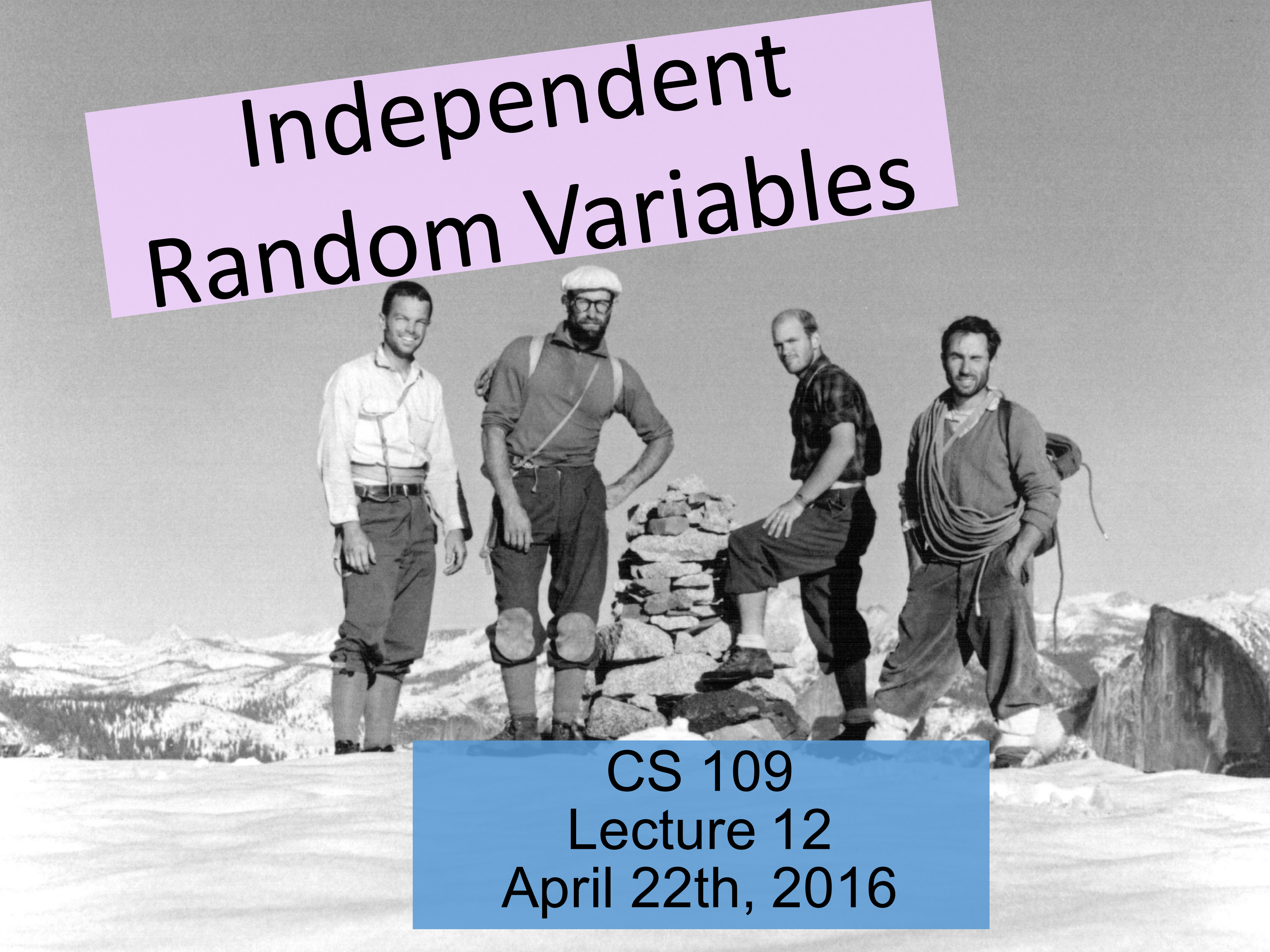
Independent Random Variables



CS 109
Lecture 12
April 22th, 2016

Today:

1. Multi variable RVs
2. Expectation with multiple RVs
3. Independence with multiple RVs

Review

Discrete Joint Mass Function

- For two discrete random variables *X* and *Y*, the **Joint Probability Mass Function** is:

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

- Marginal distributions:

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

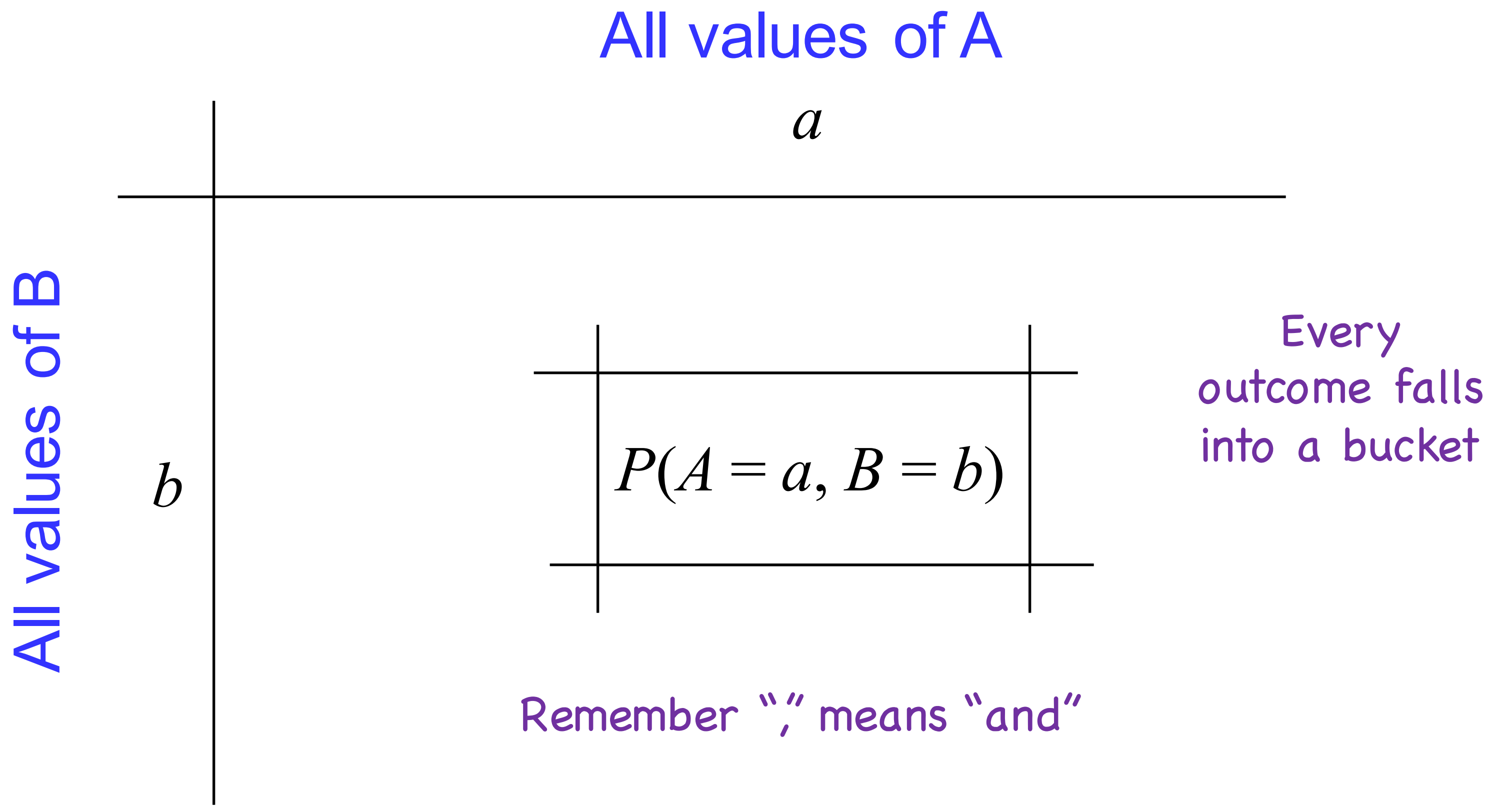
$$p_Y(b) = P(Y = b) = \sum_x p_{X,Y}(x, b)$$

