Probability:

(1) Sample space:

Random experiment = Experiment in which:
- All outcomes of the experiment are known in advance.
- Any performance of the exp yields an outcome that is not known in advance.
- The experiment can be repeated under identical conditions.

Ω: Set of all possible outcomes of the experiment.

σ-field: A collection $\mathcal{F}$ of subsets of Ω ST:
- $\Omega \in \mathcal{F}$
- $A \in \mathcal{F} \Rightarrow A^c = \Omega - A \in \mathcal{F}$
- $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

→ $\mathcal{F}$ is closed WRT countable union and complements.

Prop: • $\phi \in \mathcal{F}$ (*) $\Omega \in \mathcal{F} \Rightarrow \phi = \Omega^c \in \mathcal{F}$
• $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$ (**) $A_1, \ldots \in \mathcal{F} \Rightarrow A_1^c, \ldots \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n^c \in \mathcal{F} \Rightarrow \bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c \in \mathcal{F}$ (De Morgan)

Sample/Measurable Space = of an experiment is a pair $(\Omega, \mathcal{F})$ where: Ω = set of all possible outcome of exp, and $\mathcal{F}$ a σ-field of subsets of Ω.
↳ Measurable Sets: Elements of $\mathcal{F}$.

Sample points / elementary outcomes: Elements of Ω.

Event: Any set $A \in \mathcal{F}$ (So $A \in \Omega$) → "Event A happens if the outcome of the exp corresponds to a point in A".
↳ Simple/Elementary event: one-point set event: $A = \{\omega\}$ for $\omega \in \Omega$.

- Finite Sample Space: $(\Omega, \mathcal{F})$ where Ω contains a finite # of pts. Ex: Coin Toss: $\Omega = \{H, T\}$
- Discrete Sample Space: $(\Omega, \mathcal{F})$ where Ω contains at most a countable # of pts. Ex: Toss coin until you see H: $\Omega = \{H, TH, TTH, \ldots\}$
- Uncountable Sample Space: $(\Omega, \mathcal{F})$ where Ω contains uncountably many pts. Ex: lifetime of a randomly selected light bulb from a production line: $\Omega = (0, \infty)$.
- Continuous Sample space: $(\Omega, \mathcal{F})$ where $\Omega = \mathbb{R}^k$ or some rectangle in $\mathbb{R}^k$

Note: • Ω contains at most a countable # of pts: can take $\mathcal{F}$ = class of all subsets of Ω. ⇒ Every subset of Ω is an Event.
• Ω has uncountably many points: set of all subsets of Ω is still a σ-field, but too large

De Morgan's Law: $\boxed{\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c}$ and $\boxed{\left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c}$

Limsup/inf: $\{A_n : n \geq 1\}$ a sequence of Events:

- $\boxed{\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k}$ is a set, hence an event: $\omega \in \limsup_{n\to\infty} A_n \Rightarrow \forall n \geq 1, \exists k \geq n$ ST A_k occurs.
  → ∞ many of the A_n occur
  "$\omega \in$ at least one A_k"
- $\boxed{\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k}$ is a set, hence an event. $\omega \in \liminf_{n\to\infty} A_n \Rightarrow \exists n \geq 1$ ST A_k occurs $\forall k \geq n$
  → only finitely many of the A_n DO NOT occur.
- $\lim_{n\to\infty} A_n = A \Longleftrightarrow \limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = A$

(II) Kolmogorov's Probability axioms & Consequences:

Probability (measure): $(\Omega, \mathcal{F})$ a measurable space associated with a random exp: a set function $P: \mathcal{F} \to \mathbb{R}$ is a probability measure if:

(1) $P(A) \geq 0 \;\; \forall A \in \mathcal{F}$

(2) $P(\Omega) = 1$

(3) $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ for any sequence $\{A_n\}_{n=1}^{\infty}$ of disjoint events ($A_i \cap A_j = \phi$ if $i \neq j$) $\rightsquigarrow$ Countable additivity property!

- $P(A)$ = the probability of event A (occuring), for any $A \in \mathcal{F}$
- $(\Omega, \mathcal{F}, P)$ = Probability Space

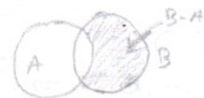
Consequences: $(\Omega, \mathcal{F}, P)$ a prob. space, $A, B \in \mathcal{F}$ events:

① $P(\phi) = 0$ $\quad$ (∵) $1 = P(\Omega) = P(\Omega \cup \phi) \overset{(3)}{=} P(\Omega) + P(\phi) = 1 + P(\phi) \Rightarrow P(\phi) = 1 - 1 = 0$

② $P(A^c) = 1 - P(A)$ $\quad$ (∵) $1 = P(\Omega) = P(A \cup A^c) \overset{(3)}{=} P(A) + P(A^c) \Rightarrow P(A^c) = 1 - P(A)$

③ $P(B - A) = P(B) - P(A)$ if $A \subseteq B$ $\quad$ (∵)

$A \subseteq B$ $\quad$ $P(B) = P(A \cup (B-A)) = P(A) + P(B-A)$



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

in general: $P(B - A) = P(B) - P(A \cap B)$

④ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ $\quad$ (∵)

$P(A \cup B) = P(①) + P(②) + P(③)$

⑤ $P(A \cup B) \leq P(A) + P(B)$ $\quad$ (∵) by ④, $P(A) + P(B) - P(A \cup B) = P(A \cap B) \geq 0$. (exists since A, B are events.)

⑥ $A \subseteq B \Rightarrow P(A) \leq P(B)$, hence $0 \leq P(A) \leq 1$ $\quad$ (∵) ③ $\Rightarrow P(B) - P(A) = P(B - A) \geq 0 \Rightarrow P(B) \geq P(A)$. $\phi \subseteq A \subseteq \Omega \Rightarrow P(\phi) \leq P(A) \leq P(\Omega)$ $\Rightarrow 0 \leq P(A) \leq 1$.

⑦ $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{\substack{i,j=1 \\ i<j}}^{n} P(A_i \cap A_j) + \sum_{\substack{i,j,k=1 \\ i<j<k}}^{n} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n+1} P(\bigcap_{k=1}^{n} A_k)$ $\rightsquigarrow$ Inclusion/Exclusion principle

(∵) $n=2$ ✓ $\rightsquigarrow$ inductio.

Note: • Ω = Sure event, ϕ = impossible event

• if $P(A) = 1$; A = Certain event, $P(A) = 0$: A = null event

⚠ • $P(A) = 1 \not\Rightarrow A = \Omega$
• $P(A) = 0 \not\Rightarrow A = \phi$

Bonferroni's ineq: $A_1, \dots, A_n \in \mathcal{F}$: $\sum_{i=1}^{n} P(A_i) - \sum_{i<j}^{n} P(A_i \cap A_j) \leq P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$

(∵) $n=2$: $P(A_1) + P(A_2) - P(A_1 \cap A_2) = P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$
↳ inductio.

Boole's ineq: $A_1, \dots, A_n \in \mathcal{F}$:

① $P(A_1 \cap A_2) \geq 1 - P(A_1^c) - P(A_2^c)$ $\quad$ (∵) by 2nd ineq

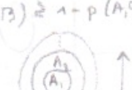
② $P(\bigcap_{i=1}^{n} A_i) \geq 1 - \sum_{i=1}^{n} P(A_i^c)$ $\quad$ (∵) $1 - P(\bigcap_{i=1}^{n} A_i) = P((\bigcap_{i=1}^{n} A_i)^c) \overset{\text{de Morgan}}{=} P(\bigcup_{i=1}^{n} A_i^c) \overset{\text{Bonferroni}}{\leq} \sum_{i=1}^{n} P(A_i^c)$

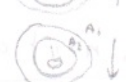
③ $P(\bigcap_{i=1}^{\infty} A_i) \geq 1 - \sum_{i=1}^{\infty} P(A_i^c)$ $\quad$ (∵) $\bigcap_{i=1}^{\infty} A_i = (\bigcup_{i=1}^{\infty} A_i^c)^c \in \mathcal{F}$ is an event.

Or $A = A_1$, $B = \bigcap_{i=2}^{\infty} A_i \Rightarrow P(A \cap B) \geq 1 - P(A_1^c) - P(\bigcap_{i=2}^{\infty} A_i^c) \geq \cdots$

Def: A sequence of events $\{A_n\}_{n=1}^{\infty}$ is: • non decreasing if $A_n \subseteq A_{n+1} \; \forall n \geq 1$. So $\lim_{n \to \infty} A_n = \bigcup_{i=1}^{\infty} A_i$

• non increasing if $A_n \supseteq A_{n+1} \; \forall n \geq 1$. So $\lim_{n \to \infty} A_n = \bigcap_{i=1}^{\infty} A_i$





thm: $(\Omega, \mathcal{F}, P)$ a prob space, $\{A_n\}_{n=1}^{\infty}$ a seq of events:

(1) $\{A_n\}$ non decreasing $\Longrightarrow P(\bigcup_{n=1}^{\infty} A_n) = P(\lim_{n \to \infty} A_n) = \lim_{n \to \infty} P(A_n)$

(2) $\{A_n\}$ non increasing $\longrightarrow P(\bigcap_{n=1}^{\infty} A_n) = P(\lim_{n \to \infty} A_n) = \lim_{n \to \infty} P(A_n)$

(∵) $A_0 = \phi$: $\bigcup_{n=1}^{\infty} A_n = A_1 \cup (A_2 - A_1) \cup \dots \cup (A_n - A_{n-1}) = \bigcup_{n=1}^{\infty} (A_n - A_{n-1}) \Rightarrow$ disjoint union

So $P(\bigcup_{n=1}^{\infty} A_n) = P(\bigcup_{n=1}^{\infty} (A_n - A_{n-1})) \overset{(3)}{=} \sum_{n=1}^{\infty} P(A_n - A_{n-1}) = \lim_{n \to \infty} \sum_{i=1}^{n} P(A_i - A_{i-1})$
$= \lim_{n \to \infty} (P(A_n) - P(A_0))$
$= \lim_{n \to \infty} P(A_n)$

(∵) A_n non increasing $\Rightarrow A_n^c$ non decreasing.

corr: $\{A_n\}_{n=1}^{\infty}$ a seq of events : $P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$

(∵) $B_n = \bigcup_{i=1}^{n} A_i \Rightarrow \{B_n\}_{n=1}^{\infty}$ = non decreasing seq, and $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$

So $P(\bigcup_{n=1}^{\infty} A_n) = P(\bigcup_{n=1}^{\infty} B_n) = \lim_{n \to \infty} P(B_n) = \lim_{n \to \infty} (\bigcup_{i=1}^{n} A_i) \leq \lim_{n \to \infty} \sum_{i=1}^{n} P(A_i) = \sum_{n=1}^{\infty} A_n$

thm: $\{A_n\}_{n=1}^{\infty}$ a seq of events, $\lim_{n \to \infty} A_n = A$: $P(\lim_{n \to \infty} A_n) = \lim_{n \to \infty} P(A_n) = P(A)$

(∵) $\lim_{n \to \infty} A_n = A \; \exists \Rightarrow \limsup A_n = \liminf A_n = A$

$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} B_n$ with B_n non decreasing $\Rightarrow P(\lim_{n \to \infty} A_n) = P(A) = P(\limsup A_n) = P(\bigcap_{n=1}^{\infty} B_n) = \lim_{n \to \infty} P(B_n) = \lim_{n \to \infty} P(\bigcup_{k=n}^{\infty} A_k) \geq \lim_{n \to \infty} P(A_n)$

$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} B_n$ with B_n non increasing $\Rightarrow P(\lim_{n \to \infty} A_n) = P(A) = P(\liminf A_n) = P(\bigcup_{n=1}^{\infty} B_n) = \lim P(B_n) = \lim_{n \to \infty} P(\bigcap_{k=n}^{\infty} A_k) \leq \lim_{n \to \infty} P(A_n)$

So $P(\lim_{n \to \infty} A_n) \leq \lim_{n \to \infty} P(A_n)$ & $P(\lim_{n \to \infty} A_n) \geq \lim_{n \to \infty} P(A_n)$, So Equality!

thm: A random experiment with M distinct elementary outcomes ; $\Omega = \{w_1, \dots, w_M\}$. If w_i's are equally likely: $P(w_i) = \frac{1}{M}$ $\Rightarrow$ Uniform discrete prob model

(pf) $1 = P(\Omega) = \sum_{i=1}^{M} P(w_i) = \sum_{i=1}^{M} c = Mc \Rightarrow c = 1/M$

thm: $(\Omega, \mathcal{F}, P)$: $\Omega = \{w_1, \dots, w_M\}$ with probability model $P(w_i) = \frac{1}{M}$; $\forall$ event $A \in \mathcal{F}$, $P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{M}$

(pf) $A = \{w_i : i \in I\}$, $|A| \leq |\Omega| = M \Rightarrow P(A) = P(\bigcup_{i \in I} \{w_i\}) = \sum_{i \in I} P(w_i) = \sum_{i \in I} \frac{1}{M} = \frac{1}{M} |A| = \frac{|A|}{|\Omega|}$

III Combinatorics

$\Omega = \{\omega_1, \dots, \omega_n\}$ finite, $\mathcal{F}$ = set of all subsets of Ω, $P: \mathcal{F} \to \mathbb{R}$ ST $P(A) = \sum_{\omega_j \in A} \underset{\text{elementary events}}{P(\omega_j)}$ $\forall A \in \mathcal{F}$ event.

Def: An assignment of probability is equally likely (or uniform) if $p(\omega_j) = \frac{1}{|\Omega|}$ $\forall 1 \le j \le |\Omega|$

Prop: $p(A) = \frac{|A|}{|\Omega|}$ $\forall A \in \mathcal{F}$ event.

Rule 1:
$A_1, \dots, A_k$ tasks ; n_i ways to perform task A_i $\Longrightarrow$ total # of ways to perform tasks $(A_1, A_2, \dots, A_k)$ together $= n_1 \cdot n_2 \cdots n_k$

Note: each $\omega_i \in \Omega$ is: $\omega_i = \{a_1^{(i)}, a_2^{(i)}, \dots, a_k^{(i)}\}$.

Ex:
- r distinguishable balls in n cells: $\omega_i = \{a_1^{(i)}, \dots, a_r^{(i)}\}$
- Toss a coin r times: $p(0 \text{ heads}) = p(TTT\cdots T) = \left(\frac{1}{2}\right)^r$
- Roll a die r times: $P(0 \text{ Six}) = P(\underset{5 \text{ choices}}{\cdot, \cdot, \cdot, \dots, \cdot}) = \left(\frac{5}{6}\right)^r$

Rule 2: (ordered samples)
Any ordered arrangement $(a_{i_1}, a_{i_2}, \dots, a_{i_r})$ of r elements taken from a set of n elements $a_1, \dots, a_n$ $\to$ r-tuples

- Sampling with replacement: (repetit° permitted) $\Longrightarrow$ $\boxed{\#\text{ samples of size } r = n^r}$
- Sampling without replacement: (no repetitions) $\Longrightarrow$ $\boxed{\#\text{ Samples of Size } r = {}_nP_r = \frac{n!}{(n-r)!} \quad (r<n)}$ Permutations

Ex:
- Sample of size r of a set of n elements with replacement: $p(\text{every elements are distinct}) = \frac{{}_nP_r}{n^r}$
- n balls randomly placed in n holes $\Rightarrow$ $p(\text{each cell is occupied}) = \frac{n!}{n^n}$

Rule 3: (different subsets)

- Subsets without replacement: $\boxed{\#\text{ different unordered populations of size } r \le n \text{ from a pop° of } n \text{ elements} = \binom{n}{r} = \frac{n!}{r!(n-r)!} = C_r^n}$ Combinations
- Subsets with replacement: $\boxed{\#\text{ different unordered samples of size } r \le n \text{ from collect° of } n \text{ objects} = \binom{n+r-1}{r}}$

Ex: r balls in n cells: $A_k^{(i)} = \{k \text{ balls in } i^{th} \text{ cell}\}$, $k = 0, \dots, r$ $\Rightarrow$ $p(A_k) = p(\text{"choose } k \text{ balls"}) \cdot p(\text{"distribute the remaining } (r-k) \text{ balls in } (n-1) \text{ cells"})$
$\Rightarrow p(A_k) = \frac{\binom{r}{k} \cdot (n-1)^{r-k}}{n^r}$

Rule 4: (different partitions)
Populat° of n elements:

$$\boxed{\#\text{ ways we can partition the pop° into } k \text{ Subpop° of Sizes } r_1, r_2, \dots, r_k \text{ respectively} = \binom{n}{r_1, \dots, r_k} = \frac{n!}{r_1! \, r_2! \cdots r_k!} \quad \begin{pmatrix} r_1 + \dots + r_k = n \\ 0 \le r_i \le n \end{pmatrix}}$$
$\to$ multinomial coeff

Ex: $p(\text{"a hand of 13 cards contains 2 spade, 7 hearts, 3 diamonds, 1 club"}) = \frac{\binom{13}{2} \cdot \binom{13}{7} \cdot \binom{13}{3} \cdot \binom{13}{1}}{\binom{52}{13}}$

Ex: Urn with 5 red, 3 Green, 2 blue, 4 white balls: $p(\text{"select } 2R, 2G, 1B, 3W\text{"}) = \frac{\binom{5}{2}\binom{3}{2}\binom{2}{1}\binom{4}{3}}{\binom{14}{8}}$

Example: (Birthday problem)

class of r students: $\begin{cases} P(\text{"All birthdays are distinct"}) = \binom{365}{r} = \frac{365!}{r!(365-r)!} \\ P(\text{"At least 2 students have same birthday"}) = 1 - \frac{365!}{r!(365-r)!} \end{cases}$

$(\because)$ $\Omega = \{\omega_i = (d_1, d_2, \dots, d_r)\}$, $d_i \in \{1, 2, \dots, 365\}$, $i = 1, 2, \dots, r$

Example: (capture/recapture problem)
N objects in a box: T are tagged, $N-T$ are untagged $\rightsquigarrow$ Random Sample of Size n

$$p(x \text{ objects are tagged}) = \frac{\binom{T}{x} \cdot \binom{N-T}{n-x}}{\binom{N}{n}}$$

$(\because)$ # ways to choose x tagged objects $= \dfrac{\left(\begin{matrix}\#\text{ ways to choose } x \\ x \text{ tagged objects from } T \text{ objects}\end{matrix}\right) \cdot \left(\begin{matrix}\#\text{ ways to choose } (n-x) \\ \text{untagged obj from } N-T \\ \text{untagged objects}\end{matrix}\right)}{\left(\begin{matrix}\#\text{ ways to chose } n \text{ objects} \\ \text{from } N \text{ objects}\end{matrix}\right)}$

disjoint events so multiply.

So $\max\{0; n+T-N\} \le x \le \min(T, n) \in \mathbb{N}$

$(\because)$ $\begin{cases} 0 \le n-x \le N-T \\ 0 \le x \le T \end{cases} \Rightarrow \begin{cases} n+a-N \le x \le n \\ 0 \le x \le T \end{cases}$

IV Conditional probabilities:

→ We know that event B has occurred: want to use the info to make a probability statement about event A.

