Convolution and Conditional

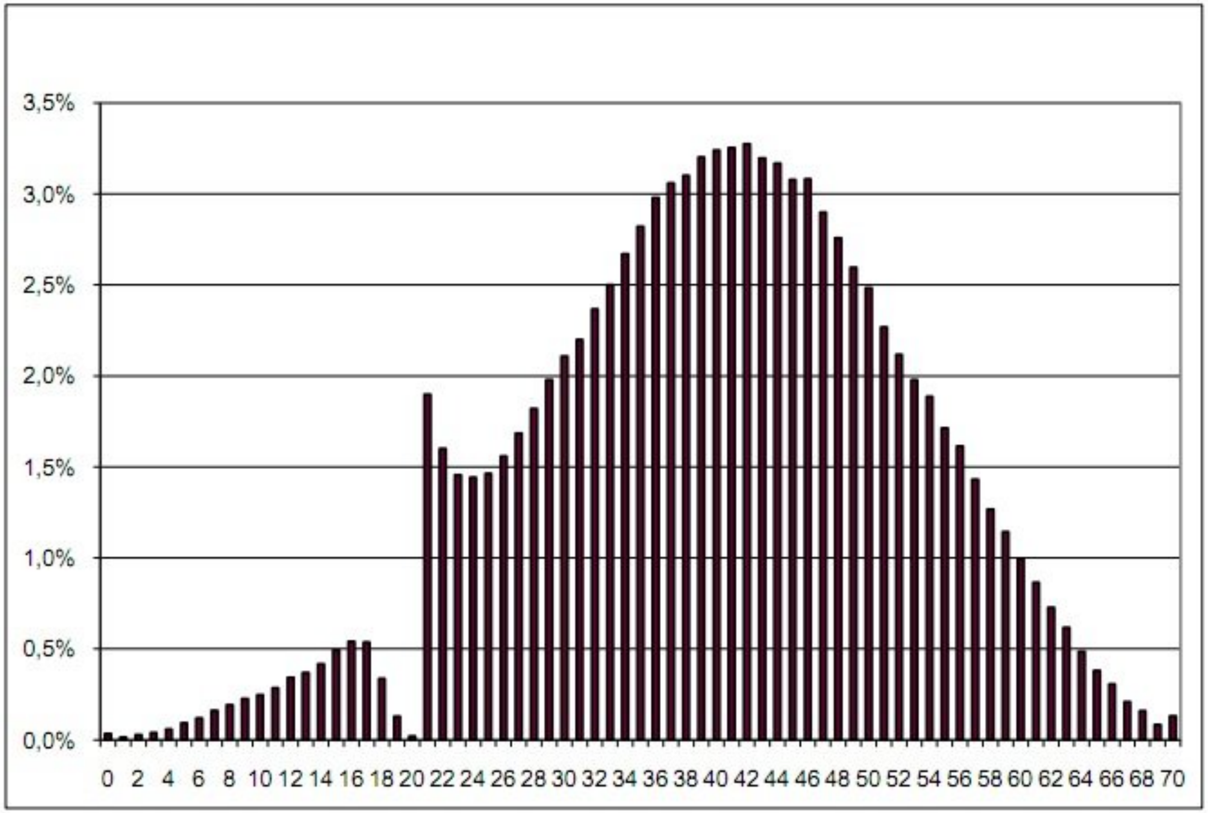
Chris Piech
CS109, Stanford University

Altruism?

Scores for a standardized test that students in Poland are required to pass before moving on in school

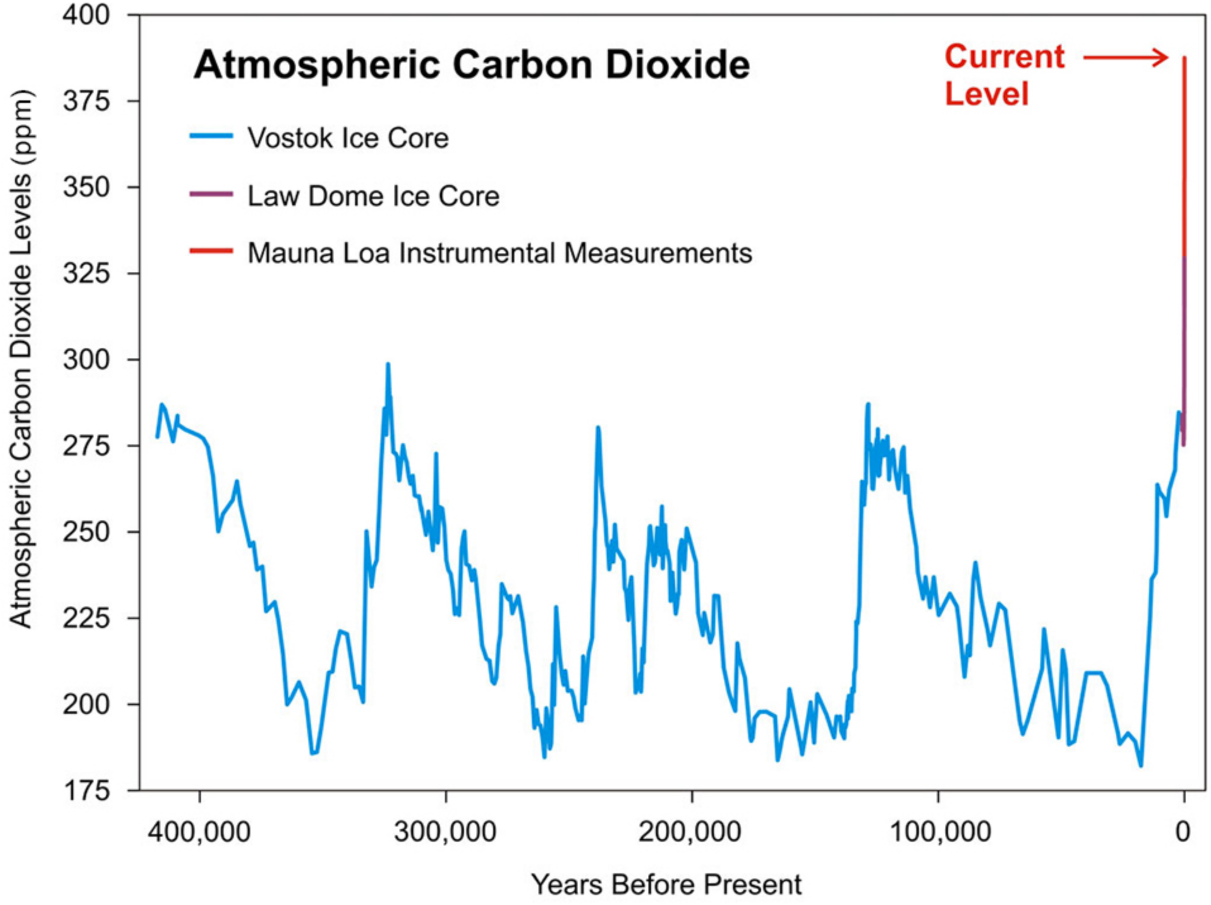
See if you can guess the minimum score to pass the test.

