Probability — Some definitions:

1. **Experiment**: A repeatable procedure with well-defined possible outcomes.
2. **Sample Space**: The set of all possible outcomes.
3. **Event**: A subset of the sample space.
4. **Probability Function**: A function giving the probability for each outcome.

Some Examples: ① **Experiment**: Toss a fair coin.

Sample Space: $\Omega = \{H, T\}$

Event: $E_1 = \{H\}$

Probability: $P(E_1) = P(H) = 1/2$

② **Experiment**: Roll a dice.

Sample Space: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Event: $E_2 = \{2, 4, 6\}$

Probability: $P(E_2) = P(2) + P(4) + P(6) = \frac{3}{6} = 1/2$

Discrete Sample Space: It is listable. It can be either finite or infinite. **Examples**: $\{H, T\}$, $\{1, 2, 3, 4, 5, 6\}$, $\{1, 2, 3, 4, \dots\}$ $\{2, 3, 7, 11, 13, 17, \dots\}$

Probability Function: For a discrete sample space $\Omega = \{w_1, w_2, \dots w_n\}$, a probability function P is assigned to each outcome w a number $P(w)$ called the probability of w satisfying the two conditions:

① $0 \leq P(w) \leq 1$ ② $\sum_{j=1}^{n} P(w_j) = 1$

$$P(E) = \sum_{w \in E} P(w)$$

Some rules of Probability: ① $P(A^c) = 1 - P(A)$

② If L and R are disjoint then $P(L \cup R) = P(L) + P(R)$

③ If L and R are not disjoint, then

$$P(L \cup R) = P(L) + P(R) - P(L \cap R)$$

This can be generalized to the inclusion-exclusion principle:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i P(A_i) - \sum_{\substack{i,j \\ i \neq j}} P(A_i \cap A_j) + \sum_{\substack{i,j,k \\ i \neq j \neq k}} P(A_i \cap A_j \cap A_k) \cdots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

Example of inclusion-exclusion principle:

Experiment: Select a number between 1 and 100 (both inclusive)

Sample Space: $\Omega = \{1, 2, \ldots 100\}$

Event: E is the event that the number is either divisible by 2 or divisible by 3.

Probability: $P(E) = ?$

Let A be the event that the number is divisible by 2.
Let B be the event that the number is divisible by 3.
Then $P(E) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= P(A) + P(B) - P(C)$ where

C is the event that the number is divisible by 6.

$$= \cancel{50 + 33} \left(\left\lfloor \frac{100}{2} \right\rfloor + \left\lfloor \frac{100}{3} \right\rfloor - \left\lfloor \frac{100}{6} \right\rfloor\right) / 100$$

$$= (50 + 33 - 16) / 100$$

$$= 67/100$$

Conditional Probability: The conditional probability of A knowing that B occurred is written as:

$$P(A|B) = P(A\cap B)/P(B), \text{ provided } P(B) \neq 0.$$

Multiplication Rule: $P(A\cap B) = P(A|B)\cdot P(B) = P(B|A)\cdot P(A)$

Example of Conditional Probability:

Experiment: Roll two dice

Sample Space: $\Omega = \{(1,1), (1,2), \ldots, (6,6)\}$.

Events: Let A be the event that the sum of two dice is even. Let B be the event that one dice is 1.

Probability: $P(A) = (\overset{2}{1} + \overset{4}{3} + \overset{6}{5} + \cancel{} + \overset{8}{5} + \overset{10}{3} + \overset{12}{1})/36 = 1/2$

$P(B) = 11/36$. $\quad P(A|B) = \dfrac{P(A\cap B)}{P(B)} = \dfrac{5/36}{1/2} = \dfrac{5}{18}$

$P(A\cap B) = 5/36$

Law of Total Probability: Suppose the sample space Ω is divided into n disjoint events $B_1, B_2, \ldots B_n$. Then for any event A: $P(A) = \sum_{i=1}^{n} P(A\cap B_i)$

$$\Rightarrow \boxed{P(A) = \sum_{i=1}^{n} P(A|B_i)\, P(B_i)}$$

Example of Law of Total Probability: **Experiment**: Two balls are drawn one after the other from an urn containing 5 Red and 2 Green balls.