Conditional probability: $(\Omega, \mathcal{F}, P)$ a prob space, $A, B \in \mathcal{F}$, $P(B) > 0$:

conditional probability of A given B: $\boxed{P(A|B) = P_B(A) = \dfrac{P(A \cap B)}{P(B)}}$ $(P(B) \neq 0)$ $\forall A \in \mathcal{F}$

Note: $P(A \cap B | B) = P(A|B)$

Prop: (multiplication rule)

- $\boxed{P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)}$ $\forall A, B \in \mathcal{F}$, $P(A), P(B) > 0$
- $\boxed{P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P\left(A_n \middle| \bigcap_{i=1}^{n-1} A_i\right)}$

(∞) induction on n: $n=2$ ✓

$P\left(\bigcap_{i=1}^{n+1} A_i\right) = P\left(\left(\bigcap_{i=1}^{n} A_i\right) \cap A_{n+1}\right)$

$= P\left(\bigcap_{i=1}^{n} A_i\right) P\left(A_{n+1} \middle| \bigcap_{i=1}^{n} A_i\right)$

Ex: $\boxed{\begin{matrix} 6R \\ 4G \end{matrix}}$ * seq of 4 marbles without replacement $\Rightarrow P(\underset{A_1}{G}\,\underset{A_2}{G}\,\underset{A_3}{R}\,\underset{A_4}{G}) = P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1) P(A_2|A_1) P(A_3|A_1 \cap A_2) P(A_4|A_1 \cap A_2 \cap A_3)$

$= \frac{4}{10} \cdot \frac{3}{9} \cdot \frac{6}{8} \cdot \frac{2}{7} = \frac{1}{35}$.

Prop: (Total probability rule) $(\Omega, \mathcal{F}, P)$ a prob space: $B_i \in \mathcal{F}$, $i = 1, 2, \cdots$, B_i pairwise disjoint ($B_i \cap B_j = \emptyset$ $(i \neq j)$), $\Omega = \bigcup_{i=1}^{\infty} B_i$

$\boxed{P(A) = \sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(A|B_i)\, P(B_i)}$ $\forall A \in \mathcal{F}$

(∞) $A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{\infty} B_i\right) = \underbrace{\bigcup_{i=1}^{\infty} A \cap B_i}_{\text{pairwise disjoint}}$

Prop: (Baye's rule) $(\Omega, \mathcal{F}, P)$ a prob space: $B_i \in \mathcal{F}$, $i = 1, 2, \ldots$, B_i pairwise disjoint

$\boxed{P(B_k|A) = \dfrac{P(A|B_k)\, P(B_k)}{\sum_{i=1}^{\infty} P(A|B_i)\, P(B_i)}, \quad k = 1, 2, \cdots}$ $\forall A \in \mathcal{F}$, $P(A) > 0$

Example:

- Drug test: 99% Sensitivity (for a drug user [event "U"], the test is ⊕ 99% of the time) → $P(+|U) = 99\%$
  98% Specificity (for a non drug user, the test is ⊖ 98% of the time) → $P(-|U^c) = 98\%$

  ↳ 0,6% of employees use the drug $\Rightarrow P(+) \overset{\text{Tot prob}}{=} P(+|U)\, P(U) + \underbrace{P(+|U^c)}_{=1-P(-|U^c)}\, \underbrace{P(U^c)}_{=1-P(U)} = 0{,}99 \cdot 0{,}6\% + 0{,}02 \cdot 99{,}4\% \Rightarrow P(+) = 2{,}582\%$

  $\Rightarrow P(U|+) = \dfrac{P(+|U)\, P(U)}{P(+)} = \dfrac{0{,}99 \cdot 0{,}06}{2{,}582\%} \Rightarrow P(U|+) = 23{,}01\%$

- Medical test: • False negative person: the person IS a disease carrier but the test is ⊖
  • False positive person: the person is NOT a disease carrier but the test is ⊕

  → 7% of pop° are disease carriers $\Rightarrow P(C) = 7\%$
  → if disease carrier → 1% have ⊖ test $\Rightarrow P(-|C) = 1\%$
  → if Not disease carrier → 2,5% have ⊕ test $\Rightarrow P(+|C^c) = 2{,}5\%$

  $P(C|+) = \dfrac{P(+|C)\, P(C)}{P(+|C)P(C) + P(+|C^c)P(C^c)} = \dfrac{0{,}99 \cdot 0{,}07}{0{,}99 \cdot 0{,}07 + 0{,}025 \cdot 0{,}93}$

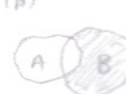
  $\Rightarrow P(C|+) = 74{,}9\%$

Prop:

- $\boxed{P(A^c|B) = 1 - P(A|B)}$

  (∞) $P(A^c|B) = \dfrac{P(A^c \cap B)}{P(B)} = \dfrac{P(B - A)}{P(B)} = \dfrac{P(B) - P(A \cap B)}{P(B)} = 1 - \dfrac{P(A \cap B)}{P(B)} = 1 - P(A|B)$

- $P(A|B^c) \neq 1 - P(A|B)$ in general



(∞) $P(A|B^c) = \dfrac{P(A \cap B^c)}{P(B^c)} = \dfrac{P(A) - P(A \cap B)}{1 - P(B)} \neq 1 - \dfrac{P(A \cap B)}{P(B)}$

V) Independence of events :

Independent events : $(\Omega, \mathcal{F}, p)$ a probability space;

the events A and B in $\mathcal{F}$ are independent if $\boxed{p(A \cap B) = p(A)\, p(B)}$ ⇝ def still valid for $p(A)=0$ or $p(B)=0$

Prop : A and B are independent if : $p(A|B) = p(A)$ for $p(B) > 0$; <u>OR</u> $p(B|A) = p(B)$ for $p(A) > 0$

Prop : A & B are indep : $p(A \cap B) = p(A)p(B)$

$\iff$ A & B^c are indep : $p(A \cap B^c) = p(A)p(B^c)$

$\iff$ A^c & B are indep : $p(A^c \cap B) = p(A^c)p(B)$

$\iff$ A^c & B^c are indep : $p(A^c \cap B^c) = p(A^c)p(B^c)$

(∵) $p(A \cap B^c) = p(A - B) = p(A - B \cap A)$
$= p(A) - p(B \cap A)$ (∵ $B \cap A \subseteq A$)
$= p(A) - p(A)p(B)$
$= p(A)(1 - p(B)) = p(A)p(B^c)$

Mutual independence : $(\Omega, \mathcal{F}, p)$: $C = \{A : A \in \mathcal{F}\}$ a collection of events

Events in C are mutually independant $\iff$ for <u>any</u> subcollection $\{A_{i_1}, \dots, A_{i_K}\}$ in C : $p\left(\bigcap_{j=1}^{K} A_{i_j}\right) = \prod_{j=1}^{K} p(A_{i_j})$

⚠ "Independent" events $\neq$ "disjoint" events !

- A, B disjoint : $A \cap B = \phi$ so $p(A \cap B) = p(\phi) = 0$, Hence $0 = p(A \cap B) = p(A)p(B) \iff \{p(A)=0$, $p(B)=0$ → A & B can only be independent if either $p(A)=0$ or $p(B)=0$
- A, B independent : $p(A \cap B) = p(A)p(B)$: $p(A \cup B) = p(A) + p(B) - p(A \cap B) = p(A) + p(B) - p(A)p(B)$

A, B disjoint $\Rightarrow$ $p(A \cup B) = p(A) + p(B)$ $\Rightarrow$ $p(A)p(B) = 0$ → A & B are disjoint only for either $p(A)=0$ or $p(B)=0$

→ $A, B \in \mathcal{F}$, $P(A) > 0$, $P(B) > 0$: A & B are disjoint ⇒ they are <u>NOT</u> independent.
A & B are independent ⇒ they are <u>NOT</u> disjoint.

Examples :

- 4 tagged marbles : W B Y WBY , E_i = "the symbol "i" appears on the marble" $(i = \{B, W, Y\})$

$P(E_W) = P(E_B) = P(E_Y) = \frac{1}{2}$

$P(E_W \cap E_B) = P(E_W \cap E_Y) = p(E_B \cap E_Y) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}$ ⇝ <u>pairwise</u> independent events !

$P(E_W \cap E_B \cap E_Y) = \frac{1}{4} \neq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = P(E_W)p(E_Y)p(E_B)$ ⇝ <u>NOT</u> mutually independent

- $\Omega = \{1, 2, 3, 4\}$: $p(1) = \frac{\sqrt{2}}{2} - \frac{1}{4}$; $p(2) = \frac{1}{4}$; $P(3) = \frac{3}{4} - \frac{\sqrt{2}}{2}$; $P(4) = \frac{1}{4}$

$E_1 = \{1, 3\}$, $E_2 = \{2, 3\}$, $E_3 = \{3, 4\}$

$P(E_1) = \frac{1}{2}$, $P(E_2) = 1 - \frac{\sqrt{2}}{2}$, $P(E_3) = 1 - \frac{\sqrt{2}}{2}$

$P(E_1 \cap E_2) = p\{3\} \neq p(E_1)p(E_2)$; $p(E_1 \cap E_3) = p\{3\} \neq p(E_1 \cap E_3)$; $p(E_2 \cap E_3) = p(3) \neq p(E_2)p(E_3)$ → <u>NOT</u> pairwise independent.

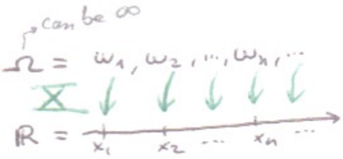
$P(E_1 \cap E_2 \cap E_3) = p(\{3\}) = \frac{3}{4} - \frac{\sqrt{2}}{2} = \frac{1}{2}\left(1 - \frac{\sqrt{2}}{2}\right)\left(1 - \frac{\sqrt{2}}{2}\right) = p(E_1)p(E_2)p(E_3)$ ⇝ "triplewise" independent

→ <u>NOT</u> mutually indep.

→ <u>NOT</u> mutually indep

Random Variables & Their Prob. distribution:

(I) Random Variables:



$\Rightarrow X : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B})$

Random Variable: $(\Omega, \mathcal{F})$ a measurable space: the finite, single-valued function $X : \Omega \to \mathbb{R}$ is a RV if: $X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F}$ for all Borel set $B \in \mathcal{B}$

Thm: X is an RV $\Leftrightarrow X^{-1}((-\infty, x]) := \{\omega : X(\omega) \leq x\} := \{X \leq x\} \in \mathcal{F} \ \forall x \in \mathbb{R}$.
→ i.e. $\mathcal{F}$ is Borel measurable.

Note: if X is an RV: the sets $\{X = x\}$, $\{X < x\}$, $\{a < X < b\}$, $\{a < X \leq b\}$, $\{a < X < b\}$, $\{a \leq X \leq b\}$ are all events!

Ex: • Indicator function (of a set A): A set $A \subseteq \Omega$: $I_A(\omega) = \begin{cases} 0, & \omega \notin A \\ 1, & \omega \in A \end{cases}$ ⇝ I_A is an RV $\Leftrightarrow A \in \mathcal{F}$

• Toss a coin once: $\Omega = \{H, T\}$, $X :=$ # of H observed ⇝ $\begin{cases} X(H) = 1 \\ X(T) = 0 \end{cases}$ is an RV   (∵) $X^{-1}(-\infty, x] = \begin{cases} \phi, & x < 0 \\ \{T\}, & 0 \leq x < 1 \\ \{H, T\}, & 1 \leq x \end{cases}$ $\Rightarrow X^{-1}(-\infty, x] \in \mathcal{F}$

• Toss a coin Twice: $\Omega = \{HH, HT, TH, TT\}$, $X :=$ # of H observed ⇝ $\begin{cases} X(HH) = 2 \\ X(HT) = X(TH) = 1 \\ X(TT) = 0 \end{cases}$ is an RV   (∵) $X^{-1}(-\infty, x] = \begin{cases} \phi, & x < 0 \\ \{TT\}, & 0 \leq x < 1 \\ \Omega - \{HH\}, & 1 \leq x < 2 \\ \Omega, & 2 \leq x \end{cases}$

Thm: • If $(\Omega, \mathcal{F})$ is a discrete measurable space with Ω a countable set of pts, $\mathcal{F}$ = class of all subsets of Ω. Then every function $X : \Omega \to \mathbb{R}$ is a RV

• X a RV on $(\Omega, \mathcal{F}) \Rightarrow aX + b$ is an RV on $(\Omega, \mathcal{F})$ ($a, b \in \mathbb{R}$: constants!), X^2 is an RV on $(\Omega, \mathcal{F})$, $\frac{1}{X}$ is an RV on $(\Omega, \mathcal{F})$ if $\{X = 0\} = \phi$

⇝ when working with an RV: only knowing the form of X is not enough ⇒ Need a prob dist of X!

(II) Probability distribution of a Random Variable:

Cumulative distribution f°: X an RV on $(\Omega, \mathcal{F}, P)$: the funct° $F(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) \ \forall x \in \mathbb{R}$ is the CDF of X → See Caratheodory's Extension Thm

Prop: (1) F is a non decreasing f°: $x < y \Rightarrow F(x) \leq F(y)$ (∵) $F(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) \leq P(\{\omega \in \Omega : X(\omega) \leq y\}) = P(X \leq y) = F(y)$ since the events: $\{\omega \in \Omega : X(\omega) \leq x\} \subseteq \{\omega \in \Omega : X(\omega) \leq y\}$ for $x \leq y$

(2) F is a right continuous f°: if $\{x_n\}$ decreases to x, then $\lim_{n\to\infty} F(x_n) = F(x)$

(3) $F(-\infty) = 0$ and $F(+\infty) = 1$

(4) F has at most a countable set of discontinuities (∵) F is monotonically increasing

Proof: (2) • $A_n = \{X \leq x_n\} = \{X \in (-\infty, x_n)\}$ where $x_n \searrow x \Rightarrow A_n$ is a decreasing seq of events $\Rightarrow \lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n = \{X \leq x\} = A$

So $\lim_{n\to\infty} A_n = A \Rightarrow \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P(X \leq x_n) = \lim_{n\to\infty} P(A_n) \underset{\text{by continuity of the prob}}{=} P(\lim_{n\to\infty} A_n) = P(A) = P(X \leq x) = F(x)$

⇒ ⚠ • If $y_n \nearrow x$: $B_n = \{X \leq y_n\} = \{X \in (-\infty, y_n)\}$ so B_n is an increasing seq of events $\Rightarrow \lim_{n\to\infty} B_n = \bigcup_{n=1}^{\infty} B_n = \{X < x\} = B$ (strict)

So $\lim_{n\to\infty} F(y_n) = P(X < x) = F(x^-) \neq F(x)$ in general (strict)

(3) • $\{x_n\}$ ST $x_n \to -\infty$: then the events $C_n = \{X \leq x_n\}$ form a ↘ seq of events $\Rightarrow \{\omega \in \Omega : X(\omega) \leq x_n\} = \bigcap_{n=1}^{\infty} C_n = \lim_{n\to\infty} C_n = \phi$

$\Rightarrow \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P(C_n) = P(\lim_{n\to\infty}(C_n)) = P(\phi) = 0$.

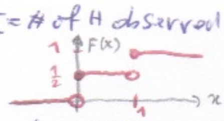
• $\{x_n\}$ ST $x_n \to +\infty$: then the events $D_n = \{X \leq x_n\}$ form a ↗ seq of events:

$\Rightarrow \{\omega \in \Omega : X(\omega) \leq x_n\} = \bigcup_{n=1}^{\infty} D_n = \lim_{n\to\infty} D_n = \{X(\omega) : X(\omega) < \infty\} = \Omega \Rightarrow \lim_{n\to\infty} F(y_n) = \lim_{n\to\infty} P(D_n) = P(\lim_{n\to\infty} D_n) = P(\Omega) = 1$

Discrete RV:

• Random exp: Toss a fair coin twice: $\Omega = \{H, T\}$, X = # of H observed.

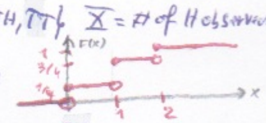
$\begin{cases} X(H) = 1 \\ X(T) = 0 \end{cases}$



$F(x) = P(X \leq x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}, & 0 \leq x < 1 \\ 1, & 1 \leq x \end{cases}$

• Random exp: Toss a coin twice: $\Omega = \{HH, HT, TH, TT\}$, X = # of H observed

$\begin{cases} X(HH) = 2 \\ X(HT) = X(TH) = 1 \\ X(TT) = 0 \end{cases}$



$F(x) = P(X \leq x) = \begin{cases} 0, & x < 0 \\ 1/4, & 0 \leq x < 1 \\ 3/4, & 1 \leq x < 2 \\ 1, & 2 \leq x \end{cases}$

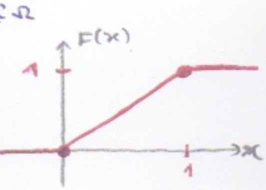
Continuous RV:

$\Omega = [0, 1]$: $P(I) = \text{Length}(I) \ \forall I \subseteq \Omega$

$X(\omega) = \omega \in [0, 1]$

$F(x) = P(X \leq x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x < 1 \\ 1, & 1 \leq x \end{cases}$

Prob to pick X in $[0, x]$ = x if $x \in [0, 1]$



Discrete Random Variables:

I Discrete RV:

Discrete RV: a RV $\mathbb{X}$ on $(\Omega, \mathcal{F}, p)$ that takes at most a countable set of values

Support of a RV: The Support of $\mathbb{X}$ is the set of possible values of $\mathbb{X}$: $S = \{x_1, x_2, \dots\} \subseteq \mathbb{R}$ must be countable!

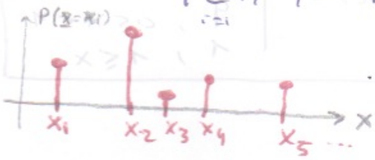
Probability mass function of a discrete RV:

the PMF of $\mathbb{X}$ is the collection of numbers $\{P_i\}$ Satisfying $\begin{cases} 0 \leq P_i \leq 1, & i=1,2,\dots \\ \sum_{i=1}^{\infty} P_i = p(S) = p(\Omega) = 1 \end{cases}$, where: $P_i = p(\mathbb{X} = x_i)$ for all $x_i \in S$.

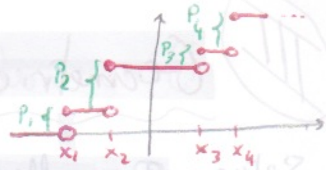
$\Rightarrow$ the PMF of $\mathbb{X}$ is:

x	x_1	x_2	...
$P(\mathbb{X}=x)$	P_1	P_2	...