2.1. Poziom podstawowy



Wykres 1. Rozkład wyników na poziomie podstawowym

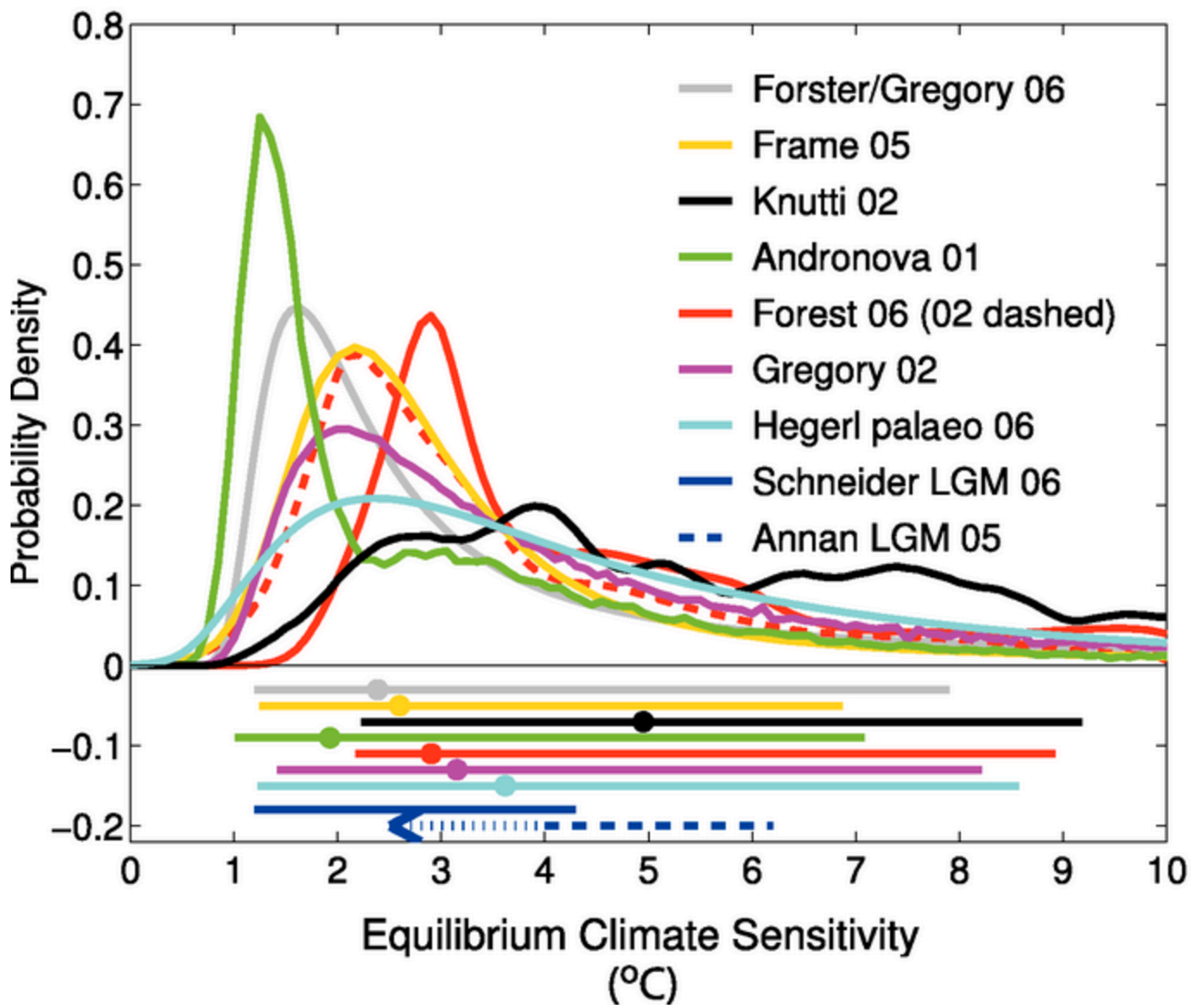
http://freakonomics.com/2011/07/07/another-case-of-teacher-cheating-or-is-it-just-altruism/comment-page-2/



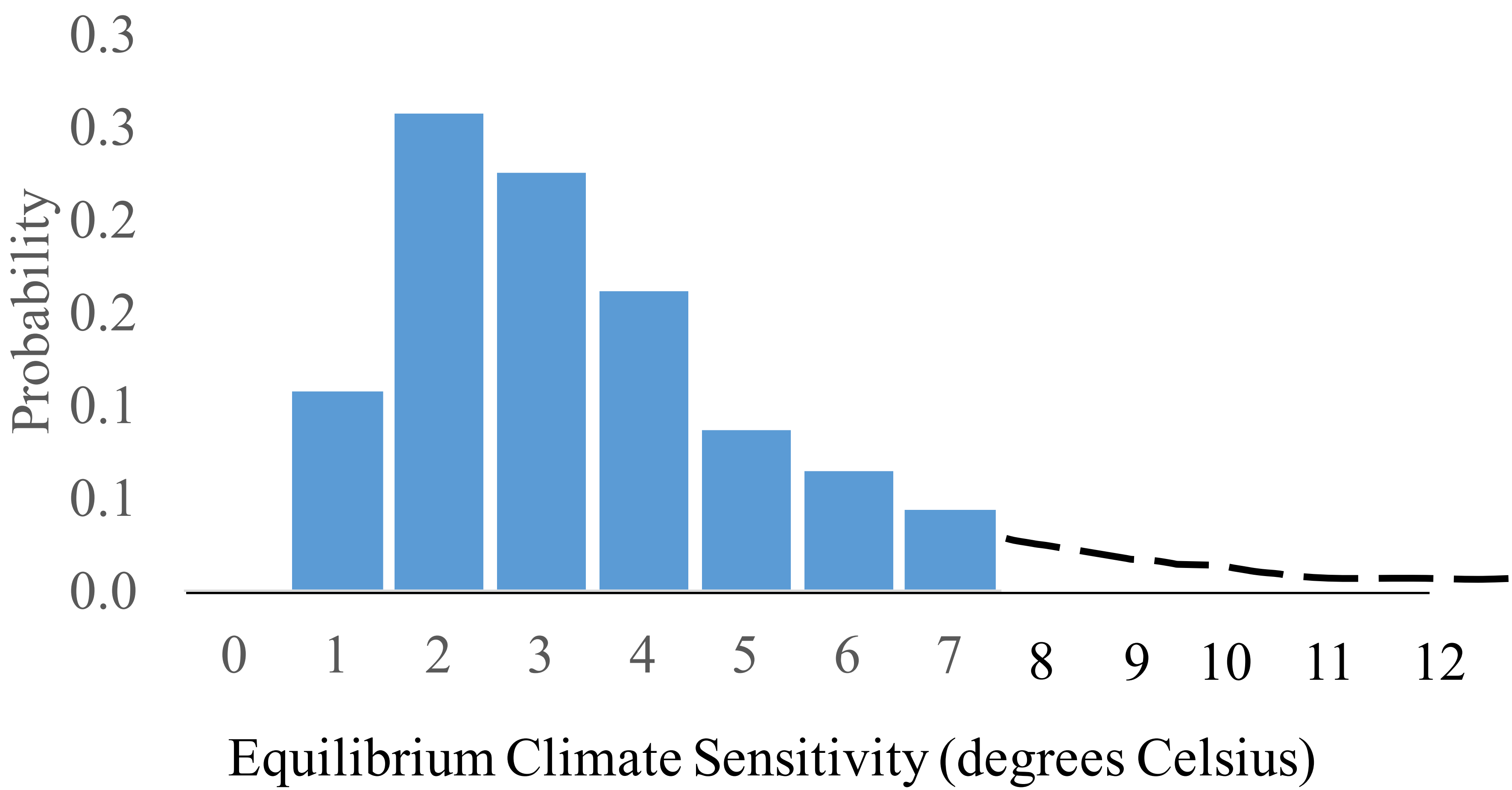
CO2 Pre-Industrial: 275 parts per million (ppm)

