Probability Review

Basic Concepts

Probability

Sample space Ω – a set of possible outcomes;

e.g. dice = {1, 2, 3, 4, 5, 6}

Event – a subset of sample space; e.g., small numbers = {1, 2, 3}

Probability of an event E, $P(E) = |E| / |S|$

Sample space S = { ♠, ♥, ♦, ♣ } x {A, 2, 3, 4, 5, 6 7, 8, 9, T, J, Q, K}

Event = { ♠9 }; *P*({♠9}) = | {♠9} | / |S| = 1 / 52

Event = a card of heart

= { ♥A, ♥1, ♥2, ♥3, ♥4, ♥5, ♥6, ♥7, ♥8, ♥9, ♥T, ♥J, ♥Q, ♥K }

= *P*(a card of heart) = 13 / 52

Event = a red card ; *P*(a red card) = 26 / 52

Probability

Example: Probability when the sum of 2 rolled fair dice is 6

Solution1:

$|S|$ = $6^2 = 36$ outcomes

E = {(1,5), (2,4), (3,3), (4,2), (5,1)}

$P(E)$ = $|E| / |S|$ = 5/36

Solution 1 is correct because
all its outcomes are equally likely

Solution2:

S be the set of 2-combinations with repetition,

$|S|$ = $C(6+2-1,2) = 21$

E = {{1,5}, {2,4}, {3,3}}

So $P(E)$ = $|E| / |S|$ = 3/21 = 1/7

$P(E) = |E| / |S|$ with the assumption
that all outcomes are equally likely. (correct?)

Formal Definition

1. Sample space Ω: possible outcomes.
   Flipping a coin: $\Omega = \{ H, T \}$

2. Collection $\mathcal{F}$ of events. An event is a subset of Ω.
   (i) $\emptyset \in \mathcal{F}$
   (ii) if $A \in \mathcal{F}$, then $(\Omega \setminus A) \in \mathcal{F}$
   (iii) if $A_i \in \mathcal{F}$ for each positive integer i, then $\cup_i A_i \in \mathcal{F}$

   $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \{H,T\} \}$.
   $\emptyset$ is an impossible event, e.g., getting a "6"

3. Probability function $Pr : \mathcal{F} \rightarrow [0,1]$
   (i) $Pr(\Omega) = 1$
   (ii) For countable number of pairwise disjoint events A_i,
   $$Pr(\cup_i A_i) = \sum_i Pr(A_i).$$

Probability on Cards

Example 1: Probability (a 5-card hand without a spade ace)

$|S| = C(52,5) = 52 \times 51 \times 50 \times 49 \times 48 / 5!$

$|E_1| = C(51,5) = 51 \times 50 \times 49 \times 48 \times 47 / 5!$

$P(E_1) = |E_1| / |S| = 47 / 52$

Are all "combinations" equally probable? Why combinations?

"Yes" - because all combinations can be obtained in 5! ways.

Another approach

Probability of choosing the first card not ♠A = 51/52

Probabilities of the 2nd, 3rd, 4th and 5th card not ♠A
= 50/51, 49/50, 48/49, 47/48

Probability that all 5 cards are not ♠A
= (51/52) (50/51) (49/50) (48/49) (47/48) = 47 / 52

Probability on cards

Ex 2: Probability (a 5-card hand without any ace)

$|E_2| = C(48,5) = 48 \times 47 \times 46 \times 45 \times 44 / 5!$

$|S| = C(52,5) = 52 \times 51 \times 50 \times 49 \times 48 / 5!$

$P(E_2) = |E_2| / |S| = C(48,5) / C(52,5)$

$= (48/52)\ (47/51)\ (46/50)\ (45/49)\ (44/48) = 0.658842$

Ex 3: Probability (a 5-card hand at least one ace)

$|E_3| = |S| - |E_2|; \quad P(E_3) = 1 - |E_2| / |S| = 1 - P(E_2)$

Ex 4: Probability (a 5-card hand with two pairs)