$\rightsquigarrow$



Note: • The CDF of $\mathbb{X}$ is given by: $F(x) = P(\mathbb{X} \leq x) = \sum_{x_i \leq x} P_i = \begin{cases} 0, & x < x_1 \\ P_1, & x_1 \leq x < x_2 \\ P_1 + P_2, & x_2 \leq x < x_3 \\ \vdots \\ 1, & x \to \infty \end{cases}$



• $P_i = P(\mathbb{X} = x_i) = P(\mathbb{X} \leq x_i) - p(\mathbb{X} < x_i) = F(x_i) - F(x_i^-)$ ↳ Left limit

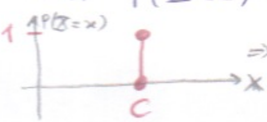
• If $A \subseteq S$ (Support of $\mathbb{X}$): $P(\mathbb{X} \in A) = \sum_{x_i \in A} P(\mathbb{X} = x_i) = \sum_{x_i \in A} P_i$

→ Ex: Toss a coin twice: $\mathbb{X}(HH) = 2$, $\mathbb{X}(HT) = \mathbb{X}(TH) = 1$, $\mathbb{X}(TT) = 0$
So $S = \{0,1,2\}$: let $A = \{0,1\} \Rightarrow P(\mathbb{X} \in A) = P(\mathbb{X}=0) + P(\mathbb{X}=1) = \frac{1}{4} + \frac{1}{2}$
$\Rightarrow P(\mathbb{X} \in A) = \frac{3}{4}$

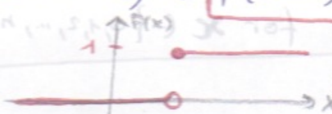
Degenerate RV:

Define: $\mathbb{X} : \Omega \to \mathbb{R}$ ST $\boxed{\mathbb{X}(\omega) = c, \forall \omega \in \Omega}$ ($c \in \mathbb{R}$ a constant)

then $P(\mathbb{X} = c) = P(\{\omega \in \Omega : \mathbb{X}(\omega) = c\}) \Rightarrow \boxed{P(\mathbb{X} = c) = 1}$ $\Rightarrow \boxed{F(x) = \begin{cases} 0, & x < c \\ 1, & c \leq x \end{cases}}$



$\Rightarrow$ PMF of $\mathbb{X}$



$\Rightarrow$ CDF of $\mathbb{X}$

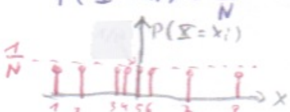
Discrete Uniform RV:

Let $\mathbb{X} : \Omega \to \mathbb{R}$, and take WLOG: $S = \{1, 2, 3, \dots, N\}$ for $N < \infty$ ST $P(\mathbb{X} = x_i) = \frac{1}{N}$, $i = 1, 2, \dots N$

PMF of $\mathbb{X}$:

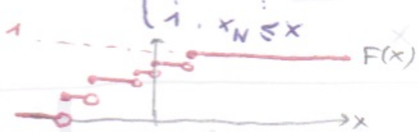
$S = \{1, 2, \dots, N\}$

$P(\mathbb{X} = x_i) = \frac{1}{N}$



CDF of $\mathbb{X}$:

$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{N}, & 0 \leq x < x_1 \\ \frac{2}{N}, & x_1 \leq x < x_2 \\ \vdots \\ 1, & x_N \leq x \end{cases}$



Discrete Uniform RV

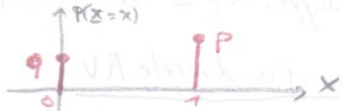
Ex: Used to catch Scientific fraud in papers (+ use Benford's Law)

II Bernoulli RV: (Jacob Bernoulli) $\longrightarrow$ $\boxed{X \sim Ber(p)}$

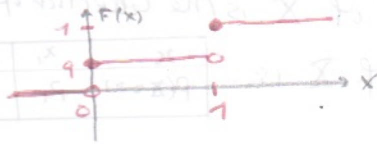
Setup: A random experiment with 2 outcomes: "success" and "failure" $\Longrightarrow$ $\boxed{\Omega = \{S, F\}}$

RV: $X : \Omega \to \mathbb{R}$, $\boxed{\begin{cases} X(S) = 1 \\ X(F) = 0 \end{cases}}$ and $\boxed{\begin{cases} P(X=1) = p \\ P(X=0) = q = 1-p \end{cases}}$ $\begin{matrix} \longrightarrow \text{prob of "success"} \\ \longrightarrow \text{prob of "failure"} \end{matrix}$

PMF of X: $\boxed{p(X=x) = p^x (1-p)^{1-x} \quad \text{for } x \in \{0,1\}}$



CDF of X: $\boxed{F(x) = \begin{cases} 0 & , x < 0 \\ q & , 0 \le x < 1 \\ 1 & , 1 \le x \end{cases}}$ $\longrightarrow (\text{as}) \; 1 = p + q$



III Binomial RV: $\longrightarrow$ $\boxed{X \sim Bin(n, p)}$

Setup: ① n independent Bernoulli experiments with outcomes: "success" or "failure".

② $\begin{cases} P(S_i) = p & , i = 1, \dots, n \\ P(F_i) = 1-p & , i = 1, \dots, n \end{cases}$ $\begin{matrix} \longrightarrow \text{Prob of "success" of the } i^{th} \text{ experiment} \\ \longrightarrow \text{prob of "failure" of the } i^{th} \text{ experiment} \end{matrix}$ $\boxed{0 < p < 1}$

③ $\boxed{X = \# \text{ of successes in } n \text{ trials}}$ $\longrightarrow S = \{0, 1, 2, \dots, n\}$

PMF of X: $\boxed{P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x \in \{0, 1, 2, \dots, n\}}$

(as) • Experiment: 1, 2, 3, ..., n, each with outcomes S or F

$P(x \text{ Succ and } (n-x) \text{ fail}) \propto p^x q^{n-x} = p^x (1-p)^{n-x}$: the proportionality factor: "success" can happen anywhere among the exp $\to$ choose x "succ" among the n experiments $\leadsto \binom{n}{x}$.

• This is a PMF: $\sum_{x=0}^{n} P(X=x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (p + (1-p))^n = 1^n = 1$

CDF of X: $\boxed{F(x) = \begin{cases} 0 & , x < 0 \\ \sum_{x=0}^{i} \binom{n}{x} p^x (1-p)^{n-x} & , i \le x < i+1 \quad \text{for } i = 0, 1, \dots, n-1 \\ 1 & , n \le x \end{cases}}$

$i \in \mathbb{N}$

Application: • A compagny sells screws in packages of 10

• If 2 or more screws in a package are defective, the compagny refunds the custommer

• P("a screw is defective") = 0,01 independently of the other screws.

Q: what proportion of packages will the compagny replace?

A: $X := \#$ of defective screws in a package of 10 screws

So $X \sim Bin(10, 0{,}01)$ where "success" = "defective screw" with prob $p = 0{,}01$.

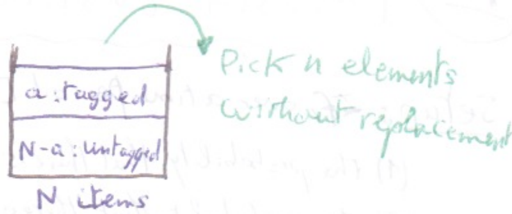
$\Rightarrow p(X = x) = \binom{10}{x} 0{,}01^x \cdot 0{,}99^{10-x}$

$\Rightarrow p(X \ge 2) = 1 - p(0 \le X < 2) = 1 - p(0 \le X \le 1) = 1 - p(X=0) - p(X=1) = 1 - 0{,}99^{10} - 10 \cdot 0{,}01 \cdot 0{,}99^9$

$\Rightarrow P(X \ge 2) = 0{,}4\%$

IV Hypergeometric RV:

Setup: Box with N items, a of which are tagged.
Take a sample of size n without replacement.



RV: $\boxed{X := \#\text{ tagged items in the Sample of size } n}$

PMF of X:
$$\boxed{P(\underset{\in \mathbb{N}}{X = x}) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}, \quad \text{for } \underbrace{\max\{0, a+n-N\} \leq x \leq \min\{a, n\}}_{\text{support of } X}}$$

CDF of X:
$$\boxed{F(x) = \begin{cases} \sum\limits_{x=m}^{m'} \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}, & \text{for } i \leq x < i+1, \ m \leq i \leq M \\ 0, & x < m \\ 1, & M \leq x \end{cases}}$$
where: $\begin{cases} m = \max\{0, a+n-N\} \\ M = \min\{a, n\} \end{cases}$

Prop: If $N \to \infty$ and $a \to \infty$, with $\underbrace{\frac{a}{N}}_{\text{proportion of tagged people}} \to p \in (0,1)$ and n fixed:
$$\Rightarrow \boxed{P(X = x) \longrightarrow \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{for } 0 \leq x \leq n} \Longrightarrow \text{Binomial Approx}$$

Conclusion: In a hypergeometric distribution: if $a, N \gg n$ then the PMF $\longrightarrow Bin(n, p = \frac{a}{N})$
→ if the Sample is huge, it does not matter if we do the sampling with or without replacement!

Application: • A large city of size $N = 10$ million: 30% of the popᵒ has a genetic defect.
• Select a sample of $n = 20$ pers. without replacement

Q: prob ("5 pers. in the Sample have the defect") = ?

A: (1) Exact: $N = 10^7, a = 3 \cdot 10^6, n = 20, x = 5 \Rightarrow P(X=5) = \frac{\binom{a}{5}\binom{N-a}{15}}{\binom{N}{20}} \simeq 0{,}1788632$

(2) Binomial approx: $P(X=5) \sim Bin(n=20, p = \frac{3 \cdot 10^6}{10^7}) = \binom{20}{5} 0{,}3^5 \cdot 0{,}7^{15} \simeq 0{,}1788631$

V Geometric RV:

Setup: Bernoulli Trial with prob of "Success" $= p$.

RV: $\boxed{X = \#\text{ of Trials needed to observe the } 1^{st} \text{ Success}}$

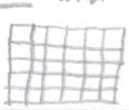
PMF of X: $\boxed{P(X = x) = p(1-p)^{x-1}, \quad x = 1, 2, 3, \ldots}$

(∞) 1 success, $(x-1)$ failures
it's a PMF: $\sum\limits_{x=1}^{\infty} p(1-p)^{x-1} = p \sum\limits_{x=0}^{\infty} (1-p)^x = p \frac{1}{1-(1-p)} = p \cdot \frac{1}{p} = 1$

Alternate Representatioᵒ: • RV: $\boxed{Y = \#\text{ of failures until we see the } 1^{st} \text{ Success}}$
• PMF: $\boxed{P(Y = y) = p(1-p)^y, \quad y = 0, 1, 2, \ldots}$ (∞) y failures and 1 Success.

Application: • An oil company digs oil wells:
• Prob of finding oil ("Success") at different localisatᵒ of an area is CST and the outcomes are indep.

Q: How many wells must the company dig to get the 1^{st} Success?



⇝ discretize the region.

VI Poisson distribution : $X \sim \text{Poiss}(\lambda)$

Recall : $o(h)$: $\lim_{h\to 0} \frac{o(h)}{h} = 0$

Setup: If, over a time period $[0,t]$, the occurrence of something verifies:

(1) the probability that there is exactly 1 occurrence in an interval of length h (small) is $\lambda h + o(h)$

(2) the probability that there are more than 1 occurences in an interval of length h (small) is $o(h)$

(3) the nb of occurrences in disjoint subintervals are independent.

RV: $\boxed{N(t) = \# \text{ of occurrences in the time interval } [0,t]}$

thm: $P(N(t)=k) = \frac{1}{k!} e^{-\lambda t} (\lambda t)^k$, $k = 0,1,2,\dots$ $\quad \lambda$ = rate at which the occurrence happens

Proof: $0, \frac{t}{n}, \frac{2t}{n}, \dots, \frac{(n-1)t}{n}, t = \frac{nt}{n}$: let $h = \frac{t}{n}$: then if $n \to \infty$, $h \to 0$

Event A = "k subintervals contain exactly 1 occurrence, and $(N-k)$ remaining subintervals contain 0 occurrences"

Event B = " $N(t)=k$ and at least 1 subinterval contains 2 or more occurrence"

$\Rightarrow p(N(t)=k) = p(A) + p(B)$

• let B_i = "ith subinterval contains 2 or more occurrences"

So $0 \le p(B) \le p(\bigcup_{i=1}^n B_i) = \sum_{i=1}^n p(B_i) = \sum_{i=1}^n o(h) = \sum_{i=1}^n o(\frac{t}{n}) = n \cdot o(\frac{t}{n}) = n \cdot (\frac{t}{n}) \cdot \frac{o(t/n)}{t/n} \to t \cdot 0 = 0$ as $n \to \infty$. $\Rightarrow P(B) \to 0$

• $P(A) = \binom{n}{k} \cdot \underbrace{[\lambda \cdot (\frac{t}{n}) + o(h)]^k}_{\text{Prob of success}} \cdot \underbrace{[1 - \lambda \cdot (\frac{t}{n}) - o(h)]^{n-k}}_{\text{Prob of failure}} \longrightarrow \frac{e^{-\lambda t}}{k!}(\lambda t)^k$ as $n \to \infty$ $\Rightarrow p(N(t)=k) = p(A) = \frac{e^{-\lambda t}}{k!}(\lambda t)^k$ as $n \to \infty$.

RV: $\boxed{X = \# \text{ of occurrences in a time interval} : S = \{0,1,2,3,\dots\}}$

→ # of earthquakes in a region of US in a month
→ # of radioactive decays in 1 second
→ # of hits in a website in 1 day
→ # of soldiers killed in Prussian war due to horse kicks

PMF: $\boxed{P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \ x = 0,1,2,\dots, \ \lambda > 0 : \text{(constant) parameter of the Poisson distribution}}$

↳ rate at which the event occurs

Ex: λ = rate at which customers enter a shop

(✓) $\sum_{x=0}^{\infty} P(X=x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} \cdot e^{\lambda} = 1$

thm: The Poisson distribution approximates the binomial distribution if $\left.\begin{matrix} n \to \infty \\ p_n \to 0 \end{matrix}\right\}$ ST $n p_n \to \lambda > 0$ (CST)

i.e. $P(X = x) = \binom{n}{x} p_n^x (1-p_n)^{n-x} \longrightarrow \frac{e^{-\lambda} \lambda^x}{x!}$ $\left(\begin{matrix} n \to \infty \\ p_n \to 0 \end{matrix} \text{ and } np_n \to \lambda > 0\right)$

Bin : $x = 0,1,2,\dots,n$

Poisson : $x = 0,1,2,\dots,\infty$ as $n \to \infty$

Application: Due to serious defects : $\left\{\begin{matrix} n = 10'000 \text{ cars are recalled} \\ p = 0{,}05\% \text{ is the prob that a car has the defect.} \end{matrix}\right.$

• Prob that more than 10 cars in the sample have the defect :

Y = # cars with the defect among the $n = 10'000$ cars

so $Y \sim \text{Bin}(10\,000; 0{,}0005) \Rightarrow P(10 \le Y \le 10\,000) = 1 - P(0 \le Y \le 9) = 1 - \sum_{y=0}^{9} P(Y=y) \simeq 3{,}179\,\%$

OR: $X \sim \text{Poiss}(\lambda = np = 5)$ (n.) np = expected nb of defective cars in sample.

so $P(10 \le Y \le 10\,000) = 1 - P(0 \le Y \le 9) \approx 1 - P(0 \le X \le 9) = 1 - \sum_{x=0}^{9} \frac{e^{-5} 5^x}{x!} \simeq 3{,}183\,\%$

• Prob that there are no defective cars in the sample :

$P(Y=0) = \binom{10'000}{0} 0{,}0005^0 \cdot (1 - 0{,}0005)^{10'000} \simeq 0{,}6729\,\%$

$P(Y=0) \simeq P(X=0) = \frac{e^{-5} 5^0}{0!} = e^{-5} \simeq 0{,}6737\,\%$

Continuous Random Variables:

I) Continuous Random Variables:

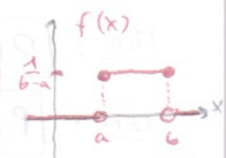
Continuous RV = a RV that takes an uncountable set of values: it can take any value in an interval.

Probability density function of a Cont. RV: (PDF) A nonnegative function $f(x)$ ST: $\begin{cases} F(x) = \int_{-\infty}^{x} f(t)\,dt, & \forall x \in \mathbb{R} \\ f(x) \geq 0, & \forall x \in \mathbb{R} \end{cases}$

Note:
- $\lim_{x \to \infty} F(x) = \int_{-\infty}^{\infty} f(x)\,dx = 1$
- $F(b) - F(a) = P(a < X \leq b) = \int_a^b f(x)\,dx = P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b)$
- $F'(x) = \frac{d}{dx} F(x) = f(x)$
- For small δ: $P(a \leq X \leq a+\delta) \simeq f(a)\cdot\delta \Rightarrow f(a) \simeq \frac{P(a \leq X \leq a+\delta)}{\delta}$ ⤳ f is a density!
- For any RV X: $P(X = x) = F(x) - F(x^-) = P(X \leq x) - P(X < x) = 0 \quad \forall x \in \mathbb{R}$.
  ↳ X is a continuous RV $\Rightarrow P(X = x) = 0 \ \forall x \in \mathbb{R}$.
- $f \in \mathbb{R}$ but $0 \leq F(x) \leq 1 \quad \forall x \in \mathbb{R}$.

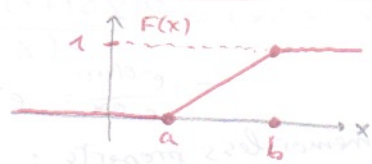
II) Continuous Uniform RV: $\boxed{X \sim U[a,b]}$

PDF = X is a cont unif RV over $[a,b]$, $-\infty < a < b < \infty$ if: $\boxed{f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}}$



(so) $\int_a^b f(x)\,dx = \int_a^b \frac{1}{b-a}\,dx = 1$ ✓ and $f(x) \geq 0$.

CDF = $\boxed{F(x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ 1, & b \leq x \end{cases}}$

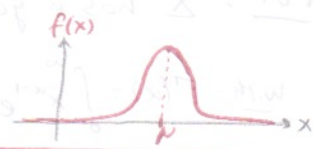


Note: $P(c \leq x \leq d) = F(d) - F(c) = \int_c^d f(x)\,dx \Rightarrow P(c \leq x \leq d) = \frac{d-c}{b-a}$ if $a < c < d < b$ are finite.

standard uniform dist: $f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases} \Rightarrow F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x < 1 \\ 1, & 1 \leq x \end{cases}$

Applications: Generate random numbers.

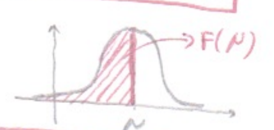
III) Gaussian/Normal distribution: $\boxed{X \sim N(\mu, \sigma^2)}$



PDF: The RV X has a normal distribution if: $\boxed{f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \ -\infty \leq x \leq \infty, \ \begin{cases} \mu \in \mathbb{R} \\ \sigma^2 > 0 \end{cases}}$

Note: $\begin{cases} \mu: \text{location parameter} \\ \sigma: \text{how fat is the curve} \end{cases}$

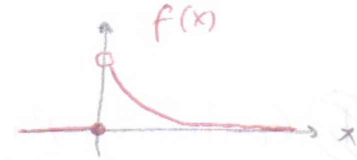
CDF: $\boxed{F(x) = \int_{-\infty}^{x} f(t)\,dt = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt \quad \forall x \in \mathbb{R}}$. Note: $F(\mu) = \frac{1}{2}$



standard Normal dist: $Z \sim N(0, 1)$: $\boxed{Z = \frac{X - \mu}{\sigma}}$ $\quad \boxed{\begin{cases} \text{PDF: } \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, & -\infty < z < \infty \\ \text{CDF: } \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\,dt, & z \in \mathbb{R}. \end{cases}}$

Note:
- $\Phi(0) = \frac{1}{2}$, $\boxed{\Phi(z) = 1 - \Phi(-z) \ \forall z \in \mathbb{R}}$
- $\int_{-1}^{1} \phi(z)\,dz = P(-1 \leq z \leq 1) \approx 0{,}683$
- $\int_{-2}^{2} \phi(z)\,dz = P(-2 \leq z \leq 2) \approx 0{,}954$
- $\int_{-3}^{3} \phi(z)\,dz = P(-3 \leq z \leq 3) \simeq 0{,}997$

IV Exponential distribution: $\boxed{X \sim EXP(\theta)}$

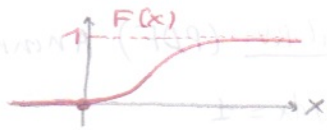


PDF: an Exponential RV X has the PDF: $\boxed{f(x) = \begin{cases} \theta e^{-\theta x}, & x>0 \\ 0, & x\le 0\end{cases}}$

θ = parameter of the exponential dist.

(ok) $f(x)\ge 0 \ \forall x\in\mathbb{R}$ and $\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty}\theta e^{-\theta x}dx = -e^{-\theta x}\Big|_{-\infty}^{\infty} = 0-(-1) = 1$ ✓

CDF: $\boxed{\underset{P(X\le x)}{\underset{\shortparallel}{F(x)}} = \int_{-\infty}^{x} f(t)dt = \begin{cases}0, & x<0\\ 1-e^{-\theta x}, & x\ge 0\end{cases}}$



Application: life-time of an object, survival time of a patient, ...