CO2 Today: 407 parts per million (ppm)

Climate Sensitivity



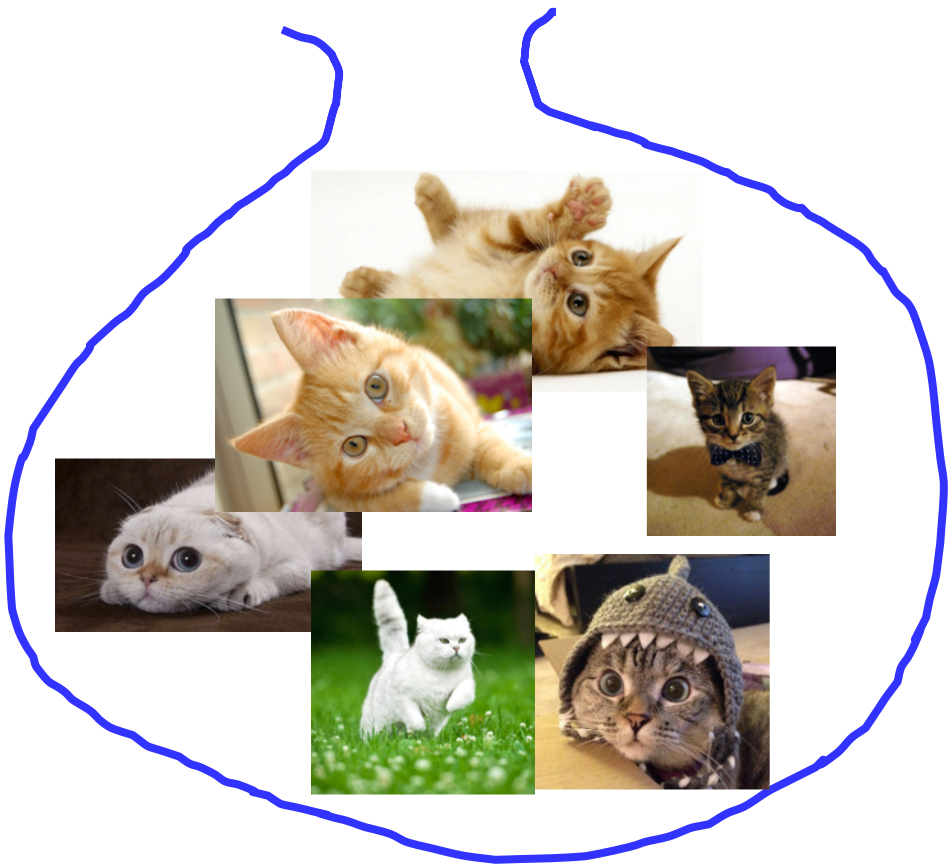
Climate Sensitivity



Algorithmic Practice

Choosing a Random Subset

Original Set (size n)



Subset (size k)

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:

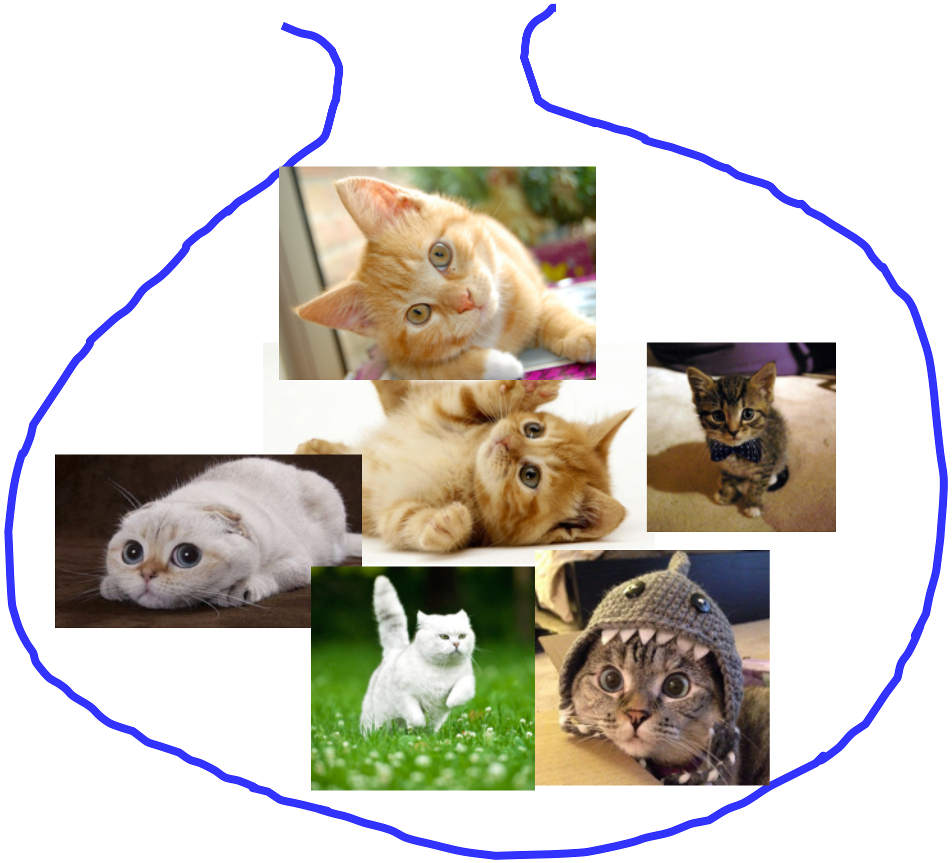
```
int indicator(double p) {
   if (random() < p) return 1; else return 0;
}

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);
}
```

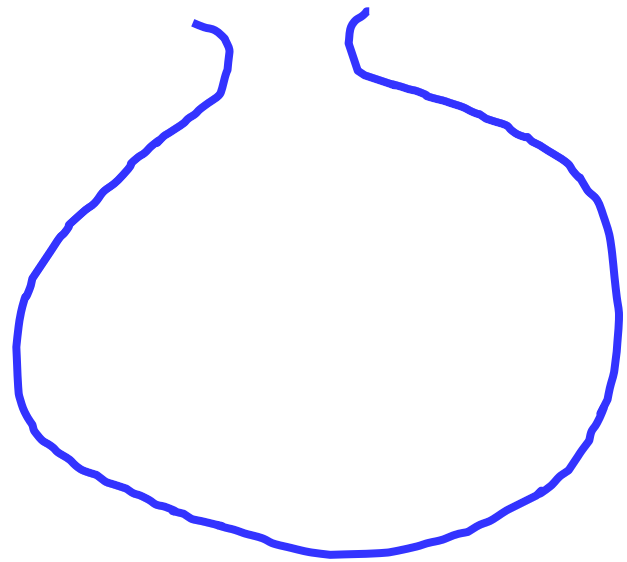
$P(I[1]=1)=\dfrac{k}{n}$ and $P(I[i+1]=1 \mid I[1],\ldots,I[i])=\dfrac{k-\sum_{j=1}^{i} I[j]}{n-i}$ where $1<i<n$

Choosing a Random Subset

Original Set (size n)



Subset (size k)

