Probability
MA 381
Prof. Gouvêa

Sep 5, 2018

- Problem Sets + Writing (5, 1-page, papers)
  - ↳ read / hear Stats/Math, think → tell ...

$\boxed{\text{Probabilistic ideas}}$ → ~1650, but games of chance had been around for a long time

- Fermat — Pascal correspondence
- EU: selling annuity
- life insurance
- Jakob Bernoulli → [Ars Conjectandi] → starting point of prob. theory

↳ $\dfrac{\text{green eyes}}{\text{all eyes}}$ ~ prob. of green eyes if "all eyes" are big → but how big?

⇒ $\boxed{\text{probability theory}}$ But what is a probability?

- subjective, uncertainty rather than randomness
- objective, empirical

$\boxed{\text{Kolmogorov}}$ 1930s → axioms → turn probability into abstract mathematics.

Axioms, Theorems ...

But the axioms are based on the real world ...

Sep 7, 2018 $\boxed{\text{To read 1.1 – 1.4}}$

$\boxed{\text{Sample Spaces . Probability Spaces}}$

Ex Roll a die

$$\Omega = \{ ⚀, ⚁, ⚂, ⚃, ⚄, ⚅ \}$$

Elements of Ω are usually called ω. ($\omega \in \Omega$)

<u>Events</u> all subsets of Ω . (careful with quantifier)

= elementary $\{⚀\} = A$, Ω (anything happens)

$B = \{⚀, ⚁, ⚂\}$, $\emptyset$ (nothing happens)

<u>All events</u> $\mathcal{F}$

<u>Probability</u> For any event A, $0 \leq P(A) \leq 1$

$$\boxed{P(\emptyset) = 0, \quad P(\Omega) = 1}$$

If A & B are disjoint, then $P(\underbrace{A \cup B}_{\text{union}}) = P(A) + P(B)$

(no outcome(s) in common) $A \cap B \neq \emptyset$

true for finite sums/unions $A_1 \cup A_2 \ldots \cup A_n \ldots$

Let die be fair $\rightarrow$ $\boxed{P(\omega) = \frac{1}{6}}$ (fair die)

("uniform")

$$\boxed{P(A) = \frac{\#A}{\#\Omega}}$$

<u>Summary</u> (1) Ω : sample space

(2) $\mathcal{F}$: collection of events (subsets of Ω)

(3) $P: \mathcal{F} \longmapsto [0,1]$

$P(\emptyset) = 0, \quad P(\Omega) = 1$

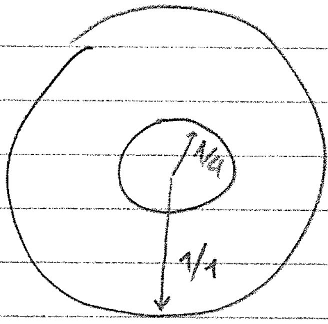
(4) If $A_1 \ldots A_n$ are events and $A_i \cap A_j = \emptyset$ if $i \neq j$ then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

Infinite sets

$$\Omega = \{(x,y) \mid x^2+y^2 \leq 1\}$$

$$A = \{(x,y) \mid x^2+y^2 \leq 1/16\}$$



Know

$$P(A) = \frac{\text{area}(A)}{\text{area}(\Omega)} = \frac{1}{16} \quad \text{But } \underline{\text{area}}???$$

Know $P(\Omega) = \sum P((x,y))$

If $P((x,y)) = 0$, then $P(\Omega)$ But $P(\Omega) = 1$
else, $P(\Omega) \to \infty$ WTF?

Additivity $A_1, A_2, A_3, \ldots A_n$ sequence of events. $A_i \cap A_j = \emptyset$ (disjoi-

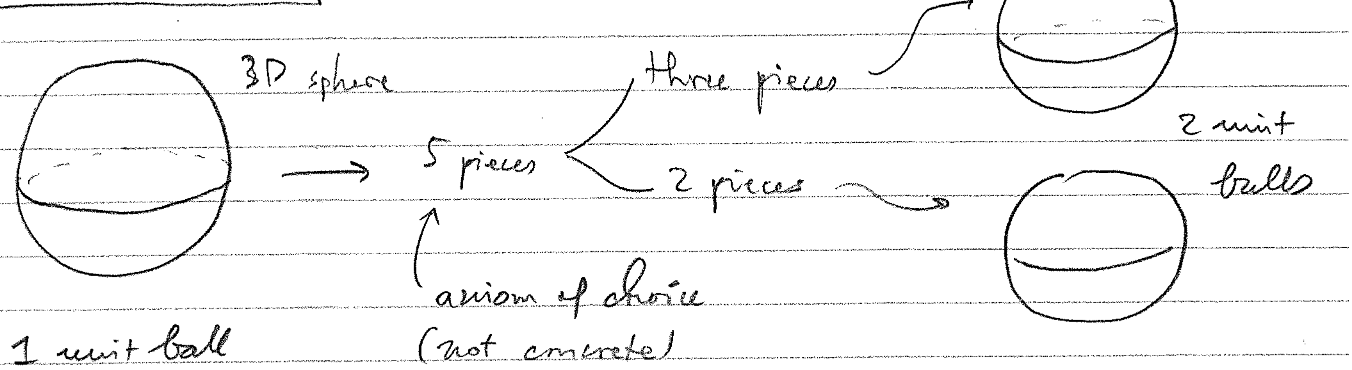
(Countability required) then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \longrightarrow$ infinite list of numbers.

$\boxed{\text{What is } \mathcal{F}?}$ (Borel sets)

- $\emptyset$ and Ω are events
- If A is an event, A^c is also an event
- A_i events $\Longrightarrow$ $\bigcup_{i=1}^{\infty} A_i$ is an event

$\bigcap_{i=1}^{\infty} A_i$ is an event

$\boxed{\text{Banach - Tarski}}$ Paradox



In practice, we rarely need to know Ω & $\mathcal{F}$

p 10, 2018

Recall probability space Ω : set of outcomes

$\mathcal{F}$: collection of "events" : subsets of Ω, including Ω, $\emptyset$, closed under complements & countable union & intersection

$P : \mathcal{F} \longmapsto [0,1]$

- $P(\Omega) = 1$
- $P(\emptyset) = 0$

if $A_1, A_2, \dots A_n \dots$ are events and $A_i \cap A_j = \emptyset$ if $i \neq j$, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

$\boxed{\text{Consequence}}$

$$\boxed{\text{If } B_1 \subset B_2 \subset B_3 \dots \text{ seq of events, then } P\left(\bigcup_{i=1}^{\infty} B_n\right) = \lim_{n\to\infty} P(B_n) \quad \text{(makes sense)}}$$

"Uniform". If Ω finite, $P(\omega) = \dfrac{1}{\#\Omega}$ $\forall \omega \in \Omega$

(1) Consequently $\boxed{P(A) = \dfrac{\#A}{\#\Omega}}$

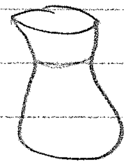
(2) $\Omega = [a, b]$ or region in plane...

$$\boxed{P(A) = \frac{size(A)}{size(\Omega)}}$$ → N.b. size of Ω has to be finite.

Assignment 1, Q4

↳ another way let $A = \{n\}$, $n \in \mathbb{N}$
↳ $P(A) = 0 \rightarrow P(\mathbb{N}) = 0$
However, $P(\mathbb{N}) = 1$ (infinite)
→ Contradiction

Sep 12, 2018



Urn with n-balls. Sample Uniform probability ≡ "choose at random"

(1) Sample k times with replacement, order matters

$$\Omega = \{(x_1, \dots x_k) \mid x_i = 1, \dots n\}$$

$\#\Omega = ?$ n choices, k-times $\Rightarrow$ $\boxed{\#\Omega = n^k}$

(2) Sample k times w/o replacement, order matters

$$\Omega = \{(x_1, \dots x_k) \mid x_i = 1, \dots n, \ x_i \neq x_j \text{ if } i \neq j\}$$

$$\#\Omega = \frac{n!}{(n-k)!} = (n)_k$$

(3) Sample k times, without replacement, order does not matter

$$\Omega = \{\{x_1, \dots x_k\} \subset \{1, 2, \dots n\}\}$$ (k-element subsets

$$\#\Omega = \frac{n!}{(n-k)!\,k!} = \binom{n}{k} = C_n^k$$

Note $\boxed{\binom{n}{k} = \binom{n}{n-k}}$

Note $\boxed{\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = 2^n}$ $\to$ # of subsets ...

Note sampling with replacement without order $\to$ Wait til Chapt. 4

~~P(at least~~ $\boxed{\text{BIRTHDAY PROBLEM}}$

$\to$ P(at least one shared birthday) = 1 − P(no shared birthday)

Let room = 18 people $\to$ P(~~at~~ shared birthday) $= \dfrac{\#A}{\#\Omega} = \dfrac{(366)_{18}}{366^{18}}$

$$= \frac{366}{366} \cdot \frac{365}{366} \cdots \frac{349}{366} \simeq 0.654$$

$\boxed{\text{Consequences}}$

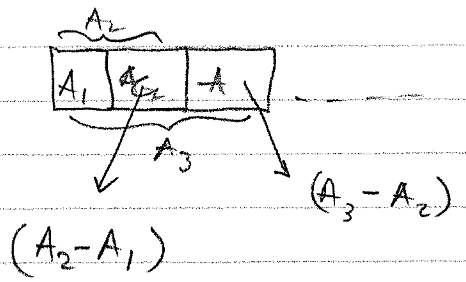
$\boxed{A \subset B \implies P(B) \geq P(A)}$ (monotonicity)

$B = A \cup B\setminus A$ it follows $P(B) = P(A) + \underbrace{P(B\setminus A)}_{\geq 0}$ So $P(B) \geq P(A)$

$A_1 \subset A_2 \subset A_3 \subset A_4 \subset \cdots \subset A_n \subset \cdots$

Let $A = \bigcup_{n=1}^{\infty} A_n$. Claim $P(A) = \lim_{n\to\infty} P(A_n)$

Proof



So $P(A) = P(A_1) + P(A_2 - A_1) + \cdots$

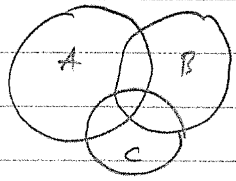
$$= \lim_{n\to\infty} P(A_n)$$

If $B_1 \supset B_2 \supset \cdots \supset B_n$

$$\boxed{B = \bigcap_{n=1}^{\infty} B_n \quad \text{then} \quad \lim_{n \to \infty} P(B_n) = P(B)}$$

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

What if there're 3 sets?



$$\boxed{\begin{aligned} P(A \cup B \cup C) = {} & P(A) + P(B) + P(C) \\ & - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ & + P(A \cap B \cap C) \end{aligned}}$$

Inclusion - Exclusion Principle

Sep 14, 2018

For n sets .. $A_1, A_2 \ldots A_n$, k-fold intersection $i_1 < i_2 < \ldots < i_k$

$$\boxed{P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) = \sum_k (-1)^{k+1} \sum_{i_1 < i_2 < i_3 \ldots < i_k} P(A_{i_1} \cap A_{i_2} \cap \ldots A_{i_k})}$$

Ex n people going to a concert. Each leaves their hat at the cloakroom. But, get a random hat back
What is the probability that no one gets the correct hat?

A_i = event that person i gets their own hat
A_i^c = person i gets the wrong hat

So $A = \bigcap_{i=1}^{n} A_i^c = \left(\bigcup_{i=1}^{n} A_i \right)^c$

k people get their own hat

$$\underbrace{\sum_{i_1 < i_2 < \ldots i_k} P(A_{i_1} \cap A_{i_2} \cap \ldots A_{i_k})}_{\text{all equal given } k} = (\#\text{ of terms})\,(\text{value of each term})$$

$$= \binom{n}{k} \frac{(n-k)!}{n!} = \frac{1}{k!}$$

$(n-k)!$ → permute other people's hat

So $P\left(\bigcup_{i=1}^{n} A_i\right)^{\circ} = \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!}$

So $P(A) = 1 - P(A^c) = 1 - \sum = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \simeq \frac{1}{e}$ (large n)

$\boxed{\text{Random variables}}$

Def A random variable X is a (nice enough) function $\Omega \mapsto \mathbb{R}$, where Ω is the probability space

Notation

$X : \Omega \longmapsto \mathbb{R}$

$X(\omega)$ = value of X at ω

Example Rolling 2 dice $\Omega = \{(x_1, x_2) \mid x_1, x_2 \in \{1 .. 6\}\}$

X_1 outcome of die # 1 , X_2 = outcome of die 2

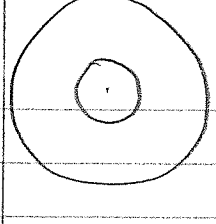
$X_1(x_1, x_2) = x_1$ (roll of 1^{st} die)

let $S = X_1 + X_2$

$\{S = 8\} = \{\omega \in \Omega \mid S(\omega) = 8\}$
$= \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$

$P(S=8) = \frac{5}{36}$

$P(1 \leq S \leq 4) \ldots \quad P(S \in B) \quad B$ is some set of real numbers

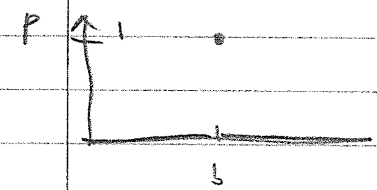


Ex

let X = distance from center...