Link with Poisson distribution: $P(N(t)=k) \approx e^{-\lambda t}\frac{(\lambda t)^k}{k!}$

- Let T = "waiting time until the 1st occurrence happens"
- Then: $\begin{cases} P(T>t) = \overset{\text{Nothing happened yet}}{P(N(t)=0)} = e^{-\lambda t}, & t>0 \\ P(T\le t) = F_T(t) = 1-e^{-\lambda t}, & t>0\end{cases}$ $\Longrightarrow$ T has an exponential distribution!

Memoriless property of the exponential distribution:

Thm: $\boxed{P(X>t+x \mid X>x) = P(X>t), \ \forall t>0}$

Corr: $\boxed{P(X>t+x) = P(X>t)\cdot P(X>x)}$ $\Rightarrow 1-F(t+x) = (1-F(t))(1-F(x))$ ⇝ functional equation!

Proof: $P(X>t+x \mid X>x) = \frac{P((X>t+x)\cap(X>x))}{P(X>x)} = \frac{P(X>x+t)}{P(X>x)} = \frac{1-F(x+t)}{1-F(x)} = \frac{1-(1-e^{-\theta(x+t)})}{1-(1-e^{-\theta x})}$
$= \frac{e^{-\theta(x+t)}}{e^{-\theta x}} = e^{-\theta t} = P(X>t)$

Note: X a RV with memoriless property:
- X is a discrete RV $\iff$ X has a geometric dist
- X is a cont. RV $\iff$ X has an exponential dist.

V Gamma distribution: $\boxed{X \sim Gamma(\alpha, \beta)}$

PDF: X has a gamma distribution if: $\boxed{f(x) = \frac{1}{\Gamma(\alpha)\cdot\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}, \ \forall x>0, \ \alpha,\beta>0}$

with: $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}e^{-x}dx$, $\Gamma(1)=1$, $\Gamma(\frac{1}{2})=\sqrt{\pi}$, $\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$, $\Gamma(n)=(n-1)! \ \forall n\in\mathbb{N}$.

Note: if $\alpha=1$: $X \sim Exp(\frac{1}{\beta})$ $\Rightarrow$ $Gamma(1,\beta) \sim Exp(\frac{1}{\beta})$

Functions of a Random Variable :

(I) Function of an RV :

Goal : • Let X be an RV over $(\Omega, \mathcal{F}, P)$ and g a measurable function : $g:(\mathbb{R}, \mathcal{B}) \longrightarrow (\mathbb{R}, \mathcal{B})$ ST $\forall B \in \mathcal{B} : g^{-1}(B) = \{x \in \mathbb{R} : g(x) \in B\} \subset \mathcal{B}$

• We know either the CDF F_X or the PMF/PDF $P(X=x)/f_X$ of X :

Q : Let Y be the RV def by : $Y = g(X) \rightarrow$ Find F_Y or f_Y or $P(Y=y)$

Prop : $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(\{x : g(x) \leq y\}) = P(\{\omega \in \Omega : X(\omega) \in g^{-1}(-\infty, y]\}) \quad \forall y \in \mathbb{R}$.

Ex : • X a RV with CDF F_X : let $Y = |X|$

$F_Y(y) = P(Y \leq y) = P(|X| \leq y) = P(-y \leq X \leq y) = P(X \leq y) - P(X < -y) = F_X(y) - F_X(-y) + P(X = -y) \quad \forall y \in \mathbb{R}$

• $Y = aX + b \quad (a, b \in \mathbb{R}, a \neq 0)$

$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(aX \leq y - b) \Rightarrow F_Y(y) = \begin{cases} P(X \leq \frac{y-b}{a}) & , a > 0 \\ P(X \geq \frac{y-b}{a}) = 1 - P(X < \frac{y-b}{a}) & , a < 0 \end{cases} \Rightarrow F_Y(y) = \begin{cases} F_X(\frac{y-b}{a}) & , a > 0 \\ 1 - F_X(\frac{y-b}{a}) + P(X = \frac{y-b}{a}) & , a < 0 \end{cases} \quad \forall y \in \mathbb{R}$

• $Y = X^+ = \max(0, X) = \begin{cases} X, & X \geq 0 \\ 0, & X < 0 \end{cases}$

$F_Y(y) = P(Y \leq y) = P(\max(0, X) \leq y) = \begin{cases} P(\emptyset) = 0 & , y < 0 \\ P(X \leq 0) & , y = 0 \\ P(X \leq y) & , y > 0 \end{cases} \Rightarrow F_Y(y) = \begin{cases} 0 & , y < 0 \\ F_X(0) & , y = 0 \\ F_X(y) & , y \geq 0 \end{cases}$

• $Y = X^- = \min(0, X) = \begin{cases} X, & X \leq 0 \\ 0, & X > 0 \end{cases}$

$F_Y(y) = P(Y \leq y) = P(\min(0, X) \leq y) = \begin{cases} P(X \leq y) & , y < 0 \\ P(X \leq 0) & , y = 0 \\ P(\Omega) & , y \geq 0 \end{cases} \Rightarrow F_Y(y) = \begin{cases} F_X(y) & , y \leq 0 \\ F_X(0) & , y = 0 \\ 1 & , y > 0 \end{cases}$

• Try also : $Y = X^k \ (k \in \mathbb{Z}, X > 0)$, $Y = e^X$, $Y = \ln X \ (X > 0)$, $Y = \text{Sgn}(X)$, ...

(II) Function of a discrete RV :

Prop : X a discrete RV, A a countable set ST $\begin{cases} P(X = x) > 0 \ \forall x \in A \\ P(A) = 1 \end{cases}$

Let $g : A \longleftrightarrow B$ be a bijection : then g^{-1} is a single valued function : $g^{-1} : B \longleftrightarrow A$

Let $Y = g(X)$ be an RV : $P(Y = y) = \begin{cases} P(g(X) = y) = P(X = g^{-1}(y)) \overset{\text{if } g \text{ is not injective}}{=} \sum_{\{x : g(x) = y\}} P(X = x) & , y \in B \\ 0 & , y \in B^c \ (y \notin B) \end{cases}$

Ex : • $X \sim Bin(n, p)$: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, 2, ..., n$: $Y := n - X$

$P(Y = y)$ for $y = 0, 1, 2, ..., n \Rightarrow P(Y = y) = P(n - X = y) = P(X = n - y) = \binom{n}{n-y} p^{n-y} (1-p)^y = \binom{n}{y} (1-p)^y p^{n-y}$

So $Y \sim Bin(n, q = 1-p) \rightsquigarrow Y = \#$ of failures, $X = \#$ of success.

• $X \sim Poisson(\lambda)$: $Y = X^2 + 3$, $P(X = k) = \begin{cases} e^{-\lambda} \frac{\lambda^k}{k!} & , k = 0, 1, 2, ... \\ 0 & , \text{otherwise} \end{cases} \quad (\lambda > 0)$

$P(Y = y) = P(X^2 + 3 = y) = P(X^2 = y - 3) = P(X = +\sqrt{y-3})$ as $X \geq 0$ by def.

So $P(Y = y) = \begin{cases} \dfrac{e^{-\lambda} \lambda^{\sqrt{y-3}}}{\sqrt{y-3}} & , \sqrt{y-3} = 0, 1, 2, ... \quad \text{i.e. } y = 3 + n^2 \ (n \in \mathbb{N}) \\ 0 & , \text{otherwise} \end{cases}$

•

$X = x$	-2	-1	0	1	2
$P(X = x)$	$1/5$	$1/6$	$1/5$	$1/15$	$11/30$

$S_X = \{-2, -1, 0, 1, 2\}$: let $Y = X^2 \Rightarrow S_Y = \{0, 1, 4\}$

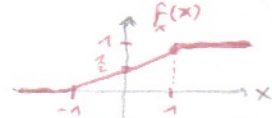
$Y = y$	0	1	4
$P(Y = y)$	$1/5$	$7/30$	$17/30$

$P(Y = 0) = P(X^2 = 0) = P(X = 0) = 1/5 = 6/30$

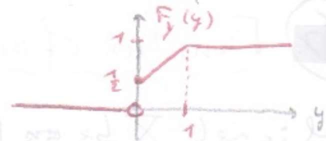
$P(Y = 1) = P(X^2 = 1) = P(X = -1) + P(X = 1) = 1/6 + 1/15 = 7/30$

$P(Y = 4) = P(X^2 = 4) = P(X = -2) + P(X = 2) = 1/5 + 11/30 = 17/30$

III Function of a Continuous RV:



Ex: • $X \sim U[-1,1]$, $Y = \max(0, X) = X^+$: $F_X(x) = \begin{cases} 0 & , x < -1 \\ \frac{x+1}{2} & , -1 \le x < 1 \\ 1 & , 1 \le x \end{cases}$



we saw that $F_Y(y) = \begin{cases} 0 & , y < 0 \\ F_X(y), & y \ge 0 \end{cases} \Rightarrow F_X(x) = \begin{cases} 0 & , x < 0 \\ \frac{y+1}{2} & , 0 \le x < 1 \\ 1 & , 1 \le x \end{cases}$

↳ F_Y has a jump discontinuity at $y=0$, but is cont elsewhere $\Rightarrow$ Y is neither a cont nor a discrete RV.

• X a cont RV with PDF : $f_X(x) = \begin{cases} \frac{x+1}{2} & , -1 \le x \le 1 \\ 0 & , |x| > 1 \end{cases}$, Let $Y = X^2 \to$ Find F_Y and f_Y

• $F_X(x) = \begin{cases} 0 & , x < -1 \\ \int_{-1}^{x} \frac{t+1}{2} dt, & -1 \le x < 1 \\ 1 & , 1 \le x \end{cases} \Rightarrow F_X(x) = \begin{cases} 0 & , x < -1 \\ \frac{1}{4} + \frac{x}{2} + \frac{x^2}{4} & , -1 \le x < 1 \\ 1 & , 1 \le x \end{cases}$

• $F_Y(y) = P(Y \le y) = P(X^2 \le y) = \begin{cases} P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) & , y \ge 0 \\ P(\phi) = 0 & , y < 0 \end{cases} \Rightarrow F_Y(y) = \begin{cases} 0 & , y < 0 \\ \sqrt{y} & , 0 \le y < 1 \\ 1 & , 1 \le y \end{cases}$

• $f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) & , 0 < y < 1 \\ 0 & , \text{otherwise} \end{cases} \Rightarrow f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} & , 0 < y < 1 \\ 0 & , \text{otherwise} \end{cases}$

Thm 1: X a cont RV with CDF F_X and PDF f_X.

g a diff function $\forall x \in \mathbb{R}$, where either : $g'(x) > 0 \ \forall x \in \mathbb{R}$ or $g'(x) < 0 \ \forall x \in \mathbb{R} \longrightarrow g$ is strictly monotone on $\mathbb{R}$

$\Rightarrow Y = g(X)$ is a cont. RV with :

CDF : $F_Y(y) = \begin{cases} F_X(g^{-1}(y)) & , \text{if } g \nearrow \\ 1 - F_X(g^{-1}(y)) & , \text{if } g \searrow \end{cases} \Rightarrow$ PDF : $f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right| & , \alpha < y < \beta \\ 0 & , \text{otherwise} \end{cases}$

where : $\alpha = \min\{g(-\infty), g(+\infty)\}$
$\beta = \max\{g(-\infty), g(+\infty)\}$

Proof : • g ST $\nearrow$:

g ST $\nearrow \Rightarrow g^{-1}(y) = x$ Exists and g^{-1} is ST $\nearrow$ and diff

$F_Y(y) = P(Y \le y) = P(g(X) \le y) = P(X \le g^{-1}(y)) = F_X(g^{-1}(y))$

So $f_Y(y) = \frac{d}{dy} F_Y(g^{-1}(y)) = F_X'(g^{-1}(y)) \cdot \frac{d}{dy}(g^{-1}(y))$

• g ST $\searrow$: Exercise

Ex: • $X \sim EXP(\theta)$: $X > 0$, $f_X(x) = \begin{cases} \theta e^{-\theta x} & , x > 0 \\ 0 & , x \le 0 \end{cases} \longrightarrow$ Let $Y = X^{1/\beta}$, $X > 0$, $\beta > 0$:

$Y = g(X) = X^{1/\beta}$: ST $\nearrow$ for $X > 0 \Rightarrow g^{-1}(y) = y^\beta \Rightarrow f_Y(y) = (\theta e^{-\theta y^\beta}) \cdot \left| \frac{d}{dy} y^\beta \right| = \theta e^{-\theta y^\beta} y^{\beta - 1} \beta$

So : $f_Y(y) = \begin{cases} \beta \theta e^{-\theta y^\beta} y^{\beta-1} & , y > 0 \\ 0 & , y \le 0 \end{cases}$ $\Longrightarrow$ Waloddi Weibull distributᵒ : param (β, θ)
→ use it in reliability

• $X \sim N(0,1)$: $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$: Let $Y = X^2$

CDF : $F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \int_{-\infty}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt - \int_{-\infty}^{-\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

PDF : $f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{2\sqrt{y}} \left(f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right) = \frac{1}{\sqrt{y}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}}$, $y > 0$

$\left(f_Y(y) = \frac{1}{\Gamma(\alpha) \beta^\alpha} y^{\alpha-1} e^{-\frac{y}{\beta}} \right)$

So $f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2} & , y > 0 \\ 0 & , y \le 0 \end{cases} \Rightarrow Y \sim \text{Gamma}(\alpha = \frac{1}{2}, \beta = 2)$

$\longrightarrow$ Chi-Square with 1 DoF

Thm 2: X a cont RV with CDF F_X and PDF f_X.

$g(x)$ a diff functᵒ $\forall x \in \mathbb{R}$ ST $g(x)$ is monotone on "K" subintervals : $I_1, I_2, \dots, I_K$ ST $\bigcup_{i=1}^{K} I_i = S_X$ (support of X), $I_i \cap I_j = \phi$ if $i \ne j$

If $Y = g(X) \Rightarrow f_Y(y) = \begin{cases} \sum_{i=1}^{K} f_X(g_i^{-1}(y)) \cdot \left| \frac{d}{dy} g_i^{-1}(y) \right| & , \alpha < y < \beta \\ 0 & , \text{otherwise} \end{cases}$
↳ $\alpha = \min\{g(-\infty), g(+\infty)\}$
↳ $\beta = \max\{g(-\infty), g(+\infty)\}$

where g_i^{-1} is the inverse of g over the i^{th} subinterval.

IV. Generating random numbers from a specified Continuous RV :

Prop : X a cont RV with CDF : $F_X(x) = P(X \leq x)$, $x \in \mathbb{R}$.

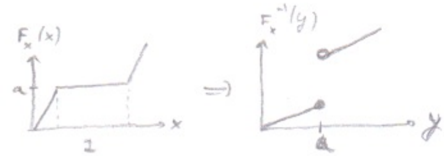
Let $Y := F_X(X)$: then $Y \sim U[0,1]$

⚠ We cannot write : $F_X(X) \cancel{=} P(X \leq X) = 1$

Proof : $F_Y(y) = P(Y \leq y) = P(F_X(X) \leq y)$

X is a cont RV $\Rightarrow F_X$ is cont + $\nearrow$ $\Rightarrow F_X^{-1}$ is cont + $\nearrow$

if F_X is flat over some interval I, then define $F_X^{-1}(y) = \inf\{x : F_X(x) \geq y\}$



So $F_Y(y) = F_X(F_X^{-1}(y)) = y$

So $F_Y(y) = \begin{cases} 0 & ; y < 0 \\ y & ; 0 \leq y < 1 \\ 1 & ; 1 \leq y \end{cases} \Rightarrow f_Y(y) = \begin{cases} 1 & , 0 < y < 1 \\ 0 & , \text{otherwise} \end{cases} \Rightarrow Y \sim U[0,1]$

Ex : $X \sim EXP(\theta)$, $\theta > 0$ and $X \geq 0$: $F_X(x) = \begin{cases} 1 - e^{-\theta x} & , x > 0 \\ 0 & ; x \leq 0 \end{cases} \Rightarrow Y = F_X(X) = 1 - e^{-\theta X}$

$\Rightarrow Y \in [0,1]$ (as $X \geq 0$)

$y \in [0,1]$: $F_Y(y) = P(Y \leq y) = P(1 - e^{-\theta X} \leq y) = P(-\theta X \geq \log(1-y)) = P\left(X \leq \frac{\log(1-y)}{-\theta}\right)$

So $F_Y(y) = F_X\Big(\underbrace{\frac{\log(1-y)}{-\theta}}_{\text{positive as } y \in [0,1)}\Big) = 1 - e^{-\theta \frac{\log(1-y)}{-\theta}} = 1 - (1-y) = y \Rightarrow F_Y(y) = \begin{cases} 0 & , y < 0 \\ y & ; 0 \leq y < 1 \\ 1 & , 1 \leq y \end{cases} \Rightarrow Y \sim U[0,1]$

Problem : Let X be a cont RV with CDF F_X : How do we generate random nb from F_X ?

1. Generate numbers from $U \sim U[0,1]$
2. Since $X \sim F_X$, then $F_X(X) \sim U[0,1] \Rightarrow U \overset{dist}{=} F_X(X)$ in the sense of distributions.

So $X \overset{dist}{=} F_X^{-1}(U)$ in the sense of distributions

Moments, Variance & expectation Values :

① Expected value :

Expected value / Mean / 1^{st} moment :

- X a discrete RV : $\mu = E(X) = \sum_{x_i \in S_X} x_i \, P(X = x_i)$ exists if $E(|X|) = \sum_{x_i \in S_X} |x_i| \cdot P(X = x_i) < \infty$
- X a cont RV with PDF $f(x)$: $\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$ exists if $E(|X|) = \int_{-\infty}^{\infty} |x| \cdot f(x) dx < \infty$

Ex : Discrete RV :

- X : discrete uniform RV with support $S_X = \{x_1, \dots, x_N\}$ $\Rightarrow E(X) = \sum_{i=1}^{N} x_i \cdot P(X = x_i) = \frac{\sum_{i=1}^{N} x_i}{N}$
- $X \sim Bin(n,p)$ with $S_X = \{0, 1, \dots, n\}$ $\Rightarrow E(X) = \sum_{x=0}^{n} x \, P(X = x) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$

  So $E(X) = \sum_{x=1}^{n} \frac{pn \cdot (n-1)!}{(n-x)!(x-1)!} p^{x-1} (1-p)^{n-x} = np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} = np \cdot 1$ $\Rightarrow E(X) = np$.

  $\boxed{X \sim Bin(n,p) \Rightarrow E(X) = np}$
- $P(X = n) = \frac{c}{n^2}$, $n = 1, 2, \dots$ and $c > 0$: $E(X) = \sum_{n=1}^{\infty} n \cdot \frac{c}{n^2} = c \sum_{n=1}^{\infty} \frac{1}{n} \longrightarrow \infty$ $\Rightarrow E(X)$ DNE !
- $P\left(X = (-1)^{x+1} \cdot \frac{3^x}{x}\right) = \frac{2}{3^x}$, $x = 1, 2, \dots$: $E(|X|) = \sum_{x=1}^{\infty} \left|(-1)^{x+1} \cdot \frac{3^x}{x}\right| \cdot \frac{2}{3^x} = \sum_{x=1}^{\infty} \frac{2}{x} \rightarrow \infty \Rightarrow E(X)$ DNE !

  (But $E(X) = \sum_{x=1}^{\infty} 2 \cdot \frac{(-1)^{n+1}}{x} = 2 \ln 2$ $\neq \infty$)

Cont RV :

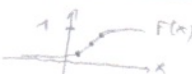
- $X \sim Exp(\theta)$, $\theta > 0$: $f(x) = \begin{cases} \theta e^{-\theta x} & , x > 0 \\ 0 & , x < 0 \end{cases}$

  So $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} \theta x e^{-\theta x} dx = \frac{1}{\theta}$ $(\theta > 0)$

  $\boxed{X \sim Exp(\theta) \underset{(\theta > 0)}{\Rightarrow} E(X) = \frac{1}{\theta}}$
- X with PDF : $\boxed{f(x) = \frac{1}{\pi(1+x^2)}, x \in \mathbb{R}}$ $\rightarrow$ Cauchy distribution ! $E(|X|) = \int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)} dx = \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx = \frac{\ln(1+x^2)}{\pi} \Big|_0^{\infty}$

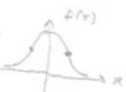
  $\Rightarrow E(|X|) \rightarrow \infty$ $\Rightarrow E(X)$ DNE !!!