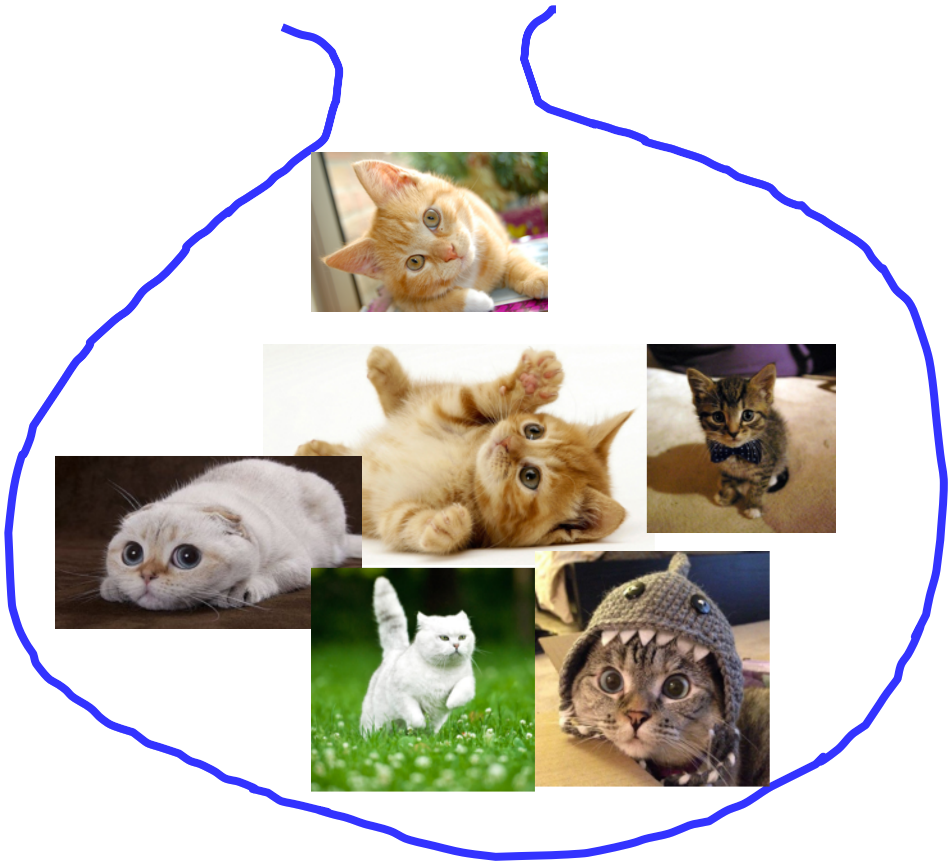


Random Subsets the Happy Way

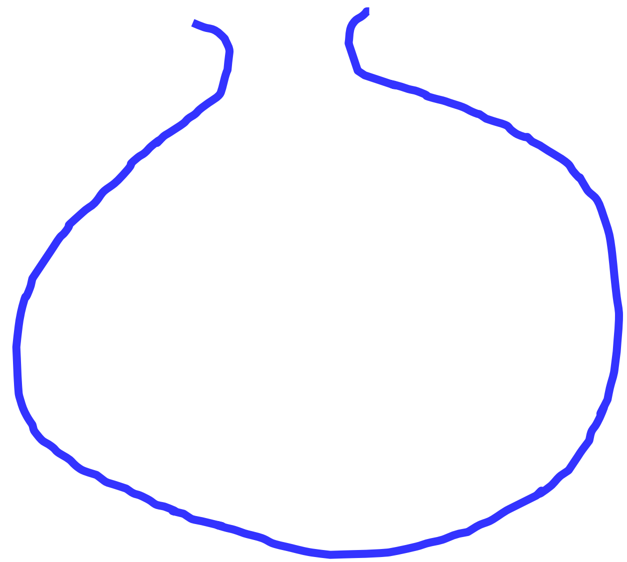
- Proof (Induction on (k + n)): (i.e., why this algorithm works)
  - Base Case: k = 1, n = 1, Set S = {*a*}, **`rSubset`** returns {*a*} with p=$1\Big/\binom{1}{1}$
  - Inductive Hypoth. (IH): for k + x ≤ c, Given set S, |S| = x and k ≤ x, **`rSubset`** returns any subset S' of S, where |S'| = k, with 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') = $\left(1-\frac{k}{n}\right)\cdot 1\Big/\binom{n-1}{k} = \left(\frac{n-k}{n}\right)\cdot 1\Big/\binom{n-1}{k} = 1\Big/\binom{n}{k}$

Induction Cases

Original Set (size *n*)

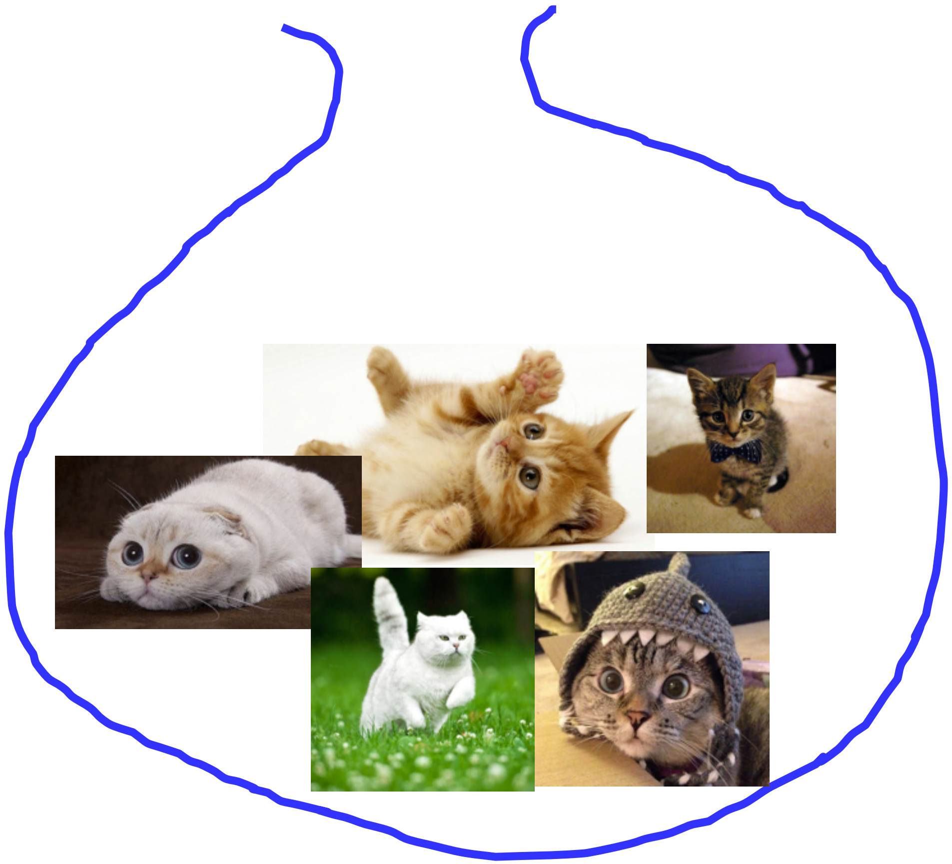


Subset (size *k*)