Ex $X: \Omega \longmapsto \mathbb{R}$ $\quad X(\omega) = b \quad \forall \omega$

Say X is a degenerate random variable if there exists $a, b \in \mathbb{R}$ such that $P(X=b)=1$



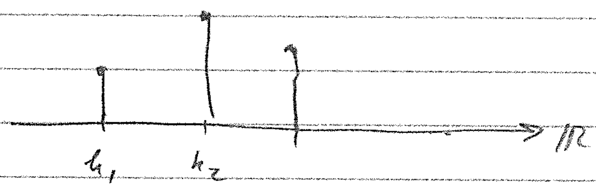
Say X is a discrete random variable if there is a set

$K = \{k_1, k_2, \ldots\} \subset \mathbb{R}$ (countable or finite) such that

$$P(X \in K) = 1 = \sum_{n=1}^{\infty} P(X = k_n)$$

(p.m.f)

(probability mass function for disc...



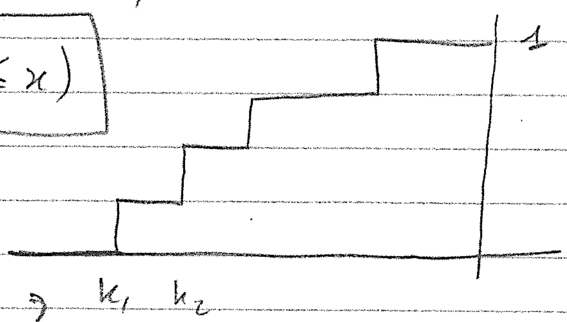
$$p_X(x) = \begin{cases} 0 & \text{if } x \notin K \\ P(X=x) & \text{if } x \in \ldots \end{cases}$$

can't do this for cont. variables...

$$\boxed{F_X(x) = P(X \le x)}$$

(c.d.f)

cumulative distribution function



pt 17, 2018 $\boxed{\text{Problem set 2}}$

1 4 balls: 1 white, 1 green, 2 red. Draw 3 balls with replacement

$\Omega = \{(a, b, c, \cancel{d}) \mid \quad \}$ = set of 3-tuples

$\# \Omega = 4^3 = 64$

A = do not see 3 colors

W = don't see white
G = don't see green
R = don't see red

$\Big\}\; A = W \cup G \cup R$

So $P(A) = P(W \cup G \cup R) = P(W) + P(G) + P(R)$ $\quad 1/64 \quad 7/64$

$- P(W \cap G) - P(G \cap R) - P(W \cap R)$

(a) $+ P(W \cap G \cap R)$

$\frac{2^3}{64} \qquad 0$

(b) A^c = see 3 colors

↳ $6 \cdot 1 \cdot 1 \cdot 2$

5 Roll 2 dice $\quad D_1$ = roll of 1^{st} die

D_2 = roll of 2^{nd} die

$X = \max(D_1, D_2)$

$Y = \min(D_1, D_2)$

$P(X \leq k)$ $\quad P(X \leq 6) = 1$

$P(X \leq 1) = P(X = 1) = \frac{1}{36} \qquad (1,1)$

$P(X \leq 5) = \frac{5}{6} \cdot \frac{5}{6} = \frac{25}{36}$

So $P(X = 6) = P(X \geq 6) - P(X \geq 5) = \frac{11}{36}$

(4) 3 players 1, 2, 3 $\quad \Omega = \{(x_1, x_2, x_3, x_4) \mid x_i \in \{1,2,3\}\}$

$\# \Omega = 3^4 = 81$ possible outcomes

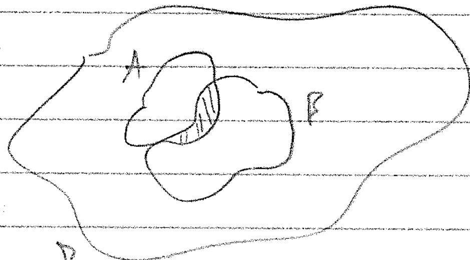
P(someone wins no games) = P(1 wins none ∪ 2 wins none ∪ 3 wins none)
= 1 − P(three different winners)

Conditioning & Independence

Conditional Probability

$P(A)$

Event B has happened..



$P(A|B)$ = probability of A given B

Know: $P(B|B) = 1$
if $\omega \notin B \quad P(\omega|B) = 0$

Def $\boxed{P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(AB)}{P(B)}}$

only makes sense if $P(B) \neq 0$

$P(*|B)$ is still a probability measure on Ω $\begin{cases} P(\Omega|B) = 1 \\ P(\emptyset|B) = 0 \\ \dots \end{cases}$

if Ω finite & uniform, then $P(A|B) = \dfrac{\# AB}{\# B}$

So

$$\boxed{\begin{aligned} P(A \cap B) &= P(A|B)P(B) \\ &= P(B|A) \cdot P(A) \end{aligned}}$$

$$\boxed{P(A_1 A_2 \ldots A_n) = P(A_1)\,P(A_2|A_1)\,P(A_3|A_1A_2) \ldots P(A_n|A_1 \ldots A_{n-1})}$$

$P(\text{pos test} \mid \text{disease}) = 0.95$

$P(\text{disease} \mid \text{pos test}) = ?$

Suppose $\Omega = B \cup B^c$ (total probability)

Let $A = (A \cap B) \cup (A \cap B^c)$

$P(A) = P(AB) + P(AB^c) = P(A|B)\,P(B) + P(A|B^c)\,P(B^c)$

Law of total probability. In general, if $\Omega = \bigcup_{i=1}^{n} B_i$ and $\emptyset = B_i \cap B_j$ if $i \neq j$, then

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

Oct 18, 2018

Recall Conditional Probability $P(A|B) = \dfrac{P(AB)}{P(B)}$

So $P(AB) = P(A|B)\,P(B)$
$= P(B|A)\,P(A)$

Today | Bayes's Formula

Law of total probability

$$P(A) = P(A|B)\,P(B) + P(A|B^c)\,P(B^c)$$

More generally if $B_1, B_2, \ldots B_n$ partition in Ω

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

Bayes' Formula

To find $P(B|A)$ from conditional info relative to B and B^c

$$\boxed{P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}}$$

More generally

$$\boxed{P(B_j|A) = \frac{P(AB_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}}$$

Note the numerator is one of the terms in denominator...

Example (Medical test)

$A = \{\text{test } (+)\}$ Test detects disease 96% of the time.
$B = \{\text{disease}\}$ Test gives false positive 2% of the time.

So $P(A|B) = 0.96$
$P(A|B^c) = 0.02$
and given $P(B) = 0.005 \rightarrow P(B^c) = 0.995$

$\boxed{\text{What is } P(B|A)?}$

$$P(B|A) = \frac{P(AB)}{P(A)}$$

$$= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

So

$$\boxed{P(B|A) = \frac{0.96 \times 0.005}{0.005 \times 0.96 + 0.02 \times 0.995} \simeq 0.194}$$

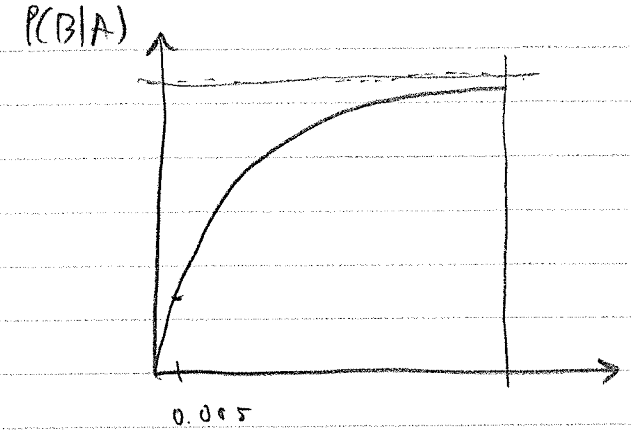
So if you get a (+) result, you have ~ 20% of really be sick

Why so low? Bcz not many ppl are not sick

→ more likely to get false (+) → ~~real~~ real (+)

$\boxed{\text{But note}}$ → despite low real positive ... this is not a problem because we're assuming $P(B) = 0.005$
BUT in reality, $P(\text{sick})$ of the person tested is high.

let $P(B) = p$ → $P(B|A) = \dfrac{96p}{96p + 2(1-p)} = \dfrac{48p}{48p + 1 - p} = \dfrac{48p}{1+47p}$



If $p = 0.5$ → $P(B|A) = 0.97$

So $P(B|A)$ very sensitive

$\boxed{P(B) = p}$ → is the prior probability
$P(B|A)$ → posterior probability

Updating knowledge
how likely sth happen
given your certainty
abt something...

Example

Someone pulls out a die, roll, tell you the answer

B_1 4-sided die
B_2 6-sided die
B_3 12-sided die

"4" → now what?

$P(B_1) = 1/3$
$P(B_2) = 1/3$
$P(B_3) = 1/3$
prior probability ...

$A = \{\text{rolled a } 4\}$

$$P(B_2|A) = \frac{P(A|B_2)P(B_2)}{P(A|B_2)P(B_2) + P(A|B_1)P(B_1) + P(A|B_3)P(B_3)}$$

$$= \frac{(1/6)(1/3)}{(1/6)(1/3) + (1/4)(1/3) + (1/12)(1/3)} = \frac{1}{3}$$

okay ... $P(B_2|A) = \frac{1}{3}$

what is? $P(B_1|A) = \dfrac{(1/4)(1/3)}{\left(\frac{1}{3}\right)\left[\frac{1}{4}+\frac{1}{6}+\frac{1}{12}\right]} = \dfrac{1}{2}$

and $P(B_3|A) = \dfrac{1}{6}$

Note

$$\boxed{\begin{array}{llll} P(B_1) = & P(B_2) = & P(B_3) = \frac{1}{3} & \text{(Prior)} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} & \text{(posterior)} \end{array}}$$

"Science is Bayesian"... develop based on Bayesian stats.
knowledge updating process...

→ Note incredibly useful ...

Friday

↪ INDEPENDENCE

Idea: $\boxed{P(A|B) = P(A)}$ (knowing B has happened has no effect on A ...)

$P(A|B) = \dfrac{P(AB)}{P(B)} = P(A)$

A ind B ⇔ B ind A

So $\boxed{P(AB) = P(A)P(B)}$ ← convenient definition, but motivation is

Questions

(1) How can I tell?

(2) What does it mean for $A_1, A_2, \dots A_n$ to be independent?

21, 2018

Independence

Definition Two events A & B are independent if

$$P(AB) = P(A)P(B)$$

Note A & B disjoint if $AB = \emptyset$ usually not independent!

Example Roll two dice 1 Blue 1 Red

A: red die shows a 4
B: sum = 7
C: sum = 8
D: blue die shows an even number

$P(A) = \frac{1}{6}$ $P(B) = \frac{1}{6}$, $P(C) = \frac{5}{36}$, $P(D) = \frac{1}{2}$

$P(AB) = \frac{1}{36}$ $P(AC) = \frac{1}{36}$ $P(AD) = \frac{3}{36} = \frac{1}{12}$

$P(BC) = 0$ $P(BD) = \frac{3}{36} = \frac{1}{12}$ $P(CD) = \frac{1}{12}$

which pairs are independent? A & B, A & D, B & D

So.. $P(ABD) \overset{?}{=} P(A)P(B)P(D)$ → **No!**

Def $A_1, A_2, A_3, \ldots A_n$ are mutually independent if for any subset $I \subset \{1, 2, 3, \ldots n\}$

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

$\boxed{\text{Sampling with replacement}}$ $A_1, A_2, A_3, \ldots$

↳ produces independent sequence of events...

$\boxed{\text{Sampling without replacement}}$ $A_1, A_2, A_3 \ldots$

↳ not independent sequence of events...

Theorem

$$\boxed{A \text{ \& } B \text{ independent} \Rightarrow \left.\begin{matrix} A, B^c \\ A^c, B \\ A^c, B^c \end{matrix}\right\} \text{ are independent pairs}}$$

Proof
$$\begin{aligned} P(AB^c) &= P(A) - P(AB) \\ &= P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) \\ &= P(A)P(B^c) \end{aligned}$$

Meta-theorem

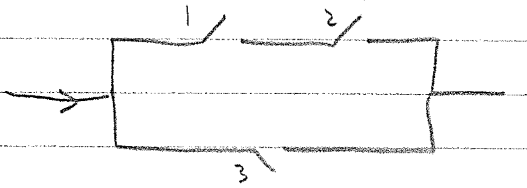
$$\boxed{\begin{matrix} A_1 \ldots A_n \\ \text{independent} \end{matrix} \Rightarrow \begin{matrix} A_1^*, A_2^*, \ldots A_n^* \text{ independent for every} \\ \text{choice of } * = c \text{ or nothing} \end{matrix}}$$

Subtlety

$$\left.\begin{matrix} A \text{ \& } B_1 \\ A \text{ \& } B_2 \\ B_1 B_2 = \emptyset \end{matrix}\right\} A \text{ independent of } B_1 \cup B_2$$

necessary → $B_1 B_2 = \emptyset$

Example



S_1 : (1) closed
S_2 : (2) closed
S_3 : (3) closed

$P(S_i) = p_i$, S_i independent...

What is $P(\text{current})$? $P(\text{current}) = P\left((S_1 S_2) \cup S_3\right)$
$= P(S_1 S_2) + P(S_3) - P(S_1 \cap S_2 \cap S_3)$

$= P(S_1)P(S_2) + P(S_3) - P(S_1)P(S_2)P(S_3)$

$= p_1p_2 + p_3 - p_1p_2p_3$ //

Independence & Random variables...

