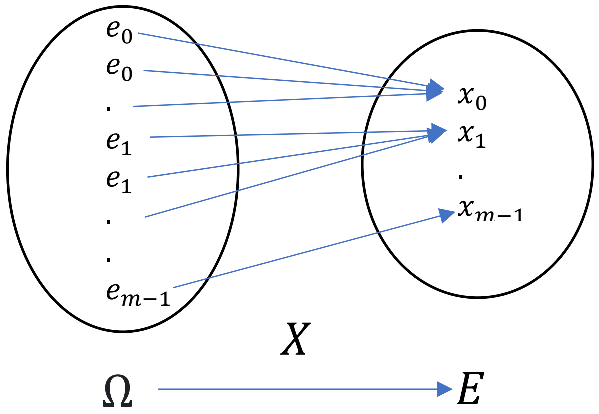
Random Variable

Definition

Random Variable X is a function that maps elements from all the possible outcomes of an experiment (called *Sample Space* Ω) to a Measurable Space E

$$X: \Omega_X \rightarrow \mathrm{E}$$

i.e., we perform an experiment which can have many outcomes (some repeating more than once), a Random Variable assigns a 'measurable number' (Integer, Boolean or Real) to each of the outcomes in Ω



Note: All repetitions from Ω_X maps to same value in range E.

Sample Space, Ω

Set of all possible occurrences in an experiment.

$$\Omega_X = \{e_0, e_0, \dots, e_1, e_1, \dots, e_{m-1}\}$$

Theoretically elements in sample space are unique even if some element can occur more than once in the experiment. But to assist in counting correct probability, we would write all repetitions. (In the end the set will eat away all the repetitions, so 'theoretical definition' of sample space remains intact).

Discrete	Continuous
When number of outcomes of the experiment are finite then sample space is called *Discrete Sample Space* $$\Omega_X = \{e_0, e_0, \dots, e_1, e_1, \dots, e_{m-1}\}$$	When number of outcomes of the experiment are countably infinite then sample space is called *Continuous Sample Space* $$\Omega_X = \{e_0, e_0, \dots, e_1, e_1, \dots, e_\infty\}$$

Measurable Space, E

Measurable space E consists of non-repeating countable values,

$$E = \{x_0, x_1, \dots, x_m\}$$

Properties:

1. Measurable Space does not have any repetitions.
2. E_{min}: Minimum value that X can take from E
3. E_{max}: Maximum value that X can take from E

Discrete	Continuous
If E = Integer (I) then X is called **Discrete Random Variable, i.e.,** $$X: \Omega_X \to I$$ For Example: $$X: e_{-m} \to -m, e_{-1} \to -1, e_0 \to 0, e_1 \to 1, \dots, e_m \to m$$ Then, $\mathrm{E} = \{-\mathrm{m}, \dots, -2, -1, 0, 1, 2, \dots, \mathrm{m}\}$ with $E_{min} = -m$ & $E_{max} = m$	If E = Real (R) then X is called **Continuous Random Variable, i.e.,** $$X: \Omega_X \to R$$ For Example: $$X: e_r \to r \text{ } where \text{ } r \in R$$ i.e., $E = \{-\infty, \dots, -2dx, -dx, 0, dx, 2dx, \dots, \infty\}$ with $E_{min} = -\infty$ & $E_{max} = \infty$

Event

Any possible subset of measurable space, E is called an Event. Represented by capital letters: A, B, C etc. There would be 2^{Ω_X} possible events. For Example

Event A: $\{x_0\}$ which can also be represented by saying outcome where $X = x_0$
Event B: $\{x_0, x_6\}$:Either outcome is x_0 or x_6. In other words, $X = x_0 \; or \; X = x_6$
Event C: $\{x_i \; to \; x_j\}$: All the continuous events from x_i till x_j, $X \in [x_i, x_j]$

Event To Random Variable

Given an Event A on sample space Ω_X, we can represent it as another random variable I_A which is equals to 1 when any of the outcome from sample space is in event A and 0 otherwise.

$$I_A = \begin{cases} 1, & e \in A \\ 0, & otherwise \end{cases}$$

Where e is any random outcome of the experiment from Ω_X This especial Random Variable which can only take two values is called Indicator Variable. Hence mathematically any occurrence of any event A can be represented by I_A

Every Indicator Variable follows Bern(p) where p is probability of event A.

For example, $\Omega_X = \{1, 2, 3, 4\}$, (all outcomes are equally alike) then

Event (A)	Corresponding Indicator Variable (I_A)'s sample space (Ω_{I_A})
$\{1, 2\}$	$P_{I_A}(0) = \frac{1}{2}$; $P_{I_A}(1) = \frac{1}{2}$; $\Omega_{I_A} = \{1, 1, 0, 0\}$
$\{2, 3\}$	$P_{I_A}(0) = \frac{1}{2}$; $P_{I_A}(1) = \frac{1}{2}$; $\Omega_{I_A} = \{0, 1, 1, 0\}$
$\{3, 1\}$	$P_{I_A}(0) = \frac{1}{2}$; $P_{I_A}(1) = \frac{1}{2}$; $\Omega_{I_A} = \{1, 0, 1, 0\}$
$\{1, 2, 3\}$	$P_{I_A}(0) = \frac{1}{4}$; $P_{I_A}(1) = \frac{3}{4}$; $\Omega_{I_A} = \{1, 1, 1, 0\}$

		Properties	1. Probability of random variable I_A taking value 1 is same as Probability of the A from Ω_X. $P_{I_A}(i_A = 1) = P(\, x \in A) = P_X(A)$ $E[A] = E[I_A] = P(A)$ $Var(A) = Var(I_A) = P(A) - P(A)^2 = P(A)Q(A)$ 2. Given two events A and B, $occurence\ of\ event\ A \cap B = I_{AB}$ $P_{I_{AB}}(i_{AB} = 1) = P(x \in A\ \&\ y \in B) = P_X(A \cap B)$ $E[A \cap B] = E[I_{AB}] = P(A \cap B)$ $Var(A \cap B) = P(A \cap B) - P(A \cap B)^2$ $Cov(A, B) = P(A \cap B) - P(A)P(B)$ 3. If n events are mutually independent then corresponding indicator variables are also mutually independent. (and vice versa is true as well)

<table>
<tr><td rowspan="4">Probability Function, $P_X(x)$</td><td colspan="3">A Function that assigns a value between $[0, 1]$ to each element in Measurable Space, E, depending upon its occurrence in Ω_X. It represents the likelihood of happening or occurrence of that element x where $x \in E$.

Probability of Random Variable X taking value x where $x \in E$, is denoted by $P_X(x)$.</td></tr>
<tr><td colspan="2">Discrete</td><td>Continuous</td></tr>
<tr><td colspan="2">Probability Mass Function (PMF), $\boldsymbol{P_X(x)}$: Probability of X taking value x. It is calculated as
$$P_X(x) = \frac{count\ of\ x\ in\ \Omega_\mathrm{X}}{total\ count\ in\ \Omega_\mathrm{X}}$$
Note: we specify all repetitions in Ω_X observed in the experiment, hence count will account for correct likelihood.</td><td>Probability Density Function (PDF), $\boldsymbol{f_X(x)}$: Probability of X taking value between x to $x + dx$ per unit dx.

Then Probability of X taking between x to $x + dx$:
$$dF_X(x) = f_X(x)dx$$
Where $F_X(x)$ is Cumulative Probably Function</td></tr>
<tr><td>Examples</td><td>Random Variable, X: Taking out two balls b1 & b2 from a bag with n red and m blue, assigning $(\mathrm{r,r}) = 0, (\mathrm{r,b}) = 1, (\mathrm{b,r}) = 2$ & $(\mathrm{b,b}) = 3$
E: $\{0, 1, 2, 3\}$
Probability:
$$\mathrm{P}_X(x) = \begin{cases} \frac{{}^{n}_{2}C}{{}^{m+n}_{\ \ 2}P}, & \mathrm{x} = 0 \\ \frac{{}^{n}_{1}C\,{}^{m}_{1}C}{{}^{m+n}_{\ \ 2}P}, & \mathrm{x} = 1 \\ \frac{{}^{m}_{1}C\,{}^{n}_{1}C}{{}^{m+n}_{\ \ 2}P}, & \mathrm{x} = 2 \\ \frac{{}^{m}_{2}C}{{}^{m+n}_{\ \ 2}P}, & \mathrm{x} = 3 \end{cases}$$
Notice number of ways of selecting two balls from $m + n$ is not the domain rather occurrences of all the duplicates are counted towards the probability of each event. (This is all because of definition of random variable counting all possible (r, r) as one unique x_i)</td><td>Random Variable, X: n points placement on circle divided by a diameter line, assuming same side = 0, different side = 1
Domain: {0, 1}
Probability:</td></tr>
</table>

		Random Variable, X: Number of heads in coin toss n times such that probability of head is p. **Domain:** $\{0, 1, 2, \ldots, n\}$ **Probability:** $$P_X(x) = {}^{n}_{x}C \times p^x \times (1-p)^{n-x}$$	
		Random Variable, X: Sum of numbers on two fair dice throws. **Domain:** $\{2, 3, 4, \ldots, 12\}$ **Probability:** $$P_X(x) = \frac{\min(x-1, 13-x)}{36}$$	

Properties

1. For any **Event** $\boldsymbol{A}$ which is a subset of Measurable Space E, its Probability, $P_X(A)$ is sum or integral of probability of all its individual elements:

Discrete	**Continuous**
Given an Event A = $\{x_0, x_6, \ldots, x_m\}$ then $$P_X(A) = \sum_{x \in A} P_X(x)$$	Given an Event A = $[x, x + \Delta x]$ then $$P_X(A) = \int_{x}^{x+\Delta x} f_X(x)dx$$

