Discussion 1 Notes

3/28/2016

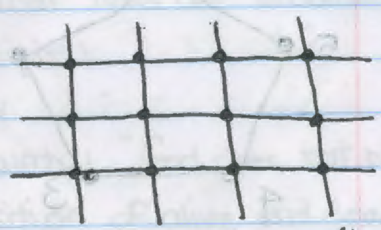
TA: Spencer Frei
Prof: Ying Nian Wu

Applications of Probability:
- Statistics
- Machine Learning
- Finance
- Probability for its own sake
- Physics

<u>Percolation</u>
Given some porous rocks, determine if water can flow completely through it.

Recall that a graph G is a collection of vertices and edges, denoted by $G=(V,E)$

Consider the lattice graph $\mathbb{Z}^2$



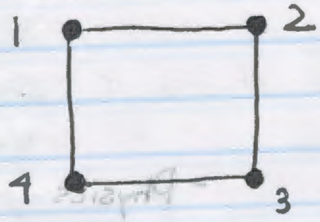
Each edge is "open" with probability p, or "closed" with probability $1-p$

Percolation occurs if there is an infinite path of open connected edges.

<u>Random Walks</u>
- Google's web crawler

Consider a random walk on a graph $G=(V,E)$.
- It is considered a simple random walk if it walks to each neighbor of a point in the walk with equal probability.

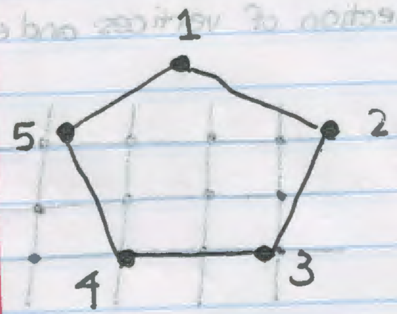
Given this square



Starting from 1, can we reach 1 in an even number of steps? How about 4?

For 1: Yes! Think of the parity of ~~the~~ bipartite graph.
For 4: No!

Answer the same question, for this graph



For 1: Yes!
For 4: Also yes!

Answer this question for any cycle with n vertices. For what values of n is this possible?

For odd values of n. Think: for an even n we will always switch parity. However, for an odd n we can change our parity be selecting which direction to go in.

Lecture 1 3/29/2016

Professor's Contact Info:
Email: ywu@stat.ucla.edu
Office Hours: TR 3:30 - 4:30 PM MS 8971

Attendance (5%)
Weekly Homework Assignments (35%)
Midterm (20%)
Final Exam (40%)

Midterm tentatively listed on Apr 28, 2016

Why do we need probability theory?

At the most fundamental level, physical laws are probabilistic.

- Heisenberg uncertainty principle
- Bohr: wave particle duality
- Einstein was against uncertainty: "God does not play dice"
- Schrödinger's cat: superposition of alive and dead
- Born: $|\text{wave function}|^2$ = probability density function
- Two slit experiment → where are we likely to find an electron?

Even if physical laws are deterministic, we still need statistics to study the motion of many elements

- Ideal gas properties: pressure, temperature, etc.
- Ferromagnetism
- Arrow of time: the ever increasing nature of entropy.
- Brownian motion: dust particles in water shows zig-zag
- Einstein: caused by bombardments from invisible water molecules
- Finance, Black-Scholes equation

Compression and error correction coding
- Images/videos seen on the internet are compressed
- Huffman encoding
- Bad signals are corrected based on a probability model.

ex: do we correct to B or D?

Speech, Vision, Language, Reasoning, Learning, AI
- Training done by probability distribution
- Generalize to testing data.

Google page-rank, random surfing model
- 85% chance that we will click on a link
- 15% that we go to a random web page
- Malkov chain → stationary distribution.

Monte Carlo Algorithm
Metropolis Algorithm

Lecture 2 3/31/16

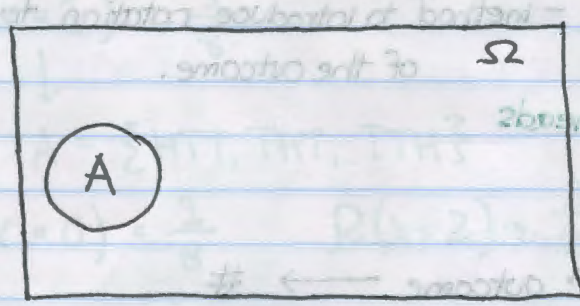
Basic Concepts

Experiment (phenomenon) $\rightarrow$ outcome (random)

Sample space: the set of all possible outcomes (represented by Ω or S)

Events: subsets of the sample space Ω
(represented by A, B, C, and sometimes E, F)
$A \subset \Omega$

Let's use a Venn Diagram representation



Probability is dependent on events
(represented by $P(A)$, or unconventionally $Pr(A)$ or $Prob(A)$)

Example 1:

Consider flipping a fair coin.