Def: Let $X_1, X_2, \dots X_n$ be random variables on the same space Ω

We say they are independent if

$$P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = P(X_1 \in B_1) \dots P(X_n \in B_n)$$

for all (Borel) subsets $B_1, B_2, \dots B_n$ of $\mathbb{R}$

Theorem: if the X_i's are discrete random variables, it's enough to check

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_1 = x_1)P(X_2 = x_2) \dots P(X_n = x_n)$$

for all x_i runs thru possible values of the random var X_i

Proof: Distributive law... (pg 56-57...)

Experiment $\begin{cases} \text{w/ probability } p \text{ success} \\ \text{failure with probability } 1-p \end{cases}$

$$X = \begin{cases} 1 \text{ if success} & P(X=1) = p \\ 0 \text{ if not success} & P(X=0) = 1-p \end{cases}$$

$\boxed{\text{Bernoulli random variable w/ success prob. } p}$

$X \sim Ber(p)$

Sept 24, 2018

HW 2.17 A, B, C mutually independent

$$P(A) = \frac{1}{2},\ P(B) = \frac{1}{3},\ P(C) = \frac{1}{4}$$

$$P(AB \cup C) = P(AB) + P(C) - P(ABC)$$
$$= \frac{1}{6} + \frac{1}{4} - \frac{1}{24} = \frac{9}{24} = \frac{3}{8}$$

Another way $P(AB \cup C) = P(ABC^c \cup C) = P(ABC^c) + P(C)$
$$= \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{3}{4} + \frac{1}{4}$$
$$= \frac{3}{8}$$

(6) Show $1 - x \le e^{-x}$ if $0 \le x \le 1$

$\hookrightarrow (1-x) + \frac{x^2}{2!} - \frac{x^3}{3!} + \dots$ (alternating series)

error controlled by $\frac{x^2}{2!}$

Another way: show $f(x) = e^{-x} - (1-x) > 0$ and increasing...

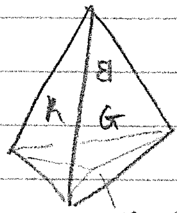
2.31 $\Omega = \{(b,b), (b,g), (g,b), (g,g)\}$

$$P(\omega) = \frac{1}{4}$$

$P(\text{other's boy} \mid \text{there at least 1 girl}) = \frac{2}{3}$

$P(\text{older is boy} \mid \text{younger} = \text{girl}) = \frac{1}{2}$

(5)



$R \cap G \cap B$

R = land on red
B = ——— blue
G = ——— green

But $P(RBG) = \frac{1}{4}$

→ R, B, G, R not jointly independent

$$P(R) = \frac{2}{4} = P(G) = P(B)$$
$$P(RG) = P(GB) = P(RB) = \frac{1}{4}$$
$\} \to$ (a)

(7) $P(A)P(B_1) = P(AB_1)$, $P(A)P(B_2) = P(AB_2)$

$B_1 B_2 = \phi$

Show $P(A \cap (B_1 \cup B_2)) = P(A)\,P(B_1 \cup B_2)$

$$P(AB_1) + P(AB_2) = P(A)\,(P(B_1) + P(B_2))$$

(since B_1, B_2 disjoint)

$\boxed{\text{Sequence of independent events}}$

Recall $X \sim Ber(p)$ (Bernoulli random variables)

$$\left\{ \begin{array}{l} P(X=1) = p \\ P(X=0) = 1-p \end{array} \right\}$$

Suppose $X_1, X_2, \ldots, X_n$ are independent random variables $Ber(p)$

Want to look at $S_n = X_1 + X_2 + \ldots + X_n$, which is a random variable, with possible values are $0, 1, \ldots, n$

- $P(S_n = 0) = P(X_1 = 0, X_2 = 0, \ldots, X_n = 0) = (1-p)^n$
- $P(S_n = n) = p^n$
- $\boxed{P(S_n = k) = p^k (1-p)^{n-k} \cdot \binom{n}{k}}$ binomial distribution...

Sanity Check $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1 = (1-p+p)^n = 1$

Short proof $(x+y)^n = (x+y)(x+y) \ldots (x+y)$

$$= \binom{n}{0} x^n + \binom{n}{1} x^{n-1} y + \binom{n}{2} x^{n-2} y^2 + \ldots + \binom{n}{0} y^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$

A random variable with probability mass function

$$\boxed{P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{is} \quad X \sim Bin(n, p)}$$

Sept 26, 2018

Recall $Ber(p)$, $Bin(n,p)$, $\underbrace{Geom(p),\ Hypergeom(?)}_{\text{Today}}$

① Geometric Sequence of independent Bernoulli trial ... $X = k$ s.t first success happens on the k^{th} trial ...

$$P(X=k) = (1-p)^{k-1} p^1 \qquad k = 1, 2, 3, \dots \infty$$

Proof $P(X=\infty) = 0$

$$\sum_{k=1}^{\infty} p(1-p)^{k-1} = p\left[1 + (1-p) + (1-p)^2 + \dots\right] = p \cdot \frac{1}{1-(1-p)} = 1$$

So $P(X=\infty) = 0$ (finite k uses up all probability)

② Hypergeometric Urn with N balls $N = N_A + N_B$ balls

N_A : azure ball

N_B : brown ball

Sample n things w/o replacement

X = # of A balls in sample

$$\boxed{P(X=k) = \frac{\binom{N_A}{k}\binom{N-N_A}{n-k}}{\binom{N}{n}}} \rightarrow Hypergeom(N, N_A, n)$$

$X \sim Hypergeom(N, N_A, n)$

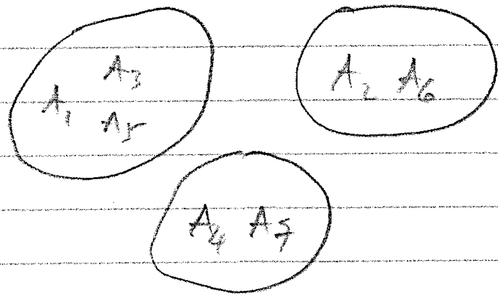
Note $a < b \Rightarrow \binom{a}{b} = 0$ by convention ...

Two final points on conditioning & independence

① Given $A_1, A_2, \dots A_n$ independent

Make
$B_1, B_2, B_3 \dots B_k$ where B_1 = made out of A_i, $i \in I_1$
B_2 = made out of A_i, $i \in I_2$
and $I_1, I_2, I_3 \dots I_k$ are partitions of $\{1, 2, 3, \dots n\}$

Then $B_1, B_2, \dots B_k$ are independent



② $A_1, A_2 \dots, A_n$ indep. $\quad P(*)$

Consider $P(*|B)$. Now, are $A_1 \dots A_n$ independent given B?

$$P\left(\cap A | B\right) = \prod_{i \in I} P(A_i | B)$$

unrelated! We don't know...

(look at Examples 2.38 – 2.40)

Example

Coins → fair $\quad P(T) = \frac{1}{2} \quad 90\%$

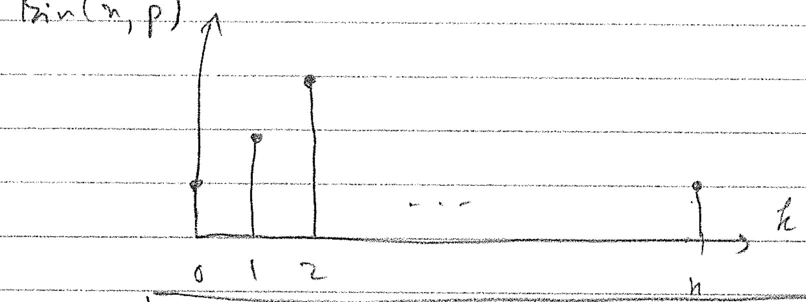
Coins → biased $\quad P(T) = \frac{3}{5} \quad 10\%$

Flip twice. A_1 1^{st} tails
A_2 2^{nd} tails

$P(A_1 | \text{fair}) P(A_2 | \text{fair}) = P(A_1 A_2 | F)$
$P(A_1 | \text{bias}) P(A_2 | \text{bias}) = P(A_1 A_2 | B)$

But $P(A_1 A_2) \neq P(A_1) P(A_2)$

$\boxed{X \text{ discrete RV}}$. Probability mass function $P_X(k) = P(X = k)$



Continuous $\boxed{X = \text{pt picked uniformly at random in } [a, b]}$

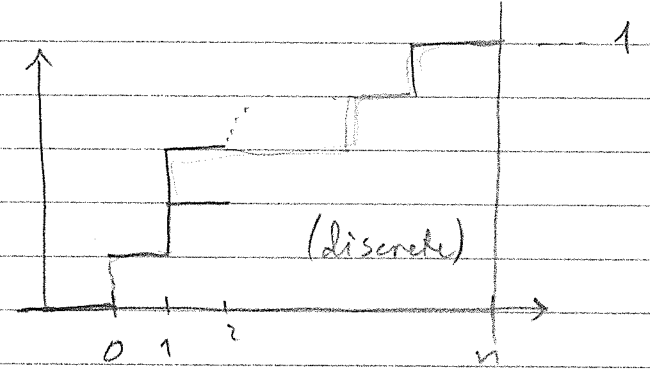
$P(X \in A) = \dfrac{\text{lngth}(A)}{\text{lngth}([a, b])}$, and $P(X = k) = 0$ (lngth of point is 0)

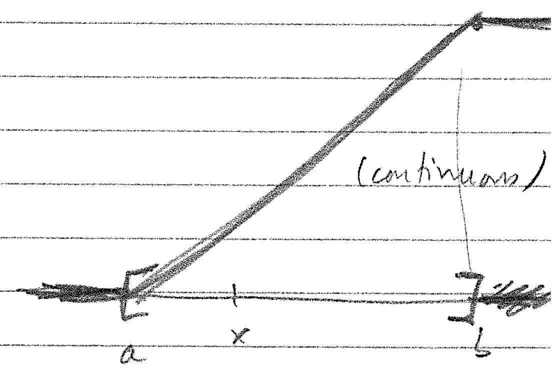
can't do probability mass function...

Rather, use

→ Cumulative Density Function (works for both)

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





$\boxed{P(l < X \le u) = F_X(u) - F_X(l)}$

$\boxed{\text{Sept 28 2018}}$

$\boxed{\text{Random Variables}}$

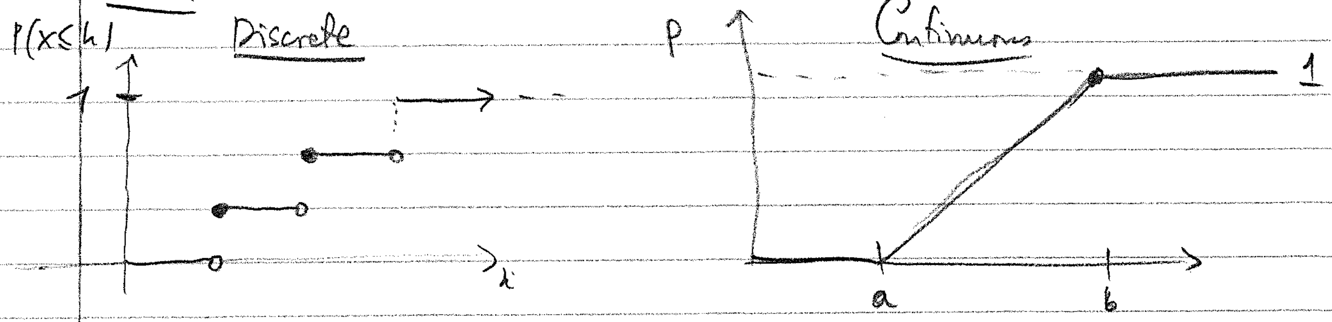
$X : \Omega \longmapsto \mathbb{R}$ $\quad P(X \in B)$, $B \subset \mathbb{R}$ (nice subset)

→ probability distribution

$F_X(x) = P(X \le x)$ → cdf for X

Theorem $\boxed{F_X \text{ determines the probability distribution of } X}$

Recall



$$P(a \le X \le b) = F_X(b) - \lim_{x \to a^-} F_X(a) \qquad P(a < X \le b) = F_X(b) - F(a)$$

$$= F_X(b) - F_X(a^-)$$

$\boxed{\text{Properties of } F_X}$

(1) F_X is positive & increasing

(2) $\lim_{x \to \infty} F(x) = 1$

(3) $\lim_{x \to -\infty} F(x) = 0$

(4) $\lim_{x \to a^+} F(x) = F(a)$ (right-continuous)

$\boxed{F_X = cdf}$

$\boxed{F_X = P(X \le x)}$

Theorem

$\boxed{\text{Any such function is the CDF of some random variable}}$

Types of random variables

Def: X is a discrete RV if there exists $\{k_1, k_2, \dots\} \subset \mathbb{R}$ such that

Discrete →

$$\sum_{i=1}^{\infty} P(X = k_i) = 1$$

k's are possible values and CDF is a step function

Continuous variable

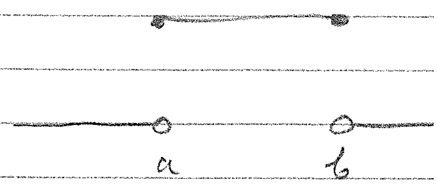
Def Continuous random variable is one such that there exists $f(x)$ with

$$F(x) = \int_{-\infty}^{x} f(\bar{x})\,d\bar{x}$$

Ex

$$X \sim \text{Unif}(a,b) \qquad F(x) = \begin{cases} 0 & x<a \\ (x-a)/(b-a) & a \le x \le b \\ 1 & \text{if } x > b \end{cases}$$

$$f(x) = \begin{cases} 0 & \text{if } x < a \\ 0 & \text{if } x > b \\ \dfrac{1}{b-a} & \text{if } a < x < b \end{cases}$$



probability density function (pdf)

Note

$$P(a \le X \le a+h) = \frac{F(a+h) - F(a)}{h}, \quad h \simeq \underbrace{f(a)}_{\text{pdf}}h$$

Note Probability of a single point $= 0$

$$P(a \le X \le b) = \int_a^b f(t)\,dt \Rightarrow P(X=a) = \int_a^a f(t)\,dt = 0$$

Note we can mix discrete & continuous. These monsters exist, but they don't matter...