- Example: X = value of die D_1, Y = value of die D_2

$$P(X = 1) = \sum_{y=1}^{6} p_{X,Y}(1, y) = \sum_{y=1}^{6} \frac{1}{36} = \frac{1}{6}$$

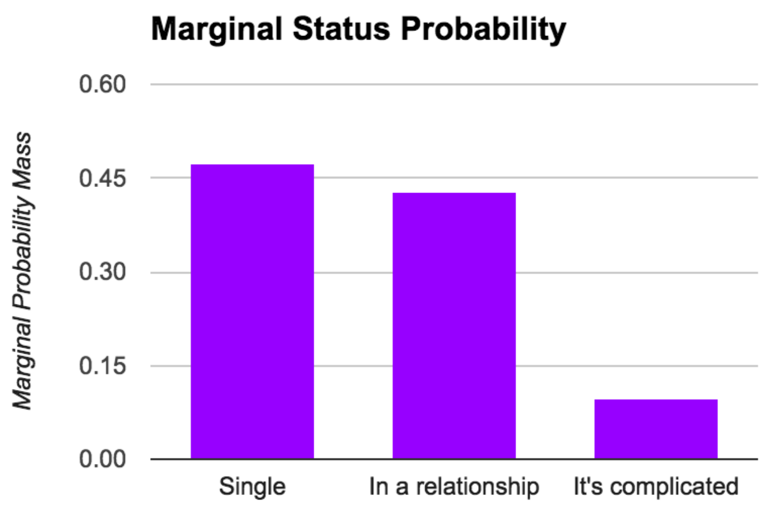
Probability Table

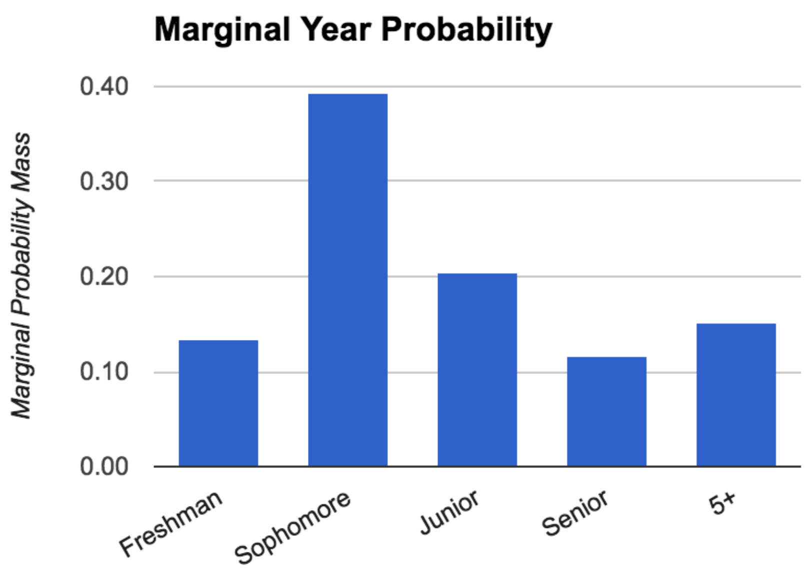
- States all possible outcomes with several discrete variables
- Often is not “parametric”
- If #variables is > 2, you can have a probability table, but you can’t draw it on a slide