$\Omega = \{H, T\}$ Events: $\{H\}$ $\{T\}$ $\{H, T\}$ $\emptyset$

$P(\{H\}) = P(H) = 1/2$
$P(\{T\}) = P(T) = 1/2$
$P(\Omega) = 1$
$P(\emptyset) = 0$

Example 2

Flip a fair coin twice independently.

$\Omega = \{HH, HT, TH, TT\}$

A = 1 head, 1 tail → statement (logical)

$= \{HT, TH\}$ → set (theoretical)

$P(A) = 1/2$

Random variables - method to introduce notation for properties of the outcome.

X = # of heads

$A \Rightarrow X = 1$

Experiment → outcome → #

ω $\quad$ $x(\omega)$

Ω	X
HH	2
HT	1
TH	1
TT	0

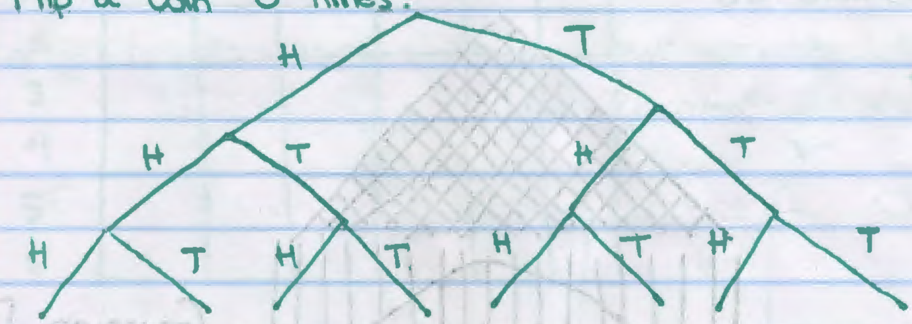
Thus, to formalize the A we call it

$$A = \{\omega : X(\omega) = 1\}$$

$$\begin{aligned} P(A) &= P(\{\omega : X(\omega) = 1\}) \\ &= P(X=1) \\ &= \frac{1}{2} \end{aligned}$$

Example 3

Flip a coin 3 times.