Discrete	Continuous
$P(a \le X \le b) = \sum_{a \le k \le b} P(X=k)$	$P(a \le X \le b) = \int_a^b f(t)\,dt$

Note these are quite the same ... $= \int_a^b dF$

So $\int_a^b dF = \int_a^b f(t)\,dt = \int_a^b F'(t)\,dt = \sum_{a \le k \le b} P(X = k) \ldots$

- 7, 2018

Recall $X \sim \text{Unif}[a, b]$

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$$

$\boxed{\text{Properties of } f(x)}$

(1) $f(x) \ge 0$

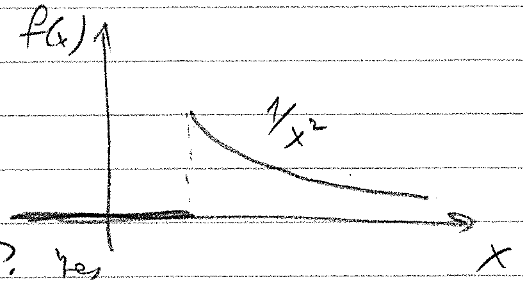
(2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Lab4

$$\boxed{\int_{-\infty}^{\infty} f(x)\,dx = \lim_{a \to \infty} \int_a^{c} f(x)\,dx + \lim_{b \to +\infty} \int_c^b f(x)\,dx \ne \lim_{n \to \infty} \int_{-n}^{n} f(x)\,dx}$$

$\boxed{Ex}$

$$f(x) = \begin{cases} 1/x^2 & \text{if } x \ge 1 \\ 0 & \text{if } x < 1 \end{cases}$$



Is $f(x)$ a pdf? $f(x)$ is positive? Yes

$$\int_{-\infty}^{\infty} f(x)\,dx = \underbrace{\int_{-\infty}^{1} f(x)\,dx}_{0} + \int_1^{\infty} f(x)\,dx$$

$$0 + \int_1^{\infty} \frac{1}{x^2}\,dx = \lim_{b \to \infty} \int_1^b \frac{1}{x^2}\,dx = \lim_{b \to \infty} \left(\frac{-1}{x}\right)_1^b$$

$$\lim_{b \to \infty} \left(\frac{-1}{b} + 1\right) = 1$$

So $\boxed{f(x) \text{ is a pdf.}}$

$\boxed{Ex}$

$$f(x) = \begin{cases} b\sqrt{a^2 - x^2} & -a \le x \le a \\ 0 \end{cases}$$

(Wigner distribution)

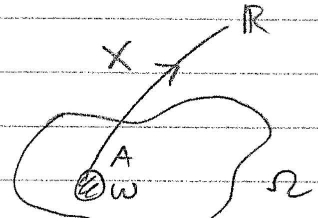
$$\int_{-a}^{a} b\sqrt{a^2 - x^2}\,dx = b\,\frac{\pi a^2}{2} = 1 \quad \text{So} \quad b = \frac{2}{\pi a^2}$$

Oct 2, 2018

$\boxed{\text{Midterm! Oct 10, Olin 1, 7:00 - 8:30pm}}$

Expectation of a Random Variable

<u>Prelude</u>: Integration



$\int_\Omega X\,dP$ what does this mean?

<u>Recall</u> $\int_a^b f(x)\,dx$ — $f(x)$: fn, dx: measures length.

measure: $P(A)$

<u>Now</u>

$$\int_\Omega X\,dP = \begin{cases} \sum_k k\,P(X=k) & \text{(discrete)} \\ \int_{-\infty}^{\infty} x\,f_X(x)\,dx & \text{(cont)} \end{cases}$$

$P(X=k)$ → probability mass function

$f_X(x)$ → probability density function

<u>Note</u> $\boxed{\text{discrete: pmf} \quad \dots \quad \text{continuous: pdf}}$

<u>Note</u> Require in every case

$$\sum_k |k|\,P(X=k) \text{ converges} \quad \underline{\text{OR}} \quad \int_{-\infty}^{\infty} |x f_X(x)|\,dx \text{ converges}$$

If X is an RV

let $$E(X) = \begin{cases} \sum_k k\,P(X=k) & \text{if } X \text{ discrete} \\ \int_{-\infty}^{\infty} x\,f_X(x)\,dx & \text{if } X \text{ continuous} \end{cases}$$

"Expected value of X" or "Expectation value of X", or "Mean"

or "first moment of X" $= \mu_X = \mu$

(1) roll one die, Z = number on top.

$$E(Z) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + \dots \frac{1}{6} \cdot 6 = (1+2+3+\dots+6)\frac{1}{6} = 3.5$$

(2) $X \sim Ber(p)$ $\quad X = \begin{cases} 1 & \text{w/ prop } p \\ 0 & \text{w/ prop } (1-p) \end{cases}$

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

(3)

$A \subset \Omega$ event. Define $I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$

$$E(I_A) = 1 \cdot P(\omega \in A) + 0 \cdot P(\omega \notin A) = P(\omega \in A) = P(A)$$

Equiv. $\int_\Omega I_A \, dP = P(A)$

(4) $X \sim Unif[a, b]$

Note $f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & x \notin [a,b] \end{cases}$

So

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

$$= \underbrace{\int_{-\infty}^{a} x f_X(x) dx}_{0} + \int_a^b x f_X(x) dx + \underbrace{\int_b^{+\infty} x f_X(x) dx}_{0}$$

$$= 0 + \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} + 0$$

$$= \left(\frac{b+a}{2}\right)$$

(makes sense)

(5) $\boxed{X \sim Bin(n, p)}$ Guess : np

$$E(X) = \sum_{k=0}^{n} k P(X=k)$$

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

$$= \sum_{k=0}^{n} \frac{k n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \quad \text{let } j = 0 \to n-1$$

$$= np \underbrace{\sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j}}_{1}$$

$$= \boxed{np}$$

(6) $\boxed{X \sim Geom(p)}$ $P(X=k) = p(1-p)^{k-1}$

$$E(X) = \sum_{k=1}^{\infty} k p (1-p)^{k-1} = p \sum_{k=1}^{\infty} k (1-p)^{k-1}$$

Scratch

$$\sum_{n=0}^{\infty} (n+1) x^n = \frac{d}{dx} \sum_{n=0}^{\infty} x^{n+1} = \frac{d}{dx}(1 + x^2 + \dots) = \frac{1}{(1-x)^2}$$

So

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

? 5, 2018

Recall X ~ random var on Ω

$$E(X) = \int_{\Omega} X\,dP = \begin{cases} \sum k P(X=k) \\ \int_{-\infty}^{\infty} x f_X(x)\,dx \end{cases}$$

$\boxed{\text{① Bad Guys}}$

↳ Flip a coin til get tails. Tails at the n^{th} flip → get \$ 2^n

X = winnings

$$E(X) = \sum_{n=1}^{\infty} 2^n \frac{1}{2^n} = \sum_{n=1}^{\infty} 1 \to +\infty$$

$\boxed{\text{② Modify to get}}$

$1+1-1+1-\dots$ has no limit at all

↳ $E(X)$ is undefined

③ $f(x) = \begin{cases} 1/x^2 & \text{if } x > 1 \\ 0 & \text{if } x \text{ elsewhere} \end{cases}$

$$E(X) = \int_1^{\infty} \frac{x}{x^2}\,dx = \lim_{b\to\infty} \int_1^b \frac{1}{x}\,dx = \lim_{b\to\infty} \ln(b) \to +\infty$$

II. Random variable X $g: \mathbb{R} \mapsto \mathbb{R}$

let $Y = g(X)$ ⟶ Y is a new random variable.

Note $\boxed{E(g(X)) \neq g(E(X))}$

Observe that $\{Y=k\} = \bigcup_{x \in g^{-1}(k)} \{X=x\}$ ← disjoint

$$E(Y) = \sum_k k P(Y=k)$$

$$= \sum_k k \sum_{x\in g^{-1}(k)} P(X=x) = \sum_k \sum_{x \in g^{-1}(k)} g(x) P(X=x) = \sum_x g(x) P(X=x$$

So $\boxed{E(Y) = E(g(X)) = \sum_x g(x) P(X=x)}$

So

$$\boxed{E(g(X)) = \begin{cases} \sum_k g(k)\, P(X=k) \\ \int_{-\infty}^{\infty} g(x) f_X(x)\, dx \end{cases}}$$

→ "Law of the Unconscious Statistician"

Example

$$E[aX+b] = \int_{-\infty}^{\infty} (ax+b) f_X(x)\, dx$$

$$= aE(X) + b$$

So $\boxed{E[g(X)] = g(E(X)) \text{ if } g(X) \text{ is linear}}$

not true in general

Now $E[X+Y]$, X, Y are RV on Ω

Well, by definition, $E[X+Y] = \int_\Omega (X+Y)\, dP = \int_\Omega X\, dP + \int_\Omega Y\, dP$

So

$$\boxed{E[X+Y] = E[X] + E[Y]}$$

Recall $\text{Bin}(n,p)$: $\mu = np$

Moments The n^{th} moment of random variable X is

$E[X^n]$. Existence if $E[X^n]$ exists for some n, then $E[X^l]$ exists if $l \le n$.

Ex $X \sim \text{Unif}[0, C]$

$$E[X^n] = \int_0^c \frac{x^n}{c}\,dx = \frac{x^{n+1}}{(n+1)c}\Big|_0^c = \frac{c^n}{n+1} \qquad \text{(trivial)}$$

Ex $\{0, 1, 2, \ldots, c\} = X \qquad P(\cancel{X} = i) = \frac{1}{c+1}$

$$\left.\begin{aligned} E[X] &= \sum_{i=0}^{c} \frac{i}{c+1} = \frac{1}{c+1}\,\frac{c(c+1)}{2} = \frac{c}{2} \\ E[X^2] &= \sum_{i=0}^{c} \frac{i^2}{c+1} = \frac{\frac{c(c+1)(2c+1)}{6}}{c+1} = \frac{c(2c+1)}{6} \end{aligned}\right\} \text{(hard)}$$

Ex

$$\begin{array}{cccccc} \frac{1}{10} & & \frac{1}{10} & \frac{1}{10} & \cdots & \\ -100 & & 0 & 1 \;\; 2 & \cdots & 8 \end{array}$$

Get $\mu = \cancel{cccc} \; \frac{-100 + 0 + 1 \ldots + 8}{10} = ?$ skewed

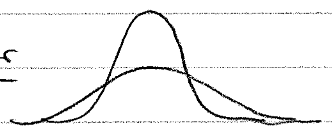
$\boxed{\text{Def}}$ A real number m is called <u>a median</u> of X if

$$P(X \le m) \ge \frac{1}{2} \quad \text{and} \quad P(X \ge m) \ge \frac{1}{2}$$

$\boxed{\text{Def}}$ x is a p^{th} quantile of X if $P(X \le x) \ge p$ and $P(X \ge x) \ge 1-p$ $\quad 0 \le p \le 1$

t 8, 2018 $\boxed{\text{Variance}}$ X random variable $E(X) = \mu$

But Consider



Same μ, but different "variation"

$\boxed{\text{Detect variation}}$

$$\boxed{E\left[(X-\mu)^2\right] = \text{Variance}}$$

Ex

X	1	2	3	...	6
$P_1(X)$	1/6	1/6	1/6	...	1/6
$P_2(X)$	1/4	1/6	1/12	...	1/4

So $Var_1(X) = \frac{1}{6}(1-3,5)^2 + \frac{1}{6}(2-3,5)^2 + \dots \frac{1}{6}(6-3,5)^2 = 2.$

$Var_2(X) = \frac{1}{4}(1-3,5)^2 + \frac{1}{6}(2-3,5)^2 + \dots \frac{1}{4}(2-3,5)^2 = 3.$

Standard deviation of X $\sqrt{Var(X)} = \sigma_X = \sigma$ So $Var(X) = \sigma_X^2$

$$\boxed{Var(X) = \begin{cases} \sum_k (k-\mu)^2 P(X=k) \\ \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx \end{cases}}$$

Ex $X \sim Ber(p)$ $\mu = p = E[X]$

So $Var(X) = (1-p)^2 p + (0-p)^2(1-p) = \boxed{p(1-p)}$

$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - \int_{-\infty}^{\infty} 2\mu x f_X(x)\,dx + \int_{-\infty}^{\infty} \mu^2 f_X(x)\,dx$$

So $Var(X) = E[X^2] - 2E[X]E[X] + E^2[X]$

$$\boxed{Var(X) = \left(\underbrace{E[X^2]}_{2^{nd}\text{ moment}} - \underbrace{E^2[X]}_{1^{st}\text{ moment}}\right)}$$ again $\boxed{\sigma_X = \sqrt{E(X^2) - E(X)^2}}$

$\boxed{\text{Can } Var(X) = 0 \text{ ?}}$ $Var(X) = \sum_k (k-\mu)^2 P(X=k) = 0 \Rightarrow (k-\mu)P(X=k) = 0$

