Probability

What is probability?

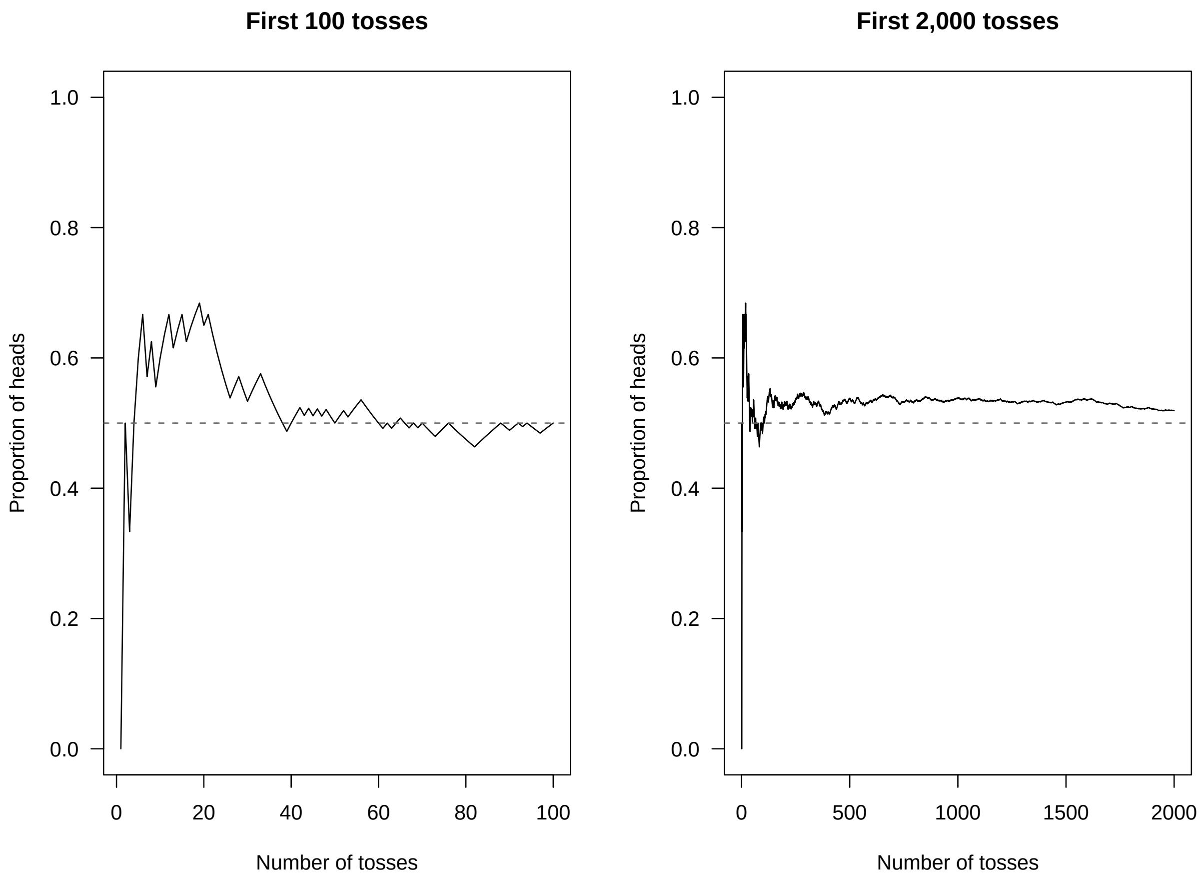
A branch of mathematics concerning the study of random processes

Random does not mean haphazard.

What do I mean when I say the following?

“The probability that he is a carrier...”
“The chance of rain tomorrow...”

- Long term frequency
- Degree of belief



Why study probability?

Probability is (in a sense) the inverse of statistics.

— Deduction vs induction.

Statistics: Given these particular data (the *realization* of a random process), what can we say about the nature of the process that generated them?

Probability: Given a complete specification of the nature of a random process, what is the chance of observing a particular result?

Mathematics

Why is mathematics useful?

- Abstraction

Why is mathematics hard?

- Abstraction
- Notation
- Lots of rules
- Culture

Mathematics is supposed to make things easier.