2. **Complement Event,** $\overline{\boldsymbol{A}}$: X takes all the values except for values in another event A:
$$1 - P_X(A)$$

3. **Conditional Event,** $\boldsymbol{\frac{A}{B}}$**:** X taking values of event A conditioned on B (i.e. B has happened, now from that what's the probability of A)
    a. Approach 1: use B as sample space and then find occurrence of A from B then use:
    $$\frac{Count\ of\ x \in A\ \cap B}{Count\ of\ x\ \in B}$$
    b. Approach 2:
    $$P_X\left(\frac{A}{B}\right) = \frac{P_X(A \cap B)}{P_X(B)}$$

4. Given probabilities of A conditioned on B_i event $P\left(\frac{A}{B_i}\right) \vee i\ \in [0, m]$ such that all values of X in B_i are mutually exclusive and exhaustive to measurable space E, then:
$$P_X(A) = \sum_{i=0}^{m} P\left(\frac{A}{B_i}\right)$$

5. **Intersection Event,** $\boldsymbol{A} \cap \boldsymbol{B}$: X taking all the values which are common in both events A or B.

    a. Approach 1
        i. Find intersection of values of X from both the event sets A & B.
        ii. Use Probability of that set.

b. Approach 2 *(Multiplication Property)*

$$P_X(A \cap B) = P_X(A)P_X\left(\frac{B}{A}\right) = P_X(B)P_X\left(\frac{A}{B}\right)$$

Given n events $E_1, \dots, E_n$

$$P_X(E_0 \cap \dots \cap E_n) = P_X(E_0) \times P_X\left(\frac{E_1}{E_0}\right) \times P_X\left(\frac{E_2}{E_0 \cap E_1}\right) \times \dots . \times P_X\left(\frac{E_{i_n}}{E_0 \cap \dots \cap E_{n-1}}\right)$$

c. **Independence of Events:** Given n events $E_1, \dots, E_n$

i. They are **pairwise independent** if

$$P_X(E_i \cap E_j) = P_X(E_i) \times P_Y(E_j) \bigvee i, j$$

ii. They are **mutually independent** if for any k sized-subset events $E_{i_0}, E_{i_1}, \dots, E_{i_{k-1}}$:

$$P_X(E_{i_0} \cap E_{i_1} \cap \dots \cap E_{i_{k-1}}) = P_X(E_{i_0}) \times P_X(E_{i_1}) \times \dots . \times P_X(E_{i_{k-1}})$$

6. **Union Event A ∪ B**: X taking all the values which are in event A or B. Also called Union property.

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

a. For n Events: $E_1, E_2, \dots, E_n$, probability of union of all the events

$$P_X\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} (-1)^{i-1} \times S_i$$

Where

S_i is sum of probabilities of all the possible i intersections, i.e.,

$$S_i = \sum P(E_{j_1} \cap E_{j_2} \cap E_{j_3} \cap \dots \cap E_{j_i})$$

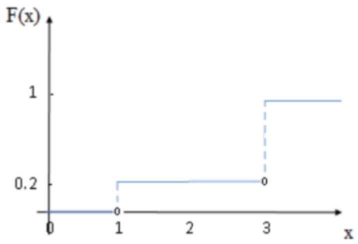
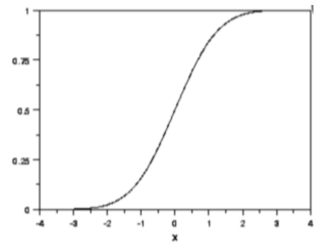
Note S_i will repeatedly include probabilities of i+1, i+2, …, so on intersections

		b. Probability of union of atleast k intersecting events Or Probability of an element e from Ω_X to be common to **at-least** any k interesting events. (its union of k intersecting events, it does not mean that elements common to just k intersecting events – union of them can make them common to k+1, k+2 and so on elements) $$P_X\left(\bigcup E_{i_1} \cap E_{i_2} \cap E_{i_3} \cap \dots \cap E_{i_k}\right) = \sum_{i=k}^{n} (-1)^{i-k} \times {}_{k-1}^{i-1}C \times S_i$$ Note: It does not mean that an element $e \in \bigcup E_{i_1} \cap E_{i_2} \cap E_{i_3} \cap \dots \cap E_{i_k}$ from Ω_X will be common to only k intersecting events, it might as well be common to k+1, k+2 and so on intersecting events – Only guarantee is that it will not be repeated twice. Example: 1: https://en.wikipedia.org/wiki/Inclusion–exclusion_principle#Counting_derangements c. Probability of an element e from Ω_X to be in **exactly** any k intersecting Events: $$\sum_{i=k}^{n} (-1)^{i-k} \times {}_{k}^{i}C \times S_i$$

Cumulative Distribution Function (CDF), $F_X(x)$

Probability of X taking value less than or equal to x, it is denoted by $F_X(x)$ and is given by

$$F_X(x) = P_X(X \le x)$$

Discrete	**Continuous**
 $$F_X(x) = \sum_{x=E_{min}}^{x} P_X(x)$$	 $$F_X(x) = \int_{x=E_{min}}^{x} dF_X(x) = \int_{x=E_{min}}^{x} f_X(x)dx$$

	Properties	1. $P_X(\ a \le x \le b) = F_X(b) - F_X(a)$: If we know cumulative distribution function of the Random variable, than directly using it for interval probability is easier way to find probability. 2. $F_X(x)$ is non-decreasing and right continuous 3. Minimum value of 0 at E_{min} and maximum value of 1 at E_{max} ○ $F_X(E_{min} - \epsilon) = 0$ for any $\epsilon > 0$ ○ $F_X(E_{max}) = 1$ 4. If output from $F_X(x)$ are considered as another sample space Ω_y then $Y = F_X(X)$ will follow Uniform Distribution with boundaries $[0, 1]$ Note, $Y \ne F_X(x)$ as $F_X(x)$ is not random at all, instead we substitute $x = X$, Y inherits the randomness of X, ofcourse through CDF transformation which makes it uniform distribution. • Given $Y \sim U(0, 1)$ generate a Exponential Distribution function $$Y = F_X(X)$$ $$X = F_X^{-1}(Y)$$ For exponential distribution,$F_X(x) = 1 - e^{-\lambda x}$, hence $$X = -\frac{1}{\lambda} \times \ln(1 - Y)$$

Moment Generating Function, $M_X(t)$

Moment Generating Function $M_X(t)$ of random variable X is an alternative specification of its probability distribution. Thus, it provides the basis of an alternative route to analytical results compared with working directly with probability density functions or cumulative distribution functions.

$$M_X(t) = E_X[e^{tX}]$$

Discrete	Continuous
$M_X(t) = \sum_{x \in E} e^{tx} P_X(x)$	$M_X(t) = \int_{x \in E} e^{tx} f_X(x) dx$

The moment-generating function is so named because it can be used to find the moments of the distribution, The series expansion of e^{tx} is

$$e^{tX} = 1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \frac{t^4X^4}{4!} + \cdots + \frac{t^nX^n}{n!}$$

Hence

$$M_X(t) = E_X[e^{tX}] = 1 + tE[X] + \frac{t^2E[X^2]}{2!} + \frac{t^3E[X^3]}{3!} + \frac{t^4E[X^4]}{4!} + \cdots + \frac{t^nE[X^n]}{n!}$$

Where $E[X^n] = n^{th} moment\ of\ X$

Properties

1. $M_{aX+b}(t) = e^{bt} M_X(at)$
2. If $M_X(t) = M_Y(t)$ for all t then $P_X(a) = P_Y(a)$ for all a, i.e. $X\ and\ Y$ have same distribution.
3. $n^{th} moment\ of\ X = \frac{d^n M_X(t)}{dt^n} |t = 0 = M_X^n(0)$
4. $M_X(t) \geq e^{\mu_X t}$
5. $\lim_{n\to\infty} M_X\left(\frac{t}{n}\right)^n = e^{t \times E[X]}$ (Hint: Use L'Hopital's rule)

Moments	
	Moments of random variable X are specific values $x_m \in E_X$ which are of statistical importance.

Mean or Average, μ_X or Expectation, $E[X]$

Value of random variable X which specify its central tendency. It’s also called Average value of X denoted by μ_X or Expectation of random variable, denoted by $E[X]$.