So $\boxed{k = \mu \text{ OR } P(X=k) = 0 \ \forall k}$

$\Rightarrow$ $\boxed{P(X=\mu)=1}$ (degenerate random variable)

$\rightarrow$ X is degenerate (<u>almost</u> constant)

//

$\boxed{X \sim \text{Binomial}(n,p)}$

<u>Recall</u> $E[X] = np$

<u>So</u> $\boxed{Var(X) = E[X^2] - E[X]^2}$

where $E[X^2] = \sum_{k=0}^{n} k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \left(n(n-1)p^2 + np\right)$

<u>So</u> (verify this)

$Var(X) = \boxed{np(1-p)}$

$\boxed{\text{Secret Theorem \# 2}}$

> if X, Y are <u>independent</u> variables on the same Ω,
>
> $$Var(X+Y) = Var(X) + Var(Y)$$

$\boxed{X \sim \text{Geom}(p)}$

$\mu = \frac{1}{p}$. $E[X^2] = \frac{2-p}{p^2}$ <u>So</u> $\boxed{Var(X) = \frac{1-p}{p^2}}$

$\boxed{X \sim \text{Unif}[a,b]}$

$\mu = \frac{a+b}{2} = E[X]$

$$E[X^2] = \int_a^b x^2 \cdot \frac{1}{b-a}\,dx = \frac{1}{b-a} \cdot \frac{1}{3}(b^3-a^3) =$$

$$= \frac{1}{3}\left(a^2 + ab + b^2\right)$$

<u>So</u> $\boxed{Var(X) = \frac{(b-a)^2}{12}}$

X random var $a, b \in \mathbb{R}$

$$\boxed{E[aX+b] = aE[X] + b}$$

What about Variance?

$$\boxed{Var(aX+b) = a^2 Var(X)}$$

Proof $Var(aX+b) = E[(aX+b)^2] - E[aX+b]^2$

$$= \int_{-\infty}^{\infty} (\underbrace{aX+b}_{var} - \underbrace{a\mu - b}_{\bar{\mu}})^2 f(x) dx$$

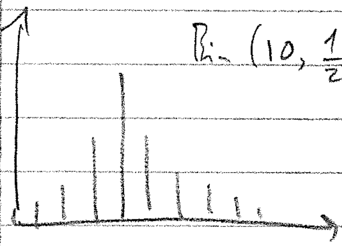
$$= a^2 \int_{-\infty}^{\infty} (X-\mu)^2 f(x) dx = a^2 Var(X)$$

Now

$X \rightsquigarrow E[X]$ } important descriptor of variable.

$\quad \rightarrow Var[X]$

Recall



$Bin(10, \frac{1}{2})$ $\quad \sigma^2 = Var(X) = 10 \cdot \frac{1}{2} \cdot \frac{1}{2} = 2.5$

$\hookrightarrow \sigma = \sqrt{2.5} \rightarrow$ 99% of values in ± 3

We'll calculate

$$\sum_{np-3\sigma \le k \le np+3\sigma} \binom{n}{k} p^k (1-p)^{n-k}$$

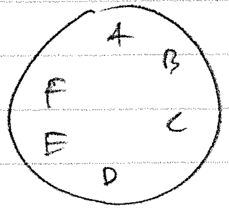
where $\sigma = \sqrt{np(1-p)}$

p.118 Table summarizing random vars, $E(X)$, $Var(X)$ ...

$E[X+Y] = E[X] + E[Y]$

$Var[X+Y] = Var[X] + Var[Y]$ if X, Y independent

$\boxed{1, 52}$ 3 married couple @ round table. P(someone next to spouse)



Total arrangements : $6! = 720$

Couples $1a, 1b, 2a, 2b, 3a, 3b$

$A_1 \ (1a, 1b)$

$A_2 \ (2a, 2b)$

$A_3 \ (3a, 3b)$

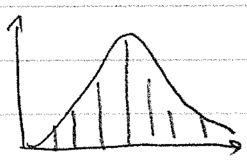
$\# A_1 = 6 \cdot 2 \cdot 4! = 288 = \# A_2 = \# A_3$

$\# A_1 A_2 A_3 = 4, 4 \cdot 2^3 \cdot 6$

12, 2018

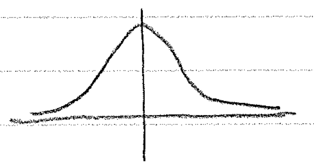
The Gaussian (Normal) Distribution

History $X \sim Bin(n, p)$



$P(a \leq X \leq b)$ is hard to find

the function e^{-x^2}



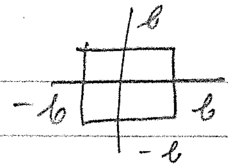
$$\int_{-\infty}^{\infty} e^{-x^2} dx = 2 \int_0^{\infty} e^{-x^2} dx$$ need to show $\int_0^{\infty} e^{-x^2} dx$ converges

Claim $\int_1^{\infty} e^{-x^2} dx$ converges well $e^{-x^2} \leq e^{-x}$ for $x \geq 1$

So $\int_1^b e^{-x^2} dx \leq \int_1^b e^{-x} dx = [-e^{-x}]\Big|_1^b = e^{-1} - e^{-b}$ converges e^{-1} as $b \to \infty$

So $\int_{-\infty}^{\infty} e^{-x^2} dx = \lim_{b \to \infty} \int_{-b}^{b} e^{-x^2} dx$ Hard Instead ...

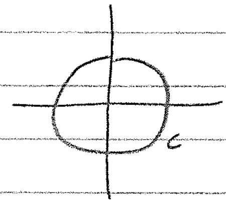
Integrate this $\displaystyle\iint_{\mathbb{R}^2} e^{-(x^2+y^2)}\,dx\,dy$



$$= \int_{-b}^{b}\int_{-b}^{b} e^{-(x^2+y^2)}\,dx\,dy = \int_{-b}^{b}\int_{-b}^{b} e^{-x^2}e^{-y^2}\,dx\,dy$$

$$= \int_{-b}^{b}\left(e^{-y^2}\int_{-b}^{b} e^{-x^2}dx\right)dy = \int_{-b}^{b} e^{-y^2}dy \int_{-b}^{b} e^{-x^2}dx$$

So $\displaystyle\iint_{\mathbb{R}^2} e^{-(x^2+y^2)}\,dx\,dy = \left(\int_{-\infty}^{\infty} e^{-x^2}dx\right)^2$



but $\displaystyle\iint_{\mathbb{R}^2} e^{-(x^2+y^2)}\,dx\,dy = \iint_{\mathbb{R}^2} e^{-r^2}\,r\,dr\,d\theta$

$$= \int_0^{2\pi}\int_0^{R} e^{-r^2}\,r\,dr\,d\theta = 2\pi\int_0^{R} re^{-r^2}dr = \pi\int_0^{c^2} e^{-u}\,du$$

$$= \pi\left(1-e^{-c^2}\right) \quad \text{As } c\to\infty \Rightarrow \iint_{\mathbb{R}^2} e^{-(x^2+y^2)}\,dx\,dy = \pi$$

So $$\boxed{\int_{-\infty}^{\infty} e^{-x^2}dx = \sqrt{\pi}}$$

Next $\displaystyle\int_{-\infty}^{\infty} e^{-x^2/2}\,dx = ?$ let $u = \dfrac{x}{\sqrt{2}} \Rightarrow du = \dfrac{1}{\sqrt{2}}dx$

$$\sqrt{2}\int_{-\infty}^{\infty} e^{-u^2}du = \sqrt{2\pi}$$

$$\Rightarrow \boxed{\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}}$$

Def A random variable Z has standard Gaussian dist. if its pdf is

$$\boxed{\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}}$$

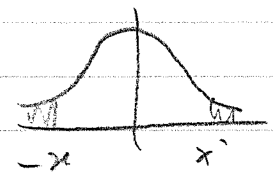
Def

$$\boxed{\Phi(x) = \int_{-\infty}^{x} \varphi(t)\,dt \Rightarrow \Phi(b) - \Phi(a) = P(a \leq X \leq b)}$$

can't be written explicitly

Note

$$\boxed{\Phi(-x) = \int_{x}^{\infty} \varphi(t)\,dt = \int_{-\infty}^{-x} \varphi(t)\,dt = 1 - \Phi(x)}$$



Note $\boxed{Z \sim \mathcal{N}(0,1)}$

expectation

variance.

$$E(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2}\,dx = 0 \quad (x \text{ odd})$$

$$E(Z^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{x}_{u} \cdot \underbrace{\left(x e^{-x^2/2}\right)}_{v'} dx = 1$$

So $\boxed{Var(X) = E(Z^2) - E(Z) = 1}$

<u>Need</u> $\mathcal{N}(\mu, \sigma^2)$

Take $X = \sigma Z + \mu$ then $E(X) = \mu$
$Var(X) = \sigma^2$

$$P(X \le x) = P(\sigma Z + \mu \le x) = P\left(Z \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

So $\boxed{P(X \le x) = \Phi\left(\frac{x-\mu}{\sigma}\right)}$

So $f_X(x) = \Phi'\left(\frac{x-\mu}{\sigma}\right) \cdot \left(\frac{1}{\sigma}\right) = \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}$

So $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2 \cdot \frac{1}{2}}$

Normal dist.

So $\boxed{X \sim \mathcal{N}(\mu, \sigma^2) \text{ if its pdf is } \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$

<u>Ex</u> To find z s.t. $P(-z \le Z \le z)$ is $\simeq \frac{2}{3}$

$\Phi(z) - \Phi(-z) \simeq \frac{2}{3}$

$\therefore \Phi(z) - 1 + \Phi(z) \simeq \frac{2}{3}$

→ about 1σ away → get $\frac{2}{3}$ the probability

$\therefore 2\Phi(z) - 1 = \frac{2}{3}$

$\therefore \Phi(z) \simeq \frac{5}{6} = 0.8333 \rightarrow z = 0.97 \simeq \sigma$

cdf

<u>Oct 17, 2018</u> <u>Recall</u> Standard Normal <u>and</u> $\Phi(-x) = 1 - \Phi(x)$ ~~and~~ general norm

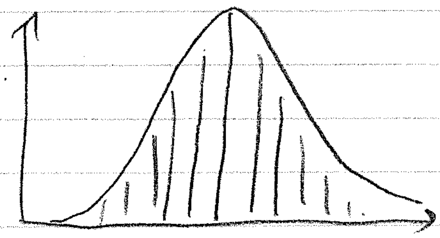
$X \sim N(\mu, \sigma^2)$ $X = \sigma Z + \mu$ $\rightarrow E[X] = \mu$, $Var(X) = \sigma^2$

<u>cdf</u> $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ <u>pdf</u> $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$E[aX+b] = aE[X]+b = a\mu+b$
$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X) = a^2\sigma^2$
$\Big\}\ aX+b \sim \mathcal{N}(a\mu+b, \sigma^2 a^2)$

$\boxed{\text{Binomial} \approx \text{Normal}}$

$\mathrm{Bin}(1000, 0.6)$
$\mu = 600 = np$
$\sigma^2 = 240 = np(1-p)$



$\mathcal{N}(600, 240)$

$\mathrm{Bin}(1000, 0.6) \approx \mathcal{N}(600, 240)$ agrees almost exactly!

Suppose $S_n \sim \mathrm{Bin}(n,p) \to \mu_{S_n} = np,\ \sigma^2 = np(1-p)$

So $\boxed{\dfrac{S_n - np}{\sqrt{np(1-p)}}}$ has $\mu = 0$, $\sigma^2 = 1$ $\to$ $P\left(a \le \dfrac{S_n-np}{\sqrt{np(1-p)}} \le b\right)$ $(-\infty \le a \le b \le \infty)$

Take limit

Central Limit Theorem for Binomial Dist

$$\boxed{\lim_{n\to\infty} P\left(a \le \frac{S_n-np}{\sqrt{np(1-p)}} \le b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx}$$

So
$$\boxed{\lim_{n\to\infty}\left(P\left(\frac{S_n-np}{\sqrt{np(1-p)}} \le x\right)\right) = \Phi(x)}$$
(cdf)
(convergence in distribution)

Remarks $\boxed{\text{Rule of Thumb}}$

$$P\left(a \le \frac{S_n-np}{\sqrt{np(1-p)}} \le b\right) \approx \Phi(b) - \Phi(a) \quad \text{when } np(1-p) \geq 10$$

Theorem

$$\left| P\left(\frac{S_n-np}{\sqrt{np(1-p)}} \le x\right) - \Phi(x)\right| \le \frac{3}{\sqrt{np(1-p)}}$$

Note $S_n \sim Bin(n,p)$. **Idea** $\frac{S_n}{n} \approx p$ but this makes no se

$\hookrightarrow$ look at $\frac{S_n - np}{n}$

Want $\boxed{\lim_{n\to\infty} P\left(\left(\frac{S_n}{n} - p\right) > \varepsilon\right) = 0}$ we'll prove this (weak law of large numbers...)