$\Omega = \{$ 8 sequences, as shown by this chart $\}$

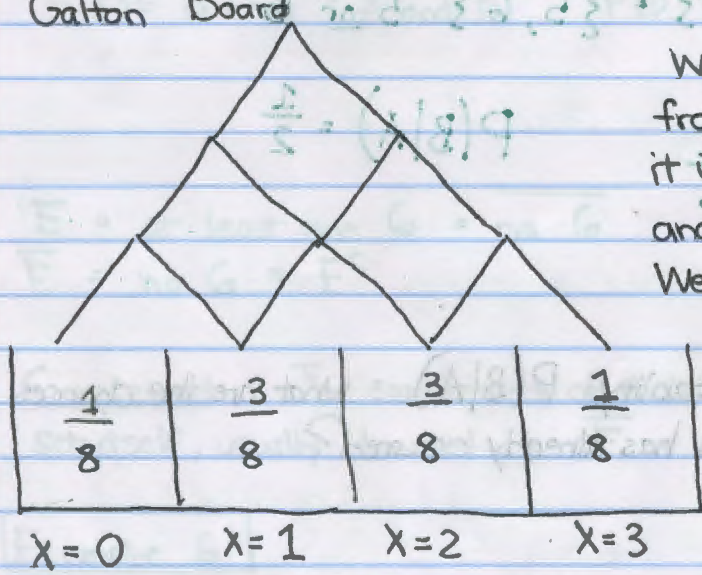
$P(X=1) = \frac{3}{8}$

$\downarrow$

$A = \{HTT, THT, TTH\}$

$P(X=0) = \frac{1}{8}$ $\quad P(X=2) = \frac{3}{8}$ $\quad P(X=3) = \frac{1}{8}$

Galton Board

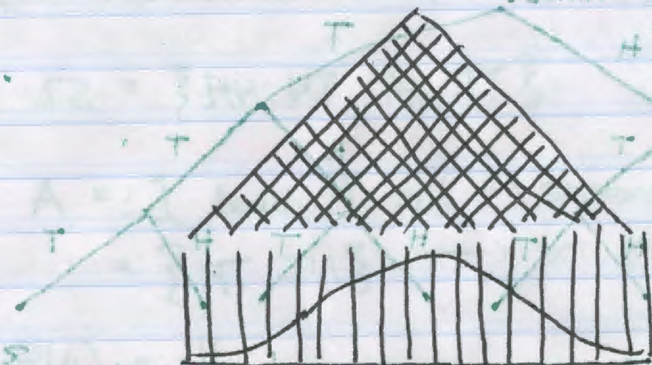


We drop a ball starting from the top. At each level it will go either left or right and eventually land in a bucket. We will see that each bucket has a possible chance, but the chance is not necessarily identical for each bucket.

X = # of right turns (analogous to the number of right turns)

Notice that the layers of the Galton Board identically represents Pascal's Triangle

Note that for n flips we have 2^n possibilities!



Gaussian Distribution

We treat a random walk as a random traversal of a graph like this.

Example 4:

Roll a fair die.
Let X be the numerical result

$A = \{X > 4\} = \{5, 6\}$ 5 or 6

$P(A) = \frac{1}{3}$

$P(X = 6) = \frac{1}{6}$
$\downarrow$
B

$P(B|A) = \frac{1}{2}$

Conditional probability $P(B|A)$ = what are the chances of B given that A has already occurred?

Example 5:

Roll a fair die twice independently
Let X be the 1st numerical result
Let Y be the 2nd numerical result

X \ Y	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

36 total outcomes

$$P(X+Y>10) = P(\{(5,6),(6,5),(6,6)\}) = \frac{1}{12}$$

Let $C = \max(X,Y) > 4$

$A = X > 4$

$Y > 4 = B$

$C = A$ or $B = A \cup B$

$\cancel{P(C) = A \cup B - A \cap B}$

$P(C) = P(A) + P(B) - P(A \cap B)$

$= \frac{1}{3} + \frac{1}{3} - \frac{1}{9} = \frac{5}{9}$

$D = \min(X,Y) > 4$

$D = A \cap B = A$ and B

$P(D) = P(A) \cdot P(B)$

$= \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$

E = at least one 6 = $\overline{\text{no } 6}$

$\overline{E}$ = no 6 = E^c

<u>Complement</u> – The complement of a set is the opposite of the set itself, usually denoted as $\overline{E}$ or E^c

$\boxed{\text{Example 6:}}$

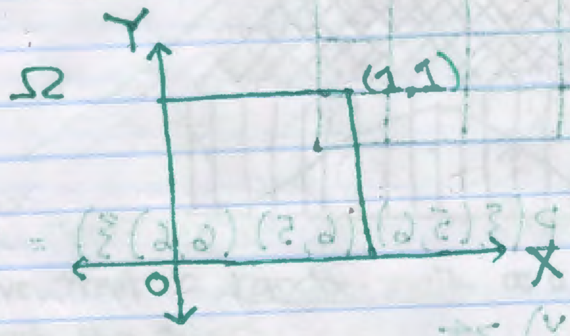
Random Number Generator

In a computer, we generate a number from 0 to 1.

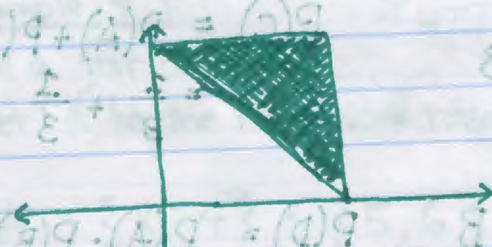
Let X be this number.