Symmetric distribution : A RV X with CDF $F(\cdot)$ is symmetric if :



$\exists \alpha \in \mathbb{R}$ ST $P(X \geq \alpha + x) = P(X \leq \alpha - x)$, $\forall x \in \mathbb{R}$

i.e. $\exists \alpha \in \mathbb{R}$ ST $F(\alpha + x) + F(\alpha - x) = 1 + \underbrace{P(X = \alpha + x)}_{= 0 \text{ if } X \text{ is cont.}}$, $\forall x \in \mathbb{R}$



Note : X cont : then X symm $\Leftrightarrow$ $\exists \alpha \in \mathbb{R}$ ST $f(\alpha + x) = f(\alpha - x)$ $\forall x \in \mathbb{R}$

Prop : X a symmetric + cont RV with PDF $f(\cdot)$ around α : If $E(|X|) < \infty$, then $E(X) = \alpha$.

Ex : Cauchy dist : $E(|X|) = \infty \Rightarrow E(X)$ DNE !

Thm : X an RV, $g : \mathbb{R} \rightarrow \mathbb{R}$ a Borel measurable function, $Y = g(X)$:

- X discrete, $E(|g(X)|) < \infty \Rightarrow E(g(X)) = \sum_{x \in S_X} g(x) \cdot P(X = x)$
- X cont, with PDF $f_X(\cdot)$ $E(|g(X)|) < \infty \Rightarrow E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$

Proof :

- X discrete with support S_X, Y cont with support S_Y

  $E(Y) = \sum_{y \in S_Y} y \cdot P(Y = y) = \sum_{g(x) \in S_Y} g(x) \cdot P(g(X) = y) = \sum_{g(x) \in S_Y} g(x) \cdot P(X = g^{-1}(y)) = \sum_{g(x) \in S_Y} g(x) \left(\sum_{\{x \in S_X : g(x) = y\}} P(X = x) \right) = \sum_{x \in S_X} g(x) P(X = x)$
- Suppose g satisfies the condit. of the thm of last chapter : then $f_Y(y) = f_X(g^{-1}(y)) \cdot \left|\frac{d}{dy} g^{-1}(y)\right|$, $\alpha < y < \beta$.

  $E(Y) = \int_{\alpha}^{\beta} y f_Y(y) dy = \int_{\alpha}^{\beta} y \cdot f_X(g^{-1}(y)) \cdot \underbrace{\left|\frac{d}{dy} g^{-1}(y)\right| dy}_{= dx} = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx$.

General proof later !

Corr :

- $E(aX + b) = a E(X) + b$ $\forall a, b \in \mathbb{R}$ if $E(|X|) < \infty$.
- $E(g_1(X) + g_2(X) + \dots + g_n(X)) = E(g_1(X)) + E(g_2(X)) + \dots + E(g_n(X))$ if $E(|g_i(X)|) < \infty$ $\forall 1 \leq i \leq n$.
- $E(c) = c$, $\forall c \in \mathbb{R}$ constant.

  ↳ X is a degenerate RV : $\begin{cases} P(X = c) = 1 \\ P(X \neq c) = 0 \end{cases}$

Ex : $\begin{cases} X \sim N(0,1) \\ Y = \mu + \sigma X \end{cases}$ $\Rightarrow E(Y) = E(\mu + \sigma X) = \mu + \sigma \cdot E(X) = \mu + 0 = \mu$.

So $E(Y) = \mu$

④ Higher order moments of an RV:

Def = X an RV, $n \in \mathbb{N}$, $\alpha > 0$: if they exist,

- $E(X^n) = n^{th}$ moment of X.
- $E(|X|^\alpha) = \alpha^{th}$ absolute moment of X.

central moments of an RV: X an RV, $k \in \mathbb{N}$, $\mu = E(X)$ exists: if they exist,

$\mu_k = E((X-\mu)^k) = k^{th}$ central moment of X.

$E(|X-\mu|^k) = k^{th}$ absolute central moment of X.

$E((X-c)^k) = k^{th}$ moment of X about the point c.

Note = $\begin{cases} E(X-\mu) = 0 \\ E(X-c) = \mu - c \end{cases}$

Variance = $Var(X) = \sigma^2 = \mu_2 = E((X-\mu)^2) \geq 0$ ↳ 2^{nd} central moment of X

Standard deviation: $\sigma = \sqrt{Var(X)}$

Prop = $\boxed{Var(X) = E(X^2) - E(X)^2 \geq 0}$

(※) $Var(X) = E(X^2 - 2X\mu + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2$

Note: $E(X^n) \in \mathbb{R}$, but $Var(X) \geq 0 \; \forall X$.

Ex:
- $X \sim Bin(n,p)$: $E(X) = np$, $E(X^2) = n(n-1)p^2 + np \Rightarrow Var(X) = npq$

(※) $E(X^2) = E(X^2 - X + X) = E(X) + E(X(X-1)) = np + \sum_{x=1}^{n} x(x-1)\binom{n}{x} p^x (1-p)^{n-x} = np + n(n-1)p^2 \underbrace{\sum_{x=2}^{n} \frac{(n-2)!}{(n-x)!\,(x-2)!} p^{x-2}(1-p)^{n-x}}_{= \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2}(1-p)^{n-x} = 1}$

- $X \sim N(0,1)$: $E(X) = 0$, $E(X^2) = 1 \Rightarrow Var(X) = 1$.

$E(X^n) = 0$ if n odd ; $E(X^n) = (n-1)\cdots 5 \cdot 3 \cdot 1$ if n even

Thm: X an RV: $E(|X|^t) < \infty$ for some $t > 0 \Rightarrow E(|X|^s) < \infty \; \forall\, 0 < s \leq t$

↳ "Higher order moments exist ⇒ Lower order moment exist"

Proof:
- X cont: $E(|X|^s) = \int_{-\infty}^{\infty} |x|^s f_X(x)\,dx = \int_{|x|<1} |x|^s f_X(x)\,dx + \int_{|x|\geq 1} |x|^s f_X(x)\,dx \leq \int_{|x|<1} 1 \cdot f_X(x)\,dx + \int_{|x|\geq 1} |x|^t f_X(x)\,dx$

So $E(|X|^s) \leq \int_{-\infty}^{\infty} f(x)\,dx + \int_{-\infty}^{\infty} |x|^t f_X(x)\,dx = 1 + E(|X|^t) < \infty$.

- X discrete: Similar but for $\sum$.

Thm: X an RV: $E(|X|^t) < \infty$ for some $0 < t < \infty \Rightarrow \lim_{n\to\infty} n^t \underbrace{P(|X| > n)}_{\text{"Tail probability"} = P(X>n) + P(X<-n)} = 0$

Proof =
- X cont: $E(|X|^t) = \int_{-\infty}^{\infty} |x|^t f_X(x)\,dx = \lim_{n\to\infty} \int_{-n}^{n} |x|^t f_X(x)\,dx < \infty \Rightarrow \lim_{n\to\infty} \int_{|x|>n} |x|^t f_X(x)\,dx = 0$

But $\int_{|x|>n} |x|^t f_X(t)\,dx \geq \int_{|x|>n} n^t f_X(x)\,dx = n^t P(|X| > n)$

So $\lim_{n\to\infty} n^t P(|X|>n) = 0$

- X discrete: similar

→ $E(|X|^t) < \infty \Rightarrow$ tail probability goes to zero faster than $1/n^t$!

⚠ Converse NOT true!! Ex: X with PMF: $\begin{cases} P(X = n) = \frac{c}{n^2 \log n} \\ x = 2, 3, \ldots \end{cases}$ $\Rightarrow P(X > n) = \sum_{x>n} \frac{c}{x^2 \log x} \simeq \int_n^{\infty} \frac{c}{x^2 \log x}\,dx \simeq \frac{c}{n \log n}$

$\Rightarrow n\,P(X>n) = \frac{c}{\log n} \to 0$ as $n \to \infty$

But: $E(X^1) = \sum_{n=1}^{\infty} \frac{c}{n \log n} \to \infty$!!!

Thm:
- $Var(X) < \infty \Leftrightarrow E(X^2) < \infty$
- $Var(X) = 0 \Leftrightarrow X$ is a degenerate RV: $X = \mu$
- $Var(aX+b) = a^2\,Var(X)$

(※) $= E[(aX+b) - E(aX+b)]^2 = E(a^2(X-\mu)^2) = a^2\,Var(X)$

- $\begin{cases} X^2 = (X - \mu + \mu)^2 \leq 2(X-\mu)^2 + 2\mu^2 \text{ : } Var(X) < \infty \Rightarrow E(X^2) < \infty \\ 0 \leq (X-\mu)^2 \leq X^2 + \mu^2 \text{ : } E(X^2) < \infty \Rightarrow Var(X) < \infty \end{cases}$
- $\begin{cases} 0 = Var(X) = \sum_{x \in S_X} (x-\mu)^2 P(X=x) \Rightarrow P(X=x) = 0 \text{ or } (x-\mu)^2 = 0 \Rightarrow x = \mu \;\forall x \in S_X \Rightarrow X = \mu. \\ X = \mu \Rightarrow Var(X) = E(x-\mu)^2 = E(0) = 0. \end{cases}$

Quantiles of a distribution = X a RV with CDF $F(\cdot)$

- x_p is the p^{th} quantile of X if: $\boxed{\begin{cases} P(X \leq x_p) \geq p \\ P(X \geq x_p) \geq 1-p \end{cases}}$ $(0<p<1)$, i.e. $\boxed{p \leq F(x_p) \leq p + P(X = x_p)}$
- $x_{0,5}$ is the median of X

Note: X cont $\Rightarrow F(x_p) = p$

Ex:
- $X \sim N(\mu, \sigma^2) \Rightarrow E(X) = \mu = x_{0,5}$  → X symmetric, $\mu = E(X)$ exists $\Rightarrow x_{0,5} = \mu$
- X discrete with PDF: $P(X=x) = \begin{cases} 1/4, & x = -2, 0 \\ 1/3, & x = 1 \\ 1/6, & x = 2 \end{cases}$

$\Rightarrow \forall\, 0 \leq x < 1$: $\begin{cases} P(X \leq x) = P(X=-2) + P(X=0) = 1/2 \\ P(X \geq x) = P(X=1) + P(X=2) = 1/2 \end{cases}$ } median NOT unique!

Moment Generating function of an RV :

I) Moment Generating function:

Moment generating funct^{o} : X an RV with CDF $F(\cdot)$: if $E(e^{tx})$ exists over some neighborhood of zero $(-t_0, t_0)$

then $\boxed{M(t) = E(e^{tX})}$ is the MGF of X $\forall t \in (-t_0, t_0)$

Note : $\boxed{M(0) = 1}$

Thm 1 = (uniqueness thm) the MGF uniquely determines the distribution of an RV

Conversely: if the MGF exists, it is unique.

Thm 2 = If $M(t)$ exists for all $t \in (-t_0, t_0)$: $\boxed{M^{(k)}(0) = \frac{d^k}{dt} M(t)\Big|_{t=0} = E(X^k)}$, $\forall k \in \mathbb{N}$

Corr : If $M(t)$ exists on $(-t_0, t_0)$: $M(t) = 1 + E(X)\, t + \frac{E(X^2)}{2!} t^2 + \dots + \frac{E(X^n)}{n!} t^n + \dots$

$= M(0) + M^{(1)}(0) t + \frac{M^{(2)}(0)}{2!} t^2 + \dots + \frac{M^{(n)}(0)}{n!} t^n + \dots$

1) Binomial distribution : $X \sim Bin(n, p)$

Prop : • $M(t) = (pe^t + q)^n$, $\forall t \in \mathbb{R}$, $(q = 1-p)$

• $E(X) = np$, $E(X^2) = np + n(n-1)p^2$

• $Var(X) = npq$

(pf) • $M(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$

$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} = (pe^t + q)^n$, $\forall t \in \mathbb{R}$

• $M(0) = (p+q)^n = 1^n = 1$

• $M'(0) = n(pe^0)(pe^0+q)^{n-1} = np(p+q)^{n-1} = np = E(X)$

• $M''(0) = \dots = np + n(n-1)p^2$

$\Rightarrow Var(X) = np + n(n-1)p^2 - (np)^2 = np(1-p) = npq$.

2) Poisson distribution: $X \sim Pois(\lambda)$

Prop : • $M(t) = e^{\lambda(e^t - 1)}$, $\forall t \in \mathbb{R}$

• $E(X) = \lambda$, $E(X^2) = \lambda^2 + \lambda$

• $Var(X) = \lambda$

So $\boxed{E(X) = \lambda = Var(X)}$ for Poisson dist.

(pf). $M(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \cdot \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t}$, $\forall t \in \mathbb{R}$

• $M(0) = e^{\lambda(e^0 - 1)} = e^{\lambda \cdot 0} = 1$

• $M'(0) = (\lambda e^0) e^{\lambda(e^0-1)} = \lambda = E(X)$

• $M''(0) = \lambda e^0 e^{\lambda(e^0-1)} + (\lambda e^0)^2 e^{\lambda(e^0-1)} = \lambda + \lambda^2 = E(X)^2$ $\Big\}$ $Var(X) = \lambda + \lambda^2 - \lambda^2 = \lambda$

3) Normal distribution : $X \sim N(\mu, \sigma^2)$

Prop : • $M(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$, $\forall t \in \mathbb{R}$

• $E(X) = \mu$, $E(X^2) = \mu^2 + \sigma^2$

• $Var(X) = \sigma^2$

(pf) • $M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} dx \overset{\text{complete the square}}{=} e^{\mu t + \frac{\sigma^2 t^2}{2}} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-(\mu+t\sigma^2))^2}{2\sigma^2}}}_{=1}$

• $M(0) = e^{\mu \cdot 0 + \frac{\sigma^2 0^2}{2}} = e^0 = 1$

• $M'(0) = (\mu + \sigma^2 \cdot 0) M(0) = \mu = E(X)$

• $M''(0) = (\mu + \sigma^2 0)^2 M(0) + \sigma^2 M(0) = \mu^2 + \sigma^2 = E(X^2)$ $\Big\}$ $Var(X) = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$

4) Exponential distribution: $X \sim EXP(\theta)$ $(\theta > 0)$: $f(x) = \begin{cases} \theta e^{-\theta x} , & x > 0 \\ 0 , & x < 0 \end{cases}$

Props : • $M(t) = \frac{\theta}{\theta - t}$, $\forall t < \theta$

• $E(X) = \frac{1}{\theta}$, $E(X^2) = \frac{2}{\theta^2}$

• $Var(X) = \frac{1}{\theta^2}$

(pf) • $M(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} \theta e^{-\theta x} dx = \theta \int_0^{\infty} e^{-(\theta - t)x} dx = \frac{\theta}{\theta - t}$ if $t < \theta$

• $M(0) = \frac{\theta}{\theta} = 1$

• $M'(0) = \frac{\theta}{(\theta - 0)^2} = \frac{1}{\theta} = E(X)$

• $M''(0) = \frac{2\theta}{(\theta - 0)^3} = \frac{2}{\theta^2} = E(X^2)$ $\Big\}$ $Var(X) = \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2}$

(II) Moment Inequalities :

→ Very useful in probability theory!

Thm : Let $h(x) \geq 0$ a Borel measurable function of an RV X.

If $E(h(X))$ exists, then : $\boxed{P(h(X) \geq M) \leq \frac{E(h(X))}{M}}$, $\forall M > 0$

↳ $h(X) \geq 0$ so $h(X) = |h(X)|$

Proof : • X discrete :

$$E(h(X)) = \sum_{x \in S_X} h(x)\, p(X=x) = \sum_{\{x : h(x) \geq M\}} h(x) \cdot p(X=x) + \underbrace{\sum_{\{x : h(x) < M\}} h(x) \cdot p(X=x)}_{\geq 0} \geq \sum_{\{x : h(x) \geq M\}} M \cdot P(X=x)$$

so $E(h(X)) \geq M \cdot P(h(X) \geq M)$

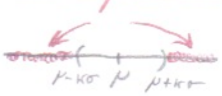
• X cont : (Similar)

Markov inequality : X an RV : if $E(|X|^k) < \infty$ for some $k > 0$:

then $\boxed{P(|X| \geq M) \leq \frac{E(|X|^k)}{M^k}}$, $\forall M > 0$

Proof : $P(X \geq M) = P(|X|^k \geq M^k) \leq \frac{E(|X|^k)}{M^k}$

Chebychev - Bienayme inequality : X a RV with $E(X) = \mu$, $Var(X) = \sigma^2$, $E(X^2) < \infty$:



then $\boxed{P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}}$, $\forall k > 0$

Proof : $P(|X-\mu| \geq k\sigma) = P(|X-\mu|^2 \geq k^2\sigma^2) \leq \frac{E((X-\mu)^2)}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$.

Ex : $X \sim N(\mu, \sigma^2)$

• Tables : $\begin{cases} P(|X-\mu| \leq \sigma) \simeq 0{,}6827 \\ P(|X-\mu| \leq 2\sigma) \simeq 0{,}9545 \\ P(|X-\mu| \leq 3\sigma) \simeq 0{,}9973 \end{cases}$

• Chebychev Bienayme : $\begin{cases} P(|X-\mu| \leq \sigma) = 1 - P(|X-\mu| \geq \sigma) \geq 1 - \frac{1}{1} = 0 \\ P(|X-\mu| \leq 2\sigma) = 1 - P(|X-\mu| \geq 2\sigma) \geq 1 - \frac{1}{2^2} = 0{,}75 \\ P(|X-\mu| \leq 3\sigma) = 1 - P(|X-\mu| \geq 3\sigma) \geq 1 - \frac{1}{3^2} \simeq 0{,}89 \end{cases}$

Multivariate Random Variables :

(I) CDF, PMF, PDF and marginals :

<u>Random vector</u>: $(\Omega, \mathcal{F}, P)$, $X = (X_1, \dots, X_n)$: $X : \Omega \longrightarrow \mathbb{R}^n$

$\omega \longmapsto X(\omega) = (X_1(\omega), \dots, X_n(\omega))$ ↳ n dim random variable

where X must be a measurable funct :

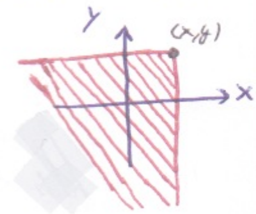
$X^{-1}(I) = \{\omega : X(\omega) \in I\} \in \mathcal{F} \quad \forall I = \{(x_1, \dots, x_n) : -\infty < x_i < a_i, i = 1, \dots, n\}$ for some $a_i \in \mathbb{R}$.