Oct 19, 2018

$$\boxed{\begin{array}{c}\text{Theorem: Given } 0 < p < 1,\ -\infty \leq a \leq b \leq \infty \\ S_n \sim Bin(n,p) \\ \lim_{n\to\infty} P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx\end{array}}$$

Proof

Let $1-p=q$ $\quad a \leq \frac{S_n - np}{\sqrt{npq}} \leq b \Rightarrow np + a\sqrt{npq} \leq S_n \leq np + b$

Probability:
$$\sum_{np + a\sqrt{npq} \leq k \leq np + b\sqrt{pqn}} \frac{n!}{(n-k)!\,k!} p^k q^{n-k}$$

Stirling's Formula $\quad n! \approx n^n e^{-n}\sqrt{2\pi n}$ where $f(x) \sim g(x)$ means

$$\lim_{n\to\infty} \frac{f(x)}{g(x)} = 1$$

So $\frac{n!}{(n-k)!\,k!} p^k q^{n-k}$

$$\simeq \frac{n^n e^{-n}\sqrt{2\pi n}}{(n-k)^{n-k} e^{-(n-k)}\sqrt{2\pi(n-k)} \cdot k^k e^{-k}\sqrt{2\pi k}} \simeq \frac{1}{\sqrt{2\pi pqn}} e^{-\frac{(k-np)^2}{2npq}}$$

So
$$\sum_{np + a\sqrt{npq} \leq k \leq np + b\sqrt{npq}} \frac{1}{\sqrt{2\pi}\sqrt{npq}} e^{\frac{-(k-np)^2}{2(\sqrt{npq})^2}} = \sum f(x_k)\,\Delta x_k = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$$

$$x_k = \frac{k-np}{\sqrt{npq}}$$

$$\int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$$ 42

So $$\boxed{\lim_{n\to\infty} P\left(a \le \frac{S_n - np}{\sqrt{npq}} \le b\right) = \Phi(b) - \Phi(a)}$$

So $$\boxed{P(k_1 \le S_n \le k_2) \simeq \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right)}$$

<u>But note</u>

$$\boxed{P(k_1 \le S_n \le k_2) = P\left(k_1 - \tfrac{1}{2} \le S_n \le k_2 + \tfrac{1}{2}\right) \simeq \Phi\left(\frac{k_2 + \frac{1}{2} - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - \frac{1}{2} - np}{\sqrt{npq}}\right)}$$

Continuity correction — better approximation (important when k_1 & k_2 close together)

$$S_n = X_1 + X_2 + \dots + X_n \qquad X_i \sim Ber(p)$$

$\mu_{S_n} = np$ whereas $\mu_{X_i} = p$

$\sigma_{S_n} = \sqrt{npq}$ $\qquad \sigma_{X_i} = \sqrt{pq}$

<u>Notation</u> $$\boxed{\frac{S_n - np}{\sqrt{npq}} = \frac{S_n - n\mu}{\sqrt{n}\,\sigma}}$$ → of the Bernoulli var

$\boxed{\text{Ex}}$

Fair coin, flip 10,000 times. What is $P(4850 \le \#\text{heads} \le 5100) = ?$

$$\sum_{4850 \le k \le 5100} \binom{10000}{k} p^k q^k = \sum_{4850 \le k \le 5100} \binom{10000}{k} (1/2)^{10000} = 0.9764817\ldots$$

$$= \Phi\left(\frac{5100 - 5000}{\sqrt{2500}}\right) - \Phi\left(\frac{4850 - 5000}{\sqrt{2500}}\right) = \Phi(2) - \Phi(-3) \simeq \boxed{0.9759}$$

$$= \Phi(2) - 1 + \Phi(3)$$

$$\simeq \Phi(2.01) - \Phi(-3.01) \simeq \boxed{0.976478}$$ → Better approx.

→ continuity correction works...

The 3-σ rule

$$P(np - 3\sigma \le S_n \le np + 3\sigma) \qquad \sigma = \sqrt{npq}$$

$$\approx \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 \approx 0.9974$$

<u>Rule of thumb</u> $npq > 10$

<u>or</u>

$np > 10$ & $nq > 10$

<u>OR</u>

$n > 9 \cdot \max\left(\frac{q}{p}, \frac{p}{q}\right)$

Oct 22, 2018

Chebysheff's inequality Let X be a discrete random variable with an expectation μ ($E[X] = \mu$) & $\varepsilon > 0$ be any real positive number. Then,

$$P(|X - \mu| \ge \varepsilon) \le \frac{Var(X)}{\varepsilon^2}$$

<u>Proof</u> Let $p_X(x)$ be p.m.f of X. Then the probability that X differs from its mean (μ) by at least ε is given

$$P(|X-\mu| \ge \varepsilon) = \sum_{|x-\mu| \ge \varepsilon} p_X(x)$$

We also know

$$Var(X) = E[(X-\mu)^2] = \sum_{x} (x-\mu)^2 p_X(x)$$

<u>Now</u>

$$Var(X) = \sum_{x} (x-\mu)^2 p_X(x) \ge \sum_{|x-\mu| \ge \varepsilon} (x-\mu)^2 p_X(x)$$

It is also true that

$$\sum_{|x-\mu| \ge \varepsilon} (x-\mu)^2 p_X(x) \ge \sum_{|x-\mu| \ge \varepsilon} \varepsilon^2 p_X(x) = \varepsilon^2 \sum_{|x-\mu| \ge \varepsilon} p_X(x) = \varepsilon^2 P(|X-\mu|$$

So $\quad Var(X) \geq \varepsilon^2 \, P(|X-\mu| \geq \varepsilon)$

OR $\quad \boxed{P(|X-\mu| \geq \varepsilon) \leq \dfrac{Var(X)}{\varepsilon^2}}$

This is particularly useful when $\varepsilon = k\sigma$ $\quad (k>0)$ and σ is the standard deviation

$$\rightarrow \boxed{P(|X-\mu| \geq k\sigma) \leq \frac{Var(X)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}}$$

$\hookrightarrow$ This works for any (discrete) random variable. & conservative

Ex

$k=2 \rightarrow$ How much of pmf falls out/in 2σ?

if $k=2$

$$\hookrightarrow P(|X-\mu| \geq 2\sigma) \leq \frac{1}{4}$$

$\rightarrow$ we will exclude at most $\frac{1}{4}$, capture at least $\frac{3}{4}$

$k=3$ $\quad 3\sigma$ away $\rightarrow$ exclude at most $\frac{1}{9}$, capture at least $\frac{8}{9}$

Weak Law of Large Numbers (WLLN)

Let $X_1, X_2 \dots X_n$ be independent & identically distributed with $E(X_i) = \mu$ & $Var(X_i) = \sigma^2 < \infty$ (finite)

Let $S_n = \sum_{i=1}^{n} X_i$. Then for any $\varepsilon > 0$

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \rightarrow 0 \text{ as } n \rightarrow \infty$$

or

$$P\left(\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right) \rightarrow 1 \text{ as } n \rightarrow \infty$$

$\boxed{\text{Proof}}$: By Chebyshoff $P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{Var\left(\frac{S_n}{n}\right)}{\varepsilon^2}$

Now since $x_1 \ldots x_n$ independent

$$Var(S_n) = Var(X_1) + Var(X_2) + \ldots + Var(X_n) = n\sigma^2$$

and

$$E(S_n) = \sum_{i=1}^{n} E(X_i) = n\mu$$

So $Var\left(\frac{S_n}{n}\right) = Var\left(\frac{S_1}{n}\right) + \ldots = \frac{1}{n^2} Var(S_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

and $E\left(\frac{S_n}{n}\right) = \frac{1}{n} E(S_n) = \frac{n\mu}{n} = \mu$

By Chebyshoff: $P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{Var\left(\frac{S_n}{n}\right)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$

Let $n \to \infty$

$$\to 0 \leq \lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq 0$$

So

$$\boxed{\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) = 0}$$ Convergence in probability $\frac{S_n}{n} \xrightarrow{n} \mu$

Or $\frac{S_n}{n} \to \frac{1}{n} E(S_n)$

Oct 24, 2018

Recall $S_n \sim Bin(n, p)$ (i.e. $S_n = X_1 + X_2 + \ldots + X_n$, where X_i are independent Bernoulli r.v. with $X_i \sim Ber(p)$ $\forall i$,

then

S_n is well-approximated by a normal distribution when n is large. Specifically,

$S_n \sim \mathcal{N}(np, np(1-p))$ where np = mean of $Bin(n,p)$

$np(1-p)$ = variance of $Bin(n,$

In fact, given $a<b$

$$\boxed{\lim_{n\to\infty} P\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right) = P(a \le Z \le b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = \Phi(b) - \Phi(a)}$$

where $Z \sim N(0,1)$

Connections/generalizations This in fact, is a special case of the Central Limit Theorem (CLT)

Theorem

Given a sequence of independent & identically (same dist) distributed (i.i.d) r.v. $\{X_i\}$ with

$$\mu = E[X_i] \text{ and } \sigma^2 = Var(X_i) = E(X_i^2) - E(X_i)^2 = E[(X_i-\mu)^2] \quad \forall i$$

finite $\forall i$, then $\forall a<b$

$$\lim_{n\to\infty} P\left(a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$$

This says that X_i doesn't have to be Bernoulli. They only need to have finite Variance. This says no matter what you start with, converge to normal

$\boxed{N(0,1) \text{ is said to be an attractor}}$ (hence, "central")

Connection to WLLN

Recall the WLLN: $\boxed{\lim_{n\to\infty} P\left(\left|\frac{S_n - n\mu}{n}\right| \le \varepsilon\right) = 1 \quad \forall \varepsilon > 0}$

Let's see why this makes sense in connection to CLT (not a proof)

Observe that $\varepsilon > 0$, $P\left(\left|\frac{S_n}{n} - \mu\right| \le \varepsilon\right) = P\left(-\varepsilon \le \frac{S_n}{n} - \mu \le \varepsilon\right) = \ldots$

$$= P\left(-\varepsilon \le \frac{S_n - n\mu}{n} \le \varepsilon\right)$$

Note: CLT doesn't allow bounds to dep. on n ...

$$= P\left(\frac{-\varepsilon\sqrt{n}}{\sigma} \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le \frac{\varepsilon\sqrt{n}}{\sigma}\right)$$

recaptured the statement

~~By CLT~~, For large n $\sqrt{n}\,\varepsilon/\sigma \to \infty$

(not exactly)

So $\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \le \varepsilon\right) = P\left(-\infty \le \frac{S_n - \mu n}{\sigma\sqrt{n}} \le \infty\right)$

In view of CLT, $\simeq \int_{-\infty}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx = 1$

So, the CLT, in some sense, captures more information about S_n than weak law of large numbers.

Not only $\frac{S_n}{n}$ approximated by μ, CLT also speaks to the nature of this approximation

<u>Application</u> Random Walks

(Brownian Motion (statistical mechanics))

Suppose that you live on $\mathbb{Z}$

$$\cdots \quad -1 \quad 0 \quad 1 \quad \cdots$$

At step 1, you stand at $x = 0$, flip an <u>unfair</u> Bernoulli coin with p (walk to $+1$) and with prob. $(1-p)$ to -1.

Each step taken is independent $+1$ for prob p, -1 for prob. q

Each step is $X_i = \begin{cases} +1 & \text{w/ } p \\ -1 & \text{w/ } q \end{cases}$

or

X_i are independent, $P(X_i = 1) = p$, $P(X_i = -1) = 1-p$

Note: My position @ time n is $S_n = X_1 + X_2 + \dots + X_n$

We could ask: After long enough (enough steps), what is the probability that I'll be between a & b?

i.e., ~~what is~~ $\cancel{P\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right)?}$

Note $E[X_i] = (1)p + (-1)(1-p)$

$= 2p - 1$

$Var[S_n] = 4np(1-p)$

By CLT, $P\left(a \le \frac{S_n - (2p-1)}{\sqrt{4np(1-p)}} \le b\right) \cong \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$

$\boxed{\text{For } p = \frac{1}{2} = 1-p}$ So

$$\boxed{P\left(a \le \frac{S_n - 0}{\sqrt{n}} \le b\right) = P\left(a \le \frac{S_n}{\sqrt{n}} \le b\right) \approx \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx}$$

Q: What is the probability that I'm back to 0?, infinitely often

Ans: $P(\text{back to } 0 \text{ i.o.}) = 1$ in $\mathbb{R}$, $\mathbb{R}^2$, not $\mathbb{R}^3$

Poisson Distribution

Oct 26, 2018

Consider pmf for binomial : $P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$

Think about what happens if

$$\begin{cases} n \to \infty \\ p \to 0 \quad \text{such that } np = \lambda \text{ (constant)} \end{cases}$$

↳ Poisson is a good model for rare events

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!} \frac{n!}{(n-k)!\, n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}$$

$$= \frac{\lambda^k}{k!} \underbrace{\left[n(n-1)\dots(n-k+1)\right] \cdot \frac{1}{n^k}} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^n} \left(1 - \frac{\lambda}{n}\right)^{-k}$$

$\boxed{\text{What happens if } n \to \infty}$

$$P(S_n = k) = \frac{\lambda^k}{k!} \cdot (1) \cdot e^{-\lambda} \cdot (1) = \frac{\lambda^k}{k!} e^{-\lambda}$$

So $$\boxed{P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \;;\; k = 0, 1, 2, \dots}$$

$X \sim \text{Poisson}(\lambda)$. ↳ This is the probability that we observe k events in some time interval.

Intervals are independent

$\boxed{\text{What is } E(X)\,?}$ $\quad \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^x}{(x-1)!}$

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

λ is the mean number of events occuring in some time period.