Case 1

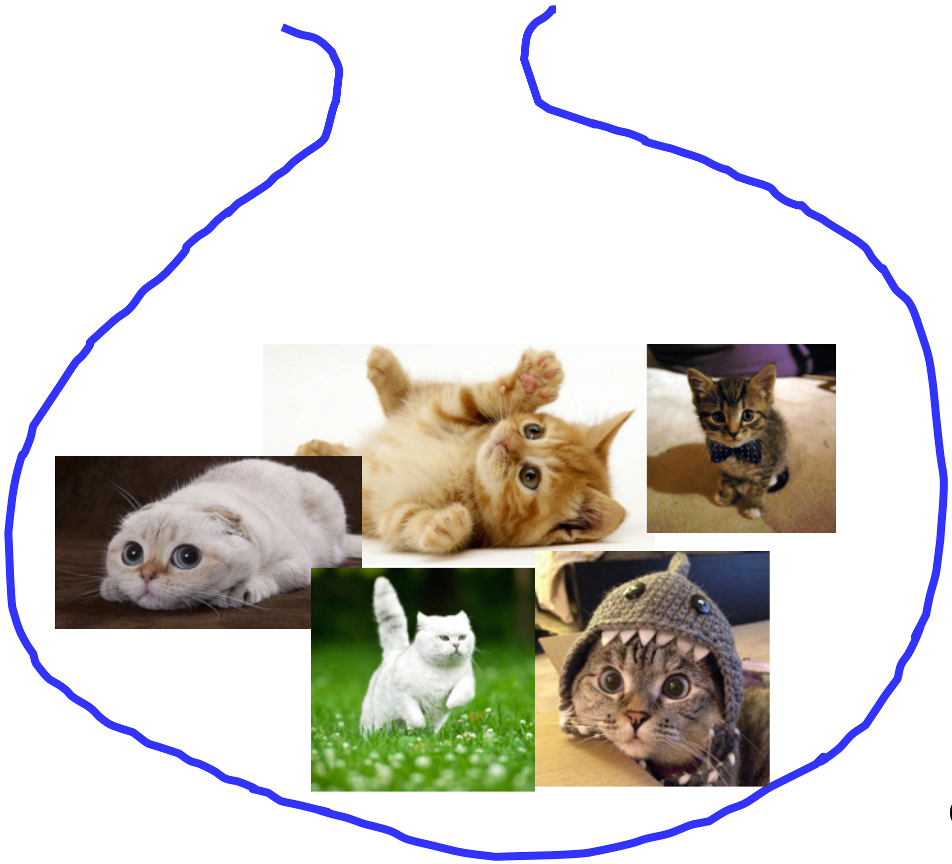
Original Set (size n)



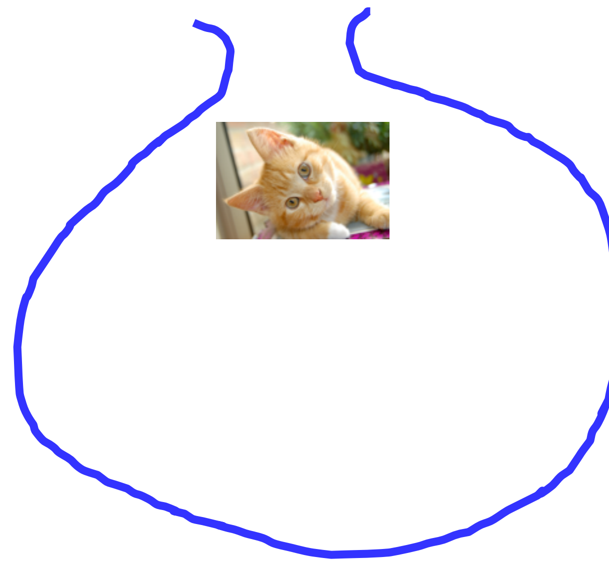
Subset (size k)

Case 1

Original Set (size *n-1*)



Subset (size *k-1*)

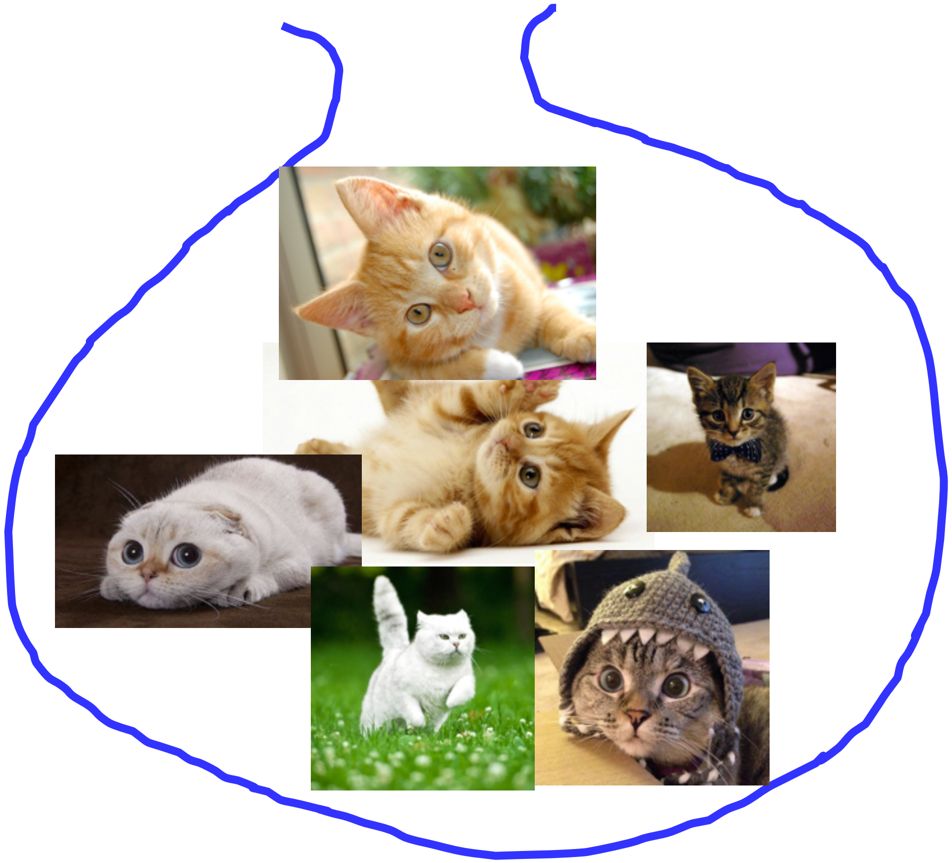


By induction we know that all subsamples of size k-1 from n-1 items are equally likely

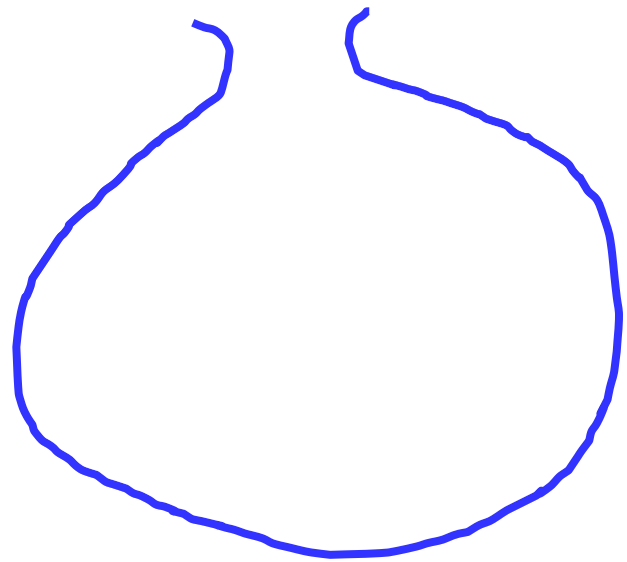
$$\mathrm{P(subset)} = \frac{k}{n} \cdot 1 \Big/ \binom{n-1}{k-1} = 1 \Big/ \binom{n}{k}$$

Choosing a Random Subset

Original Set (size *n*)