Sample Space: $\Omega = \{rr, rg, gr, gg\}$

Events: R_1: First ball is red, R_2: Second ball is red. G_1: First ball is green, G_2: Second ball is green.

Probability: $P(R_1) = \frac{5}{7}$, $P(G_1) = 2/7$,

$P(R_2|R_1) = 4/6$, $P(R_2|G_1) = 5/6$

$P(R_2) = P(R_2|R_1)\,P(R_1) + P(R_2|G_1)\,P(G_1) = \frac{4}{6}\cdot\frac{5}{7} + \frac{5}{6}\cdot\frac{2}{7}$

$= 30/42 = 5/7$

Independent Events: Two events A and B are independent if $P(A \cap B) = P(A)\,P(B) \iff P(A|B) = P(A) \iff P(B|A) = P(B)$

Example of independent events: **Experiment**: Toss a fair coin and roll a die.

Sample Space: $\Omega = \{ (H,1), (H,2) \ldots, (T,6) \}$

Events: A: Coin shows H, B: Die shows 6

$A \cap B = \{(H,6)\}$

Probabilities: $P(A) = 1/2$, $P(B) = 1/6$, $P(A \cap B) = 1/12 = P(A)\,P(B)$

A and B are independent events

Example of non independent events. **Experiment**: Toss a fair coin and roll a die.

Sample Space: $\Omega = \{ (H,1), (H,2) \ldots, (T,6) \}$.

Events: $C: \{(H,6)\}$, $D: \{(T,1)\}$

$C \cap D = \phi$

Probabilities: $P(C) = 1/12$, $P(D) = 1/12$, $P(C \cap D) = 0$

$P(C \cap D) \neq P(C)\,P(D)$

C and D are not independent events

Occurance of A has no effect on occurance of B.

Occurance of C implies non-occurance of D.

Occurance of D implies non-occurance of C.

Bayes Theorem: $P(B/A) = \dfrac{P(A/B).\,P(B)}{P(A)}$

Example of Bayes Theorem: **Experiment:** Toss a coin 5 times.

Sample Space: $\Omega = \{ HHHHH, \ldots \}$

Events: H_1: First toss is H, H_A: all 5 tosses are H.

Probabilities: $P(H_1) = 1/2$, $P(H_1/H_A) = 1$, $P(H_A/H_1) = 1/16$,

$$P(H_1/H_A) = \frac{P(H_A/H_1).\,P(H_1)}{P(H_A)} = \frac{(1/16)(1/2)}{(1/32)} = 1$$

Random Variables: A discrete random variable is a function $X: \Omega \to R$ that takes a discrete set of values. For any value a we write $X = a$ to mean the event consisting of all outcomes ω with $X(\omega) = a$.

Example: **Experiment**: Roll two dice.

Sample Space: $\Omega = \{ (i,j) \mid 1 \le i \le 6, 1 \le j \le 6 \}$

Random Variable: For $\omega = (i,j)$, define $X(i,j) = i+j$

Event: $X = 6$ denotes the event $\{(1,5), (2,4), (3,3), (4,2), (5,1)\}$.

Probability: $P(X=6) = 5/36$.

Probability Mass Function: The probability mass function (pmf) of a discrete random variable is the function $p(a) = P(X=a)$.

Example: The pmf in the previous example is:
$p(a) = 0$ for $a \notin \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
$p(2) = 1/36$, $p(3) = 2/36$, $p(4) = 3/36$, $p(5) = 4/36$, $p(6) = 5/36$, $p(7) = 6/36$, $p(8) = 5/36$, $p(9) = 4/36$, $p(10) = 3/36$, $p(11) = 2/36$, $p(12) = 1/36$.