$P\left(X > \frac{1}{2}\right) = \frac{1}{2}$, $P\left(X > \frac{2}{3}\right) = \frac{1}{3}$

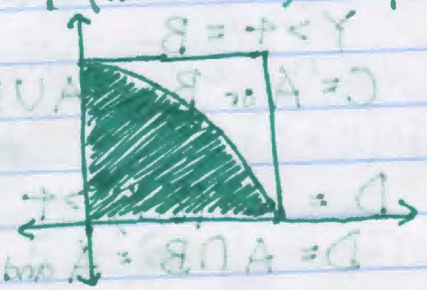
Let's generate two numbers independently using this RNG.
We call them (X, Y)



$P(X+Y \geq 1) = \frac{1}{2}$



$P(X^2+Y^2<1) = \frac{\pi}{4}$



Discussion 2 4/4/2016

Sample Spaces

① Deck of cards (52)

a) Drawing one card?

Ω = Set of size 52 that consists of each individual playing card.

$= \{\heartsuit, \spadesuit, \clubsuit, \diamondsuit\} \times \{2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A\}$

b) Drawing n cards, with replacement?

Ω = set of size 52^n cards that consists of every possible combination of drawings for each of the n cards?

$= \Omega^n = \underbrace{\Omega \times \Omega \ldots\ldots \times \Omega}_{n \text{ times}}$

$\{(\omega_1^1, \omega_2^1), (\omega_1^2, \omega_2^2) \ldots\ldots\}$

$\omega_1^i \in \{\diamondsuit, \heartsuit, \clubsuit, \spadesuit\}$

$\omega_2^i \in \{2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A\}$

c) Draw cards (with replacement) until we get a spade? → specifically, an ace of spades?

$\Omega^{\mathbb{N}} = (x_1, x_2, x_3, x_4, \ldots\ldots)$

$x_i \in \Omega$

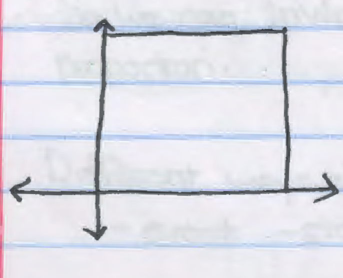
This is an infinitely large sample space.

Probability

② a) Probability of getting a black number ≤ 3? 1/13 if Ace is not less

b) Probability of getting a red face card (including Ace)? 2/13

③ $X \sim$ uniform $(0, 1)$



$P(x \leq t) = 0$ if $t \leq 0$

$P(x \leq t) = 1$ if $t \geq 1$

$P(x \leq t) = t$ if $0 \leq t \leq 1$

$P(X = a) = 0$

$\{X \leq t\} = \{X = t\} \cup \{x < t\}$

a) If $a > 0$, $b \in \mathbb{R}$

What do you think the distribution of $Y = a \cdot X + b$ is?

Let's look at what happens when $X = 0$ and when $X = 1$

If $a = 1$, then $x+b \sim \text{uniform}(b, 1+b)$

$b = 0$: then $aX \sim \text{uniform}(0, a)$

$aX + b \sim \text{uniform}(b, a+b)$

b) What if $a < 0$?

$aX + b \sim \text{uniform}(b+a, b)$

Lecture 3 4/5/2016

Experiment $\longrightarrow$ Outcome $\longrightarrow$ number(s)

$\omega \in \Omega$ random variables (functions/mappings)

$\downarrow$ $X(\omega)$

sample space

Prototype I: Sampling a population (e.g. U.S. population)

Ω	height (X)	weight (Y)	gender (Z)
1	$X(1)$	$Y(1)$	$Z(1)$
2	$X(2)$	$Y(2)$	$Z(2)$
3	$X(3)$	$Y(3)$	$Z(3)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
ω	$X(\omega)$	$Y(\omega)$	$Z(\omega)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
M	$X(M)$	$Y(M)$	$Z(M)$

M = 300,000,000 . the U.S. population

$$\underset{A}{P(\text{person is male})} = \frac{|A|}{|\Omega|} = \frac{\#\text{ of males}}{\text{total population}}$$

$|\Omega| = M$

So we can trivially see that probability is just a population proportion.

Different ways of looking at A:

- event - statement about outcome - subset (subpopulation) event $\quad$ random variable

$$A = \{\omega : Z(\omega) = \text{male}\}$$

This is the mathematical representation of the event.

If we define Z as a random variable then we can write this as

$$Z = \text{male}$$

and so...

$$P(A) = P(Z = \text{male})$$

* recall that $|X|$ is the cardinality, or size, of the set X.

Another event...