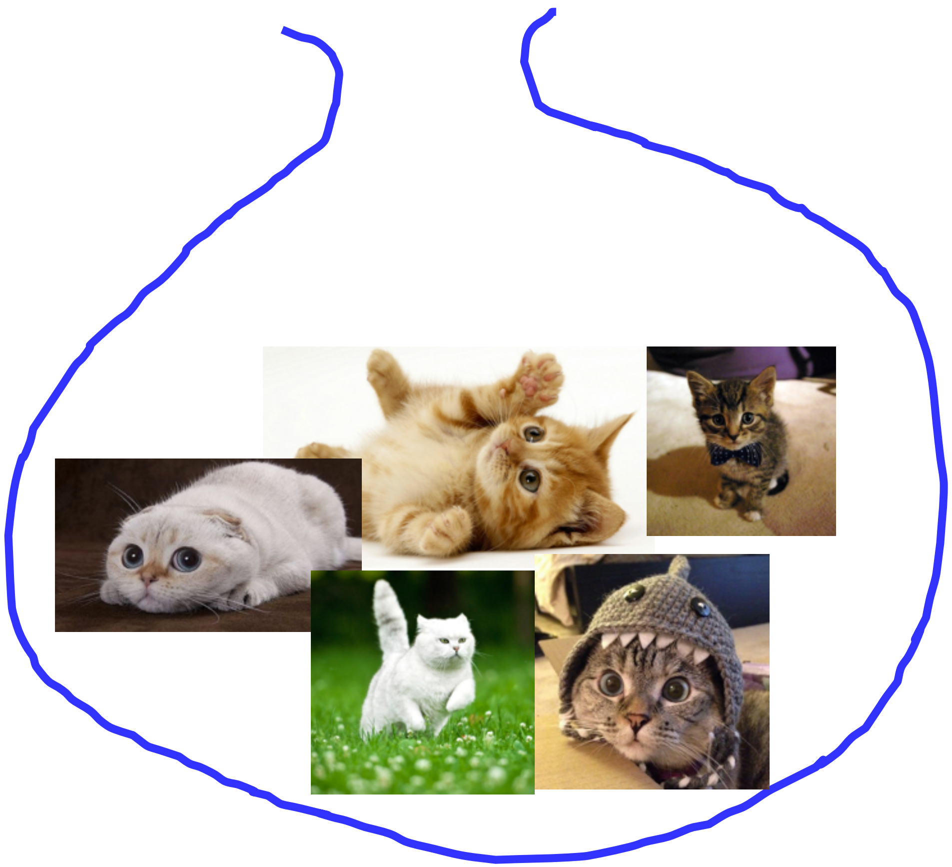


Subset (size *k*)

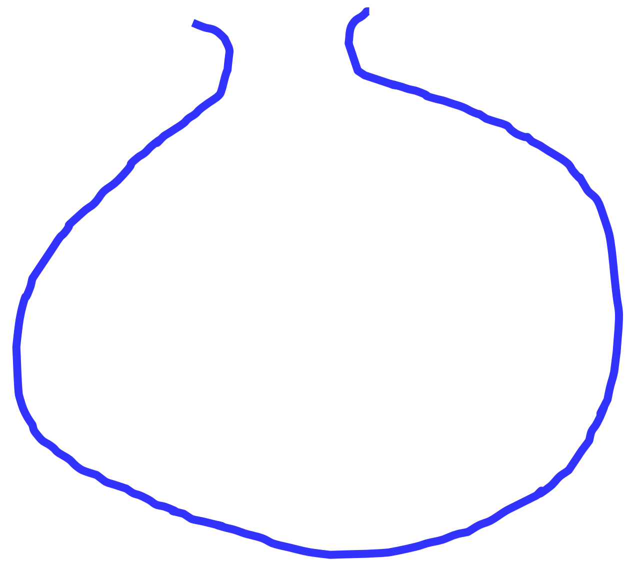


Case 2

Original Set (size *n-1*)



Subset (size *k*)



By induction we know that all subsamples of size k from n-1 are equally likely

$$P(\text{subset}) = \left(1-\frac{k}{n}\right)\cdot 1\Big/\binom{n-1}{k} = \left(\frac{n-k}{n}\right)\cdot 1\Big/\binom{n-1}{k} = 1\Big/\binom{n}{k}$$

All combinations are in either case.
Each combination in the cases are
equally likely

The Story so Far

Independence

Expectation

Joint Random Variables
(in discrete and in
continuous world)

Conditionals

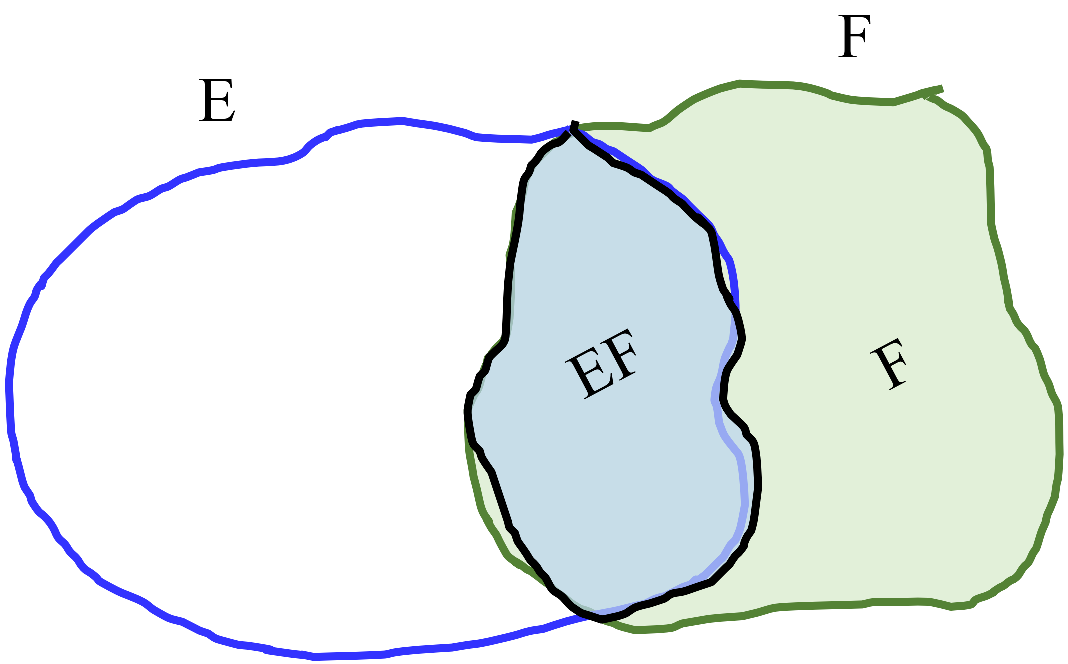
Adding

Conditionals with multiple variables

Discrete Conditional Distribution

- Recall that for *events* E and F:

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



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 $p_Y(y) > 0$):