<u>What about the variance of X?</u>

$\hookrightarrow$ $Var(X) = E(X^2) - E(X)^2$

<u>Well</u> $E(X^2) = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!}$ ... (less fun)

$\boxed{\text{another way to try this}}$

$E(X^2) = ?$ <u>Well</u> $E[X(X-1)] = E(X^2) - E(X)$

We want $E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda}\lambda^x}{x!}$

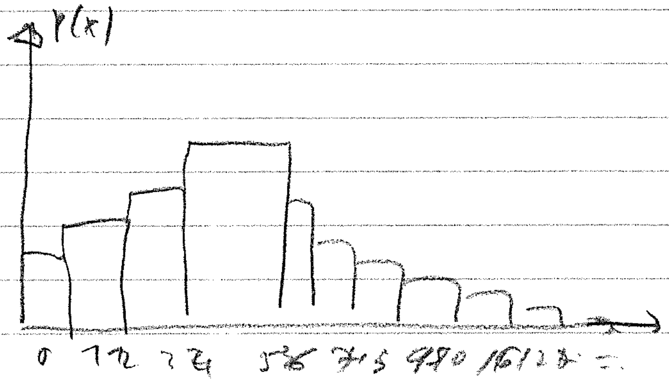
$$= \left(\sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^{x-2}}{(x-2)!}\right) \cdot \lambda^2 = \lambda^2$$

<u>So</u>

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

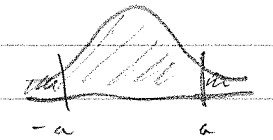
<u>So</u> $\boxed{Var(X) = (\lambda^2 + \lambda) - \lambda^2 = \lambda}$.

$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$



$P(|X| < a) = \phi(a) - \phi(-a)$

$= \Phi(a) - (1 - \Phi(a))$



Oct 29, 2018

Recall WLLN

$$\boxed{\text{Chebychev's Ineq: } P(|X-\mu| \geqslant \varepsilon) \leq \frac{Var(X)}{\varepsilon^2} \quad \text{where } \begin{cases} X \text{ random var} \\ \mu, \sigma^2 \text{ finite} \end{cases}}$$

Note: if $\varepsilon \leq \sigma \Rightarrow$ this gives no information

$\boxed{\text{Random Walks}}$ dimensional controlled ... $\boxed{\text{Poisson Distribution}}$

(Confidence Intervals & related things)

$X \sim Bin(n, p)$

Then $E(X) = np$, and by WLLN $\frac{X}{n}$ close to p

Suppose p is unknown & $S_n \sim Bin(n,p)$. How do I estimate p

Guess: $\hat{p} = \frac{S_n}{n}$ should be close for large n

Try to compute $P(|\hat{p} - p| < \varepsilon)$

Note $|\hat{p} - p| < \varepsilon \Rightarrow \left|\frac{S_n}{n} - p\right| < \varepsilon \Rightarrow \left|\frac{S_n - np}{n}\right| < \varepsilon$

So $-\varepsilon < \frac{S_n - np}{n} < \varepsilon$

So $-n\varepsilon < S_n - np < n\varepsilon$

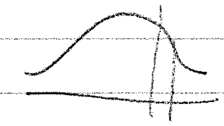
$$\rightarrow \frac{-n\varepsilon}{\sqrt{np(1-p)}} \leqslant \frac{S_n - np}{\sqrt{np(1-p)}} \leqslant \frac{n\varepsilon}{\sqrt{np(1-p)}}$$

So $$\frac{-\varepsilon}{\left(\sqrt{\frac{pq}{n}}\right)} \leqslant \frac{S_n - np}{\sqrt{npq}} \leqslant \frac{\varepsilon}{\left(\sqrt{\frac{pq}{n}}\right)}$$

i.i.d : independent
identically distributed

So $P(|\hat{p}-p|<\varepsilon) \cong \Phi\left(\dfrac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}}\right) - \Phi\left(\dfrac{-\varepsilon}{\sqrt{\frac{p(1-p)}{n}}}\right)$

$\simeq 2\Phi\left(\dfrac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}}\right) - 1$



<u>Note</u> $\max_{0\le p\le 1} \sqrt{p(1-p)} = 1/2$

So $\Phi\left(\dfrac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}}\right) \geqslant \Phi\left(\dfrac{\varepsilon}{1/2\sqrt{n}}\right)$

So $\boxed{P(|\hat{p}-p|<\varepsilon) \geqslant 2\Phi\left(\dfrac{\varepsilon}{1/2\sqrt{n}}\right)-1}$

Two ways to use this: know n → can find $P(\,)$
know ε & $P(\,)$, can find n

<u>Supposed we want</u> $P(|\hat{p}-p|<0.05) \geqslant 0.99$

<u>Can compute n</u>: $P(|\hat{p}-p|<0.05) \geqslant 2\Phi\left(\dfrac{0.05}{1/2\sqrt{n}}\right)-1 \geqslant 0.99$

<u>So</u> choose n so that

(ε : margin of error)

$$2\Phi\left(\frac{0.05}{1/2\sqrt{n}}\right)-1 \geqslant 0.99$$

<u>or</u> $\Phi\left(\dfrac{\sqrt{n}}{10}\right) \geqslant 0.995 \to$ find $n \geqslant 665.54$ ...

$\boxed{\text{For confidence intervals}}$ Want ε so that $p \in (\hat{p}-\varepsilon, \hat{p}+\varepsilon)$ with probability $\geqslant r$.

<u>Ex</u> $n = 1000$ $S_n = 400$
99% CI for p = ?
What ε do we choose?

$P(|\hat{p}-p|<\varepsilon) \geqslant 0.99$

$\to 2\Phi\left(\dfrac{\varepsilon}{1/2\sqrt{n}}\right)-1 \geqslant 0.99 \Rightarrow \boxed{\text{solve for } \varepsilon}$

Oct 31, 2012

Def $\boxed{\text{X is an r.v. with values in } \mathbb{N} = \{0,1,2,\dots\} \text{ has a Poisson distribution with parameter } \lambda \text{ if } P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}}$

$E(X) = \lambda$, $Var(X) = \lambda$

Theorem

Approximate Bin with Poisson →

> Let $\lambda > 0$, and $n \in \mathbb{N}$ such that $\frac{\lambda}{n} < 1$. Suppose $S_n \sim Bin(n, \frac{\lambda}{n})$, (so that $np = \lambda$), then
>
> $$\lim_{n\to\infty} P(S_n = k) = e^{-\lambda}\frac{\lambda^k}{k!} = P(X=k)$$
>
> $X \sim Poisson(\lambda)$

approximate $P(S_n = K)$ by $Poisson(np)$

Theorem

> If $S_n \sim Bin(np)$ & $X \sim Poisson(np)$, then
>
> $$\left|P(S_n \in A) - P(X \in A)\right| \le np^2 = \lambda p$$

→ good if p small

$\boxed{\text{Poisson good for Rare events}}$ (p small)

rule of thumb: If $X \sim Poisson(\lambda)$, it counts the number of occurrences of a rare event with average # of occurrences $= \lambda$ and not strongly dependent

Note $\boxed{P(X=0) = e^{-\lambda}}$ → can get λ from P(not happening)

#

Exponential Distribution

Def X is an r.v with values in $[0,\infty)$ has an exp. dist with param (rate) λ if its density fn

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & \text{if } x\geq 0\\ 0 & \text{if } x<0\end{cases}$$

(1) Cumulative density fn: $F(t)=P(X\leq t)=\int_{-\infty}^{t} f(x)\,dx$

$$=\int_0^t f(x)\,dx = -e^{-\lambda x}\Big|_0^t = 1-e^{-\lambda t}$$

So $$\boxed{F(t)=\begin{cases}1-e^{-\lambda t} & t\geq 0\\ 0 & t<0\end{cases}}$$

So $P(X>t)=e^{-\lambda t}$; $\lim_{t\to\infty} F(t)=1$

(2) $E[X]=\int_0^\infty \lambda x f(x)\,dx = \int_0^\infty \lambda x e^{-\lambda x}\,dx = \frac{1}{\lambda}$

So $\boxed{E[X]=\frac{1}{\lambda}}$

(3) $E(X^2)=\int_0^\infty x^2 e^{-\lambda x}\,dx = \boxed{\frac{2}{\lambda^2}}$

(4) $\boxed{Var(X)=\frac{2}{\lambda^2}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}}$

Theorem $\boxed{\text{if } X\sim Exp(\lambda)\text{, then } P(X>t+s\mid X>t)=P(X>s)}$

(memoryless property)

Proof $P(X > t+s \mid X > t) = \dfrac{P(X > t+s \ \& \ X > t)}{P(X > t)} = \dfrac{P(X > t+s)}{P(X > t)}$

$= \dfrac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P(X > s)$

Theorem

if X is a continuous r.v. w/ positive values and has memoryless property then $X \sim Exp(\lambda)$ for some λ

$\boxed{\text{Exp dist} \iff \text{memoryless w/ cont. r.v}}$

The Gamma Function:

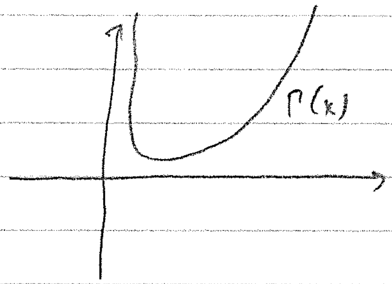
$n!$ works when $n \in \{0, 1, 2, \dots\}$

Can I define $x!$ for $x \notin \mathbb{N}$?

Yes! Requirements: $x! = x(x-1)!$ should stay true.

↳ It can be done such that the function is also differentiable.

$$\boxed{\begin{array}{l} \Gamma(n) = (n-1)! \quad \text{so} \quad \Gamma(x+1) = x\Gamma(x) \\ \Gamma \text{ is defined on } \mathbb{R} \setminus \{0, -1, -2, \dots\} \end{array}}$$



→ grows faster than e^x

$\log(\Gamma(x))$ is still concave up!

...r 2, 2018

1 $Exp(\lambda)$

Theorem Fix λ, choose n such that $\frac{\lambda}{n} < 1$. Suppose $nT_n \sim Geom(\frac{\lambda}{n})$ $\sim p$

Then $\lim_{n \to \infty} P(T_n > t) = e^{-\lambda t} = P(X > t)$
$X \sim Exp(\lambda)$

Cor $\boxed{P(T_n \le t) \to P(X \le t)}$

Proof $P(T_n > t) = P(nT_n > nt)$

$$= \sum_{k > nt} (1 - \lambda/n)^{k-1} (\lambda/n)$$

$\boxed{\lfloor x \rfloor = \text{largest int} \le x}$

$$= \sum_{k \ge \lfloor nt \rfloor + 1} (1 - \lambda/n)^{k-1} \frac{\lambda}{n}$$

$$= \left(1 - \frac{\lambda}{n}\right)^{\lfloor nt \rfloor} \frac{\lambda}{n} \sum \left(1 - \frac{\lambda}{n}\right)^{k} = \left(1 - \frac{\lambda}{n}\right)^{\lfloor nt \rfloor} \cancel{\frac{\lambda}{n}} \frac{1}{\cancel{1 - (1 - \frac{\lambda}{n})}}$$

$$= \left(1 - \frac{\lambda}{n}\right)^{\lfloor nt \rfloor} = \left(1 - \frac{\lambda}{n}\right)^{nt} \left(1 - \frac{\lambda}{n}\right)^{\lfloor nt \rfloor - nt}$$

Note $-1 < \lfloor nt \rfloor - nt \le 0$

$$= \left(1 - \frac{\lambda t}{nt}\right)^{nt} \left(1 - \frac{\lambda}{n}\right)^{\lfloor nt \rfloor - nt}$$

So $\lim_{n \to \infty} P(T_n > T) = e^{-\lambda t} \cdot 1$

So $\boxed{\lim_{n \to \infty} P(T_n > T) = e^{-\lambda t} = P(X > t)}$

2 The Gamma Function

Def $r > 0,\ r \in \mathbb{R}$. Define $\boxed{\Gamma(r) = \int_0^\infty x^{r-1} e^{-x}\,dx}$

At ∞ $\cancel{x^{r-1}e^{-x}}$ $x^{r-1} < e^{x/2}$ for large x

So $x^{r-1} \cdot e^{-x} < e^{-x/2}$ & $\int_0^\infty x^{r-1} e^{-x}\,dx < \int_0^\infty e^{-x/2}\,dx$ converge

At 0 $(r < 1)$ **Check** $\int_0^1 \frac{1}{x^r}\,dx$ converges when $r < 1$

Let $r = 1$

$$\boxed{\begin{aligned} \Gamma(1) &= \int_0^\infty e^{-x}\,dx = 1 \\ \Gamma(r+1) &= r\,\Gamma(r) \end{aligned}}$$

So $\Gamma(2) = 1\,\Gamma(1) = 1$

$\Gamma(3) = 2 \cdot \Gamma(2) = 2.1.1$

$\Gamma(4) = 3\,\Gamma(3) = 3.2.1.1 \ldots$

So

$\Gamma(n) = (n-1)!$ if $n \in \mathbb{N}$

Fact $\boxed{\Gamma(r) \text{ is infinitely differentiable}}$

Another proof $\boxed{\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin(\pi x)}}$ true if $0 < x < 1$

$$\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right) = \pi = \Gamma\left(\frac{1}{2}\right)^2$$

So $\boxed{\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}}$

3 The Gamma Distribution

Def: $r, \lambda > 0$. X continuous RV with nonnegative values. X has Gamma (r, λ) distribution if its ~~the~~ pdf is

$$f(x) = \begin{cases} \dfrac{\lambda^r x^{r-1}}{\Gamma(r)} e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

<u>Note</u> if $r = 1 \to f(x) = \lambda e^{-\lambda x}$ (exp)

Check that $\displaystyle\int_0^\infty \frac{\lambda^r x^{r-1}}{\Gamma(r)} e^{-\lambda x}\,dx = 1$

well $\displaystyle = \frac{\lambda^r}{\Gamma(r)} \int_0^\infty x^{r-1} e^{-\lambda x}\,dx$