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

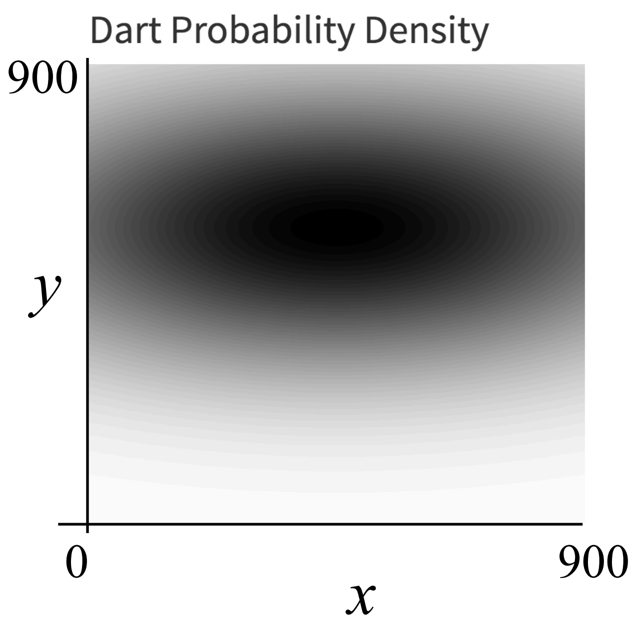




Jointly Continuous

- Random variables *X* and *Y*, are **Jointly Continuous** if there exists PDF $f_{X,Y}(x, y)$ defined over $-\infty < x, y < \infty$ such that:

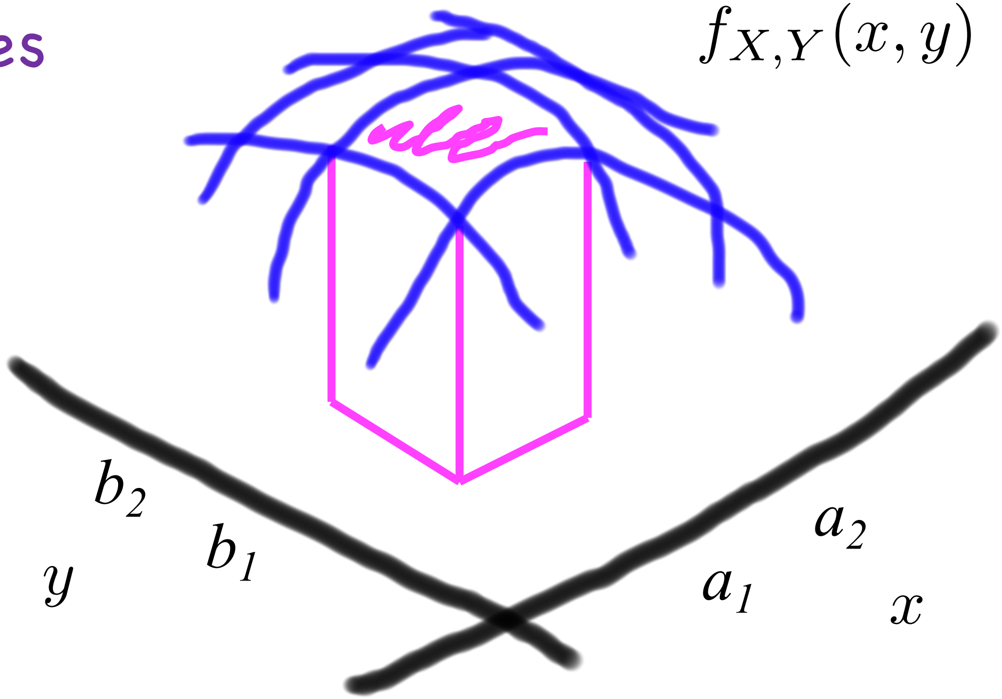
$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x, y)\, dy\, dx$$