<u>Bivariate RV</u>: 2 dim RV : (X, Y)

<u>Goal</u> : Find $P((X,Y) \in B)$ $\forall B \in \mathcal{B}^2$, the probability distribution of (X,Y)

<u>Caratheodory extension thm</u>: there is a unique prob dist corresponding to the specification :

$\forall (x,y) \in \mathbb{R}^2$: $P[X \in (-\infty, x], Y \in (-\infty, y]) = P(\{\omega \in \Omega : X(\omega) \le x \text{ and } Y(\omega) \le y\})$

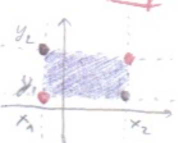


<u>Joint CDF</u> : The joint CDF of the bivariate RV (X, Y) :

$\forall (x,y) \in \mathbb{R}^2 : F(x,y) = P(X \le x, Y \le y) = P(\{\omega \in \Omega : X(\omega) \le x \text{ AND } Y(\omega) \le y\})$

↳ $F(\cdot,\cdot)$ uniquely determines the full prob dist of (X, Y)

<u>Prop</u> :
- $\lim\limits_{(x,y) \to (\infty,\infty)} F(x,y) = 1$ ↳ $P(\Omega) = 1$, $\lim\limits_{x \to -\infty} F(x,y) = 0$ $\forall y \in \mathbb{R}$ ↳ $P(\phi \cap Y \le y) = P(\phi) = 0$, $\lim\limits_{y \to -\infty} F(x,y) = 0$ $\forall x \in \mathbb{R}$ ↳ $P(X \le x \cap \phi) = P(\phi) = 0$



- $F(x,y)$ is nondecreasing & right continuous WRT both arguments
- $\forall (x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ ST $\begin{cases} x_1 \le x_2 \\ y_1 \le y_2 \end{cases}$: $0 \le P(x_1 < X \le x_2, y_1 < Y \le y_2) = F(x_2, y_2) + F(x_1, y_1) - F(x_2, y_1) - F(x_1, y_2)$

- <u>Discrete Bivariate RV</u> :

<u>Def</u> : the joint support $S \subseteq \mathbb{R}^2$ of (X, Y) is countable. <u>Note</u> : $\begin{cases} S_x = \{x_1, x_2, \dots\}, S_y = \{y_1, y_2, \dots\} \\ S = S_x \times S_y = \{(x_i, y_j) : i, j = 1, 2, \dots\} \end{cases}$

<u>Joint PMF</u> : $\begin{cases} P_{ij} = P(X = x_i, Y = y_j) \quad \forall (x_i, y_j) \in S \\ \sum\limits_{i,j} P_{ij} = 1, \quad 0 < P_{ij} < 1 \quad \forall i, j = 1, 2, \dots \end{cases}$

<u>Note</u> : $P((x,y) \in A) = \sum\limits_{(x_i, y_j) \in A} P_{ij}$, $\forall A \subseteq S$

$\forall x, y \in \mathbb{R} : F(x,y) = P(X \le x, Y \le y) = \sum\limits_{\{(x_i,y_j) \in S : x_i \le x, y_j \le y\}} P_{ij}$

<u>Marginal PMF</u> :
- of X : $\{P_{i\cdot} : i = 1, 2, \dots\}$, where $P_{i\cdot} = P(X = x_i) = \sum\limits_{j=1}^{\infty} P(X = x_i, Y = y_j) = \sum\limits_{j=1}^{\infty} P_{ij}$
- of Y : $\{P_{\cdot j} : j = 1, 2, \dots\}$, where $P_{\cdot j} = P(Y = y_j) = \sum\limits_{i=1}^{\infty} P(X = x_i, Y = y_j) = \sum\limits_{i=1}^{\infty} P_{ij}$

<u>Prop</u> : $\sum\limits_{i=1}^{\infty} P_{i\cdot} = \sum\limits_{i,j=1}^{\infty} P_{ij} = \sum\limits_{j=1}^{\infty} P_{\cdot j} = 1$ → Marginal PMFs are PMFs

- <u>Continuous Bivariate RV</u> :

<u>Def</u> : the RV (X, Y) with CDF $F(\cdot)$ is continuous if $\exists f(x,y) \ge 0$ ST : $F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\, dv\, du$

<u>Joint PDF</u> : $f(\cdot,\cdot)$ ST : $f(x,y) \ge 0$ $\forall (x,y) \in \mathbb{R}^2$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dy\, dx = 1$

<u>Note</u> : If $F(\cdot,\cdot)$ is absolutely continuous : $f(x,y) = \dfrac{\partial^2}{\partial x \partial y} F(x,y)$

<u>Marginal PDF</u> :
- of X : $f_x(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$
- of X : $f_y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$

<u>Prop</u> : $\begin{cases} f_x(x) \ge 0 \\ f_y(y) \ge 0 \end{cases}$ and $\begin{cases} \int_{-\infty}^{\infty} f_x(x)\, dx = 1 \\ \int_{-\infty}^{\infty} f_y(y)\, dy = 1 \end{cases}$ ⟹ Marginal PDFs are PDFs

- **Marginal CDF:** • of X: $F_X(x) = P(X \le x) = \lim_{y \to \infty} F(x,y), \ \forall x \in \mathbb{R}$
  • of Y: $F_Y(y) = P(Y \le y) = \lim_{x \to \infty} F(x,y), \ \forall y \in \mathbb{R}$

$f_X(x) \overset{?}{=} \frac{\partial}{\partial x} F_X(x)$

$f_Y(y) \overset{?}{=} \frac{\partial}{\partial y} F_Y(y)$

  ↳ **Lower dimensional marginal CDF:** $(X_1, \dots, X_n)$ an RV with CDF $F(\cdot, \dots, \cdot)$ → So $F(x_1, \dots, x_n) = P(X_1 \le x_1, \dots, X_n \le x_n)$

Then $F_{x_1, \dots, x_k}(x_1, \dots, x_k) = \lim_{\substack{x_{k+1} \to \infty \\ \vdots \\ x_n \to \infty}} F(x_1, \dots, x_n) = P(X_1 \le x_1, \dots, X_k \le x_k), \ \forall (x_1, \dots, x_n) \in \mathbb{R}^n$

Note: Here, $(x_1, \dots, x_k)$ can be any subset of k elements of $(x_1, \dots, x_n)$

- **Function of a RV:** (X, Y) an RV with CDF $F(\cdot,\cdot)$, $g: \mathbb{R}^2 \to \mathbb{R}$ a Borel measurable function

then $g(X,Y)$ is a univariate RV: $E(g(X,Y)) = \begin{cases} \sum_{(x,y) \in S} g(x,y) \cdot P(X = x, Y = y) \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) \cdot f(x,y)\, dy\, dx \end{cases}$

Prop: $\begin{cases} E(X_i) \text{ exists} \\ 1 \le i \le n \end{cases} \Rightarrow E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$
↳ fixed, not random!

Proof: • cont RV: $E(X+Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f(x,y)\, dy\, dx = \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f(x,y)\, dy\right) dx + \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} f(x,y)\, dx\right) dy$

• discrete RV: Similar $= \int_{-\infty}^{\infty} x f_X(x)\, dx + \int_{-\infty}^{\infty} y f_Y(y)\, dy = E(X) + E(Y)$

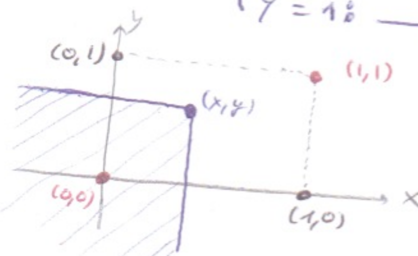
II Examples:

1 Discrete RV:

Ex: (X, Y) describes the presence of a disease in twins:

$\begin{cases} X = 0: & 1^{st} \text{ twin healthy} \\ X = 1: & \text{— has the disease} \\ Y = 0: & 2^{nd} \text{ twin healthy} \\ Y = 1: & \text{— has the disease} \end{cases}$

$Y \backslash X$	0	1	$P(Y=y)$
0	$P_{00} = 0{,}8$	$P_{10} = 0{,}02$	$P(Y=0) = 0{,}82$
1	$P_{01} = 0{,}05$	$P_{11} = 0{,}13$	$P(Y=1) = 0{,}17$
$P(X=x)$	$P(X=0) = 0{,}83$	$P(X=1) = 0{,}15$	Total: 1

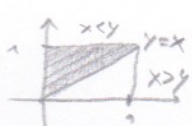


$F(x,y) = \begin{cases} 0 & , x < 0 \text{ or } y < 0 \\ 0{,}8 & , 0 \le x < 1, 0 \le y < 1 \\ 0{,}8 + 0{,}05 = 0{,}85 & , 0 \le x < 1, 1 \le y \\ 0{,}8 + 0{,}02 = 0{,}82 & , 1 \le x, 0 < y < 1 \\ 1 & , 1 \le x, 1 \le y \end{cases}$

$P(0 < X \le 1, 0 < Y \le 1) = F(1,1) - F(1,0) - F(0,1) + F(0,0)$
$= 1 - 0{,}82 - 0{,}53 + 0{,}8 = 0{,}13 = P_{11}$ ✓

2 Continuous RV:

Ex 1: (Final 2012) (X,Y) a cont RV with joint PDF: $f(x,y) = \begin{cases} c, & 0 < x < y < 1 \\ 0, & \text{otherwise} \end{cases}$



• Find c: $f(x,y) \ge 0 \Rightarrow c \ge 0$; $1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, dy\, dx = \int_0^1 \int_x^1 c\, dy\, dx = \int_0^1 cx\, dx = \frac{c}{2} \Rightarrow \boxed{c = 2}$

• Find f_X: $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy = \int_x^1 2\, dy = 2(1-x)$ if $0 < x < 1$

• Find f_Y: $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx = \int_0^x 2\, dx = 2x$ if $0 < y < 1$

$\Rightarrow f_X(x) = \begin{cases} 2(1-x), & 0 < x < 1 \\ 0, & \text{o/wise} \end{cases}$

$\Rightarrow f_Y(y) = \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{o/wise} \end{cases}$

Ex 2: (X,Y) with joint CDF:

$F(x,y) = \begin{cases} \min(x,y) & , 0 \le x < 1, 0 \le y < 1 \\ x & , 0 \le x < 1, 1 \le y \\ y & , 1 \le x, 0 \le y < 1 \\ 1 & , 1 \le x, 1 \le y \\ 0 & , \text{otherwise} \end{cases}$

• find F_X: $F_X(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x < 1 \\ 1, & 1 \le x \end{cases}$

• find F_Y: $F_Y(y) = \begin{cases} 0, & y < 0 \\ y, & 0 \le y < 1 \\ 1, & 1 \le y \end{cases}$

• find f_X: $f_X(x) = \begin{cases} 1, & 0 < y < 1 \\ 0, & \text{o/wise} \end{cases}$

• find f_Y: $f_Y(y) = \begin{cases} 1, & 0 < y < 1 \\ 0, & \text{o/wise} \end{cases}$

III Multinomial distribution : $\underline{X} = (X_1, \dots, X_{k-1}) \sim$ multinomial $(m, p_1, \dots, p_{k-1})$

Setup : A random experiment with k possible outcomes :

outcome	1	2	---	k
Prob of outcome	p_1	p_2	---	p_k

where $\sum_{j=1}^{k} p_j$, so $p_k = 1 - \sum_{j=1}^{k-1} p_j$

That experiment is repeated m times independently and under the same condit°.

$X_j :=$ # times the outcome is "j" $(j = 1, 2, \dots, m)$ $\Rightarrow \sum_{j=1}^{m} X_j = 1$, so $X_m = 1 - \sum_{j=1}^{m-1} X_j$

RK : The k^{th} RV depends on the first $(k-1)$ RVs

so take : $\underline{X} = (X_1, \dots, X_{k-1})$

$$P(X_1 = x_1, \dots, X_{k-1} = x_{k-1}) = \binom{m}{x_1, \dots, x_k} \prod_{j=1}^{k} p_j^{x_j} \quad \text{where } \binom{m}{x_1, \dots, x_k} = \frac{m!}{x_1! \cdots x_k!}$$

with : $p_k = 1 - \sum_{j=1}^{k-1} p_j$, $x_k = m - \sum_{j=1}^{k-1} x_j$

Note : Each $X_j \sim Bin(m, p_j)$

IV Conditional distributions / RV :

- **Discrete RV** : discrete RV (X, Y) with joint PMF : $\begin{cases} p_{ij} = P(X = x_i, Y = y_j) & , i, j = 1, 2, \dots \\ \sum_{i,j} p_{ij} = 1 \end{cases}$
  - Conditional PMF of X given $Y = y_j$: $P(X = x_i \mid Y = y_j) = \frac{p_{ij}}{p_{\cdot j}} \quad \forall i = 1, 2, \dots$; $p_{\cdot j} = P(Y = y_j) \neq 0$.

    Note : $\left\{ \frac{p_{ij}}{p_{\cdot j}} : i = 1, 2, \dots \right\}$ is the cond PMF of X given $Y = y_j$ for each $j = 1, 2, \dots$ fixed (as) $\sum_{i=1}^{\infty} \frac{p_{ij}}{p_{\cdot j}} = 1$
  - Cond PMF of Y given $X = x_i$: $P(Y = y_j \mid X = x_i) = \frac{p_{ij}}{p_{i\cdot}} \quad \forall j = 1, 2, \dots$; $p_{i\cdot} = P(X = x_i) \neq 0$

    Note : $\left\{ \frac{p_{ij}}{p_{i\cdot}} : j = 1, 2, \dots \right\}$ is the cond PMF of Y given $X = x_i$ for each $i = 1, 2, \dots$ fixed (as) $\sum_{j=1}^{\infty} \frac{p_{ij}}{p_{i\cdot}} = 1$

  Ex : Twins : $i \in \{0, 1\}$: $P(X = i \mid Y = 0) = \frac{P(X = i, Y = 0)}{P(Y = 0)} = \frac{p_{i0}}{p_{\cdot 0}} = \frac{p_{i0}}{0{,}82} \Rightarrow P(X = i \mid Y = 0) = \begin{cases} 80/82 & , i = 0 \\ 2/82 & , i = 1 \end{cases}$

- **Cont RV** : cont RV (X, Y) with joint PDF : $f(x, y)$
  - Cond CDF of X given $Y = y$: $F_{X|Y=y}(x) = P(X \leq x \mid Y \leq y) = \int_{-\infty}^{x} f_{X|Y=y}(t)\, dt$
  - Cond PDF of X given $Y = y$: $f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)}$ $\rightarrow \int_{-\infty}^{\infty} f_{X|Y=y}(x)\, dx = \frac{\int f(x, y)\, dx}{f_Y(y)} = \frac{f(y)}{f(y)} = 1$
  - Cond CDF of Y given $X = x$: $F_{Y|X=x}(y) = P(Y \leq y \mid X \leq x) = \int_{-\infty}^{y} f_{Y|X=x}(t)\, dt$
  - Cond PDF of Y given $X = x$: $f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)}$

  Ex : $f(x, y) = \begin{cases} 2 & , 0 < x < y < 1 \\ 0 & , \text{o'wise} \end{cases} \Rightarrow \begin{cases} f_X(x) = \begin{cases} 2(1-x) & , 0 < x < 1 \\ 0 & , \text{o'wise} \end{cases} \\ f_Y(y) = \begin{cases} 2y & , 0 < y < 1 \\ 0 & , \text{o'wise} \end{cases} \end{cases} \Rightarrow f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)} = \begin{cases} \frac{1}{y} & , 0 < x < y < 1 \\ 0 & , \text{o'wise} \end{cases} \Rightarrow$ Uniform distributo (Y fixed)

$$\boxed{\text{Cond PMF} = \frac{\text{Joint PMF}}{\text{Marginal PMF} > 0}} \qquad \boxed{\text{Cond PDF} = \frac{\text{Joint PDF}}{\text{Marginal PDF} > 0}}$$

- **Expectation Values:** (X,Y) an RV with joint CDF $F_{X,Y}(x,y)$

  - $E[X|Y=y] = \begin{cases} \sum_{x\in S_x} x\, P(X=x|Y=y) & , (X,Y) \text{ discrete} \\ \int_{-\infty}^{\infty} x\, f_{X|Y=y}(x)\, dx & , (X,Y) \text{ cont} \end{cases}$

    Note: $E[X|Y=y]$ is a function of y.
    ↳ it's an RV!

  - h: a Borel measurable f°: $E[h(X)|Y=y] = \begin{cases} \sum_{x\in S_x} h(x)\, P(X=x|Y=y) & , (X,Y) \text{ discrete} \\ \int_{-\infty}^{\infty} h(x)\, f_{X|Y=y}(x)\, dx & , (X,Y) \text{ cont} \end{cases}$

Prop: • $E(X) = E(E(X|Y))$

$(E(|h(X)|)<\infty)$ • $E(h(X)) = E(E(h(X)|Y))$

(pf) $g(y) = E(X|Y=y)$

So $E[E(X|Y)] = E(g(Y)) = \int_{-\infty}^{\infty} g(y) f_Y(y)\,dy = \int_{-\infty}^{\infty} E(X|Y=y) f_Y(y)\,dy = \int_{-\infty}^{\infty} f_Y(y) \int_{-\infty}^{\infty} x\, f_{X|Y=y}(x)\,dx\,dy$

$= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_Y(y) f_{X|Y=y}(x)\,dy\,dx = \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,dx = \int_{-\infty}^{\infty} x f_X(x)\,dx = E(X)$.

- **Variance:** $Var[X|Y=y] = E[(X-\mu_{X|Y=y})^2|Y=y]$

  $Var[X|Y=y] = E[X^2|Y=y] - (E[X|Y=y])^2$

  Where: $\mu_{X|Y=y} = E[X|Y=y]$

  Prop: $Var[X] = Var[E(X|Y)] + E[Var(X|Y)] \implies Var[E(X|Y)] \leq Var[X]$

Ex:  $f_{X,Y}(x,y) = \begin{cases} 2, & 0<x<y<1 \\ 0, & \text{o/wise} \end{cases}$ $\quad f_X(x) = \begin{cases} 2(1-x), & 0<x<1 \\ 0 & \text{o/wise} \end{cases}$ $\quad f_Y(y) = \begin{cases} 2y, & 0<y<1 \\ 0, & \text{o/wise} \end{cases}$ $\quad$ claim: $\begin{cases} E(X)=\frac{1}{3}, \; E(Y)=\frac{2}{3} \\ E(X|Y=y)=\frac{y}{2} \end{cases}$

Method 1: $E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^1 2x(1-x)\,dx = \frac{1}{3}$, $E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \int_0^1 2y^2\,dy = \frac{2}{3}$

Method 2: $f_{X|Y=y}(x) = \begin{cases} \frac{1}{y}, & 0<x<y<1 \\ 0, & \text{o/wise} \end{cases} \implies E(X|Y=y) = \int_{-\infty}^{\infty} x f_{X|Y=y}(x)\,dx = \int_0^y x\cdot\frac{1}{y}\,dx = \frac{y}{2}$

So $E(X) = E(E(X|Y=y)) = E(\frac{Y}{2}) = \frac{1}{2}E(Y) \implies E(X) = \frac{1}{2}\cdot\frac{2}{3} = \frac{1}{3}$.

V Independent RV:

Independance: RV (X,Y) with joint CDF $F(x,y)$ & marginals $F_X(x)$ & $F_Y(y)$

X,Y are indep $\iff F_{X,Y}(x,y) = F_X(x) F_Y(y), \; \forall (x,y)$

Notation: $X \perp\!\!\!\perp Y$

$\iff \begin{cases} P(X=x, Y=y) = P_X(X=x) P_Y(Y=y) & , \forall (x,y), (X,Y) \text{ discrete} \\ f_{X,Y}(x,y) = f_X(x) f_Y(y) & , \forall x,y, (X,Y) \text{ cont} \end{cases}$

$\iff P(X\in B_1, Y\in B_2) = P(X\in B_1) P(Y\in B_2) \; \forall B_1, B_2 \in \mathcal{B}(\mathbb{R})$.