The set-up

	Example
Random "experiment" A well-defined process with an uncertain outcome.	Draw 2 balls *with replacement* from an urn containing 4 red and 6 blue balls.
Sample space, $\mathcal{S}$ The set of possible outcomes.	$\{$ RR, RB, BR, BB $\}$
Event A set of outcomes (a subset of $\mathcal{S}$).	$\{$the first ball is red$\}$ = $\{$RR, RB$\}$

Events are said to occur if one of the outcomes they contain occurs. Probabilities are assigned to events.

Probability rules

- $1 \geq \Pr(A) \geq 0$ for any event A
- $\Pr(\mathcal{S}) = 1$ (where $\mathcal{S}$ is the sample space).
- If A and B are *mutually exclusive*,
  Pr(A or B) = Pr(A) + Pr(B).

Complement rule: Pr(not A) = 1 - Pr(A)

Example

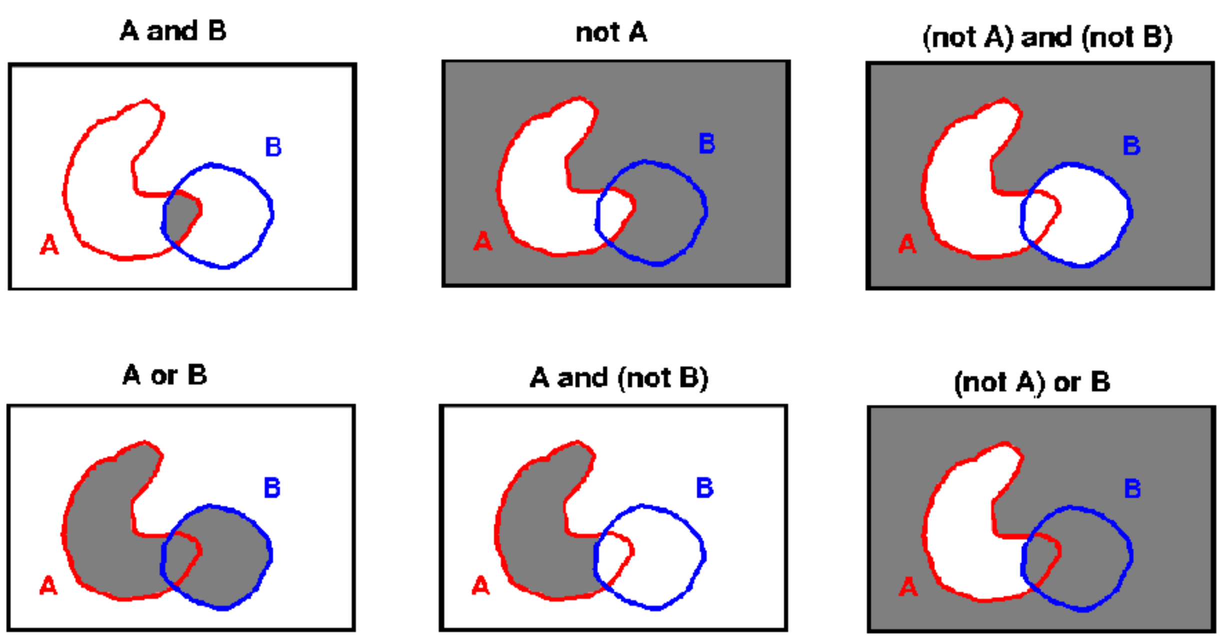
Cage with 10 rats:

- 2 infected with virus X (only)
- 1 infected with virus Y (only)
- 5 infected with both X and Y
- 2 infected with neither

Experiment: Draw one rat at random (each equally likely).

