X + Y = n) \sim Bin\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$$

Continuous Conditional Distributions

- Let X and Y be continuous random variables
  - Conditional PDF of X given Y (where $f_Y(y) > 0$):

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

$$f_{X|Y}(x \mid y)\, dx = \frac{f_{X,Y}(x, y)\, dx\, dy}{f_Y(y)\, dy}$$

$$\approx \frac{P(x \le X \le x + dx, y \le Y \le y + dy)}{P(y \le Y \le y + dy)} = P(x \le X \le x + dx \mid y \le Y \le y + dy)$$

  - Conditional CDF of X given Y (where $f_Y(y) > 0$):

$$F_{X|Y}(a \mid y) = P(X \le a \mid Y = y) = \int_{-\infty}^{a} f_{X|Y}(x \mid y)\, dx$$

  - Note: Even though P(Y = *a*) = 0, can condition on Y = *a*
    - Really considering: $P(a - \frac{\varepsilon}{2} \le Y \le a + \frac{\varepsilon}{2}) = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f_Y(y) dy \approx \varepsilon f(a)$

Let's Do an Example

- X and Y are continuous RVs with PDF:

$$f(x,y) = \begin{cases} \frac{12}{5}x(2-x-y) & \text{where } 0 < x,y < 1 \\ 0 & \text{otherwise} \end{cases}$$

  - Compute conditional density: $f_{X|Y}(x \mid y)$

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{\int_0^1 f_{X,Y}(x,y)\,dx}$$

$$= \frac{\frac{12}{5}x(2-x-y)}{\int_0^1 \frac{12}{5}x(2-x-y)\,dx} = \frac{x(2-x-y)}{\int_0^1 x(2-x-y)\,dx} = \frac{x(2-x-y)}{\left[x^2 - \frac{x^3}{3} - \frac{x^2 y}{2}\right]_0^1}$$

$$= \frac{x(2-x-y)}{\frac{2}{3}-\frac{y}{2}} = \frac{6x(2-x-y)}{4-3y}$$