Events and inequalities: Inequalities with random variables describe events. For example $X \le a$ is the set of all outcomes ω such that $X(\omega) \le a$.

Example: In the previous example, the event $X \le 4$ is: $\{(1,1), (1,2), (2,1), (1,3), (2,2), (3,1)\}$

<u>The cumulative distribution function (cdf)</u>: The cdf of a random variable X is the function F given by $F(a) = P(X \le a)$. <u>Example:</u> The cdf for the previous example is:

$F(a) = 0$ for $-\infty < a < 2$, $F(a) = 1/36 \ \forall a \in [2,3)$,
$F(a) = 3/36 \ \forall a \in [3,4)$, $F(a) = 6/36 \ \forall a \in [4,5)$,
$F(a) = 10/36 \ \forall a \in [5,6)$, $F(a) = 15/36 \ \forall a \in [6,7)$,
$F(a) = 21/36 \ \forall a \in [7,8)$, $F(a) = 26/36 \ \forall a \in [8,9)$,
$F(a) = 30/36 \ \forall a \in [9,10)$, $F(a) = 33/36 \ \forall a \in [10,11)$,
$F(a) = 35/36 \ \forall a \in [11,12)$, $F(a) = 1 \ \forall a \in [12,\infty)$

<u>Properties of pmf</u>: $0 \le p(a) \le 1$

<u>Properties of cdf</u>: ① $a \le b \Rightarrow F(a) \le F(b)$.
F is monotonically increasing (non-decreasing).
② $0 \le F(a) \le 1$
③ $\lim_{a\to\infty} F(a) = 1$, $\lim_{a\to-\infty} F(a) = 0$

<u>Bernoulli Distribution</u>: It models one trial in an experiment that can result in either success or failure. A random variable X has a Bernoulli distribution with parameter p (Bernoulli(p)) if:
① X takes the values 0 or 1.
② $P(X=1) = p$ and $P(X=0) = 1-p$

<u>pmf:</u> $p(a) = 0$ for $a \notin \{0,1\}$, $p(0) = 1-p$, $p(1) = p$

<u>cdf:</u> $F(a) = 0 \ \forall a \in (-\infty, 0)$, $F(a) = 1-p \ \forall a \in [0,1)$,
$F(a) = 1 \ \forall a \in [1,\infty)$

Binomial Distribution: Binomial(n,p) models the number of successes in n independent Bernoulli(p) trials. Binomial(n,p) random variable X can take values $0,1,2,\cdots,n$. Binomial$(1,p)$ = Bernoulli(p).

pmf: $p(a) = 0 \ \forall a \notin \{0,1,2,\cdots,n\}$,

$$p(a) = {}^nC_a\, p^a (1-p)^{n-a} \ \forall a \in \{0,1,2,\cdots,n\}.$$

cdf: $F(a) = 0 \ \forall a \in (-\infty, 0)$,

$$F(a) = \sum_{i=0}^{\lfloor a \rfloor} {}^nC_i\, p^i (1-p)^{n-i} \ \forall a \in [\lfloor a \rfloor, \lfloor a \rfloor + 1)$$

$$F(a) = 1 \ \forall a \in [n, \infty)$$

Geometric Distribution: A random variable X has a geometric distribution with parameter p (Geometric(p)) if it takes the values $\cancel{0},1,2,3,\ldots$ and its pmf is given by $p(k) = P(X=k) = (1-p)^{k-1}p$. It models the number of ~~failures before~~ trials to get the first success.