Events:	A = {rat is infected with X}	Pr(A) = 7/10
	B = {rat is infected with Y}	Pr(B) = 6/10
	C = {rat is infected with only X}	Pr(C) = 2/10

Sets

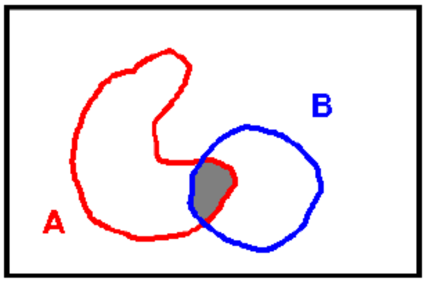


Conditional probability

$$\Pr(A \mid B) = \text{“Probability of A given B”}$$
$$= \Pr(A \text{ and } B) / \Pr(B)$$

Rat example:

[2 w/ X only, 1 w/ Y only; 5 w/ both; 2 w/ neither]



A = {infected with X}

B = {infected with Y}

$\Pr(A \mid B) = (5/10) / (6/10) = 5/6$

$\Pr(B \mid A) = (5/10) / (7/10) = 5/7$

Multiplication rule:

$$\Pr(A \text{ and } B) = \Pr(A) \times \Pr(B \mid A)$$

Independence:

A and B are said to be independent if:

$$\Pr(A \text{ and } B) = \Pr(A) \times \Pr(B)$$

Note: If A and B are independent, then

$$\Pr(A \mid B) = \Pr(A)$$
$$\Pr(B \mid A) = \Pr(B)$$

Examples

1. **Draw two cards (without replacement) from a well-shuffled deck. Calculate the following.**

   (a) **Pr(2nd card is an ace)**

   Pr(2nd card is an ace) = Pr(1st card is an ace) = 4/52 = 1/13 $\approx 7.7\%$.

   (b) **Pr(2nd card is an ace | 1st card is an ace)**

   Pr(2nd card is an ace | 1st card is an ace) = 3/51 $\approx 5.9\%$.

   (c) **Pr(2nd card is an ace | 1st card is not an ace)**

   Pr(2nd card is an ace | 1st card is not an ace) = 4/51 $\approx 7.8\%$.

   (d) **Pr(both cards are aces)**

   Pr(both cards are aces) = Pr(1st card is ace) $\times$ Pr(2nd card is ace | 1st card is ace)
   = (4/52) $\times$ (3/51) $\approx 4.5/1000$.

   (e) **Pr(2nd card is an ace | 1st card is a heart)**

   Pr(2nd card is an ace | 1st card is a heart) = 4/52 = 1/13 $\approx$ 7.7%
   (These are independent.)

   To spell it out completely:
   Pr(2nd card is an ace | 1st card is a heart)
   = Pr(1st card is a heart and 2nd card is an ace) / Pr(1st card is a heart)
   = [Pr(1st card is ace of hearts and 2nd card is an ace)
   + Pr(1st card is heart but not ace and 2nd card is an ace)]
   / Pr(1st card is a heart)
   = [(1/52) $\times$ (3/51) + (12/52) $\times$ (4/51)]/(1/4)
   = 1/13 $\approx$ 7.7%.

   (f) **Pr(get a pair)** (i.e., the two cards have the same face).

   Pr(get a pair) = Pr(AA or 22 or 33 or 44 or . . . or QQ or KK)
   = Pr(AA) + Pr(22) + Pr(33) + . . . Pr(KK)
   = 13 [(4/52) $\times$ (3/51)]
   = 3/51 $\approx$ 5.9%.

   Alternatively, Pr(get a pair) = Pr(2nd card has same face as first card) = 3/51 $\approx$ 5.9%.

2. **The seeds in Mendel's pea plants were either smooth or wrinkled, the result of a single gene with the smooth allele (*S*) dominant to the wrinkled allele (*s*). Similarly, the seeds were either yellow or green, with yellow (*Y*) dominant to green (*y*).**

   The F_1 hybrid seeds produced from crossing two pure-breeding lines (one with smooth, yellow seeds; the other with wrinkled, green seeds) are all smooth and yellow.

   Consider growing up an F_1 seed, selfing it, and the taking a random F_2 seed.

   (a) **Pr(F_2 seed is smooth)**

   Pr(F_2 seed is smooth) = 1 – Pr(F_2 seed is wrinkled) = 1 - (1/2) $\times$(1/2) = 3/4 = 75%.

   (b) **Pr(F_2 seed is green)**

   Pr(F_2 seed is green) = 1/4 = 25%.