Let $B = \{\omega : X(\omega) > 6 \text{ ft}\}$

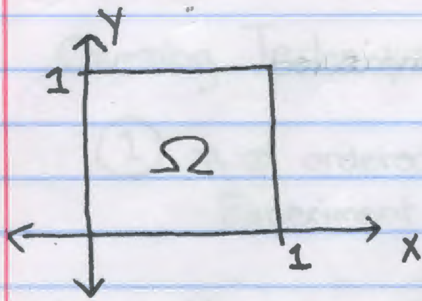
$$P(B) = P(X > 6) = \frac{|B|}{|\Omega|}$$

Recall from the previous lecture that $P(B|A)$ represents the probability that event B happens given that A has already occurred

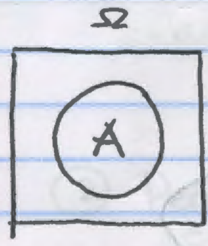
Thus, $P(B|A)$ = proportion of tall males in our example.

$$= \frac{|B \cap A|}{|A|}$$

$\boxed{\text{Prototype II}}$ randomly throw a point into a unit square

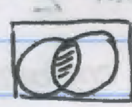
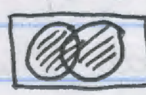


$$\omega \in \Omega \;\to\; (X(\omega), Y(\omega)) \;\to\; (x, y)$$



$$P(A) = P(\text{falls into } A)$$
$$= P((X(\omega), Y(\omega)) \in A) = P((X,Y) \in A)$$
$$= \frac{|A|}{|\Omega|} = \frac{\text{area of } A}{\text{area of } \Omega} = |A|$$

Relations

logical	AND	OR	NOT
set-theoretical	$A \cap B$	$A \cup B$	A^c $(\bar{A})$
venn diagram			

Probability is a size (measure).

Axiom 1: $P(\Omega) = 1$

2: $P(A) \geq 0$

3: If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$ additivity

Infinite additivity → if we have an infinite number of subsets

$$A_1, A_2, A_3 \ldots\ldots A_i \ldots\ldots$$
$$A_i \cap A_j = \emptyset$$
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Enhanced Axiom 3

Not all $A \subset \Omega$ can be measured.

Consequences of the axioms

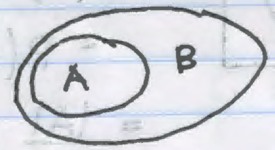
(1) $P(A^c) = 1 - P(A)$

Proof:

$$A^c \cap A = \emptyset$$
$$P(A^c \cup A) = P(\Omega)$$
$$P(A^c) + P(A) = P(\Omega)$$
$$P(A^c) = 1 - P(A)$$

(2) If $A \subset B$

$$P(A) \leq P(B)$$



In B but out of $A = B \cap A^c$

Proof:

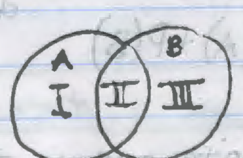
$$P(B) = P(A \cup (B \cap A^c))$$
$$= P(A) + \underbrace{P(B \cap A^c)}_{\geq 0 \text{ by Axiom 2}}$$

so $P(B) \geq P(A)$

(3) General form of Axiom 3, the <u>Inclusive/Exclusive Formula</u>

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Proof: (loosely)



$$P(A) = P(\mathrm{I}) + P(\mathrm{II})$$
$$P(B) = P(\mathrm{III}) + P(\mathrm{II})$$
$$P(A \cup B) = P(\mathrm{I}) + P(\mathrm{II}) + P(\mathrm{III})$$
$$P(\mathrm{II}) = P(A \cap B)$$

It's clear to see why the general form is what it is.

Counting Techniques

(1) # of ordered pairs $= n_1 \times n_2$

Experiment 1 $\rightarrow n_1$ outcomes

2 $\rightarrow n_2$ outcomes

Notice that for k experiments the number of k-tuples is simply $\prod_{i=1}^{k} n_i$

(2) Permutations

Suppose we have n cards and sequentially pick k cards without replacement. How many possible sequences can we get? Order matters.

$$n(n-1)(n-2)(n-3)\cdots(n-k+1) = P_{n,k}$$

Notice that if $n=k$ then $P_{n,k} = n!$ which is the number of ways to ~~include~~ order n cards.

(3) Combinations (will be covered next time!)

4/7/2016 Lecture 4

Recall from last lecture...