pmf: $p(a) = 0 \ \forall a \notin \cancel{(-\infty, 0)} \{\cancel{0},1,2,3,\ldots\}$

$$p(a) = (1-p)^{a-1} p \ \forall a \in \{\cancel{0},1,2,3,\ldots\}$$

cdf: $F(a) = 0 \ \forall a \in (-\infty, \cancel{0}1)$

$$F(a) = p\sum_{i=0}^{\lfloor a \rfloor} (1-p)^i = 1-(1-p)^{\lfloor a \rfloor + 1} \ \forall a \in [\lfloor a \rfloor, \lfloor a \rfloor + 1)$$

Uniform Distribution: Uniform(N) models any situation where all the outcomes are equally likely. X takes values $1,2,3,\cdots,N$, each with probability $1/N$.

pmf: $p(a) = 0 \ \forall a \notin \{1,2,3\cdots N\}$

$$p(a) = 1/N \ \forall a \in \{1,2,3,\cdots,N\}.$$

cdf: $F(a) = 0 \ \forall a \in (-\infty, 1)$,

$$F(a) = \lfloor a \rfloor / N \ \forall a \in [\lfloor a \rfloor, \lfloor a \rfloor + 1),$$

$$F(a) = 1 \ \forall a \in [N, \infty)$$

Expectation of a Random Variable: Suppose X is a discrete random variable that takes values $x_1, x_2, \ldots, x_n$ with probabilities $p(x_1), p(x_2), \ldots, p(x_n)$. The expected value of X is denoted $E(X)$ and defined by:

$$E(X) = \sum_{j=1}^{n} p(x_j)\, x_j = p(x_1)\, x_1 + p(x_2)\, x_2 + \cdots + p(x_n)\, x_n.$$

Example: **Experiment**: Roll a die. $\Omega = \{1, 2, 3, 4, 5, 6\}$

Random variable: $X(\omega) = \omega$

Expectation: $E(X) = \frac{1}{6}(1+2+3+4+5+6) = 3.5$

Properties of $E(X)$: ① Linearity of expectation:
If X and Y are random variables on a sample space Ω then $E(X+Y) = E(X) + E(Y)$. This can be generalized to n random variables $X_1, X_2 \ldots X_n$:

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i).$$

② If a and b are constants then $E(aX+b) = a\,E(X) + b$

Proof:

Outcome w:	w_1	w_2	$\ldots$	w_n
Value of X:	x_1	x_2	$\ldots$	x_n
Value of Y:	$\cancel{x_1}\, y_1$	y_2	$\ldots$	y_n
Value of $X+Y$:	x_1+y_1	x_2+y_2	$\ldots$	x_n+y_n
Probability $P(w)$:	$P(w_1)$	$P(w_2)$	$\ldots$	$P(w_n)$

$$E(X+Y) = \sum (x_i + y_i) P(w_i) = \sum x_i P(w_i) + \sum y_i P(w_i) = E(X) + E(Y)$$

$$E(aX+b) = \sum p(x_i)(a x_i + b) = a \sum p(x_i)\, x_i + b \sum p(x_i) = a\,E(X) + b$$

Example of linearity of expectation: Roll two dice and let X be the sum. Find $E(X)$.

Let X_1 be the value on the first die and let X_2 be the value on the second die. We have: $X = X_1 + X_2$.

$$E(X) = E(X_1 + X_2) = E(X_1) + E(X_2) = 3.5 + 3.5 = 7.$$

We can also verify this as follows:

$$E(X) = \frac{1}{36}(2) + \frac{2}{36}(3) + \frac{3}{36}(4) + \frac{4}{36}(5) + \frac{5}{36}(6) + \frac{6}{36}(7) + \frac{5}{36}(8) + \frac{4}{36}(9) + \frac{3}{36}(10) + \frac{2}{36}(11) + \frac{1}{36}(12)$$

$$= (2 + 6 + 12 + 20 + 30 + 42 + 40 + 36 + 30 + 22 + 12)/36$$

$$= 252/36 = 7.$$

Expectation of Bernoulli random variable: Let X be a Bernoulli(p) random variable. Then

$$E(X) = p \cdot 1 + (1-p) \cdot 0 = p.$$

Expectation of Binomial random variable: Let X be a Binomial(n, p) random variable. It is equivalent to n Bernoulli(p) random variable. Let X_i be i^{th} Bernoulli(p) random variable.