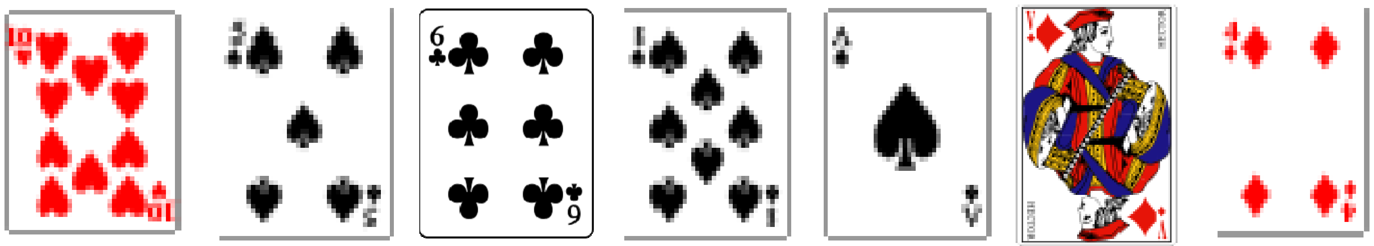
$|E_4| = C(13,2)\ C(4,2)\ C(4,2)\ C(44,1)$

Ex 5: Probability (a 5-card hand with one pair)

$|E_5| = C(13,1)\ C(4,2)\ C(12,3)\ C(4,1)\ C(4,1)\ C(4,1)$

Card Game Puzzles

Consider 7 cards of different face values from a pack of cards (no pairs) as below, **BET** a randomly picked card from the remaining pack will form a pair with one of the 7 cards? YES or NO



Let P = probability that it will form a pair

= $(7 \times 3) / (52 - 7) = 21 / 45$

ANSWER : Bet NO

Card Game Puzzles

Select 6 random cards from a pack of cards, **BET** at least 2 of them form a pair? YES or NO

Let P' = probability that all the 6 cards are different

= (48/51) (44/50) (40/49) (36/48) (32/47)

= 0.345247

P = prob (at least 2 of them form a pair)

= $1 - P' = 0.654753 > 0.5$

ANSWER : Bet YES

Application on Quality Control

Checking a shipment of thousands products is long and expensive.

A solution is to select samples at random and inspect them, instead of checking the whole lot (random sampling).

But how many samples to select? Too few might miss the defective ones.

Suppose that there are two defective pens in a box of 12.

Suppose 3 pens are sampled for testing, what is the probability that a defective pen is not selected?

Sample space = S = all groups of 3 pens

Event = E = all groups of 3 good pens

P(E) = |E| / |S| = C(10,3) / C(12,3) = 120 / 220 = 6/11

That is, more than 50% of the cases that the defective pens are not detected.

What is the sample size if we can have > 90% success rate?

That is, what is the value of x such that C(10,x) / C(12,x) < 10% ?

When x=8, C(10,8) / C(12,8) = 12/132, the sample size should be 8.

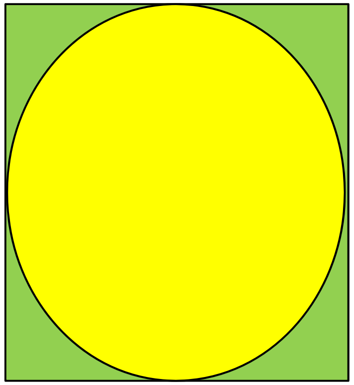
What is the sample size with > 90% success rate for a shipment of 120 pens with 20 defective?

The sample size is 11 as C(100,11) / C(120,11) = 9.963%

Application - probability

Example: Throw randomly 100,000 darts at the target.,
ignore those outside the square,
let C = number of darts inside the circle
T = number of darts inside the square

What is the expected value of 4C/T?



Sample space
– all points inside the square

Event
– all points inside the circle

Probability { a point inside the square and also inside the circle }
= area of the circle / area of the square = π / 4 = C / T.

The Prize is Right

There are 3 doors, a grand prize is behind one of them.
After you make your choice,
the host always open the losing door.
Will you change your mind and switch door?