Discrete	**Continuous**
$\mu_X = \ E[X] = \sum_{x \in E} x P_X(x)$	$\mu_X = \ E[X] = \int_{x \in E} x f_X(x) dx$

Properties

1. Given $X \geq 0$ then $E[X] \geq 0$
2. $E_Y\big[E_X[X|Y]\big] = X$
3. If given $M_X(x)$ then

$$\mu_X = \frac{dM_X(t)}{dt}\Big|_{t=0} = M_X'(0)$$

4. Expectation of a function of $X, E[g(X)]$

Discrete	Continuous
$E[g(X)] = \sum_{x \in E} g(x)P_X(x)$	$E[g(X)] = \int_{x \in E} g(x)f_X(x)dx$

   a) $g(X) = aX$ then $E[aX] = aE[X]$: If given expectation of a random experiment then expectation of same experiment with scaled *Measurable Space* is scale times expectation of that random experiment.
   b) $g(X) = X + b$ then If $E[X + b] = E[X] + b$
   c) $g(X) = |X|$ then If $E[|X|] = 0$ then X = 0
   d) $g(X) = X^2$ then If $E[X^2] = Var[X] + E[X]^2$
   e) If $P_X(x)$ is symmetrical about mean μ_X i.e. an even function $P_X(\mu_X - x) = P_X(\mu_X + x)$ and $g(X)$ is odd function around μ_X i.e., $g(\mu_X - X) = -g(\mu_X + X)$ then

$$E[g(X)] = 0$$

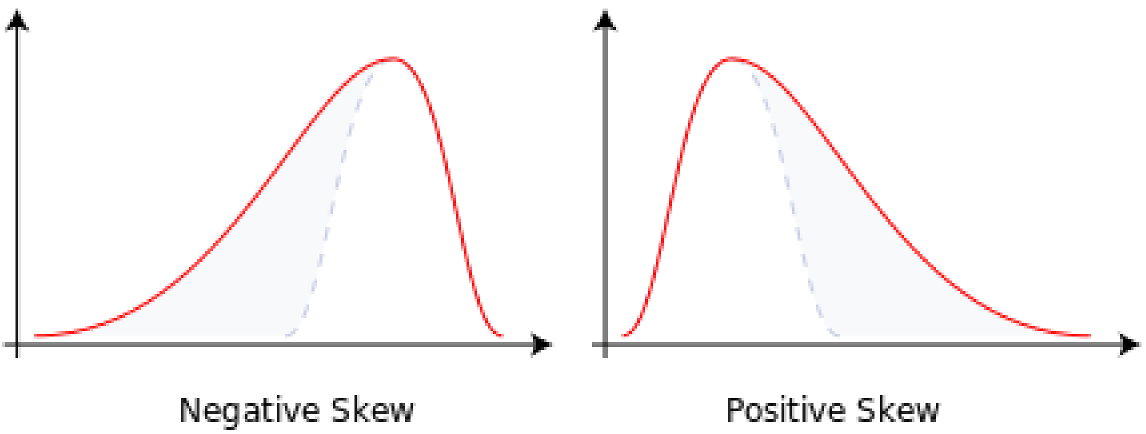
5. Expectation of Indicator Variable is Probability of corresponding event happening:

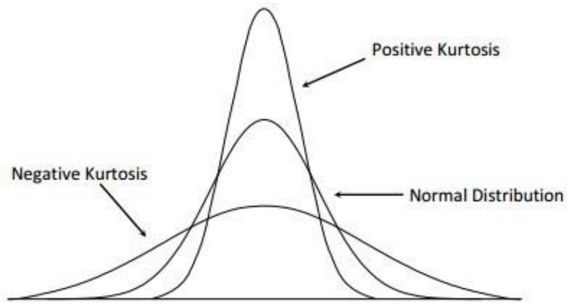
$$E[I_A] = P_X(A)$$

6. For a random variable X taking on non-negative integer values, then its expectation is given by

$$E[X] = \sum_{i=1}^{\infty} P_X(x \geq i)$$

Note: probability of **greater than or equal** to i

Variance		Expectation of squared deviation of X from its mean, its denoted by σ_X^2 $$Var(X) = E[(X-\mu_X)^2] = E[X^2] - E[X]^2$$ **Discrete**: $Var(X) = \sum_{x \in E}(x-\mu_X)^2 P_X(x) = \sum_{x \in E} x^2 P_X(x) - \mu_X^2$ **Continuous**: $Var(X) = \int_{x \in E}(x-\mu_X)^2 f_X(x)dx = \int_{x \in E} x^2 f_X(x)dx - \mu_X^2$
	Properties	1. $Var(X) \geq 0$ 2. If $X = c$ then $Var(X) = 0$ & vice versa 3. $Var(X+c) = Var(X)$ 4. $Var(cX) = c^2 Var(X)$ 5. If given $M_X(x)$ then $$\sigma_X^2 = M_X''(0) - M_X'(0)^2$$
Skew		Asymmetric leaning of distribution to either left or right.  $$Pearson's\ Moment\ Coefficient\ of\ skewness = E\left[\frac{(X-\mu_X)^3}{\sigma^3}\right] = \frac{E[X^3] - 3\mu_X\sigma_X^2 - \mu_X^3}{\sigma_X^3}$$

	Kurtosis	Degree of fatness of outer tails.  $$Kurtosis = E\left[\left(\frac{X-\mu_X}{\sigma}\right)^4\right]$$

Quantiles

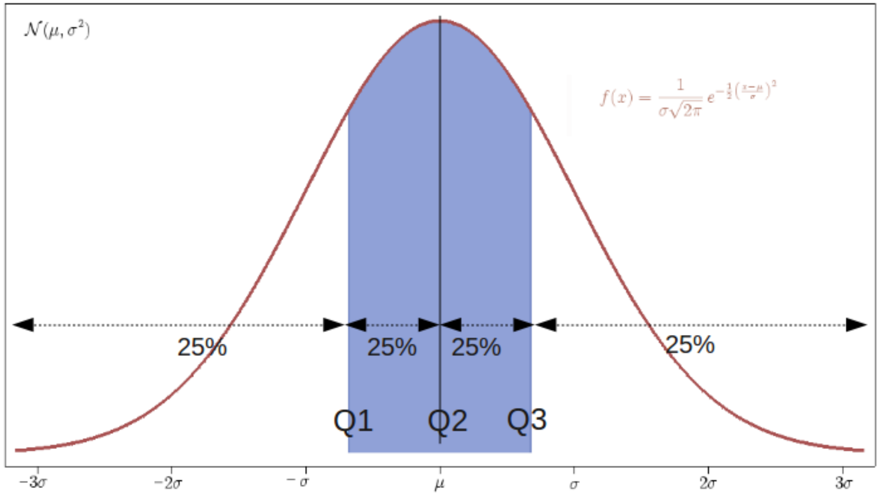
Quantiles (q) are values which divide E into q groups : $\{G_0, G_1, G_2, \dots, G_{q-1}\}$ such that i^{th} group (G_i) has all the values x_i ($\in G_i$) whose output from cumulative probability function $F_X(x_i)$ lies between $\left[\frac{i}{q}, \frac{i+1}{q}\right]$. For a q-quantiles, we would need $q-1$ values/boundaries: $\{Q_0, Q_1, \dots, Q_{q-2}\}$.

$$\frac{i}{q} \leq F_X(x_i) \leq \frac{i+1}{q} \bigvee x_i \in G_i$$

Second definition: If we map all the elements from Ω_X to corresponding E (without removing duplicates) denoting it by Ω_E then $q-1$ points $\{Q_0, Q_1, \dots, Q_{q-2}\}$ are the cut points dividing Ω_E into q equal subsets.

$$P(X < Q_i) \leq \frac{i+1}{q}$$

Third definition: For q-quantile, i^{th} boundary Q_i is the value from E before which there are $\frac{i+1}{q} \times 100\ \%$ values.



Some special Quantiles:

Name	q	Boundaries	
Median	2	$\{Q_0\}$	There are 50% value before Q_0 from Ω_E
Quartiles	4	$\{Q_0, Q_1, Q_2\}$	There are 25% values before Q_0, 50% before Q_1 & 75% before Q_2 from Ω_E
Percentiles	100	$\{Q_0, Q_1, Q_2, \dots, Q_{98}\}$	There are 1% values before Q_1, 2% before Q_2 & so on till, 99% before Q_{98} from Ω_E

- Median $(q = 2)$: $\{Q_0\}$

- Quartiles ($q = 4$): $\{Q_0, Q_1, Q_2\}$
- Percentiles ($q = 100$): $\{Q_0, Q_1, Q_2, \dots, Q_{98}\}$

Median

Median m is the value of X which divides the E in two subsets, first half having CDF_X less than half while others greater than half, such that

$$P(X \leq m) \geq \frac{1}{2} \; and \; P(X \geq m) \geq \frac{1}{2}$$

<table>
<tr><th>Discrete</th><th>Continuous</th></tr>
<tr><td>
6. Map all the elements from Ω_X to corresponding E (without removing duplicates, let's call it Ω_E
7. Sort Ω_E in ascending order.
8. If number of elements in Ω_E (=n) is even then there will be two medians:
$$m1 = \Omega_E\left[\frac{n}{2} - 1\right] and \; m2 = \Omega_E\left[\frac{n}{2}\right]$$
where
• $P(X \leq m1) = \frac{1}{2}$ and $P(X \geq m1) \geq \frac{1}{2}$
• $P(X \leq m2) \geq \frac{1}{2}$ and $P(X \geq m2) = \frac{1}{2}$
9. else if n Is odd, then there will be one median, at
$$m = \Omega_E\left[\frac{n-1}{2}\right]$$
where
• $P(X \leq m) \leq \frac{1}{2}$ and $P(X \geq m) \geq \frac{1}{2}$
</td><td>
1. Find the value $X = m$ such that $F_X(m) = \frac{1}{2}$
i.e.
$$P(X \leq m) = \frac{1}{2} \; and \; P(X \geq m) = \frac{1}{2}$$
</td></tr>
</table>

		Quartiles	A **quartile** is a type of quantile which divides the E into four parts, or *quarters*, of more-or-less equal size. The Ω_E must be ordered from smallest to largest to compute quartiles. The three main quartiles are as follows: • The first quartile (Q_0) is defined as the middle number between the E_{min} and the median of the X. It is also known as the *lower* or *25th empirical* quartile, as 25% of the data is below this point. • The second quartile (Q_2) is the median of a data set; thus 50% of the data lies below this point. • The third quartile (Q_3) is the middle value between the median and the highest value (maximum) of the data set. It is known as the *upper* or *75th empirical* quartile, as 75% of the data lies below this point.[1]
		Percentile	k-th percentile (percentile score or centile) is a score below which a given percentage k of scores in its frequency distribution falls (exclusive definition) or a score at or below which a given percentage falls (inclusive definition).