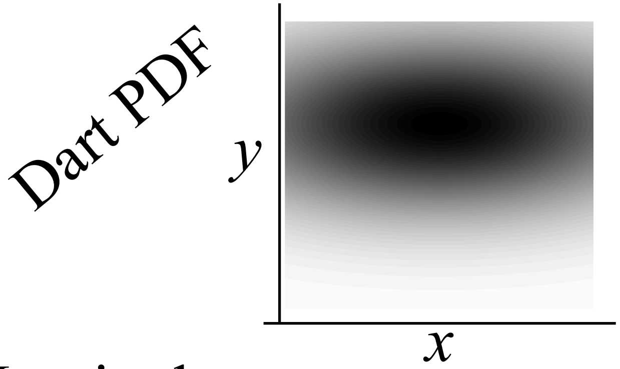
Jointly Continuous

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

Can calculate probabilities

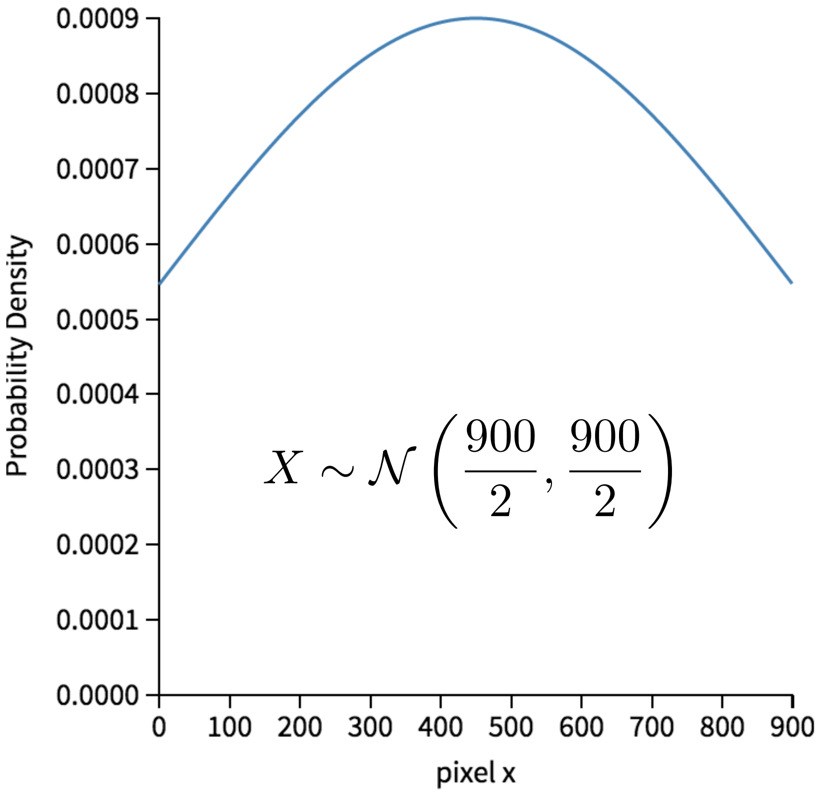


Darts!

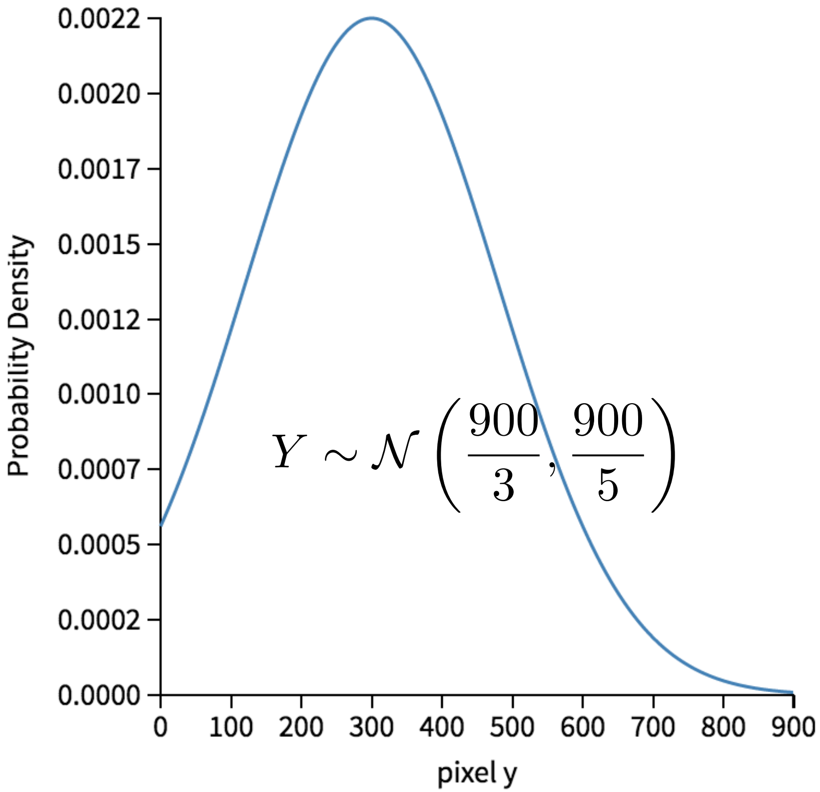


Can calculate marginal probabilities

X-Pixel Marginal



Y-Pixel Marginal



Transfer Learning

Way Back

Permutations

How many ways are there to order n distinct objects?