Will you switch door?

No switch – chance of winning is 1/3

Switch – LOSE if your choice is originally correct;
WIN if your choice is originally wrong.

Since originally the probability of losing is 2/3, by switching door, the probability of winning will be 2/3.

Another view of the solution

S = {100, 010, 001}

WLOG the first door is selected,

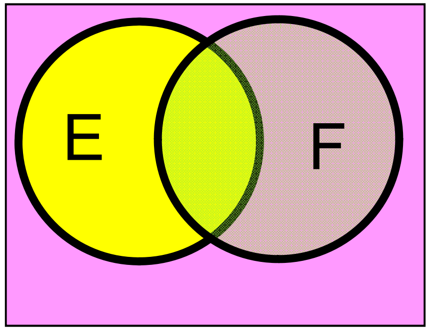
Lose win

S' = S after one door is opened = {10x or 1x0, 01x, 0x1}

Assume 100 doors with one prize door, will you change after opened 98 doors?

Conditional Probabilities

Probability of an event E, $P(E) = |E| / |S|$



$P(E \mid F)$

= conditional probability of E when given F

$= |E \cap F| / |F| = P(E \cap F) / P(F)$

F - sum of 2 dices is 4 = {(1,3),(2,2),(3,1)}

$P(F) = 3/36 = 1/12$

P(*one dice is 2 and sum of 2 dices is 4*) = $P(E \cap F) = 1/36$

P(*one dice is 2 | sum of 2 dices is 4*)

$= P(E \mid F) = P(E \cap F) / P(F) = 1/3$

Another view: There is only one (2,2) in F = {(1,3),(2,2),(3,1)}

P(*one dice is 3 and sum of 2 dices is 4*) = 2/36

P(*one dice is 3 | sum of 2 dices is 4*) = 2/3

P(*one dice is 4 given sum of 2 dices is 4*) = 0

Conditional Probabilities

$P(E \mid F) = |E \cap F| / |F| = P(E \cap F) / P(F)$

$P(E \cap F) = P(E \mid F)\, P(F) = P(F \mid E)\, P(E)$

Example: Prob (sum of 2 dices >9 given one of them is 6)

Wrong Argument:

If one of them is 6, we'll get >9 as long as the other is 4,5 or 6.

So the probability is ½

E = roll of two dices whose sum >9

= {(6,4), (6,5), (6,6), (5,5), (5,6), (4,6)} ;

$|E| = 6$; $P(E) = 1/6$

F = roll of two dices, one of which is a 6

= {(6,1),(6,2),(6,3),(6,4),(6,5),(6,6),(5,6),(4,6),(3,6),(2,6),(1,6)};

$|F| = 11$; $P(F) = 11/36$

$E \cap F$ = {(6,4), (6,5), (6,6), (5,6), (4,6)} ;

$|E \cap F| = 5$, $P(E \cap F) = 5/36$

$P(E \mid F) = 5/11$; $P(F \mid E) = 5/6$

Conditional Probabilities

Example: Given a random bit-string of length 4, what is the probability that it contains two consecutive 0's given that the first bit is 0?

S	=	sample space of all bit-strings of length 4; $\|S\| = 16$
E	=	all bit-strings with two consecutive 0's
F	=	all bit-strings with the first bit 0;
	=	{0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111};
$P(F)$	=	$\|F\| / \|S\| = 8/16$
$E\cap F$	=	{0000, 0001, 0010, 0011, 0100}; $P(E\cap F) = 5/16$

So $P(E \mid F)= P(E\cap F)/P(F) = 5/8$

Another approach by rule of sum

prob (2nd bit is 0) + prob (2nd bit is 1 and the last two bits are 0s)
$= ½ + ½ \times ¼ = 5/8$

Two Boys Paradox

Example: Given one of the 2 children is a boy, what is probability that the other child is also a boy?