<table>
<tr><td rowspan="5">Distribution</td><td>Definition</td><td colspan="7">A distribution D specifies

1. Measurable Space E and
2. Probability Function, $P_X(x)$ (PMF for discrete E and PDF for continuous E).

Thus all other quantity are automatically derived, For example, Cumulative Distribution Function $F_X(x)$, Moment Generating Function $M_X(t)$, Expectation, Variance and other moments

When we say a random variable follow a specific distribution or $X \sim D$, then X can any value from $x \ \in E$ with probability $P_X(x)$.</td></tr>
<tr><td></td><td></td><td>Measurable Space, E</td><td>Probability Function, $P_X(x)$ where $x \in E$</td><td>Expectation / Average/ Mean, μ</td><td>Variance, σ</td><td>Cumulative Distribution Function, $F_X(x)$</td><td>Moment Generating Function, $M_X(t)$</td></tr>
<tr><td rowspan="3">Discrete</td><td>Bernoulli, $Bern(p)$</td><td>$\{0, 1\}$</td><td>$\begin{cases} p, & x = 1 \\ q, & x = 0 \end{cases}$

where $q = 1 - p$</td><td>p</td><td>pq</td><td>$\begin{cases} q, & x = 0 \\ 1, & x = 1 \end{cases}$</td><td>$pe^t + q$</td></tr>
<tr><td rowspan="2">Binomial, $Bin(n, p)$</td><td colspan="6">Definition 1: This distribution defines probability of success x times out of n trials assuming each trials has uniform/constant/same probability of success p.

Definition 2: Sum of sequence of n i.i.d. random variables $S_n = X_1 + X_2 + \cdots + X_n$ such that $X_i \sim Bernoulli(p)$.

• This is n times Sample Mean of n IIDs, which will start to follow Normal Distribution for large n (Central Limit Theorem) with mean $np \ (= \mu_{X_i})$ and variance $npq \ (= n\sigma_{X_i})$.</td></tr>
<tr><td>$\{0, 1, 2, \dots, n\}$</td><td>${}^n_xC \times p^x \times q^{n-x}$

where $q = 1 - p$</td><td>np</td><td>npq</td><td>NA</td><td>$(pe^t + q)^n$</td></tr>
</table>

			Example 1: Given probability of at least one success in n trials to be A, find probability of at least one success in k trials assuming uniform probability of success across n trials. $$A = 1 - probablity\ of\ no\ success = 1 - q^n \Longrightarrow q = 1 - \sqrt[n]{A}$$ Hence, at-least one success in k trials = $1 - q^k = 1 - (1 - A)^{\frac{k}{n}}$					
		Negative Binomial/Pascal, NB(r, p)	This distribution defines **probability of success** x times before r^{th} failure assuming each trials has uniform/constant/same probability of success p (probability of success is uniformly distributed across n trials). • Note, this is defining X = number of successes, if we want number of trials (Let's say Y), then $Y = X + r$. i.e. Expected number of trials before r failures: $E[Y] = E[X] + r = \frac{1}{1-p}$					
			$Natural\ Number\ (N) = \{0, 1, 2, \dots, \infty\}$	${}^{x+r-1}_{x}C \times p^x \times q^r$	$r \times \frac{p}{1-p}$	$r \times \frac{p}{(1-p)^2}$	NA	$\left(\frac{1-p}{1-pe^t}\right)^r$ $for\ t < -logp$
		Geometric, Geo(p)	$Geo(p) = NB(1, p)$					
			$Natural\ Number\ (N) = \{0, 1, 2, \dots, \infty\}$	$p^x \times q$	$\frac{p}{1-p}$	$\frac{p}{(1-p)^2}$	NA	$\frac{1-p}{1-pe^t}$ $for\ t < -logp$

<table>
<tr>
<td rowspan="3"></td>
<td rowspan="3"></td>
<td rowspan="3">Hypergeometric, HG(N, K, n)</td>
<td colspan="6">This distribution defines probability of success x times out of n trials, where success is drawing item with specific features, whose initial counts are K out of N.(No replacement after selection in each trial).

Important: This distribution assumes that all the items are distinct, even if they have same feature. For example if there are K green balls and N-K red balls, after drawing any n balls, we can arrange then in $n!$ ways, irrespective of how many of them are red vs green.</td>
</tr>
<tr>
<td>$\{0, 1, 2, \ldots, n\}$</td>
<td>$\frac{{}^{K}_{x}C \times {}^{N-K}_{n-x}C}{{}^{N}_{n}C}$</td>
<td>$n \times \frac{K}{N}$</td>
<td>$n \times \frac{K}{N} \times \frac{N-K}{N} \times \frac{N-n}{N-1}$</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td colspan="6">Example 1: Given selection without replacement for n trials, probability of something happening in last trial is same as probability of all other selections in first n-1 trials which is same as probably of that happening in first trial itself.
Example 2: Given two exclusive events A and B happening across n trials, probability of happening of A before B, its same as probably of A's happening at first place.
For example: 2 red and 3 green : taking out one ball at a time, probability of red coming before green: 4/10
1. rrggg
2. rgrgg
3. rggrg
4. rgggr
5. grrgg
6. grgrg
7. ggrrg
8. gggrr
9. grggr
10. ggrgr</td>
</tr>
</table>

<table>
<tr>
<td rowspan="3"></td>
<td rowspan="3"></td>
<td rowspan="3">Negative Hypergeometric, NHG(N, K,r)</td>
<td colspan="6">This distribution defines probability of success x times until r^{th} failure, where success is drawing item with specific features, whose initial counts are K out of N.(No replacement after selection in each trial).

Important: This distribution assumes that all the items are distinct, even if they have same feature. For example if there are K green balls and N-K red balls, after drawing any n balls, we can arrange then in $n!$ ways, irrespective of how many of them are red vs green.</td>
</tr>
<tr>
<td>$Natural\ Number\ (N) = \{0, 1, 2, \dots, \infty\}$</td>
<td>$\frac{N-K}{N} \times \frac{{}^{K}_{x}C \times {}^{N-K-1}_{r-1}C}{{}^{N}_{x+r-1}C}$

Note, last place will not be permutated, as that for sure should be r^{th} failure, so probability of that last place to be failure is $\frac{N-K}{N}$ and then remaining place will be arranged with $(x + r - 1)!$ Ways but both in numerator or denominator.</td>
<td>$r \times \frac{K}{N-K+1}$</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td colspan="6">Example 1: Expected Number of cards that needs to be turned in order to see first ace.

Note above distribution defined number of success, which is exactly is asked

$$E[X] = r \times \frac{K}{N-K+1}$$
$N = 52, K = 48, r = 1$ (assuming ace is failure)

$$E[Y] = \frac{48}{5}$$</td>
</tr>
</table>

		Poisson, $Pos(\lambda)$	This distribution defines **probability of success** x times in unit time (with infinite trials) given average number of success in unit time (again with infinite trials) is λ, assuming each trials has uniform/constant/same probability of success. 1. Poisson's distribution is same as Binomial distribution with $n \to \infty$ and $np = \lambda$ 2. $\lambda = probablity\ of\ success\ for\ shortest\ possible\ time\ unit\ dt\ per\ unit\ dt$ 3. $dt = \lim_{n\to\infty} \frac{1}{n}$					
			$Natural\ Number\ (N) = \{0, 1, 2, \dots, \infty\}$	$\frac{\lambda^x e^{-\lambda}}{x!}$	λ	λ	NA	$e^{\lambda(e^t - 1)}$
		Uniform, $unif(a, b)$	This distribution defines probability of any integer between a and b (both included) to be constant.					
			$\{a, a+1, a+2, \dots, b\}$ $where\ a\ \&\ b\ both\ are\ integers$ Number of elements in $E, (n) = b - a + 1$	$\begin{cases} \frac{1}{n}, & a \le x \le b \\ 0, & otherwise \end{cases}$	$\frac{a+b}{2}$	$\frac{n^2 - 1}{12}$	$\begin{cases} 0, & x \le a \\ \frac{x - a + 1}{b - a + 1}, & a \le x \le b \\ 1, & b \le x \end{cases}$	
	Continuous	Uniform, $U(a, b)$	This distribution defines probability of any **Real** between a and b (both included) to be constant.					
			$R \in [a, b]$	$\begin{cases} \frac{1}{(b-a)}, & a \le x \le b \\ 0, & otherwise \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \end{cases}$	

		Normal, $N(\mu,\sigma^2)$	This distribution defines probability of any **Real Number** following a symmetric bell shape around mean μ and variance σ^2 Properties: 1. if $X \sim N(\mu_1,\sigma_1^2)$ and $Y \sim N(\mu_2,\sigma_2^2)$ then $X+Y \sim N(\mu_1+\ \mu_2,\sigma_1^2+\ \sigma_2^2)$					
			$R\ \in (-\infty,\infty)$	$\frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2	$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$ $= \phi\left(\frac{x-\mu}{\sigma}\right)$ where $\phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}x^2} dx$ Properties: 1. $\phi(-x) = 1-\phi(x)$ 2. $\phi(\mu-3\sigma) =\ 0.13\%$ 3. $\phi(\mu-2\sigma)-\phi(\mu-3\sigma)\ = 2.14\%$ 4. $\phi(\mu-\sigma)-\phi(\mu-2\sigma) = 13.6\%$ 5. $\phi(\mu)-\phi(\mu-\sigma) == 34.13\%$	$e^{\mu t+\frac{1}{2}\sigma^2 t^2}$

		Log Normal, $LN(\mu,\sigma^2)$	$if\ Y \sim LN(\mu,\sigma^2)$ then $Y = e^X$ where $X \sim N(\mu,\sigma^2)$					
			$Real\ \in [0,\infty)$	$\frac{1}{x\sqrt{2\pi\sigma^2}}\ e^{-\frac{1}{2}\left(\frac{lnx-\mu}{\sigma}\right)^2}$	$e^{\mu+\frac{\sigma^2}{2}}$	$e^{2\mu+\sigma^2} \times (e^{\sigma^2} - 1)$	$\phi\left(\frac{lnx-\mu}{\sigma}\right)$	NA

Exponential, $E(\theta)$

This distribution represents waiting time t until first event happens given

1. probability of an event happening is uniform/constant (each trial/time is independent)
2. average waiting time is θ or average number of events that happen in unit time is λ, where

$$\theta = \frac{1}{\lambda}$$

If $T \sim E(\theta)$ then

1. $f_T(t)$ is probability of first event to happen at between t to t+dt (also known as waiting time) per unit dt, given average number of event that happen in unit time interval is λ.
2. $F_T(t)$ is probability of first event happening before t = probability of atleast one event happening before t. (it's fine if after first event happened at u ($u < t$), there are more events happening between u and t => we need not worry about them.)