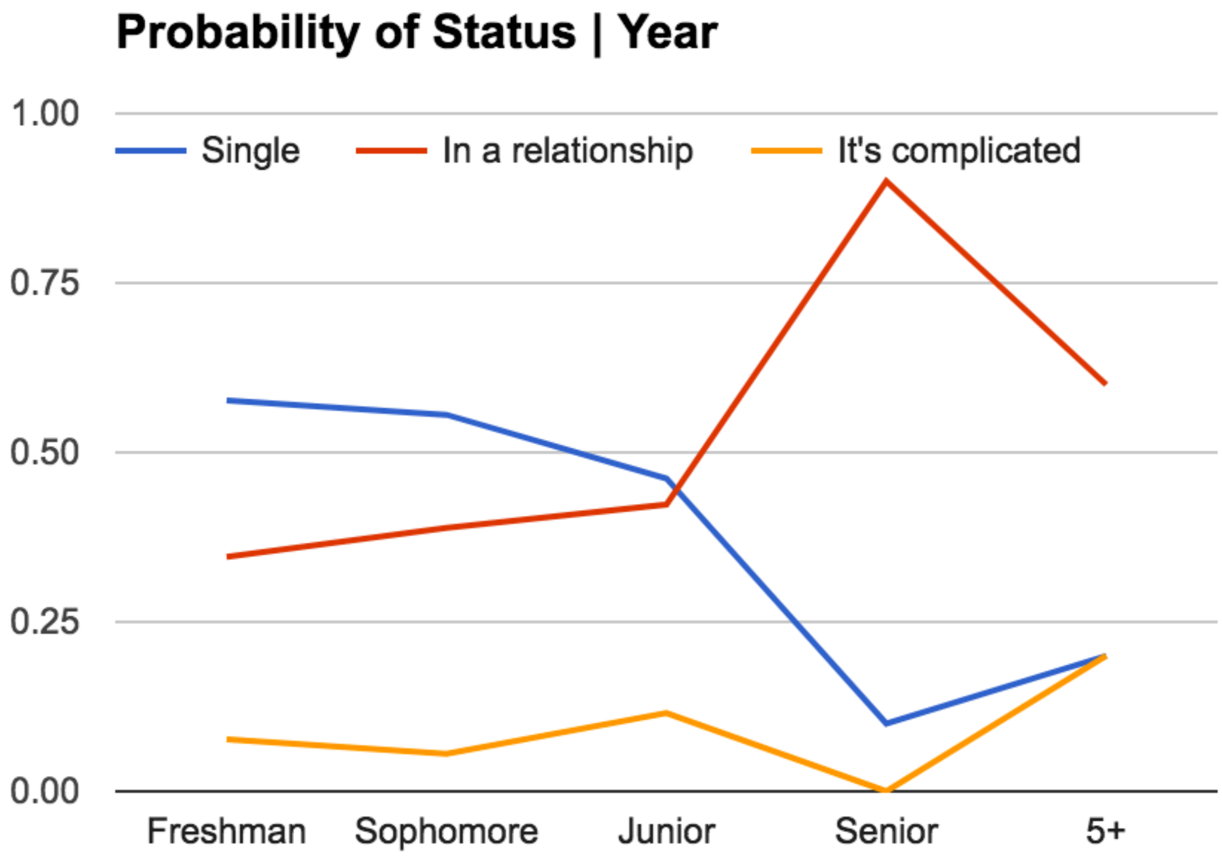
$$F_{X|Y}(a \mid y) = P(X \le a \mid Y = y) = \frac{P(X \le a, Y = y)}{P(Y = y)}$$

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

Conditional Probability?

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



Operating System Loyalty

- Consider person buying 2 computers (over time)
  - X = 1st computer bought is a PC (1 if it is, 0 if it is not)
  - Y = 2nd computer bought is a PC (1 if it is, 0 if it is not)
  - Joint probability mass function (PMF):

Y \ X	0	1	$p_Y(y)$
0	0.2	0.3	0.5
1	0.1	0.4	0.5
$p_X(x)$	0.3	0.7	1.0

  - What is P(Y = 0 | X = 0)?

$$P(Y=0 \mid X=0) = \frac{p_{X,Y}(0,0)}{p_X(0)} = \frac{0.2}{0.3} = \frac{2}{3}$$

  - What is P(Y = 1 | X = 0)?

$$P(Y=1 \mid X=0) = \frac{p_{X,Y}(0,1)}{p_X(0)} = \frac{0.1}{0.3} = \frac{1}{3}$$

  - What is P(X = 0 | Y = 1)?

$$P(X=0 \mid Y=1) = \frac{p_{X,Y}(0,1)}{p_Y(1)} = \frac{0.1}{0.5} = \frac{1}{5}$$

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: $P(a - \frac{\varepsilon}{2} \le Y \le a + \frac{\varepsilon}{2}) = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f_Y(y) dy \approx \varepsilon f(a)$

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: $f_{X|Y}(x\mid y)$

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

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

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

What happens when you add random variables?

Sum of Independent Binomials

- 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 $1 \le i \le N$

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

Sum of Independent Poissons

- 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)

Reference: Sum of Independent RVs

- Let X and Y be independent Binomial RVs
  - $X \sim \text{Bin}(n_1, p)$ and $Y \sim \text{Bin}(n_2, p)$
  - $X + Y \sim \text{Bin}(n_1 + n_2, p)$
  - More generally, let $X_i \sim \text{Bin}(n_i, p)$ for $1 \leq i \leq N$, then

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

- Let X and Y be independent Poisson RVs
  - $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$
  - $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$
  - More generally, let $X_i \sim \text{Poi}(\lambda_i)$ for $1 \leq i \leq N$, then

$$\left(\sum_{i=1}^{N} X_i\right) \sim \text{Poi}\left(\sum_{i=1}^{N} \lambda_i\right)$$

If only it were always that simple

Convolution of Probability Distributions

We talked about sum of Binomial and Poisson…who's missing from this party?

Uniform.

Summation: not just for the 1%

Dance, Dance Convolution

- Let X and Y be independent random variables
  - Cumulative Distribution Function (CDF) of X + Y:

$$F_{X+Y}(a) = P(X+Y \le a)$$

$$= \iint_{x+y \le a} f_X(x) f_Y(y)\,dx\,dy = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{a-y} f_X(x)\,dx\; f_Y(y)\,dy$$

CDF of X

$$= \int_{y=-\infty}^{\infty} F_X(a-y)\, f_Y(y)\,dy$$

PDF of Y

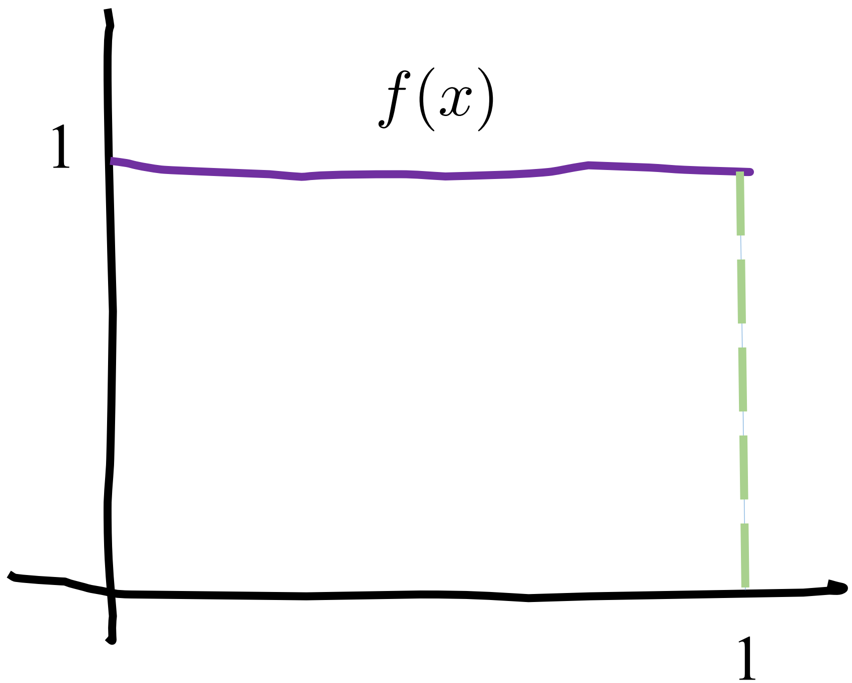
  - F_{X+Y} is called **convolution** of F_X and F_Y
  - Probability Density Function (PDF) of X + Y, analogous:

$$f_{X+Y}(a) = \int_{y=-\infty}^{\infty} f_X(a-y)\, f_Y(y)\,dy$$

  - In discrete case, replace $\int_{y=-\infty}^{\infty}$ with $\sum_y$, and $f(y)$ with $p(y)$

Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \le x \le 1$



For both X and Y

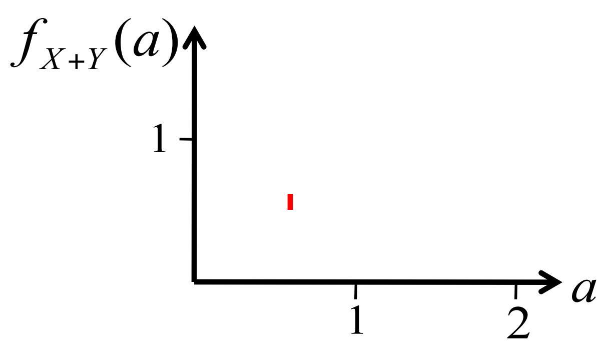
Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \le x \le 1$
  - What is PDF of X + Y?

$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y)\, f_Y(y)\, dy = \int_{y=0}^{1} f_X(a-y)\, dy$$

When $a = 0.5$:

$$\begin{aligned} f_{X+Y}(0.5) &= \int_{y=?}^{y=?} f_X(0.5-y)dy \\ &= \int_{0}^{0.5} f_X(0.5-y)dy \\ &= \int_{0}^{0.5} 1dy \\ &= 0.5 \end{aligned}$$



Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \le x \le 1$
  - What is PDF of X + Y?

$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y)\, f_Y(y)\, dy = \int_{y=0}^{1} f_X(a-y)\, dy$$

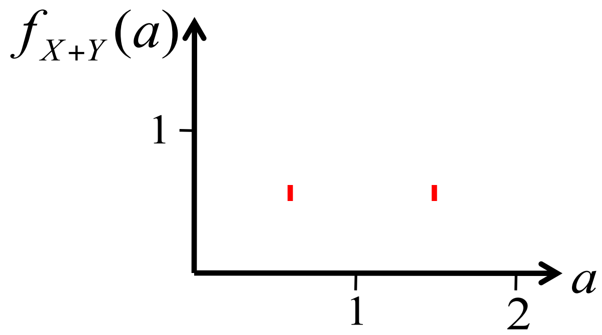
When $a = 1.5$:

$$f_{X+Y}(1.5) = \int_{y=?}^{y=?} f_X(1.5-y)dy$$

$$= \int_{0.5}^{1} f_X(1.5-y)dy$$

$$= \int_{0.5}^{1} 1dy$$

$$= 0.5$$



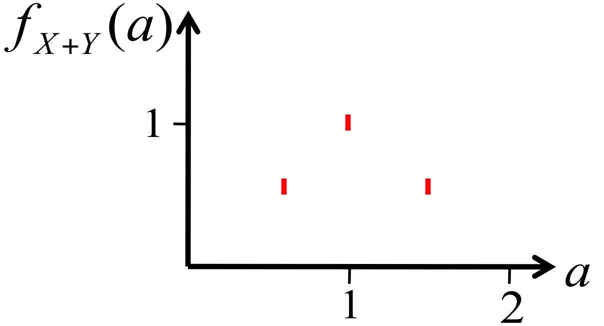
Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \leq x \leq 1$
  - What is PDF of X + Y?

$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y)\, f_Y(y)\, dy = \int_{y=0}^{1} f_X(a-y)\, dy$$

When $a = 1$:

$$\begin{aligned} f_{X+Y}(1) &= \int_{y=?}^{y=?} f_X(1-y)dy \\ &= \int_0^1 f_X(1-y)dy \\ &= \int_0^1 1dy \\ &= 1 \end{aligned}$$



Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \le x \le 1$
  - What is PDF of X + Y?

$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y)\, f_Y(y)\, dy = \int_{y=0}^{1} f_X(a-y)\, dy$$

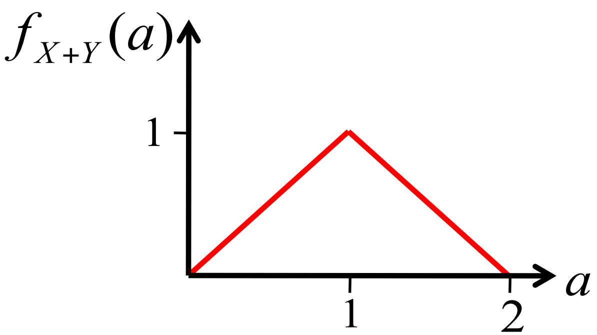
  - When $0 \le a \le 1$ and $0 \le y \le a$, $0 \le a - y \le 1$ → $f_X(a-y) = 1$

$$f_{X+Y}(a) = \int_{y=0}^{a} dy = a$$

  - When $1 \le a \le 2$ and $a-1 \le y \le 1$, $0 \le a - y \le 1$ → $f_X(a-y) = 1$

$$f_{X+Y}(a) = \int_{y=a-1}^{1} dy = 2 - a$$

  - Combining: $f_{X+Y}(a) = \begin{cases} a & 0 \le a \le 1 \\ 2-a & 1 < a \le 2 \\ 0 & \text{otherwise} \end{cases}$



Sum of Independent Normals

- Let X and Y be independent random variables
  - $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$
  - $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

- Generally, have *n* independent random variables $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2, ..., n$:

$$\left(\sum_{i=1}^{n} X_i\right) \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Virus Infections

- Say you are working with the WHO to plan a response to a the initial conditions of a virus:
  - Two exposed groups
  - P1: 50 people, each independently infected with $p = 0.1$
  - P2: 100 people, each independently infected with $p = 0.4$
  - Question: Probability of more than 40 infections?

Sanity check: Should we use the Binomial Sum-of-RVs shortcut?

A. YES!
B. NO!
C. Other/none/more

Virus Infections

- Say you are working with the WHO to plan a response to a the initial conditions of a virus:
  - Two exposed groups
  - P1: 50 people, each independently infected with $p = 0.1$
  - P2: 100 people, each independently infected with $p = 0.4$
  - A = # infected in P1 $A \sim \text{Bin}(50, 0.1) \approx X \sim N(5, 4.5)$
  - B = # infected in P2 $B \sim \text{Bin}(100, 0.4) \approx Y \sim N(40, 24)$
  - What is P(≥ 40 people infected)?
  - $P(A + B \geq 40) \approx P(X + Y \geq 39.5)$
  - $X + Y = W \sim N(5 + 40 = 45, 4.5 + 24 = 28.5)$

$$P(W \geq 39.5) = P\left(\frac{W-45}{\sqrt{28.5}} > \frac{39.5-45}{\sqrt{28.5}}\right) = 1 - \Phi(-1.03) \approx 0.8485$$

End sum of independent vars

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*)?

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

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

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

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

Course Mean

$$E[\mathrm{CS109}]$$

This is actual midpoint of course
(Just wanted you to know)