Prop: • $X \perp\!\!\!\perp Y \implies \forall g_1, g_2$ Borel measurable f°: $g_1(X) \perp\!\!\!\perp g_2(Y)$

(pf) $P(g_1(X)\leq x, g_2(Y)\leq y) = P(X\in g_1^{-1}(-\infty,x], Y\in g_2^{-1}(-\infty,y]) = P(X\in g_1^{-1}(-\infty,x])\cdot P(Y\in g_2^{-1}(-\infty,y]) = P(g_1(X)\leq x)\cdot P(g_2(Y)\leq y)$

• $X \perp\!\!\!\perp Y \implies \begin{cases} P(X=x|Y=y) = P(X=x) & , \text{discrete } (X,Y) \\ f_{X|Y}(x|y) = f_X(x) & , \text{cont } (X,Y) \end{cases}$

Thm: • (X,Y) indep RV, $E(|X|)<\infty$, $E(|Y|)<\infty$: $\begin{cases} E(XY) = E(X)E(Y) \\ E(g_1(X)g_2(Y)) = E(g_1(X))E(g_2(X)) \end{cases}$ $\forall$ Borel meas. f° g_1, g_2.

• $(X_1, \dots, X_n)$ mutually indep, $E(|X_i|)<\infty$: $E(\prod_{i=1}^n X_i) = \prod_{i=1}^n E(X_i)$

(pf) $E(g_1(X), g_2(Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x) g_2(y) f(x,y)\,dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x) g_2(y) f_X(x) f_Y(y)\,dx\,dy = \left(\int_{-\infty}^{\infty} g_1(x) f_X(x)\,dx\right)\left(\int_{-\infty}^{\infty} g_2(y) f_Y(y)\,dy\right)$

$= E(g_1(X))E(g_2(Y))$

Identically distributed: RVs X & Y are Ident. dist $\iff F_X(x) = F_Y(x) \; \forall x\in\mathbb{R}$ $\rightsquigarrow$ Notation: $X \stackrel{d}{=} Y$

↳ $X_1, \dots, X_n$ are Ident. dist $\iff F_{X_i}(x) = F_{X_j}(x) \; \forall x\in\mathbb{R}, \forall i,j\in\{1,\dots,n\}$ $\rightsquigarrow X_i \stackrel{d}{=} X_j$

Indep Ident dist (IID): $X_1, \dots, X_n$ are indep ⊕ Ident. dist.

Ex: Toss a fair coin twice: $\Omega = \{HH, HT, TH, TT\} \implies \begin{cases} X = \# H \rightarrow S_X = \{0,1,2\} \\ Y = \# T \rightarrow S_Y = \{0,1,2\} \end{cases} \implies X+Y=2$ So $X \neq Y$

But: $F_X(x) = \begin{cases} 0 & , x<0 \\ 1/4 & , 0\leq x<1 \\ 3/4 & , 1\leq x<2 \\ 1 & , 2\leq x \end{cases}$ $\quad F_Y(y) = \begin{cases} 0 & , y<0 \\ 1/4 & , 0\leq y<1 \\ 3/4 & , 1\leq y<2 \\ 1 & , 2\leq y \end{cases} \implies F_X(x) = F_Y(x) \; \forall x\in\mathbb{R} \implies X \stackrel{d}{=} Y$

VI. Transformation of RV :

Goal : • $\vec{X} = (X_1, \dots, X_n)$ an n-dimensional Random Vector with CDF : $F_{\vec{X}}(X_1, \dots, X_n)$

• $\vec{Y} = g(\vec{X})$, $g : \mathbb{R}^n \to \mathbb{R}^m$ Borel measurable $\leadsto$ $\vec{Y} = (g_1(X_1, \dots, X_n), \dots, g_m(X_1, \dots, X_n)) = g(X_1, \dots, X_n)$

$\longrightarrow$ Find the dist of $\vec{Y}$

Ex : $Y = X_1 \pm X_2$, $Y = X_1 \overset{\times}{\div} X_2$ $(g : \mathbb{R}^2 \to \mathbb{R}^1)$, $Y = X_1 + \dots + X_n$ $(g : \mathbb{R}^n \to \mathbb{R}^1)$

Discrete case = X_i's discrete $\Longrightarrow$ Y_i's discrete : $P(Y_1 = y_1, \dots, Y_m = y_m) = \sum\limits_{\substack{\vec{x} = (x_1, \dots, x_n) \\ \text{ST } g(\vec{x}) = \vec{y} = (y_1, \dots, y_m)}} P(X_1 = x_1, \dots, X_n = x_n)$

Cont case = $\vec{X} = (X_1, \dots, X_n)$ a cont RV, with PDF : $f(x_1, \dots, x_n)$

• Joint CDF of $\vec{Y} = g(\vec{X})$: $F_{\vec{Y}}(y_1, \dots, y_m) = P(Y_1 \le y_1, \dots, Y_m \le y_m) = \int\limits_{\{\vec{x} : g_i(x_1, \dots, x_n) \le y_i, 1 \le i \le m\}} f(x_1, \dots, x_n)\, dx_1 \dots dx_n$

• Joint PDF of $\vec{Y} = g(\vec{X})$:

Thm : $\vec{Y} = (Y_1, \dots, Y_n) = (g_1(X_1, \dots, X_n), \dots, g_n(X_1, \dots, X_n))$, $g : \mathbb{R}^n \to \mathbb{R}^n$ a cont 1-1 mapping.

Let $h = g^{-1}$ be cont : $\begin{cases} X_1 = h_1(Y_1, \dots, Y_n) \\ \vdots \\ X_n = h_n(Y_1, \dots, Y_n) \end{cases}$ and Suppose $\frac{\partial h_i}{\partial y_j}$, i.e. $\frac{\partial X_i}{\partial y_j}$ are cont and $\exists$.

Then $\boxed{f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = f_{\vec{X}}(h_1(y_1, \dots, y_n), \dots, h_n(y_1, \dots, y_n)) \cdot |J|}$

where :
$$J = \det \begin{pmatrix} \frac{\partial X_1}{\partial y_1} & \cdots & \frac{\partial X_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial X_n}{\partial y_1} & \cdots & \frac{\partial X_n}{\partial y_n} \end{pmatrix} \neq 0$$
(Since)

Ex : • $n = m = 2$:

$F_Y(y_1, y_2) = P(Y_1 \le y_1, Y_2 \le y_2) = P(g_1(X_1, X_2) \le y_1, g_2(X_1, X_2) \le y_2) = \iint\limits_{\{(x_1, x_2) : g_1(x_1, x_2) \le y_1, \, g_2(x_1, x_2) \le y_2\}} f(x_1, x_2)\, dx_1\, dx_2$

• X_1, X_2 Indep with PDF : $f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{o'wise} \end{cases}$; $Y_1 = X_1 + X_2$, $Y_2 = X_1 - X_2$ $\leadsto$ Find joint PDF of (Y_1, Y_2)

$\begin{cases} y_1 = x_1 + x_2 \\ y_2 = x_1 - x_2 \end{cases} \Longleftrightarrow \begin{cases} x_1 = \frac{1}{2} y_1 + \frac{1}{2} y_2 \\ x_2 = \frac{1}{2} y_1 - \frac{1}{2} y_2 \end{cases} \Longrightarrow J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2} \Longrightarrow |J| = +\frac{1}{2}$.

So $f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right) \cdot \left|-\frac{1}{2}\right| \underset{\text{indep}}{=} \frac{1}{2} \cdot f_{X_1}\left(\frac{y_1 + y_2}{2}\right) \cdot f_{X_2}\left(\frac{y_1 - y_2}{2}\right) = \begin{cases} \frac{1}{2}, & \text{if } 0 < \frac{y_1 + y_2}{2} < 1 \text{ and } 0 < \frac{y_1 - y_2}{2} < 1 \\ 0, & \text{o'wise.} \end{cases}$

VII. Multivariate MGF :

MGF = MGF of $\vec{X} = (X_1, \dots, X_n)$ is : $M_{\vec{X}}(t_1, \dots, t_n) = E\left(e^{t_1 X_1 + \dots + t_n X_n}\right)$ for $|t_i - 0| < \delta$, $i = 1, \dots, n$

$\longrightarrow$ Need $E(|e^{\cdots}|) < \infty$

So $M_{\vec{X}} = \begin{cases} \sum\limits_{(x_1, \dots, x_n)} e^{t_1 x_1 + \dots + t_n x_n} P(X_1 = x_1, \dots, X_n = x_n) & , \vec{X} \text{ discrete} \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{t_1 x_1 + \dots + t_n x_n} f(x_1, \dots, x_n)\, dx_1 \dots dx_n & , \vec{X} \text{ continuous} \end{cases}$

Prop : • $M_{\vec{X}}(0, \dots, 0) = 1$; $M_{\vec{X}}(0, \dots, t_i, \dots, 0) = M_{X_i}(t_i) = E(e^{t_i X_i})$

• $M_{\vec{X}}(\cdot, \dots, \cdot)$ uniquely determine the dist of $\vec{X}$:

• If $n = 2$ and $M_{\vec{X}}(t_1, t_2)$ $\exists$: $E\{X_1^m X_2^n\} = \frac{\partial^{m+n} M_{\vec{X}}(t_1, t_2)}{\partial t_1^m \partial t_2^n}$ $\longrightarrow$ $(m+n)^{th}$ moment of (X_1, X_2)

• $X_1, \dots, X_n$ are indep $\Longleftrightarrow$ $M_{\vec{X}}(t_1, \dots, t_n) = \prod\limits_{i=1}^{n} E(e^{t_i X_i}) = \prod\limits_{i=1}^{n} M_{X_i}(t_i)$

• $X_1, \dots, X_n$ are IID $\Longleftrightarrow$ $M_{\vec{X}}(t_1, \dots, t_n) = \prod\limits_{i=1}^{n} M_{X_1}(t_i)$; $M_{\vec{X}}(t_1, \dots, t_1) = (M_{X_1}(t_1))^n$

VIII Covariance & Correlation:

Cauchy-Schwartz: $0 \leq E(|XY|) \leq \sqrt{E(X^2)E(Y^2)}$, as long as $E(X^2)<\infty$, $E(Y^2)<\infty$.

Proof: Young's ineq: $|ab| \leq \frac{|a|^2}{2}+\frac{|b|^2}{2}$ $\forall a,b \in \mathbb{R}$

Take $A=\frac{X}{\sqrt{E(|X|^2)}}$, $B=\frac{Y}{\sqrt{E(|Y|^2)}}$ $\Rightarrow \frac{|XY|}{\sqrt{E(|X|^2)E(|Y|^2)}} \leq \frac{|X|^2}{2E(|X|^2)}+\frac{|Y|^2}{2E(|Y|^2)} \xrightarrow{E(\dots)\leq E(\dots)} \frac{E(|XY|)}{\sqrt{E(|X|^2)\cdot E(|Y|^2)}} \leq \underbrace{\frac{E(|X|^2)}{2E(|X|^2)}+\frac{E(|Y|^2)}{2E(|Y|^2)}}_{=\frac{1}{2}+\frac{1}{2}=1}$

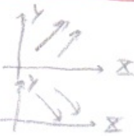
Hölder: $p,q>1$ st $\frac{1}{p}+\frac{1}{q}=1$: $0 \leq E(|XY|) \leq (E(|X|^p))^{1/p}(E(|Y|^q))^{1/q}$, as long as $E(|X|^p)<\infty$, $E(|Y|^q)<\infty$

Proof: Young $\Rightarrow |ab| \leq \frac{|a|^p}{p}+\frac{|b|^q}{q}$ $\forall p,q>1, \frac{1}{p}+\frac{1}{q}=1$; take $A=\frac{X}{(E(|X|^p))^{1/p}}$, $B=\frac{Y}{(E(|Y|^q))^{1/q}}$

Covariance: the covariance (if it exists) between 2 RV X and Y: $\boxed{Cov(X,Y)=\sigma_{X,Y}=E[(X-\mu_X)(Y-\mu_Y)]=E[XY]-E(X)E(Y)}$

$\sigma_{X,Y}$ measures the joint variat° between X & Y:

$\begin{cases} \sigma_{X,Y}>0 : X \text{ \& } Y \text{ are concordant } (X\uparrow \Leftrightarrow Y\uparrow) \\ \sigma_{X,Y}<0 : X \text{ \& } Y \text{ are discordant } (X\uparrow \Leftrightarrow Y\searrow) \\ \sigma_{X,Y}=0 : X \text{ \& } Y \text{ are uncorrelated} \end{cases}$



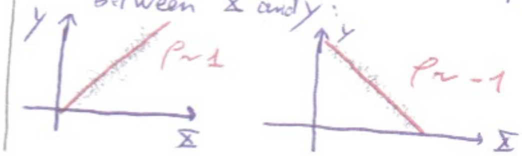
Prop:
- $Cov(X,X)=Var(X)$, i.e. $\sigma_{X,X}=\sigma_X^2$
- $Cov(aX+b, cY+d)=ac\cdot Cov(X,Y)$; i.e. $\sigma_{aX+b,cY+d}=ac\,\sigma_{X,Y}$
- X,Y indep $\Rightarrow Cov(X,Y)=0$
- $Cov(X,Y)=0 \not\Rightarrow X,Y$ indep

Correlation: the correlat° between 2 RV X and Y is: $\boxed{\rho_{X,Y}=Corr(X,Y)=\frac{Cov(X,Y)}{\sqrt{\underbrace{Var(X)}_{=\sigma_X^2}\cdot\underbrace{Var(Y)}_{=\sigma_Y^2}}}=\frac{\sigma_{X,Y}}{\sigma_X\cdot\sigma_Y}}$

↳ Need: $0<Var(X)<\infty$, $0<Var(Y)<\infty$

Prop:
- $|\rho_{X,Y}| \leq 1$, $\rho_{X,X}=1$, $\rho_{X,Y}=\rho_{Y,X}$
- $Y=aX+b \Leftrightarrow \rho_{X,Y}=\begin{cases}+1, & a>0\\ -1, & a<0\end{cases}$
- $\rho_{aX+b,cX+d}=Sign(ac)\cdot\rho_{X,Y}$
- X,Y indep $\Rightarrow \rho_{X,Y}=0$, but $\rho_{X,Y}=0 \not\Rightarrow X,Y$ indep

$\rho_{X,Y}$ measures the Linear relationship between X and Y:



Proof: (1) Proof 1: $\sigma_{X,Y} \leq |\sigma_{X,Y}|=|E[(X-\mu_X)(Y-\mu_Y)]| \leq E[|X-\mu_X|\cdot|Y-\mu_Y|] \underset{\text{Cauchy-Schwartz}}{\leq} \sqrt{E[(X-\mu_X)^2]\cdot E[(Y-\mu_Y)^2]}=\sqrt{\sigma_X^2\sigma_Y^2}$

$\Rightarrow |\sigma_{X,Y}| \leq \sigma_X\sigma_Y$ $\Rightarrow \left|\frac{\sigma_{X,Y}}{\sigma_X\sigma_Y}\right| \leq 1 \Rightarrow |\rho_{X,Y}| \leq 1$

Proof 2: $h(t)=E[((X-\mu_X)+t(Y-\mu_Y))^2] \geq 0$ $\forall t \in \mathbb{R}$ $\Rightarrow h(t)=\sigma_X^2+2t\,\sigma_{X,Y}+t^2\sigma_Y^2$

Need $\Delta=4\sigma_{X,Y}^2-4\sigma_X^2\sigma_Y^2 \leq 0 \Rightarrow \sigma_{X,Y}^2 \leq \sigma_X^2\sigma_Y^2 \Rightarrow |\rho_{X,Y}| \leq 1$

(2) Equality happens $\Leftrightarrow (X-\mu_X)+t(Y-\mu_Y)=0 \Leftrightarrow Y=aX+b$

(4) Ex: $\begin{cases}X\sim N(0,1)\\ Y=X^2\end{cases} \Rightarrow \begin{cases}E(XY)=E(X\cdot X^2)=E(X^3)=0 \text{ (odd f°)}\\ E(X)=0\end{cases} \Rightarrow \sigma_{X,Y}=E(XY)-E(X)E(Y)=0-0\cdot E(Y)=0$

$\rightarrow X,Y$ uncorrelated BUT $Y=X^2$ so not indep! $(f_{X,Y}\neq f_X\cdot f_Y)$

IX Linear Combination of RV:

Lin Comb: $X_1, \dots, X_n$ RV: $Y = \sum_{i=1}^n a_i X_i$, $a_i \in \mathbb{R}$ is a lin comb of $\{X_i\}$.

Ex: ① $a_i = 1$: $Y = \sum_{i=1}^n X_i$ → SUM

② $a_i = \frac{1}{n}$: $Y = \frac{1}{n}\sum_{i=1}^n X_i$ → AVG

Goal:
- Find $E(Y)$, $Var(Y)$
- Find $M_Y(t)$
- Find the dist of Y

① How To find $E(Y)$:

thm: $E(Y) = E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i)$

② How To find $Var(Y)$:

thm: $Var\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j Cov(X_i, X_j) = \sum_{i=1}^n a_i^2 Var(X_i) + \sum_{i \neq j} a_i a_j Cov(X_i, X_j)$

Proof: $Var(Y) = E\left((Y-\mu_Y)^2\right) = E\left[\left(\sum_{i=1}^n a_i (X_i - \mu_i)\right)^2\right] = E\left[\sum_{i=1}^n \sum_{j=1}^n a_i a_j (X_i - \mu_i)(X_j - \mu_j)\right]$

$= \sum_{i=1}^n \sum_{j=1}^n a_i a_j E\left[(X_i - \mu_i)(X_j - \mu_j)\right] = \sum_{i=1}^n \sum_{j=1}^n a_i a_j Cov(X_i, X_j)$

Corr: $X_1, \dots, X_n$ are uncorrelated (or indep):

- $Var\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 Var(X_i)$, so $\begin{cases} Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i) \\ Var\left(\sum_{i=1}^n \frac{1}{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^n Var(X_i) \end{cases}$

- if $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ $\forall i = 1, \dots, n$:

$\begin{cases} E\left(\sum_{i=1}^n X_i\right) = n\mu \\ Var\left(\sum_{i=1}^n X_i\right) = n\sigma^2 \end{cases}$ and $\begin{cases} E\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \mu \\ Var\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{\sigma^2}{n} \end{cases}$

③ How To find the dist of Y:

In general, find CDF.

If X_i's are indep: use MGF and recognize a famous dist:

thm: X_i's indep and $Y = \sum_{i=1}^n a_i X_i$: $M_Y(t) = \prod_{i=1}^n M_{X_i}(t a_i)$

Proof: $M_Y = E(e^{tY}) = E\left(e^{t a_1 X_1 + \dots + t a_n X_n}\right) \underset{\text{indep}}{=} E(e^{t a_1 X_1}) \dots E(e^{t a_n X_n}) = \prod_{i=1}^n M_{X_i}(t a_i)$

Ex: $X_i \sim N(\mu_i, \sigma_i^2)$ are indep:

$Y = \sum_{i=1}^n X_i \Rightarrow Y \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$

Proof: $M_Y(t) = \prod_{i=1}^n M_{X_i}(1 \cdot t) = \prod_{i=1}^n e^{t\mu_i + \frac{t^2 \sigma_i^2}{2}} \Rightarrow M_Y(t) = e^{t\sum_{i=1}^n \mu_i + \frac{t^2}{2}\sum_{i=1}^n \sigma_i^2} \Rightarrow Y \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$

IX Linear Combination of RV:

Lin Comb: $X_1, \dots, X_n$ RV: $Y = \sum_{i=1}^{n} a_i X_i$, $a_i \in \mathbb{R}$ is a lin comb of $\{X_i\}$.