Approach 1: If probability of an event happening is uniform then we know, probability of x events happening in unit time is given by Poisson's distribution, i.e.

$$P(x\ events\ happening\ in\ unit\ time) = \frac{\lambda^x \times e^{-\lambda}}{x!}$$

$$\Rightarrow P(0 \text{ event happening in unit time interval}) = e^{-\lambda}$$
$$\Rightarrow P(0 \text{ event happening in t time interval}) = e^{-\lambda} \times e^{-\lambda} \times \ldots t\ times = e^{-\lambda t}$$
$$\Rightarrow P(at\ least\ one\ event\ happening\ in\ t\ time) = F_T(t) = 1 - e^{-\lambda t}$$
$$\Rightarrow Probablity\ of\ first\ event\ happening\ at\ t\ \big(between\ (t, t+dt)\big) per\ unit\ dt = f_T(t) = \frac{dF_T(t)}{dt} = \lambda e^{-\lambda t}$$

Approach 2:

$$f_X(x) = \text{P(no event to happen till } x \text{ unit time)} \times \text{P(1 event happening in } dt \text{ interval at t time per unit } dt)$$
$$= e^{-\lambda x} \times \text{P(1 event happening in } dx \text{ interval at } x \text{ per unit } dx)$$

Now we know Poisson's distribution is binomial distribution with infinite trials, hence probability of event happening is an even in smallest possible trial dx is p, hence

$$\text{P(1 event happening in } dt \text{ interval at t per unit } dt) = \frac{p}{dx}$$

We also know,

			$1\ unit\ time = \lim_{n\to\infty} n \times dt$ $\frac{1}{dt} = n$ $\text{P(1 event happening in } dt \text{ interval at t per unit } dt) = \frac{p}{dt} = np = \lambda$ Hence $f_X(t) = e^{-\lambda t} \times \lambda$					
			$Real \in [0,\infty)$	$\lambda e^{-\lambda t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$1 - e^{-\lambda x}$	NA
			Property: 1. if $\tau \sim Exp(\lambda)$ where τ waiting till first event occurs/arrives from t = 0 then a. $(\tau - c) \sim Exp(\lambda)$, i.e. standing at c, waiting time still follow same distribution (hence same expectation and variance), This property of exponential distribution is called memorylessness property. b. $(c - \tau) \sim Exp(\lambda)$, i.e. standing at c, waiting time for seeing last occurred/arrived event also follows exact same distribution. (time reversal property).					

Sequence (S_n)

Definition	
Definition	Collection of Random Variables is called Sequence, i.e. $$S_n = \{X_0, X_1, \dots, X_{n-1}\} = \{ X_i \mid i \geq 1\}$$ There is no bound on n to be finite, it can be infinite as well. Sequence represents mapping of multiple experiments which are not necessarily independent to each other to an n-dimensional measurable space E^n,i.e., a Sequence S maps Ω_S (Sample space of Sequence) to n-dimensional measurable space E^n. Each of X_i can have same or different or even interdependent sample space Ω_i.

<table>
<tr><td>Sample Space, Ω_{S_n}</td><td>Cartesian product of Sample Space of each experiment
$$\Omega_{\mathrm{S_n}} = \Omega_{X_0 X_1 \ldots X_{n-1}} = \Omega_{X_0} \times \Omega_{\frac{X_1}{X_0}} \times \Omega_{\frac{X_2}{X_0,X_1}} \times \ldots \Omega_{\frac{X_{n-1}}{X_0,X_1,X_2,\ldots,X_{n-2}}}$$
Note: sample space of X_i depends upon the values all the random variables takes from $X_0 till\ X_{i-1}$</td></tr>
</table>

<table>
<tr><td rowspan="3">Measurable Space,</td><td colspan="2">Measurable Space of $S_n = \{X_0, X_1, \ldots, X_{n-1}\}$ would be E^n, such that:
$$E^n = \{(x_0^0, x_0^1, \ldots, x_0^{n-1}), \ldots, (x_{n-1}^0, x_{n-1}^1, \ldots, x_{n-1}^{n-1})\}$$
$$= \{s_0, s_1, \ldots, s_{(m-1)\times(n-1)}\}$$</td></tr>
<tr><td>Discrete</td><td>Continuous</td></tr>
<tr><td>$E^n = \{0,1,2,\ldots,m\} \times \{0,1,2,\ldots,m\} \times \ldots \times \{0,1,2,\ldots,m\}_{n\ times}$</td><td>$E^n = (R \times R \times \ldots \times R)_{n\ times}$</td></tr>
</table>

<table>
<tr><td>Event</td><td>Any possible subset of measurable space, E^n is called an Event. Represented by capital letters: A, B, C etc. For Example

Event A: Random Variable X_0 taking value x_0 and X_1 taking x_1 from sample $S = \{X_1, X_2\} = \{x_0, x_1\}$</td></tr>
</table>

Joint Probability Function, $P_{X_0X_1\dots X_{n-1}}(x_0, x_1, \dots, x_{n-1})$

Probability of Sequence S taking value s where $s \in E^n$, also called **Joint Probability (Distribution) Function** of all X_i is denoted by $P_{s_n}(s)$ or $P_{X_0X_1\dots X_{n-1}}(x_0, x_1, \dots, x_{n-1})$.

Discrete	Continuous
Probability of $X_1 \dots X_n$taking value $x_1, \dots, x_n$ is called **Joint Probability Mass Function,** $P_{X_1,X_2,\dots X_n}(x_1, \dots, x_n)$ and is represented as: $$P_{S_n}(s_n) = P_{X_1,X_2,\dots X_n}(x_1, \dots, x_n) = P_{X_1,X_2,\dots X_{n-1}}(x_0, \dots, x_{n-1}) \times P_{\frac{X_n}{X_1,X_2,\dots X_{n-1}}}\left(\frac{x_n}{x_0, \dots, x_{n-1}}\right) = P_{S_{n-1}}(s_{n-1}) \times P_{\frac{X_n}{X_1,X_2,\dots X_{n-1}}}\left(\frac{x_n}{x_0, \dots, x_{n-1}}\right)$$ where $P_{\frac{X_n}{X_1,X_2,\dots X_{n-1}}}\left(\frac{x_n}{x_0,\dots,x_{n-1}}\right)$ is the **Conditional Probability Mass Function** of X_n i.e. Its probability of it taking value x_n given $X_1, X_2, \dots X_{n-1}$ have taken values $x_0, \dots, x_{n-1}$. Alternatively, it can also be presented as $$P_{S_n}(s_n) = F_{S_n}(S_n) - F_{S_n}(S_n - 1) = F_{X_1,X_2,\dots X_n}(x_1, \dots, x_n) - F_{X_1,X_2,\dots X_n}(x_1 - 1, \dots, x_n - 1)$$	Probability of $X_1 \dots X_n$ taking value between $x_1, \dots, x_n$ to $x_1 + dx_1, \dots, x_n + dx_n$ per unit $dx_1 \dots dx_n$, is called **Joint Probability Density Function,** $f_{X_0X_1\dots X_{n-1}}(x_0, x_1, \dots, x_{n-1})$: $$f_S(s) = f_{X_1\dots X_n}(x_1, \dots, x_n) = f_{X_1\dots X_{n-1}}(x_1, \dots, x_{n-1}) \times f_{\frac{X_n}{X_1,X_2,\dots X_{n-1}}}\left(\frac{x_n}{x_0, \dots, x_{n-1}}\right) = f_{S_{n-1}}(s_{n-1}) \times f_{\frac{X_n}{X_1,X_2,\dots X_{n-1}}}\left(\frac{x_n}{x_0, \dots, x_{n-1}}\right)$$ Where $f_{\frac{X_n}{X_1,X_2,\dots X_{n-1}}}\left(\frac{x_n}{x_0,\dots,x_{n-1}}\right)$is **Conditional Probability Density Function** of X_n , i.e. prob of it taking value between $(x_n, x_n + dx_n)$ per unit dx_n given $X_1, X_2, \dots X_{n-1}$ have taken values between $x_1, \dots, x_n$ to $x_1 + dx_1, \dots, x_{n-1} + dx_{n-1}$ per unit $dx_1 \dots dx_{n-1}$. Probability of S taking value between s to $s + ds$ per unit ds, i.e. $$f_S(s) = \frac{dF_S(s)}{ds} = f_{X_0X_1\dots X_{n-1}}(x_0, x_1, \dots, x_{n-1}) = \frac{\partial^n F_{X_0X_1\dots X_{n-1}}(x_0, x_1, \dots, x_{n-1})}{\partial x_0 \partial x_1 \dots \partial x_{n-1}}$$ $$dF_S(s) = f_{X_0X_1\dots X_{n-1}}(x_0, x_1, \dots, x_{n-1}) dx_0 dx_1 \dots dx_{n-1}$$ Where $F_S(s)$ is Cumulative Probably Function