$$X = \sum_{j=1}^{n} X_j \Rightarrow E(X) = E\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} E(X_j)$$

$$= \sum_{j=1}^{n} p = np.$$ We can also compute it directly:

$$E(X) = \sum_{k=0}^{n} k\, p(k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} = np(p + 1 - p)^{n-1} = np.$$

Expectation of a Geometric random variable:

Let X be a Geometric(p) random variable. X takes values $k = \not{0}, 1, 2, \dots$ with probabilities $p(k) = (1-p)^{k-1}p$.

$E(X) = \sum_{k=\not{0}1}^{\infty} k(1-p)^{k-1}p$. $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$. Differentiating both sides we get: $\sum_{k=0}^{\infty} k x^{k-1} = \frac{1}{(1-x)^2}$.

Multiplying by x, we get: $\sum_{k=0}^{\infty} k x^k = \frac{x}{(1-x)^2}$.

Putting $x = (1-p)$, we get: $\sum_{k=0}^{\infty} k(1-p)^k = \frac{1-p}{p^2}$

Multiplying by $\frac{p}{1-p}$, we get: $\sum_{k=\not{0}1}^{\infty} k(1-p)^{k-1}p = \frac{1\cancel{-p}}{p} = E(X)$

Expectation of a Uniform random variable:

Let X be a uniform(N) random variable. X takes values $1, 2, \dots N$ with probability $1/N$.

$$E(X) = \sum_{i=1}^{N} \frac{i}{N} = \frac{N(N+1)}{2N} = \frac{N+1}{2}.$$

Variance of a random variable:

The expectation of a random variable is a measure of central tendency. The variance is a measure of how much the probability mass is spread out around expected value.

$$\mathrm{Var}(X) = E((X - E(X))^2)$$

Properties of Variance:

① If X and Y are independent, then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

② For constants a and b, $\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$.

③ $\mathrm{Var}(X) = E(X^2) - E(X)^2$

① Can be generalized to: For $X_1, X_2 \dots X_n$ mutually independent random variables: $\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$

Proof of linearity of expectation:

$$E(X+Y)=\sum_k k\,P(X+Y=k)=\sum_i\sum_j (i+j)\,P((X=i)\cap(Y=j))$$

$$=\sum_i\sum_j i\,P((X=i)\cap(Y=j))+\sum_i\sum_j j\,P((X=i)\cap(Y=j))$$

$$=\sum_i i\sum_j P((X=i)\cap(Y=j))+\sum_j j\sum_i P((X=i)\cap(Y=j))$$

$$=\sum_i i\,P(X=i)+\sum_j j\,P(Y=j)=E(X)+E(Y).$$

If X and Y are independent random variables, then
$$E(XY)=E(X)\,E(Y).$$

Proof: $E(X.Y)=\sum k\,P(X.Y=k)=\sum_i\sum_j (ij)\,P((X=i)\cap(Y=j))$

$$=\sum_i\sum_j (i.j)\,P(X=i)\,P(Y=j)$$

$$=\Big(\sum_i i\,P(X=i)\Big)\Big(\sum_j j\,P(Y=j)\Big)=E(X)\,E(Y).$$

Proof of 1: $Var(X+Y)=E((X+Y-E(X+Y))^2)$

$$=E(((X-E(X))+(Y-E(Y)))^2)=E((X-E(X))^2)+E((Y-E(Y))^2)$$

$$+E(2(X-E(X))(Y-E(Y)))=Var(X)+Var(Y)$$

$$+2E(XY-YE(X)-XE(Y)+E(X)E(Y))$$

$$=Var(X)+Var(Y)+2E(XY)-2E(Y)E(X)-2E(X)E(Y)+2E(X)E(Y)$$

$$=Var(X)+Var(Y)+2E(X)E(Y)-2E(X)E(Y)$$

$$=Var(X)+Var(Y).$$

Proof of ②: $Var(aX+b) = E\left(\left((aX+b)-(aE(X)+b)\right)^2\right)$
$= E\left(\left(a(X-E(X))\right)^2\right) = E\left(a^2(X-E(X))^2\right) = a^2 E\left((X-E(X))^2\right)$
$= a^2 Var(X)$.