Ex: ① $a_i = 1$: $Y = \sum_{i=1}^{n} X_i$ → SUM
② $a_i = \frac{1}{n}$: $Y = \frac{1}{n} \sum_{i=1}^{n} X_i$ → AVG

Goal:
- Find $E(Y)$, $Var(Y)$
- Find $M_Y(t)$
- Find the dist of Y

① How to find $E(Y)$:

thm: $E(Y) = E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$

② How to find $Var(Y)$:

thm: $Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j Cov(X_i, X_j) = \sum_{i=1}^{n} a_i^2 Var(X_i) + \sum_{i \neq j} a_i a_j Cov(X_i, X_j)$

Proof: $Var(Y) = E(Y - \mu_Y)^2 = E\left[\left(\sum_{i=1}^{n} a_i (X_i - \mu_i)\right)^2\right] = E\left[\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j (X_i - \mu_i)(X_j - \mu_j)\right]$
$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j E[(X_i - \mu_i)(X_j - \mu_j)] = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j Cov(X_i, X_j)$

Corr: $X_1, \dots, X_n$ are uncorrelated (or indep):

- $Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i)$, so $\begin{cases} Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) \\ Var\left(\sum_{i=1}^{n} \frac{1}{n} X_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) \end{cases}$
- if $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ $\forall i = 1, \dots, n$:

$\begin{cases} E\left(\sum_{i=1}^{n} X_i\right) = n\mu \\ Var\left(\sum_{i=1}^{n} X_i\right) = n\sigma^2 \end{cases}$ and $\begin{cases} E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \mu \\ Var\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{\sigma^2}{n} \end{cases}$

③ How to find the dist of Y:

In general, find CDF.

If X_i's are indep: use MGF and recognize a famous dist:

thm: X_i's indep and $Y = \sum_{i=1}^{n} a_i X_i$: $M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t a_i)$

Proof: $M_Y = E(e^{tY}) = E(e^{t a_1 X_1 + \dots + t a_n X_n}) \underset{\text{indep}}{=} E(e^{t a_1 X_1}) \dots E(e^{t a_n X_n}) = \prod_{i=1}^{n} M_{X_i}(t a_i)$

Ex: $X_i \sim N(\mu_i, \sigma_i^2)$ are indep:

$Y = \sum_{i=1}^{n} X_i \Rightarrow Y \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$

Proof: $M_Y(t) = \prod_{i=1}^{n} M_{X_i}(1 \cdot t) = \prod_{i=1}^{n} e^{t\mu_i + \frac{t^2 \sigma_i^2}{2}} \Rightarrow M_Y(t) = e^{t \sum_{i=1}^{n} \mu_i + \frac{t^2}{2} \sum_{i=1}^{n} \sigma_i^2} \Rightarrow Y \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$

Convergence & limit Theorems :

(I) Convergence :

Let $\begin{cases} (X_n)_{n=1}^{\infty} \text{ be a seq of RV with CDF } (F_n)_{n=1}^{\infty} , & X_n : \Omega \longrightarrow \mathbb{R} , X_n(\omega) \in \mathbb{R} \\ X \text{ be a RV with CDF } F(\cdot) , & X : \Omega \longrightarrow \mathbb{R} , X(\omega) \in \mathbb{R} \end{cases}$

- Convergence in probability =

$(X_n)_{n=1}^{\infty}$ Converges to X in probability if : $\forall \epsilon > 0 , \lim_{n \to \infty} P(|X_n - X| > \epsilon) = 0$

i.e. $P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}) \longrightarrow 0$ as $n \to \infty$ Notation = $X_n \xrightarrow{P} X$

Prop = As $n \to \infty$:

- $X_n \xrightarrow{P} X \Rightarrow X_n - X \xrightarrow{P} 0$ look at it as a degenerate RV
- $X_n \xrightarrow{P} 1 \Rightarrow \frac{1}{X_n} \xrightarrow{P} 1$ $(X_n \neq 0 \ \forall n)$
- $\left.\begin{matrix} X_n \xrightarrow{P} X \\ Y_n \xrightarrow{P} Y \end{matrix}\right\} \Rightarrow \begin{cases} X_n \pm Y_n \xrightarrow{P} X \pm Y \\ X_n Y_n \xrightarrow{P} XY \end{cases}$
- g a cont f^o : $X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$

Ex : $X_n \xrightarrow{P} c \in \mathbb{R} \Rightarrow g(X_n) \xrightarrow{P} g(c)$

Weak Law of Large nb = $(X_n)_{n=1}^{\infty}$ a Seq of IID RV , $\begin{cases} E(X_i) = \mu < \infty \\ Var(X_i) = \sigma^2 < \infty \end{cases}$ $\forall i \in \mathbb{N}$

Then : $\overline{X_n} := \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{P} \mu = E(X_i)$, as $n \to \infty$ $\Rightarrow$ Need : $E(X_i^2) < \infty$

$\hookrightarrow$ Look at is as a degenerate RV !

Sample average $\xrightarrow{P}$ Population average

Proof = $\left.\begin{matrix} E(\overline{X_n}) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \mu n = \mu \\ Var(\overline{X_n}) = \sigma^2/n \end{matrix}\right\} \Rightarrow 0 \leq P(|\overline{X_n} - \mu| > \epsilon) \underset{\text{chebychev}}{\leq} \frac{E(\overline{X_n} - \mu)^2}{\epsilon^2} = \frac{Var(\overline{X_n})}{\epsilon^2} = \frac{\sigma^2}{n \epsilon^2} \underset{n \to \infty}{\longrightarrow} 0$

Monte-Carlo Approx. for integrals = $g : [0,1] \longrightarrow \mathbb{R}$ a cont function :

Then : $\frac{1}{n} \sum_{i=1}^{n} g(U_i) \xrightarrow{P} \int_0^1 g(t)\,dt$, where $(U_i)_{i=1}^{\infty}$ is a seq of IID RV generated from Unif$([0,1])$

Quality of Approx : $P\left(\left| \frac{1}{n} \sum_{i=1}^{n} g(U_i) - \int_0^1 g(t)\,dt \right| < \epsilon\right) \geq 1 - \frac{\|g^2\|}{n \epsilon^2}$ $\left(\|g^2\| = \sup_{u \in [0,1]} g^2(u)\right)$

Proof = $\begin{cases} \int_0^1 g(t)\,dt = \int_0^1 1 \cdot g(t)\,dt = E[g(U)] , U \sim Unif([0,1]) \\ Var(g(U)) \leq E(g^2(U)) = \int_0^1 g^2(t)\,dt \leq \sup_{u \in [0,1]} g^2(u) = \|g^2\| < \infty \end{cases}$ because g cont $\Rightarrow$ Sup $\exists$

so $Var(g(U)) < \infty \Rightarrow$ Weak Law of Large nb : $\frac{1}{n} \sum_{i=1}^{n} g(U_i) \xrightarrow{P} \int_0^1 g(t)\,dt$

chebychev's ineq $\Rightarrow P\left(\left|\frac{1}{n} \sum_{i=1}^{n} g(U_i) - \int_0^1 g(t)\,dt\right| < \epsilon\right) \geq 1 - \frac{\|g^2\|}{n \epsilon^2}$

- Convergence in distribution:

$(X_n)_{n=1}^{\infty}$ Converges to X in distribution if: $\exists$ an RV X with CDF F, ST: $\lim_{n\to\infty} F_n(x) = F(x)$ at every continuity pt of F.

Notation: $X_n \xrightarrow{d} X$, or $X_n \xrightarrow{L} X$ ↳ in Law, or $F_n \xrightarrow{w} F$ ↳ weak conv

Ex:

- ⚠ It can happen that $F_n \to F$, F_n is a CDF, but F is NOT a CDF!

$F_n(x) = \begin{cases} 0, & x < n \\ 1, & x \geq n \end{cases} \Rightarrow F_n \longrightarrow F(x) = 0 \ \forall x \in \mathbb{R}$, So F_n is a CDF BUT NOT F!

- $F_n(x) = \begin{cases} 0 & , x < 0 \\ \left(\frac{x}{\theta}\right)^n & , 0 \leq x < \theta \\ 1 & , \theta \leq x \end{cases}$: $\lim_{n\to\infty} F_n(x) = F(x)$ where $F(x) = \begin{cases} 0, & x < \theta \\ 1, & x \geq \theta \end{cases} \Rightarrow X_n \to X$ where X is a degenerate RV $P(X=\theta)=1, P(X\neq\theta)=0$

Prop: As $n \to \infty$:

- $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$ BUT: $X_n \xrightarrow{d} X \not\Rightarrow X_n \xrightarrow{P} X$
- $X_n \xrightarrow{d} c \in \mathbb{R} \Rightarrow X_n \xrightarrow{P} X$

↳ degenerate RV (i.e. $c = cst$)

Slutsky's thm: $(X_n)_{n=1}^{\infty}, (Y_n)_{n=1}^{\infty}$: 2 Seq of RV

(1) $\begin{cases} X_n \xrightarrow{d} X \\ Y_n \xrightarrow{P} c \in \mathbb{R} \end{cases} \Rightarrow \begin{cases} X_n + Y_n \xrightarrow{d} X + c \\ X_n Y_n \xrightarrow{d} cX \end{cases}$ as $n \to \infty$

(2) $\begin{cases} X_n \xrightarrow{d} X \\ Y_n \xrightarrow{P} 0 \end{cases} \Rightarrow X_n Y_n \xrightarrow{P, d} 0$ as $n \to \infty$

(3) $\begin{cases} X_n \longrightarrow X \\ Y_n \longrightarrow c \neq 0 \end{cases} \Rightarrow \dfrac{X_n}{Y_n} \xrightarrow{d} \dfrac{X}{c}$ as $n \to \infty$

Proof: (1) Let x and $x+\varepsilon$ be 2 continuity points of $F(\cdot)$ → CDF of X (※) F is monotone ↗ ⇒ F has a Countable nb of discont

Let $c=0$: $Y_n \xrightarrow{P} 0 \Rightarrow P(|Y_n - 0| > \varepsilon) \to 0$

- Total prob ⇒ $P(X_n + Y_n \leq x) = P(X_n + Y_n \leq x, |Y_n - 0| > \varepsilon) + P(X_n + Y_n \leq x, |Y_n - 0| \leq \varepsilon)$ ↳ $-\varepsilon \leq Y_n \leq \varepsilon$

$\leq \underbrace{P(|Y_n| > \varepsilon)}_{\to 0 \text{ as } n\to\infty} + P(X_n \leq x + \varepsilon)$

$= F_{X_n}(x+\varepsilon) \longrightarrow F(x+\varepsilon)$ as $n \to \infty$.

So $\limsup_{n\to\infty} P(X_n + Y_n \leq x) \leq F(x+\varepsilon)$

- Total prob ⇒ $P(X_n + Y_n \leq x) = P(X_n + Y_n \leq x, |Y_n - 0| > \varepsilon) + P(X_n + Y_n \leq x, |Y_n - 0| \leq \varepsilon)$ ↳ $-\varepsilon \leq Y_n \leq \varepsilon$

$\geq \underbrace{-P(|Y_n| > \varepsilon)}_{\to 0 \text{ as } n\to\infty} + P(X_n \leq x - \varepsilon)$.

$= F_{X_n}(x-\varepsilon) \longrightarrow F(x-\varepsilon)$ as $n \to \infty$.

So $\liminf_{n\to\infty} P(X_n + Y_n \leq x) \geq F(x-\varepsilon)$

⇨ let $\varepsilon \to 0$: $\lim_{n\to\infty} F_{X_n+Y_n}(x) = F(x)$ $\forall$ continuity pt x of $F(\cdot)$ $\Rightarrow X_n + Y_n \xrightarrow{d} X + 0$ ✓

↳ CDF of $X_n + Y_n$

if $c \neq 0$:

II) Central limit theorem :

CLT : $(X_n)_{n=1}^{\infty}$ a Seq of IID RV with $\begin{cases} E(X_i) = \mu < \infty \\ Var(X_i) = \sigma^2 < \infty \end{cases}$

Then : $$Z_n := \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1) \text{ as } n \to \infty$$

where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ → Sample mean

RECALL : $\begin{cases} E(\overline{X}_n) = \mu \\ Var(\overline{X}_n) = \frac{\sigma^2}{n} \end{cases}$ ⇒ To get Z, we standardize $\overline{X}_n$ → $\begin{cases} E(Z) = 0 \\ Var(Z) = 1 \end{cases}$

Proof :

> RECALL : If $(Y_n)_{n=1}^{\infty}$ is a seq of RV with MGF $(M_n)_{n=1}^{\infty}$, ST $M_n(t)$ $\exists$ for $|t-0| < t_0$, $\forall n \in \mathbb{N}$
> and if $M_n(t) \longrightarrow \underset{\text{MGF of some RV } Y}{M(t)}$, $\forall |t-0| \leq t_1 \leq t_0$, $t_1 > 0$
> Then : $Y_n \xrightarrow{d} Y$ as $n \to \infty$

Assume $\mu = 0$, $\sigma^2 = 1$: So $Z_n = \sqrt{n}\,\overline{X}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$

So $M_{Z_n}(t) = E\left[e^{tZ_n}\right] = E\left[e^{\frac{t}{\sqrt{n}}\sum_{i=1}^{n} X_i}\right] = E\left[e^{\frac{t}{\sqrt{n}}X_1} \cdots e^{\frac{t}{\sqrt{n}}X_n}\right] \overset{IID}{=} \prod_{i=1}^{n} E\left[e^{\frac{t}{\sqrt{n}}X_i}\right] = \prod_{i=1}^{n} M_{X_i}(t/\sqrt{n})$

$\Rightarrow M_{Z_n}(t) = \left(M_{X_i}(t/\sqrt{n})\right)^n$ for any $1 \leq i \leq n$.

So $Log\left[M_{Z_n}(t)\right] = n \, Log\, M_{X_i}(t/\sqrt{n}) = \frac{Log(M_{X_i}(t/\sqrt{n}))}{1/n} \xrightarrow[n\to\infty]{\text{L'Hopital (twice)}} \frac{t^2}{2}$

(**) $\begin{cases} M_{X_i}(\frac{t}{\sqrt{n}}) \to M_{X_i}(0) = 1 \\ M'_{X_i}(t/\sqrt{n}) \to E_{X_i}(0) = \mu = 0 \\ M''_{X_i}(\frac{t}{\sqrt{n}}) \to E_{X_i^2}(0) = \sigma^2 - \mu^2 = 1 - 0 \end{cases}$

$\Rightarrow M_{Z_n}(t) \xrightarrow[n\to\infty]{} e^{\frac{t^2}{2}} \Rightarrow Z_n \xrightarrow{d} Z \sim N(0,1)$

Alternate forms : $$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} N(0,1)$$ (**) $\begin{cases} E\left[\sum_{i=1}^{n} X_i\right] = n\mu \\ Var\left[\sum_{i=1}^{n} X_i\right] = n\sigma^2 \end{cases}$

In practice : If $n \gtrsim 30$, then the following linear combination of any distribution is "well" approximated by a Normal dist :

$$\left.\begin{array}{l} \overline{X}_n \approx N(\mu, \frac{\sigma^2}{n}) \\ \text{or} \quad \sum_{i=1}^{n} X_i \simeq N(n\mu, n\sigma^2) \end{array}\right\} \text{ when } n \text{ is } \underline{\text{large}} \quad (n \gtrsim 30 \checkmark)$$

Ex : Let $\begin{cases} X_i \overset{IID}{\sim} Bernoulli(p) \\ Y_n = \sum_{i=1}^{n} X_n \end{cases}$:
1) Find the exact dist of Y_n → USE MGF !!!
2) Approximate the dist of Y_n by a normal dist → USE CLT !!!

1] $M_{Y_n}(t) = E\left[e^{tY_n}\right] = E\left[e^{t\sum_{i=1}^{n} X_i}\right] \overset{IID}{=} \left[M_{X_i}(t)\right]^n \overset{Bernoulli}{=} \left[1-p+pe^t\right]^n$ → Binomial !

$\Rightarrow Y_n \sim Bin(n, p)$

(**) Y_n Tells you How the # of Success because $Y_n = \sum x_i = 1+0+1+0+0+\cdots = \sum 1 = \sum \text{Success}$

2] Let $Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow[CLT]{d} N(0,1) \Rightarrow Y_n \underset{\text{in practice}}{\simeq} N(np, np(1-p))$ for large n.

Ex: Let $(X_n)_{n=1}^{\infty}$ be a Seq of IID RV : $\begin{cases} E(X_i) = \mu < \infty \\ Var(X_i) = \sigma^2 < \infty \end{cases}$, $E(|X_i|^4) < \infty$

- Let $S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$: show that $S_n^2 \xrightarrow{P} \sigma^2$ as $n \to \infty$

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

$\hookrightarrow$ "sample variance"

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2) = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{2\overline{X}_n}{n-1}\underbrace{\sum_{i=1}^{n} X_i}_{= n\overline{X}_n} + \frac{n\overline{X}_n^2}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\overline{X}_n^2$$

Weak Law of Large nb: $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[\text{(*)}\, E(X_i^2)<\infty]{P} E(X_i) = \mu \Rightarrow \overline{X}_n^2 \xrightarrow{P} \mu^2 \Rightarrow \frac{n}{n-1}\overline{X}_n^2 \to \frac{n}{n+1}\mu^2$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow[\text{(**)}\, E(X_i^4)<\infty]{P} E(X_i^2) = Var(X_i) + E(X_i)^2 = \sigma^2 + \mu^2 \Rightarrow \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 \to \frac{n}{n-1}(\sigma^2+\mu^2)$$

$\cdot \frac{n}{n}$ to get the $\frac{1}{n}$ term

$\Rightarrow S_n^2 \xrightarrow{P} \frac{n}{n+1}(\sigma^2+\mu^2) - \frac{n}{n+1}\mu^2 = \frac{n}{n+1}\sigma^2 \xrightarrow{P} \sigma^2$ as $n \to \infty$.

So $\begin{cases} S_n^2 \xrightarrow{P} \sigma^2 \\ \overline{X}_n \xrightarrow{P} \mu \end{cases}$ as $n \to \infty$

- Find the limiting distribution of : $\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}$ as $n \to \infty$ $\Rightarrow$ USE CLT!!!

$\hookrightarrow$ i.e. CDF

$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$

$S_n \xrightarrow{P} \sigma$

So $\frac{S_n}{\sigma} \xrightarrow{P} 1$

Slutsky's thm: $\frac{Z_n}{(S_n/\sigma)} = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \xrightarrow{P} \frac{N(0,1)}{1}$

So $\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \longrightarrow N(0,1)$

Also: $\frac{1}{n-1}\sum_{i=1}^{n} X_i^2 = \frac{n}{n-1}\cdot\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow[\text{(weak law of large nb)}]{P} 1\cdot E(X_i^2)$ (which exists by assumpt.)

So: $S_n^2 \xrightarrow[n\to\infty]{P} E(X_i^2) - \mu^2 = Var(X_i) = \sigma^2$

(as $Var(X_i) = E[(X_i-\mu)^2] = \sigma^2$

So $\begin{cases} S_n^2 \xrightarrow{P} \sigma^2 & \text{as } n\to\infty \\ \overline{X}_n \xrightarrow{P} \mu & \text{as } n\to\infty \end{cases}$

(2) Find the limiting distribution of: $\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}$ as $n\to\infty$
(i.e. find the CDF.

$\begin{cases} S_n \xrightarrow{P} \sigma \\ Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1) \end{cases} \xRightarrow{\text{slutsky thm \#3}}$

$$\frac{(\overline{X}_n - \mu)/\sigma}{(S_n/\sigma)/\sqrt{n}} \xrightarrow[\text{Slutsky thm}]{d} N(0,1)$$

$\to 1$

$\hookrightarrow 1$

Bivaritate normal dist:

X, Y indep $\Leftrightarrow$ X, Y are uncorrelated. is only true for (X,Y) = a bivariate normal dist.

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)\right]}$$

if $\rho = 0 \Rightarrow f(x,y) = f_X(x)\, f_Y(y)$

univariate normal.

In General: multivariate normal

$$f(x_1, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})}$$

$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$

$\hookrightarrow$ det of Var/Covar Matrix: $\Sigma = \begin{pmatrix} \sigma_1^2 & & \sigma_{ij} \\ & \ddots & \\ \sigma_{ij} & & \sigma_n^2 \end{pmatrix}$