Two random variables: (X, Y)	$P_{XY}(x,y) = P_X(x) \times P_{\frac{Y}{X}}\left(\frac{x}{y}\right)$	$f_{XY}(x,y) = f_X(x) \times f_{\frac{Y}{X}}\left(\frac{x}{y}\right)$
Three Random Variables $(X, Y Z)$	$P_{XYZ}(x,y,z) = P_X(x) \times P_{\frac{Y}{X}}\left(\frac{x}{y}\right) \times P_{\frac{Z}{X,Y}}\left(\frac{z}{x,y}\right)$	$f_{XYZ}(x,y,z) = f_X(x) \times f_{\frac{Y}{X}}\left(\frac{x}{y}\right) \times f_{\frac{Z}{X,Y}}\left(\frac{z}{x,y}\right)$

Properties:

- Marginal probability functions can be obtained by summing/integrating the joint probability function across E_{X_i} where X_i is the random variable we want to remove:

	Discrete	Continuous
	$P_{X_1,X_2,\ldots X_{n-1}}(x_0, \ldots x_{k-1}, x_{k+1}, \ldots, x_n) = \sum_{x_k} P_{X_1,X_2,\ldots X_n}(x_1, \ldots, x_n)$	$f_{X_1,X_2,\ldots X_{n-1}}(x_0, \ldots x_{k-1}, x_{k+1}, \ldots, x_n) = \int_{x_k \in E_k} f_{X_1,X_2,\ldots X_n}(x_1, \ldots, x_n)\,dx_k$
Two random variables: (X, Y)	$P_X(x) = \sum_y P_{XY}(x,y)$	$f_X(x) = \int_{y \in E_y} f_{XY}(x,y)\,dy$
Three Random Variables $(X, Y Z)$	$P_{XY}(x,y) = \sum_z P_{XYZ}(x,y,z)$	$f_{XY}(x,y) = \int_{z \in E_z} f_{XYZ}(x,y,z)\,dz$

- Given n random variables: $X_1, \ldots, X_n$, probability distribution of another random variable Y, such that

$$Y = \sum X_i$$

  a. If each $X_i \sim N(\mu, \sigma\ )$ then $Y \sim N(\sum \mu_i, \sum \sigma_i^2)$
  b. If each $X_i \sim U(0, 1)$ then Y ~ Irwin Hall Distribution

$$f_Y(y) = \frac{1}{2(n-1)!} \sum_{i=0}^{n} (-1)^n \ \times {}^n_i C \times (y - i)^{n-1} sgn(y - i)$$

$$F_Y(y) = \int_0^y f_Y(y) dy$$

    i. Special case: for $y \leq 1$

$$F_Y(y) = \frac{y^n - n(y-1)^n}{n}$$

    i.e., probability of sum of n uniform distribution to be less than 1 is $\frac{1}{(n)!}$

- $Y = Max(\vee X_i)$
  a. If X_i are IIDs with each $X_i \sim U(0, 1)$ then $F_Y(y) = y^n$ and $f_Y(y) = ny^{n-1}$
     Proof:

$$F_Y(y) = P(Y \leq\ y) = P(Max(\vee X_i) \leq\ y)$$

     If max of all X_i is less than y then that mean all the X_i are also independently less than y

$$= \ P(\ X_1 \leq y\ \ \cap X_2 \leq y\ \ \cap \ldots \cap X_n < y)$$
$$= \ P(X_1 \leq y) P(X_2 \leq y) \ldots P(X_n \leq y\ )$$
$$= yy \ldots n\ times = y^n$$

- Given n random variables IID, what probably of a specific random variable to be minimal of all:
  a. Since all are IID, every possible value $x_1, x_2, \ldots, x_n$ is equally likely
  b. Out of all possible scenarios probably that a specific random variable will be minimum will be $\frac{(n-1)!}{n!} = \ \frac{1}{n}$ (Answer)

Joint Cumulative Distribution Function (CDF), $F_{X_1,X_2,\dots X_n}(x_1,\dots,x_n)$

Probability of $X_1 \le x_1 and\ \dots and\ X_n \le x_n$ is called **Joint Cumulative Distribution Function,** $F_{X_1,X_2,\dots X_n}(x_1,\dots,x_n)$ and is represented as:

$$F_{S_n}(s) = P_{S_n}(S_n \le s)$$

$$F_{X_1,X_2,\dots X_n}(x_1,\dots,x_n) = P(X_1 \le x_1 and\ \dots and\ X_n \le x_n)$$

Discrete	Continuous
$F_{X_1,X_2,\dots X_n}(x_1,\dots,x_n) = \sum_{x_1=E_{x_1 min}}^{x_1} \dots \sum_{x_n=E_{x_n min}}^{x_n} P_{X_1,X_2,\dots X_n}(x_1,\dots,x_n)$	$F_{X_1,X_2,\dots X_n}(x_1,\dots,x_n) = \int_{E_{x_1 min}\dots E_{x_n min}}^{x_1,\dots,x_n} dF_{X_1,X_2,\dots X_n}(x_1,\dots,x_n)$ $= \int_{E_{x_1 min}}^{x_1} \dots \int_{E_{x_n min}}^{x_n} f_{X_0 X_1 \dots X_{n-1}}(x_0,x_1,\dots,x_{n-1}) dx_0 dx_1 \dots dx_{n-1}$

Moment

Mean or Average, μ_X or Expectation, $E[X]$

Expected number of “experiments” (trials) until an “event” happens, assuming happening of other events does not change probability of “event”:

a. If “event” consists of only one element of an experiment then:
   i. Find the probability of that element, let’s call it p.
   ii. Using Geometric Brownian Motion, we know expectation of first occurrence of that event is

$$\frac{1}{p}$$

   Q1: What is the expected number of rolls of a fair die **until** a 6 turns up?
   A1: 6

b. If “event” consists of more than one elements (let’s say n elements) then:
   i. Let’s say Y_n is number of “experiments” until we will see encounter of all the n elements from “event”
   ii. Then we define X_i which represents number of “experiments” we will need to perform before we see **any** i^{th} element from “event” given $i-1$ elements of “event” already occurred.

$$X_i = Y_i - Y_{i-1}$$

   iii. Now we compute $P(X_i)$ which is probability of occurrence of i^{th} element from “event” given $i-1$ elements of “event” already occurred, since this only consists of single outcome, we can use (a) and say, Expected number of trials before we see any of i^{th} element from “event” given $i-1$ elements of “event” already occurred, is:

$$E[X_i] = \frac{1}{P(X_i)}$$

   iv. We know:

$$Y_n = X_1 + \cdots + X_n$$

   Hence,

$$E[Y_n] = \sum E[X_i] = \sum \frac{1}{P(X_i)}$$

   Q1: What is the expected number of rolls of a fair die until all 6 numbers turn up?
   A1: 14.7

2. Expected number of “experiments” (trials) until an “event” happens, assuming happening of other events does not change probability of “event”)

$$r\frac{K}{(N-K+1)}$$

3. Given an Experiment conducted N times, Expected number of experiment when an event will occur:
   a. Approach 1:
      Let's say k times that event occurs, then probability of that event occurring exactly k times:
      $$P(E_{k\ times}) = \sum_{i=k}^{n} (-1)^{i-k} \times {}_{k}^{i}C \times S_i$$
      Where S_i is sum of probabilities of that event happening any i times from n trials. (will include repeated the count for i+1, i+2, and so on times)
      $$S_i = \sum P(E_{j_1} \cap E_{j_2} \cap E_{j_3} \cap \ldots \cap E_{j_i})$$
      Then use definition of Expectation:
      $$E[Number\ of\ experiment\ with\ that\ event] = \sum_{k=1}^{N} k \times P(E_{k\ times})$$
   b. Approach 2:
      Let's defined n indicator variables, $I_1, I_2, \ldots . I_N$, where
      $$I_i = \begin{cases} 1, & if \text{ "event" } occus\ in\ \ i^{th}\ experiment \\ 0, & otherwise \end{cases}$$
      Hence if a random variable Y represent number of times event occurs in the experiment then
      $$Y = \sum_{i=1}^{N} I_i$$
      Applying Expectation both side & using
      $$E\left[\sum X_i\right] = \sum E[X_i]$$
      We get,
      $$E[Y] = \sum_{i=1}^{N} E[\, I_i] = \sum_{i=1}^{N} P(\, i^{th}\ experiment\ resulting\ in\ occurenc\ of \text{ event})$$
      Note: No need for any i^{th} and j^{th} experiments to be independent.