   (c) **Pr(F_2 seed is smooth and green)**

   Pr(F_2 seed is smooth and green) = (3/4) $\times$ (1/4) = 3/16 $\approx$ 19%.
   [Since the two genes are unlinked, "the F_2 seed is smooth" and "the F_2 seed is green" are independent.]

   (d) **Pr(F_2 seed has genotype SS)**

   Pr(F_2 seed has genotype SS) = 1/4 = 25%.

   (e) **Pr(F_2 seed has genotype SS | it is smooth)**

   Pr(F_2 seed has genotype SS | it is smooth)
   = Pr(F_2 seed has genotype SS and is smooth) / Pr(F_2 seed is smooth)
   = (1/4) / (3/4) = 1/3 = 33%.

3. **Consider an urn with 3 red balls and 2 blue balls.**

   (a) **Draw 2 balls *with replacement*.**

   i. **Pr(both balls are red)**

   Pr(both balls are red) = (3/5) $\times$ (3/5) = 9/25 = 36%.

   ii. **Pr(the two balls are the same color)**

   Pr(the balls are the same color) = Pr(both balls are red) + Pr(both balls are blue)
   = (3/5) $\times$ (3/5) + (2/5) $\times$ (2/5) = 9/25 + 4/25
   = 13/25 = 52%

   (b) **Repeat the above when the draws are *without replacement*.**

   i. **Pr(both balls are red)**

   Pr(both balls are red) = (3/5) $\times$ (2/4) = 6/20 = 30%.

   ii. **Pr(the two balls are the same color)**

   Pr(the balls are the same color) = Pr(both balls are red) + Pr(both balls are blue)
   = (3/5) $\times$ (2/4) + (2/5) $\times$ (1/4) = 6/20 + 2/20
   = 8/20 = 40%

4. **Roll two fair, six-sided dice (or roll one fair, six-sided dice twice).**

   (a) **Pr(both dice are 1)**

   Pr(both are 1) = Pr(1st die is 1 and 2nd die is 1)
   = Pr(1st die is 1) $\times$ Pr(2nd die is 1) (*They are independent.*)
   = (1/6) $\times$ (1/6) = 1/36 $\approx$ 2.8%.

   (b) **Pr(at least one die is a 1)**
   We consider three different approaches to calculating this probability.

   i. There are 36 different possible outcomes and they are equally likely. 11 of these outcomes satisfy the condition, "at least one die is a 1." Thus the probability is 11/36 $\approx$ 31%.

   ii. Let A = $\{$ 1st die is a 1 $\}$ and B = $\{$ 2nd die is a 1 $\}$.

   Pr(A or B) = Pr(A) + Pr(B) – Pr(A and B)
   = 1/6 + 1/6 – 1/36
   = 11/36 $\approx$ 31%.

   iii. Pr(at least one 1) = 1 – Pr(no 1's)
   = 1 – (5/6) $\times$ (5/6) = 1 – 25/36
   = 11/36 $\approx$ 31%.

5. (A sort of famous 17th century French gambling problem.)
   Which of the following is more likely "at least one 1 in 4 rolls of a six-sided die" or "at least one pair of 1's (snake-eyes) in 24 rolls of a pair of six-sided dice?" Or are they equally likely?

   (a) **Pr(at least one 1 in 4 rolls of a 6-sided die)**

   Pr(at least one 1 in 4 rolls of a 6-sided die) = 1 – Pr(no 1's in 4 rolls)
   = 1 – Pr(1st is not 1) $\times$ Pr(2nd is not 1) $\times$ Pr(3rd is not 1) $\times$ Pr(4th is not 1)
   = 1 – (5/6) $\times$ (5/6) $\times$ (5/6) $\times$ (5/6)
   = 1 – $(5/6)^4$ $\approx$ 52%.

   (b) **Pr(at least one pair of 1's in 24 rolls of a pair of 6-sided dice)**

   Pr(at least one pair of 1's in 24 rolls of a pair of 6-sided dice)
   = 1 – Pr(no pairs of 1's in 24 rools of a pair of dice)
   = 1 – $\{$Pr(1st roll is not a pair of 1's)$\}^{24}$
   = 1 – $(35/36)^{24}$ $\approx$ 49%.