$$n!$$

Multinomial

How many ways are there to order n objects such that:

n_1 are the same (indistinguishable)

n_2 are the same (indistinguishable)

$\ldots$

n_r are the same (indistinguishable)?

$$\frac{n!}{n_1!n_2!\ldots n_r!} = \binom{n}{n_1, n_2, \ldots, n_r}$$

Called the “multinomial” because of something from Algebra

Binomial

How many ways are there to make an unordered selection of r objects from n objects?

How many ways are there to order n objects such that:

r are the same (indistinguishable)

$(n-r)$ are the same (indistinguishable)?

$$\frac{n!}{r!(n-r)!} = \binom{n}{r}$$

Called the Binomial (Multi -> Bi)

Binomial Distribution

- Consider *n* independent trials of Ber(p) rand. var.
  - X is number of successes in *n* trials
  - X is a **Binomial** Random Variable: X ~ Bin(n, p)

Binomial # ways of ordering the successes

$$P(X = i) = p(i) = \binom{n}{i} p^i (1 - p)^{n-i} \quad i = 0,1,\ldots,n$$

Probability of exactly *i* successes

Probability of each ordering of *i* successes is equal + mutually exclusive

End Review

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

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

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 is a 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

It's a Multinomial!

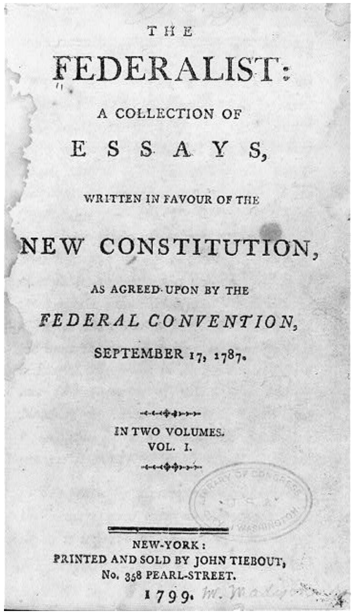
$$P\left(\begin{array}{l}\text{Viagra} = 2 \\ \text{Free} = 2 \\ \text{Risk} = 1 \\ \text{Credit-card: } 2 \\ \dots \\ \text{For} = 2\end{array} \middle| \text{spam}\right) = \frac{n!}{2!2!\dots2!} p_{\text{viagra}}^2 p_{\text{free}}^2 \dots p_{\text{for}}^2$$

Probability of seeing this document | spam

The probability of a word in spam email being viagra

Old and New Analysis

- Authorship of "Federalist Papers"
  - 85 essays advocating ratification of US constitution
  - Written under pseudonym "Publius"
    - Really, Alexander Hamilton, James Madison and John Jay
  - Who wrote which essays?
    - Analyzed probability of words in each essay versus word distributions from known writings of three authors



- Filtering Spam
  - P(word = "Viagra" | writer = you)

  << P(word = "Viagra" | writer = spammer)

Expectation with Multiple Variables?

Joint Expectation

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

- 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 **<u>independent</u>** if:

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

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

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*)

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

  - 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:

  $P(X \le a, Y \le b) = P(X \le a)\, P(Y \le b)$ 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 $X_1, X_2, \ldots, X_n$ 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 $P(A_{n+1}) = 1/(n+1)$
    - $P(A_n A_{n+1}) = (1/n)(1/(n+1)) = P(A_n)P(A_{n+1})$

Earth Day

Choosing a Random Subset

- From set of *n* elements, choose a subset of size *k* such that all $\binom{n}{k}$ possibilities are <u>equally</u> likely
  - Only have **`random()`**, which simulates X ~ Uni(0, 1)
- Brute force:
  - Generate (an ordering of) all subsets of size *k*
  - Randomly pick one (divide (0, 1) into $\binom{n}{k}$ intervals)
  - Expensive with regard to time and space
  - Bad times!

