Probability

The setup

Random "experiment": A well-defined process with an uncertain outcome (for example, toss three fair coins)

Sample space, $\mathcal{S}$: The set of possible outcomes (e.g., $\{$ HHH, HHT, HTH, THH, HTT, THT, TTH, TTT $\}$)

Event: A set of outcomes (a subset of $\mathcal{S}$) (e.g., A = $\{$exactly one head$\}$ = $\{$ HTT, THT, TTH $\}$)

An event is said to have occurred if one of the outcome it contains occurs.

Basic rules of probability

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

More rules

Pr(not A) = 1 – Pr(A)

Pr(A or B) = Pr(A) + Pr(B) – Pr(A and B)

Conditional probability

Pr(A | B) = "probability of A given B" = Pr(A and B) / Pr(B), provided Pr(B) > 0.

Independence

Events A and B are *independent* if Pr(A and B) = Pr(A) Pr(B); equivalently, Pr(A | B) = Pr(A) and Pr(B | A) = Pr(B).

Still more rules

Pr(A and B) = Pr(B) Pr(A | B) = Pr(A) Pr(B | A)

Pr(A) = Pr(A and B) + Pr(A and not B)
= Pr(B) Pr(A | B) + Pr(not B) Pr(A | not B)

Bayes rule

Pr(A | B) = Pr(A) Pr(B | A) / Pr(B)
= Pr(A) Pr(B | A) / [Pr(A) Pr(B | A) + Pr(not A) Pr(B | not A)]

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%.