F = one of the 2 children is a boy

$P(F) = \frac{3}{4}$

E = two boys

$P(E) = P(E \cap F) = \frac{1}{4}$

F = { BB, BG, GB }
E = { BB }

$P(E \mid F) = P(E \cap F) / P(F) = 1/3$

Note that $P(E \cap F) = P(E \mid F)\, P(F)$

$= P(F \mid E)\, P(E)$

An Application based on Bayes' Theorem

Known facts:

- 1 out of 100,000 person got the disease
- diagnostic test is 99% correct

What is the probability that the person got the disease when the result is positive?

F = the person got the disease

$P(F) = 0.00001$

E = the test is positive

$P(E \mid F)$ = accuracy of the test = 0.99

$P(E \mid \overline{F})$ = error of the test = 0.01

In general, there are two types of errors:

$P(\overline{E} \mid F)$ - false negative

$P(E \mid \overline{F})$ - false positive

By Bayes' Theorem, we can derive $P(F \mid E)$.

Bayes' Theorem

$P(F \mid E) = P(E \cap F) / P(E)$

where $P(E \cap F) = P(E \mid F)\, P(F)$

$P(E) = P(E \cap F) + P(E \cap \overline{F}) = P(E \mid F)\, P(F) + P(E \mid \overline{F})\, P(\overline{F})$

Known facts: $P(F) = 0.00001$

$P(E \mid F)$ = accuracy of the test = 0.99

$P(E \mid \overline{F})$ = error of the test = 0.01

Thus

$P(E \mid F)\, P(F) = 0.99 \times 0.00001 = 0.0000099 \sim 0.00001$

$P(E \mid \overline{F})\, P(\overline{F}) = 0.01 \times 0.99999 = 0.0099999 \sim 0.01$

$P(E \cap F) = P(E \mid F)\, P(F) \sim 0.00001$

$P(E) = P(E \mid F)\, P(F) + P(E \mid \overline{F})\, P(\overline{F}) \sim 0.00001 + 0.01 \sim 0.01$

$P(F \mid E) = P(E \cap F) / P(E) \sim 0.00001/0.01 \sim 0.001$

The probability that *the person got the disease when the result is positive* is quite low.

Independence

Events E and F are independent events if $P(E \mid F) = P(E)$

i.e., $P(E \cap F) = P(E) \bullet P(F)$

Intuitively: E has nothing to do with F, probability of E would not be affected by F.

Example: A coin is tossed twice. What is the probability that the first toss is head given that the second toss is head?

S = {(H,H)}, (H,T), (T,H), (T,T)}

E = First toss is head = {(H,H)}, (H,T)}; $P(E) = ½$

F = Second toss is head = {(H,H)}, (T,H)}; $P(F) = ½$

$E \cap F$ = both tosses are head = {(H,H)}, $P(E \cap F) = ¼$

$P(E \mid F) = P(E \cap F) / P(F) = ¼ / ½ = ½$

(conclusion: independent)

Independence

Events E and F are independent events if $P(E \mid F) = P(E)$
i.e., $P(E \cap F) = P(E) \cdot P(F)$

Example: Remove two balls one by one (without replacing) from an urn containing 2 black balls and 2 white balls.

What is $P(1B|2B)$, the probability that the first ball is black given that the second is black?

Note that $P(1B \mid 2B) = P(1B \cap 2B) / P(2B)$

$|S|$ = size of sample space = P(4,2) = 12

$2B = \{(b_1, b_2), (w_1, b_2), (w_2, b_2), (b_2, b_1), (w_1, b_1), (w_2, b_1)\}$; $P(2B) = ½$

$1B \cap 2B = \{(b_1, b_2), (b_2, b_1)\}$; $P(1B \cap 2B) = 1/6$

$P(1B \mid 2B) = P(1B \cap 2B) / P(2B) = 1/3$

$1B = \{(b_1, b_2), (b_1, w_1), (b_1, w_2), (b_2, b_1), (b_2, w_1), (b_2, w_2)\}$; $P(1B) = ½$

As $P(1B \mid 2B) \neq P(1B)$ (conclusion: not independent)