$\boxed{\text{Permutations}}$

- Order ~~maste~~ matters!
- $P(n,k)$ = Given n items, pick k items w/o replacement
  $= n(n-1)(n-2)\ldots(n-k+1)$ times.

$$\begin{array}{cccc} \uparrow & \uparrow & \uparrow & \uparrow \\ \text{1st} & \text{2nd} & \text{3rd} & \text{kth} \end{array}$$

$\llcorner P_{n,n} = n! = n \times n-1 \times n-2 \times \ldots \times 1$

We can visualize it this way

$$P_{n,k} = \frac{n \times (n-1) \times (n-2) \times \ldots \times (n-k+1) \times (n-k) \times \ldots \times 1}{(n-k) \times (n-k-1) \times \ldots \times 1}$$

$$= \frac{n!}{(n-k)!}$$

Birthday Problem

n people
a random person has an equal chance of having each of the 365 birthdays $\{1, 2, 3 \ldots 365\}$

Prob (At least 2 people have the same birthday) = Prob of A
Prob (All people have different birthdays) = Prob of A^c

$$P(A^c) = \frac{P_{365,n}}{365^n} = \frac{|A^c|}{365^n} = \frac{|A^c|}{|\Omega|} = \frac{365 \times 364 \times \ldots \times 365-n+1}{365 \times 365 \times \ldots \times 365}$$

As it turns out, we only need $n = 24$ in order for $P(A) \approx 1/2$.

$\boxed{\text{Combinations}}$

- Order does NOT matter!

We have a box of n balls, $\{1, 2, 3 \ldots\ldots n\}$

We pick k of them without replacement

\# of combinations $C_{n,k}$ $\binom{n}{k}$

$$C_{n,k} = \frac{n!}{(n-k)!\,k!} = \frac{P_{n,k}}{k!} = \frac{P_{n,k}}{P_{k,k}}$$

Why do we divide by k!? This is because for every combination of k items there are k! ways to permute them, but we only want to ~~call them~~ count one time!

Coin Flipping

Flip a fair coin n times

$$\text{prob (exactly } k \text{ heads)} = \frac{\binom{n}{k}}{2^n} = \frac{|A|}{|\Omega|}$$

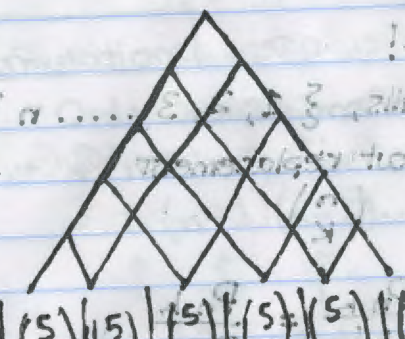
$A = \{\text{sequences with exactly } k \text{ heads}\}$

Why is the numerator what it is?

Well, we can assign a labeling to each of the n flips, and then pick k of these to be our heads. The order that we pick the flips in doesn't matter so we opt to use combinations instead of permutations.

How do we map this to the Galton Board??

(HTTHT) $\rightarrow$ (RLLRL)



$\binom{5}{0}$ $\binom{5}{1}$ $\binom{5}{2}$ $\binom{5}{3}$ $\binom{5}{4}$ $\binom{5}{5}$ # of paths that end in each bucket

0 1 2 3 4 5

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1} \qquad \binom{n}{k} = \binom{n}{n-k}$$

Given R red balls, B blue balls, $N = R+B$

pick n balls

$P(r \text{ red}, b \text{ blue}, n = r+b)$ $\quad r \leq R$, $b < B$, $n < N$

$$= \frac{|A|}{|\Omega|} = \frac{\binom{R}{r} \times \binom{B}{b}}{\binom{N}{n}}$$

Hypergeometric probability/distribution

Survey sampling N big

sample $n = 1{,}000$

$p = \frac{R}{N}$ = population proportion

$$P(r \text{ red}) = \frac{\frac{R(R-1)\ldots(R-r+1)}{r!} \; \frac{B(B-1)\ldots(B-b+1)}{b!}}{\frac{N \times (N-1) \times \ldots \times (N-n+1)}{n!}}$$

$$\binom{n}{r} p^r (1-p)^{n-r}$$

. This is known as the binomial probability formula

p = probability of event

r = # items selected

n = total # of items