      Special case, if all experiments have same probability, p: (similar to Binomial Distribution)
      $$E[Y] = Np$$

4. Expected number of Experiments, to reach:

		• If $\{X_1, \dots, X_n\}$ are mutually independent then $$E[X_1 \times \dots \times X_n] = E[X_1] \times \dots \times E[X_n]$$ • If $\{X_1, \dots, X_n\}$ are IID with X: $$E[X_1 \times \dots \times X_n] = E[X]^n$$
	Variance	• Variance of each X_i in $S_n = \{X_1, \dots, X_n\}$ can be computed similar to single random number case. • Variance of sum of Random Variables $$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j) = \sum_{i=1}^{n} Var(X_i) + 2 \times \sum_{i=1}^{n}\sum_{j=i+1}^{n} Cov(X_i, X_j)$$ Special case: If $X_1, \dots, X_n$ are mutually independent then $$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var(X_i)$$

Covariance

Given $S_n = \{X_1, \ldots, X_n\}$, i.e. n random variables, we can define a covariance as measure of association or dependence between any two random variables. Unlike Variance, covariance can be either positive or negative.

$$Cov(X_i, X_j) = E\left[(X_i - \mu_{X_i})(X_j - \mu_{X_j})\right] = E[X_i X_j] - E[X_i][X_j]$$

where if $i == j$ then

$$Cov(X_i, X_i) = Var(X_i)$$

For S_n, we can define a covariance matrix, such that i^{th} row and j^{th} column represents their correlation:

$$\begin{array}{cc} & \begin{array}{cc} x & y \end{array} \\ \begin{array}{c} x \\ y \end{array} & \begin{bmatrix} var(x) & cov(x,y) \\ cov(x,y) & var(y) \end{bmatrix} \end{array} \quad \begin{array}{cc} & \begin{array}{ccc} x & y & z \end{array} \\ \begin{array}{c} x \\ y \\ z \end{array} & \begin{bmatrix} var(x) & cov(x,y) & cov(x,z) \\ cov(x,y) & var(y) & cov(y,z) \\ cov(x,z) & cov(y,z) & var(z) \end{bmatrix} \end{array}$$

Properties:

1. Covariance matrices are symmetric & positive semi definite.
2. If X_i and X_j are pairwise independent then
$$Cov(X_i, X_j) = 0$$
as $E[X_i X_j] = E[X_i][X_j]$
3. $Cov(X, X) = Var(X)$
4. $Cov(X, Y) = Cov(Y, X)$
5. $Cov(aX + b,\ pY + q) = apCov(X, Y)$
6. $Cov(aX + bY, cW + dV) = acCov(X, W) + adCov(X, V) + bcCov(Y, W) + bdCov(Y, V)$
7. $Cov(X, constant) = 0$

	Correlation	Given $S_n = \{X_1, \dots, X_n\}$, i.e. n random variables, Correlation between two random variables is the measure of linear association between X and Y. iT is given by the covariance, scaled by overall variability in two random variables. Its represented as $\rho_{X_i X_j}$ $$\rho_{X_i X_j} = \frac{Cov(X_i, X_i)}{\sqrt{Var(X_i)Var(X_j)}}$$ Properties: 1. If two random variables are pair-wise independent then their correlation is zero (converse is not true).

Classification Of Sequence

Independent

Pairwise Independent

$S_n = \{X_0, X_1, \dots, X_{n-1}\}$ is **pairwise independent** if either

1. Joint probability function is multiplication of residual probability functions:

Discrete	**Continuous**
$\left(P_{X_i,X_j}(x_i,x_j) = P_{X_i}(x_i) \times P_{X_j}(x_j) \bigvee x_i, x_j\right) \bigvee i,j$	$\left(f_{X_i,X_j}(x_i,x_j) = f_{X_i}(x_i) \times f_{X_j}(x_j) \bigvee x_i, x_j\right) \bigvee i,j$

2. Or Joint cumulative probability of any two random variables is equals to multiplication of their marginal cumulative probably:

$$\left(F_{X_i,X_j}(x_i,x_j) = F_{X_i}(x_i).F_{X_j}(x_j) \bigvee x_i, x_j\right) \bigvee i,j$$

Where $F_{X_i}(x_i)$ is Cumulative Distribution Function of X_i at x_i
$F_{X_j}(x_j)$ is Cumulative Distribution Function of X_j at x_j and
$F_{X_i,X_j}(x_i,x_j)$ is Joint Cumulative Distribution Function of X_i and X_j taking values x_i and x_j respectively.

Properties:

1. If $\{X_1, \dots, X_n\}$ is **pairwise independent** then

$$E[X_i \times X_j] = E[X_i] \times E[X_j] \bigvee i,j$$

Mutually Independent

$S_n = \{X_1, \dots, X_n\}$ is **mutually independent** if either

1.

Discrete	**Continuous**
$P_{\{X_1,\dots,X_n\}}(x_1, \dots, x_n) = P_{X_1}(x_1) \times \dots \times P_{X_n}(x_n) \bigvee x_1, \dots x_n$	$f_{\{X_1,\dots,X_n\}}(x_1, \dots, x_n) = f_{X_1}(x_1) \times \dots \times f_{X_n}(x_n) \bigvee x_1, \dots x_n$

2. $F_{\{X_1,\dots,X_n\}}(x_1, \dots, x_n) = F_{X_1}(x_1) \times \dots \times F_{X_n}(x_n)$

Properties:

1. If $\{X_1, \dots, X_n\}$ is **mutually independent** then

$$E[X_1 \times \dots \times X_n] = E[X_1] \times \dots \times E[X_n]$$

2. Even if the set of random variables is pairwise independent, it is not necessarily mutually independent.

			Independent and Identically Distributed	When all the random variables in $S_n = \{X_0, X_1, \dots, X_{n-1}\}$ are both 1. Mutually Independent and 2. Identically distributed, i.e. **Discrete:** $P_{X_{i_0}}(x_{i_0}) = P_{X_{i_1}}(x_{i_1}) = \cdots = P_{X_{i_{n-1}}}(x_{i_{n-1}}) \bigvee x_{i_0}, \dots, x_{i_{n-1}}$ **Continuous:** $f_{X_{i_o}}(x_{i_0}) = f_{X_{i_1}}(x_{i_1}) = \cdots = f_{X_{i_{k-1}}}(x_{i_{n-1}}) \bigvee x_{i_0}, \dots, x_{i_{n-1}}$ Then Sequence is called Independent and Identically distributed. This property is usually abbreviated as *i.i.d.* or *iid* or *IID*. Properties: Example: 1. Flipping coin n times and recoding number of times head appears. 2. Choosing a card from a deck of 52 n times and recording number of times king appears. 3. Binomial Distribution.

Estimators

When all the random variables in $S_n = \{X_1, \dots, X_n\}$ are IID to X $E[X] = \mu_X$ and $Var[X] = \sigma_X^2$ then we can defined Estimators for them

Sample Mean $\overline{X}$

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Properties:

1. $E[\overline{X}] = \mu_X$
2. $Var[\overline{X}] = \frac{\sigma_X^2}{n}$
3. For large value of n, $\overline{X}$ (itself a random variable) tends to Normal distribution with μ_X mean and $\frac{\sigma_X^2}{n}$ variance, i.e.
$$\overline{X} = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{n} = N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$$
    a. It will follow all properties of Normal distribution, i.e.
$$P\left(-\infty < \overline{X} < \mu_X - \frac{3\sigma_X}{\sqrt{n}}\right) = 0.13\%$$
$$P\left(\mu_X - \frac{3\sigma_X}{\sqrt{n}} < \overline{X} < \mu_X - \frac{2\sigma_X}{\sqrt{n}}\right) = 2.14\%$$
$$P\left(\mu_X - \frac{2\sigma_X}{\sqrt{n}} < \overline{X} < \mu_X - \frac{\sigma_X}{\sqrt{n}}\right) = 13.6\%$$
$$P\left(\mu_X - \frac{\sigma_X}{\sqrt{n}} < \overline{X} < \mu_X\right) = 34.13\%$$
    And so on.
    b. Given a distribution which can be broken into n independent variables, use it to approximate to Normal and then solve. If its discrete distribution use constant correction. (l-0.5, l+0.5)
4. For large numbers $\overline{X}$ $tends\ to$ μ_x (weak law of large number & strong law of large number).

Sample Variance $\overline{S}^2$

a. Biased:

$$\overline{S}^2 = \frac{1}{n} \times \sum_{i=1}^{n} (X_i - \overline{X})^2$$

b. Unbiased:

$$\overline{S}^2 = \frac{1}{n-1} \times \sum_{i=1}^{n} (X_i - \overline{X})^2$$

Sample Std Deviation $\overline{S}$

$$\overline{S} = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (X_i - \overline{X})^2}$$

Sample Covariance $\overline{q}$:

Given two IIDs: $\{X_1, \dots, X_n\}$ and $\{Y_1, \dots, Y_n\}$

$$\overline{q} = \frac{1}{n} \times \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})$$

Sample Covariance Matrix:

Given k IIDs: $\{X_1^1, \dots, X_n^1\} \dots \{X_1^k, \dots, X_n^k\}$, let's denote them by sample vectors $\{\overrightarrow{X^1}, \dots, \overrightarrow{X^k}\}$, each vector with sample mean vector $\overline{\vec{X}} = \{\overline{\overrightarrow{X_1}}, \dots, \overline{\overrightarrow{X_k}}\}$

$\overline{Q}$ is $(k \times k)$ matrix with element $\overline{q_{ij}}$:

Creating M $(n \times k)$ matrix each column representing $\left[\overrightarrow{X^1}, \dots, \overrightarrow{X^k}\right]$ then

					$$Q = \frac{1}{n} \times \left(M - 1_{N\times1}\overline{\vec{X}}\right)^T \times \left(M - 1_{N\times1}\overline{\vec{X}}\right)$$ Property: Q is positive semidefinte, i.e. for any $(k \times 1)$, $\vec{v} = [v_1 \dots v_k]$ vector: $$\vec{v}^T Q \vec{v} \geq 0 \; for \; all \; \vec{v}$$ $$= \vec{v}^T \times \frac{1}{n} \times \left(M - 1_{N\times1}\overline{\vec{X}}\right)^T \times \left(M - 1_{N\times1}\overline{\vec{X}}\right) \times \vec{v}$$ $$= \frac{1}{n} \times \left(\vec{v}^T \times \left(M - 1_{N\times1}\overline{\vec{X}}\right)^T \times \left(M - 1_{N\times1}\overline{\vec{X}}\right) \times \vec{v}\right)$$ $$= \frac{1}{n} \times \left(\left(\left(M - 1_{N\times1}\overline{\vec{X}}\right) \times \vec{v}\right)^T \times \left(M - 1_{N\times1}\overline{\vec{X}}\right) \times \vec{v}\right)$$ $$= Sum \; of \; square \; of \; elements \; which \; is \; always \geq 0$$

Martingale

The definition of a martingale has three items. A sequence (Xn)n≥1 is said to be martingale with respect to a filtration (Fn)n≥1 if

- (Xn)n≥1 is adapted to (Fn)n≥1, i.e. Xn is Fn-measurable for each n,
- XnXn is integrable, i.e. E[|Xn|]<∞ for each nn,
- and XnXn satisfies the martingale condition, i.e. E[Xn+1|Fn]=Xn for each n.