Proof of ③: $Var(X) = E\left((X-E(X))^2\right) = E\left(X^2 - 2XE(X) + E(X)^2\right)$
$= E(X^2) - 2E(X)^2 + E(X)^2 = E(X^2) - E(X)^2$.

Variance of Bernoulli(p) random Variable X:

$Var(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p)$.

Variance of Binomial(n,p) random Variable X:

$X = \sum_{i=1}^{n} X_i$ where each X_i is Bernoulli(p) and are mutually independent. $\Rightarrow$ $Var(X) = \sum_{i=1}^{n} Var(X_i) = np(1-p)$.

We can also compute it directly as:

$$E(X^2) = \sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-j} \cdot j^2$$

$$= \sum_{j=0}^{n} \frac{n!}{(n-j)!\, j!} p^j (1-p)^{n-j} \left((j^2-j)+j\right)$$

$$= \sum_{j=0}^{n} \frac{n!\,(j^2-j)}{(n-j)!\, j!} p^j (1-p)^{n-j} + \sum_{j=0}^{n} \frac{n!\, j}{(n-j)!\, j!} p^j (1-p)^{n-j}$$

$$= n(n-1)p^2 \sum_{j=2}^{n} \frac{(n-2)!}{(n-j)!\,(j-2)!} p^{j-2} (1-p)^{n-j}$$

$$+ np \sum_{j=1}^{n} \frac{(n-1)!}{(n-j)!\,(j-1)!} p^{j-1} (1-p)^{n-j}$$

$$= n(n-1)p^2 + np$$

$$Var(X) = E(X^2) - E(X)^2 = n(n-1)p^2 + np - n^2p^2$$
$$= np - np^2 = np(1-p).$$

<u>Variance of a Geometric (p) random Variable X:</u>

$$\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i.$$ Differentiating both sides, we get:

$$\frac{1}{(1-x)^2} = \sum_{i=0}^{\infty} (i+1) x^i.$$ Differentiating both sides, we get:

$$\frac{2}{(1-x)^3} = \sum_{i=0}^{\infty} (i+1)(i+2) x^i.$$

$$\sum_{i=1}^{\infty} i^2 x^i = \sum_{i=0}^{\infty} i^2 x^i = \sum_{i=0}^{\infty} (i+1)(i+2) x^i - 3\sum_{i=0}^{\infty} (i+1) x^i$$
$$+ \sum_{i=0}^{\infty} x^i = \frac{2}{(1-x)^3} - \frac{3}{(1-x)^2} + \frac{1}{1-x} = \frac{x^2+x}{(1-x)^3}$$

$$\not\!E(X^2) = \sum_{i=1}^{\infty} p(1-p)^{i-1} i^2 = \frac{p}{1-p} \sum_{i=1}^{\infty} (1-p)^i i^2$$

$$= \frac{p}{1-p} \cdot \frac{(1-p)^2 + (1-p)}{p^3} = \frac{2-p}{p^2}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

<u>Variance of a Uniform(n) random Variable X:</u>

$$Var(X) = E(X^2) - E(X)^2 = \sum_{i=1}^{n} \frac{i^2}{n} - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{n(n+1)(2n+1)}{6n} - \frac{(n+1)^2}{4} = \frac{n+1}{12} \cdot (4n+2-3n-3)$$

$$= \frac{n^2-1}{12}.$$