(Happily) Choosing a Random Subset

- Good times:

```
int indicator(double p) {
   if (random() < p) return 1; else return 0;
}

// array I[] indexed from 1 to n
subset rSubset(k, set of size n) {
   subset_size = 0;
   I[1] = indicator((double)k/n);
   for(i = 1; i < n; i++) {
      subset_size += I[i];
      I[i+1] = indicator((k - subset_size)/(n - i));
   }
   return (subset containing element[i] iff I[i] == 1);
}
```

$P(I[1]=1)=\frac{k}{n}$ and $P(I[i+1]=1 \mid I[1],\ldots,I[i])=\frac{k-\sum_{j=1}^{i} I[j]}{n-i}$ where $1<i<n$

Random Subsets the Happy Way

- Proof (Induction on (k + n)): (i.e., why this algorithm works)
  - Base Case: k = 1, n = 1, Set S = {*a*}, `rSubset` returns {*a*} with p= $1 \Big/ \binom{1}{1}$
  - Inductive Hypoth. (IH): for k + x ≤ c, Given set S, |S| = x and k ≤ x, `rSubset` returns any subset S' of S, where |S'| = k, with p = $1 \Big/ \binom{x}{k}$
  - Inductive Case 1: (where k + n ≤ c + 1) |S| = n (= x + 1), `I[1] = 1`
    - Elem 1 in subset, choose k – 1 elems from remaining n – 1
    - By IH: `rSubset` returns subset S' of size k – 1 with p = $1 \Big/ \binom{n-1}{k-1}$
    - P(I[1] = 1, subset S') = $\frac{k}{n} \cdot 1 \Big/ \binom{n-1}{k-1} = 1 \Big/ \binom{n}{k}$
  - Inductive Case 2: (where k + n ≤ c + 1) |S| = n (= x + 1), `I[1] = 0`
    - Elem 1 not in subset, choose k elems from remaining n – 1
    - By IH: `rSubset` returns subset S' of size k with p = $1 \Big/ \binom{n-1}{k}$
    - P(I[1] = 0, subset S') = $\left(1 - \frac{k}{n}\right) \cdot 1 \Big/ \binom{n-1}{k} = \left(\frac{n-k}{n}\right) \cdot 1 \Big/ \binom{n-1}{k} = 1 \Big/ \binom{n}{k}$

Sum of Independent Binomial RVs

- Let X and Y be independent random variables
  - $X \sim \text{Bin}(n_1, p)$ and $Y \sim \text{Bin}(n_2, p)$
  - $X + Y \sim \text{Bin}(n_1 + n_2, p)$
- Intuition:
  - X has n_1 trials and Y has n_2 trials
    - Each trial has same "success" probability p
  - Define Z to be $n_1 + n_2$ trials, each with success prob. p
  - $Z \sim \text{Bin}(n_1 + n_2, p)$, and also $Z = X + Y$
- More generally: $X_i \sim \text{Bin}(n_i, p)$ for $1 \leq i \leq N$

$$\left(\sum_{i=1}^{N} X_i\right) \sim \text{Bin}\left(\sum_{i=1}^{N} n_i,\ p\right)$$

Sum of Independent Poisson RVs

- Let X and Y be independent random variables
  - X ~ Poi(λ_1) and Y ~ Poi(λ_2)
  - X + Y ~ Poi(λ_1 + λ_2)
- Proof: (just for reference)
  - Rewrite (X + Y = n) as (X = k, Y = n – k) where 0 ≤ k ≤ n

$$P(X+Y=n)=\sum_{k=0}^{n}P(X=k,Y=n-k)=\sum_{k=0}^{n}P(X=k)P(Y=n-k)$$

$$=\sum_{k=0}^{n}e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}=e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!}=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k}$$

  - Noting Binomial theorem: $(\lambda_1+\lambda_2)^n=\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k}$
  - $P(X+Y=n)=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}\left(\lambda_1+\lambda_2\right)^n$ so, X + Y = n ~ Poi(λ_1 + λ_2)