So in order for you to answer the question of when (Sn)n≥1 is a martingale you need to address the first two bullets first. Let us therefore assume that all variables are integrable, and that the filtration we are working with is indeed the natural filtration, i.e.

Fn=σ(S0,Z1,...,Zn),n≥1.

Then you're correct that you just have to show when E[Sn+1|Fn]=Sn for all nn. You correctly calculated that

E[Sn+1|Fn]=E[Sn|Fn]+E[Zn+1|Fn]=Sn+E[Zn+1|Fn]

and hence [Sn+1|Fn]=Sn if and only if [Zn+1|Fn]=0. All that is left is to recognize E[Zn+1|Fn] as E[Zn+1] due to the independence between Zn+1 and Fn.

Symmetric random walk is martingale

$$S_n = X_1 + X_2 + \cdots + X_n$$

Where each $X_i \sim 2 \times Bern\left(\frac{1}{2}\right) - 1$, i.e. takes ± 1 each with half probability.

Properties:

1. $E[S_{n+1}|F_n] = S_n$, i.e. S_n is martingale.
2. $E[S_{n+1}^2 - (n+1)] = S_n^2 - n$

3. If τ is stopping time for random walk, i.e. $S_\tau = X_0 + \cdots + X_\tau$ then still S_τ is martingale, i.e.

   a. If we are given that random walk with stop at either $S_\tau = a$ or $S_\tau = b$ (starting from 0, if its not zero then subtract initial state from both a and b) then we can assume that p is probability that either of one states will be reached (let's say a) then

   $$E[S_\tau] = p \times a + (1-p) \times b = 0$$

   b. Expected Number of step taken to reach there (i.e. $E[\tau]$)

					$E[S_\tau^2 - \tau] = 0$ $E[\tau] = E[S_\tau^2]$

Dependent

$S_n = \{X_0, X_1, \dots, X_{n-1}\}$ are called **dependent random variables** if occurrence of one random variable impacts the sample space of rest of random variables (and hence the probability), If random variables are dependent then:

	Discrete	Continuous
	$P_{\{X_{i_0}, X_{i_1}, \dots, X_{i_{k-1}}\}}(x_{i_0}, \dots, x_{i_{k-1}}) = P_{\{X_{i_0}, X_{i_1}, \dots, X_{i_{k-2}}\}}(x_{i_0}, \dots, x_{i_{k-2}}) \times P_{\frac{X_{i_{k-1}}}{X_{i_0}, X_{i_1}, \dots, X_{i_{k-2}}}}\left(\frac{x_{i_{k-1}}}{x_{i_0}, \dots, x_{i_{k-2}}}\right)$ $\bigvee x_{i_0}, \dots, x_{i_{k-1}} \bigvee k$	$f_{\{X_{i_0}, X_{i_1}, \dots, X_{i_{k-1}}\}}(x_{i_0}, \dots, x_{i_{k-1}}) = f_{\{X_{i_0}, X_{i_1}, \dots, X_{i_{k-2}}\}}(x_{i_0}, \dots, x_{i_{k-2}}) \times f_{\frac{X_{i_{k-1}}}{X_{i_0}, X_{i_1}, \dots, X_{i_{k-2}}}}\left(\frac{x_{i_{k-1}}}{x_{i_0}, \dots, x_{i_{k-2}}}\right)$ $\bigvee x_{i_0}, \dots, x_{i_{k-1}} \bigvee k$
For 2 random variables: X, Y	$P_{XY}(x, y) = P_X(x) \times P_{\frac{Y}{X}}\left(\frac{x}{y}\right)$	$f_{XY}(x, y) = f_X(x) \times f_{\frac{Y}{X}}\left(\frac{x}{y}\right)$

		Markov Chain	The sequence of dependent random variables satisfying the Markov property is called the **Markov chain**. A Markov Chain is a random process that moves from one state to another such that the next state of the process depends only on where the process is at present. Each transition is called a step. In a Markov chain, the next step of the process depends only on the present state and it does not matter how the process reaches the current state. In other words, it is "memoryless." An absorbing state is a state that is impossible to leave once reached. A state that is not absorbing is called a transient state. If every state of a Markov chain can reach an absorbing state, this Markov chain is called an absorbing Markov chain. Markov Property: to make predictions of the behaviour of a system in the future, it suffices to consider only the present state of the system and not the past history, i.e.

Discrete	Continuous
$P_{S_n}(s_n) = P_{X_1,X_2,\ldots X_n}(x_1,\ldots,x_n)$ $= P_{X_1,X_2,\ldots X_{n-1}}(x_0,\ldots,x_{n-1}) \times P_{\frac{X_n}{X_{n-1}}}\left(\frac{x_n}{x_{n-1}}\right)$ $= P_{S_{n-1}}(s_{n-1}) \times P_{\frac{X_n}{X_{n-1}}}\left(\frac{x_n}{x_{n-1}}\right)$	$f_S(s) = f_{X_1\ldots X_n}(x_1,\ldots,x_n)$ $= f_{X_1\ldots X_{n-1}}(x_1,\ldots,x_{n-1}) \times f_{\frac{X_n}{X_{n-1}}}\left(\frac{x_n}{x_{n-1}}\right)$ $= f_{S_{n-1}}(s_{n-1}) \times f_{\frac{X_n}{X_{n-1}}}\left(\frac{x_n}{x_{n-1}}\right)$

Solving Probability, Expectation, Variance, Covariance Problems:

1. Distributions
2. Sequences: $X_1 + X_2 + \cdots + X_n$ or $X_1 \times X_2 \times \ldots \times X_n$
   a. If sequence can be a symmetric random walk then use martingale + reflection.

Example: At a theater ticket office, 2n people are waiting to buy tickets. n of them have only \$5 bills and the other n people have only \$10 bills. The ticket seller has no change to start with. Ifeach person buys one \$5 ticket, what is the probability that all people will be able to buy their tickets without having to change positions?

   b. Let's say we are asked, Y = number of trials to contain K distinct items then

$$Y = X_1 + X_2 + \cdots + X_k$$

   Where X_i is number of trials to contain i^{th} item given $i-1$ items already found.

Example 1: Coupon collection problem: There are N distinct types ofcoupons in cereal boxes and each type, independent ofprior selections, is equally likely to be in a box. If a child wants to collect a complete set of coupons with at least one of each type, how many coupons (boxes) on average are needed to make such a complete set?

   c. $Y = number\ of\ distinct\ items\ found\ in\ n\ trials\ out\ of\ K\ distinct\ items$

$$Y = I_1 + \cdots + I_K$$

   Where I_i is indicator variable that indicates that i^{th} item found in n trials..

$$E[Y] = \sum P(ith\ item\ found\ in\ n\ trials)$$

$$P(ith\ item\ found\ in\ n\ trials) = out\ of\ n\ trials, ith\ item\ was\ there\ in\ at\ least\ one\ trial$$
$$= 1 - P(no\ trial\ had\ ith\ item)$$

   Example 1: If the child has collected n coupons, what is the expected number of distinct coupon types?

   Example 2: Number of times 2 appears in n trials is X and number of times 3 appears is Y then find cov between X and Y, now we know individually they follow binomial distribution but it will be computationally heavy to get their joint distribution, so best to break down binomial to sum of n Bernoulli distribution (indicator variables).

   Example: N people sit around a round table, where N > 5. Each person tosses a coin. Anyone whose outcome is different from his/her two neighbours will receive a present. Let X be the number of people who receive presents. Find E[X] and Var(X).
   https://www.probabilitycourse.com/chapter6/6_1_2_sums_random_variables.php

d. Y = Starting with \$1, if we get heads we double the money and tails then halve, expectation of payoff after n trials.
   Notice even though payoff is dependent upon prev accumulated payoff, tosses are still independent, so we can design random variables as

$$Y = S_0 \times X_1 \times X_2 \times \dots \times X_n$$

   Such that X_i presents either 2 or 1/2 and $S_0 = 1$, now since each coin tosses are independent we can use

$$E[Y] = E[X_0] \times E[X_1] \times \dots \times E[X_n]$$

3. Markov (both probability + expectation).
4. Martingale
5. Bayer's Theorem + Inclusion exclusion principle.
   a. If two random variables are independent then
      i. $P\left(\frac{X}{Y}\right) = P(X)$
      ii. $E[XY] = E[X]E[Y]$
      iii. $Var(X + Y) = Var(X) + Var(Y)$

Process or Stochastic Process

Collection of continuous Random Variables is called Process:

$$P_t = \{ X_t |\, t \in R \text{ and } t \geq 0\}$$

Random Process P maps Ω_P (*Sample Space* of Process) to t-dimensional measurable space E^t

Sample Space, Ω

Cartesian product of *Sample Space* of each Experiment

$$\Omega_P = \left\{ \prod \Omega_t \,|\, t \geq 0 \right\}$$

Measurable Space, E

$$\omega_t = \{ E^t |\, t \in R \text{ and } t \geq 0\}$$