<u>First note</u> $\displaystyle\Gamma(r) = \int_0^\infty x^{r-1} e^{-x}\,dx$

Make $u = \lambda x \Rightarrow du = \lambda\,dx \to \displaystyle = \cancel{\int_0^\infty (du\,(\lambda x)^{r-1} e^{-\lambda x} \frac{\lambda}{\Gamma(r)}}$

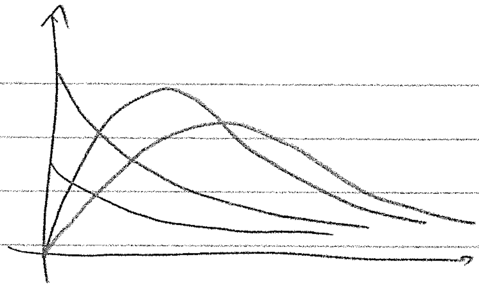
$$= \frac{1}{\Gamma(r)} \int_0^\infty (\lambda^{r-1}) \left(\frac{u}{\lambda}\right)^{r-1} e^{-u}\,du$$

t

putation

$$= \frac{1}{\Gamma(r)} \int_0^\infty u^{r-1} e^{-u}\,du = \frac{\Gamma(r)}{\Gamma(r)} = 1$$

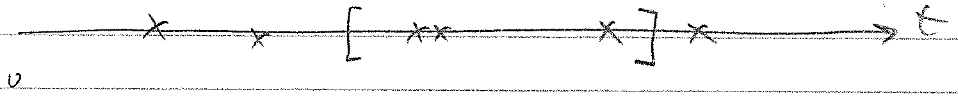
$\boxed{\text{Next}}$ $\boxed{\begin{aligned} E[X] &= \frac{r}{\lambda} \\ Var[X] &= \frac{r}{\lambda^2} \end{aligned}}$



Note

$$\chi^2(n) = \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$$

Example $\boxed{\text{Poisson Process}}$



intensity

(1) Distinct random points
(2) for any bounded interval I, $N(I) \sim \text{Poisson}\left(\lambda |I|\right)$
(3) for non-overlapping bounded intervals, $N(I)$ are independent

$$N_t = N\left([0, t]\right)$$

let T_1 = position of 1^{st} point $\rightarrow P(T_1 > t) = e^{-\lambda t}$

So $T_1 \sim Exp(\lambda)$
$T_2 \sim \text{Gamma}(2, \lambda)$
$\vdots$
$T_n \sim \text{Gamma}(n, \lambda)$

5, 2018

Moment generating Functions

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\,dx$$

function of t: $E\left[(xt)^n\right] = \int_{-\infty}^{\infty} (xt)^n f(x)\,dx$

& $\frac{t^n E(X^n)}{n!} = \int_{-\infty}^{\infty} \frac{(xt)^n}{n!} f(x)\,dx$

$$\sum_{n=0}^{\infty} \frac{t^n E(X^n)}{n!} = \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} \frac{(xt)^n}{n!} f(x)\,dx = \int_{-\infty}^{\infty} e^{xt} f(x)\,dx$$

So, definition

similar to laplace transform

Def: X is a random variable. $t \in \mathbb{R}$. The moment generating fn of X is

$$M_X(t) = E\left[e^{tX}\right] = \int_{-\infty}^{\infty} e^{xt} f(x)\,dx \quad \text{(for cont)}$$

$$= \sum_x e^{xt} P(X=x) \quad \text{(for discrete)}$$

which we hope converges for $t \in (-\delta, \delta)$

Characteristic function of X → another version of moment generating fn.

$$\varphi_X(t) = E\left[e^{itX}\right]$$

$i^2 = -1$

this is the Fourier transform ...

- If I know $M_X(t)$, how much do I know about X?

$$M_X(t) = \int_{-\infty}^{\infty} e^{xt} f(x)\,dx = \sum_{n=0}^{\infty} t^n \int_{-\infty}^{\infty} \frac{x^n}{n!} f(x)\,dx = \sum_{n=0}^{\infty} \frac{E[X^n]}{n!} t^n$$

Recall

$$f(t) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} t^n$$

So

$$M_X(t) = \sum_{n=0}^{\infty} \frac{M^{(n)}(0)}{n!} t^n = \sum_{n=0}^{\infty} \frac{E[X^n]}{n!} t^n$$

So $\boxed{M_X^{(n)}(0) = E[X^n]}$

$\boxed{\text{Ex}}$ Suppose X r.v. with $P(X=2) = \frac{1}{3}$

$P(X=0) = \frac{1}{6}$ $\quad P(X=k) = 0$

$P(X=1) = \frac{1}{2}$ $\quad (k \neq 1, 2, 0)$

$$\sum_k e^{kt} P(X=k) = \boxed{\frac{1}{3}e^{2t} + \frac{1}{6}e^{0t} + \frac{1}{2}e^{t}} = M_X(t)$$

$\boxed{\text{Ex}}$

$X \sim \text{Poisson}(\lambda)$ $\quad P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$

So

$$M_X(t) = \sum_{k}^{\infty} e^{kt} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_k \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} e^{\lambda e^t}$$

So $\boxed{M_X(t) = e^{\lambda(e^t - 1)}}$

So $M_X^{(0)}(0) = e^{\lambda(e^0-1)} = 1 = E(X^0) = E(1)$ ✓

$M_X'(0) = \lambda(e^0 \cancel{\quad}) e^{\lambda(1-1)} = \lambda = E(X)$ ✓

$M_X''(0) = \lambda + \lambda^2 = E(X^2)$ ✓

Recall $\boxed{M_X(t) = E(e^{tX})}$ **hope** $M_X(t)$ is well-defined in some interval $(-\delta, \delta)$ (containing 0)

$$M_X(t) = \sum_{n=0}^{\infty} \frac{E(X^n)}{n!} t^n \quad \text{if it's well-defined}$$

$$\boxed{M^{(n)}(0) = E[X^n]}$$

Recall $X \sim \text{Poisson}(\lambda)$ $\quad M(t) = e^{\lambda(e^t - 1)}$

Laplace Transform

Theorem X, Y are r.v. if $M_X(t)$ and $M_Y(t)$ are defined on some interval $(-\delta, \delta)$ and are equal on that interval, then

$$P(X \le x) = P(Y \le x) \quad \forall x$$

GF can [d]etermine [prob]ability dist.

Caution $E[X^n] = E[Y^n]$ does not imply equality in distribution...

Example

$X \sim \text{Exp}(\lambda)$

$$M_X(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} \underbrace{\lambda e^{-\lambda x} dx}_{f_X(x)dx}$$

So $$M_X(t) = \lambda \int_0^{\infty} e^{x(t-\lambda)} dx = \frac{\lambda}{t-\lambda} e^{x(t-\lambda)} \Big|_0^{\infty}$$

So $$\boxed{M_X(t) = \frac{\lambda}{\lambda - t} \quad \text{if } t - \lambda < 0}$$

$\boxed{\text{Example}}$ $X \sim N(0,1)$ <u>or</u> $Z \sim N(0,1)$ $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$$M_Z(t) = E\left(e^{tx}\right) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - x^2/2}\, dx$$

<u>Note</u>

$$tx - \frac{x^2}{2} = -\frac{x^2 - 2tx}{2} = -\frac{x^2 - 2tx + t^2}{2} + \frac{t^2}{2}$$

$$= \frac{t^2}{2} - \frac{(x-t)^2}{2}$$

$$M_Z(t) = \frac{1}{\sqrt{2\pi}} e^{t^2/2} \int_{-\infty}^{\infty} e^{-(x-t)^2/2}\, dx = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}}\, dx = e^{t$$

<u>So</u> $\boxed{M_X(t) = e^{t^2/2}}$

$\boxed{\text{Example}}$

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

$X = \sigma Z + \mu.$ $\quad M_X(t) = E\left[e^{tX}\right] = E\left[e^{t\mu} e^{t\sigma Z}\right]$

$$= e^{t\mu} E\left[e^{t\sigma Z}\right]$$

$$= e^{t\mu} \cancel{E}\, e^{t^2\sigma^2/2}$$

<u>So</u> $M_X(t) = e^{t\mu} e^{t^2\sigma^2/2}$

$\boxed{\text{Example}}$ $X \sim Bin(n,p)$

$$M(t) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^{n} (pe^t)^k \binom{n}{k} (1-p)^{n-k}$$

$$= (1-p+pe^t)^n$$

$\boxed{\text{Add} \Rightarrow \text{Multiply } M}$

$\boxed{\text{Example}}$ $Ber(p) = e^{t \cdot 0}(1-p) + e^t p = (1-p+pe^t)$

$\boxed{\text{Example}}$ $Geom(p)$ $M(t) = \dfrac{pe^t}{1-(1-p)e^t}$

$\boxed{\text{Example}}$ $X \sim Unif(a,b)$

$$M(t) = \int_a^b e^{tx} \frac{1}{b-a} dx = \frac{1}{t(b-a)} e^{tx} \Big|_a^b = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

$$\begin{cases} = \dfrac{e^{tb} - e^{ta}}{t(b-a)} & \text{if } t \neq 0 \\ = 1 & \text{if } t = 0 \end{cases}$$

#

v 9, 2012

$\boxed{\text{Let } X \text{ be r.v. with CDF } F(x). \ g: \mathbb{R} \mapsto \mathbb{R}. \text{ Let } Y = g(X). \text{ What is CDF of } Y?}$

X discrete $\quad p_X(x) = \begin{cases} 1/2 & x=1 \\ 1/3 & x=-1 \\ 1/6 & x=2 \\ 0 & \text{otherwise} \end{cases}$ $\quad Y = X^2$

so $P(Y=k) = P(X^2 = k) = P(X = \sqrt{k}) + P(X = -\sqrt{k})$

So $\boxed{P(Y=k) = \sum_{\ell \in g^{-1}(k)} P(X=\ell)}$

X continuous

Example let $U \sim Unif(0,1) \rightarrow F_u(x) = \begin{cases} 0 & \text{if } x<0 \\ x & \text{if } x \in [0,1] \\ 1 & \text{if } x>1 \end{cases}$

Let $g(x) = \frac{-1}{\lambda} \ln(1-x)$. What is CDF of Y? $(\lambda > 0)$

So $\boxed{Y = g(U) \geq 0}$

So $F_y(y) = 0$ if $y \leq 0$

$$P(Y \leq y) = P\left(-\frac{1}{\lambda}\ln(1-x) \leq y\right)$$

$$= P(\ln(1-x) \geq -\lambda y)$$
$$= P(1-x \geq e^{-\lambda y})$$
$$= P(x \leq 1-e^{-\lambda y}) = 1-e^{-\lambda y}$$, since $X \sim Unif[$

So $F_y(y) = \begin{cases} 0 & \text{if } y \leq 0 \\ 1-e^{-\lambda y} & \text{if } y \geq 0 \end{cases}$ So $\boxed{Y \sim Exp(\lambda)}$

Consider

$Y \sim Exp(\lambda)$, so $F_y(y) = 1-e^{-\lambda y}$

Let $x = 1-e^{-\lambda y} \Rightarrow y = \frac{-1}{\lambda}\ln(1-x) = g(x)$

Example let $Z \sim N(0,1)$. let $Y = Z^2$

Find F_Y & f_y of Y

$$\boxed{F_y(y) = \begin{cases} 0 & \text{if } y \le 0 \\ ? = 2\Phi(\sqrt{y}) - 1 & \text{if } y > 0 \end{cases}}$$

Recall $P(Y \le y) = P(Z^2 \le y) = P(-\sqrt{y} \le z \le \sqrt{y})$

$$= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1$$

Can find

$$f_y(y) = \frac{d}{dy}\left(2\Phi(\sqrt{y}) - 1\right) = 2\varphi(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{\varphi(\sqrt{y})}{\sqrt{y}}$$

$$= \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-y/2}$$

$$= \frac{1}{\sqrt{2\pi y}} e^{-y/2}$$

So

$$\boxed{f_y(y) = \begin{cases} 0 & \text{if } y \le 0 \\ \frac{1}{\sqrt{2\pi y}} e^{-y/2} & \text{if } y > 0 \end{cases}}$$

Note

$X \sim \text{Gamma}(r, \lambda)$ $\quad f_x(x) = \dfrac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)} = \dfrac{(1/2)^{1/2} x^{-\frac{1}{2}} e^{-x/2}}{\Gamma(1/2)}$

if $r = \lambda = \frac{1}{2}$

So $\boxed{\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}}$ $\quad$ **Note** $\boxed{\text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right) = \chi^2 \text{ dist}}$

X: $\quad g: \mathbb{R} \mapsto \mathbb{R}$ diff. increasing, $g' \neq 0$, except at finitely many points
$Y = g(X)$

Find $P(Y \leq y) = P(g(x) \leq y)$

g increasing $\Rightarrow$ 1-to-1

$= P(x \leq g^{-1}(y))$

$$\boxed{P(Y \leq y) = F_X(g^{-1}(y))}$$

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

If g is finite-to-1, then

$$\boxed{f_Y(y) = \sum_{x \in g^{-1}(y)} f_X(x) \frac{1}{|g'(x)|}}$$

Generating r.v. from Unif. Let $U \sim \text{Unif}[0,1]$. X cont. r.v.
What I want: find g s.t. $g(U) \overset{d}{=} X$.

Want $P(g(U) \leq x) = P(X \leq x)$

Take $g(u) = F_X^{-1}(u)$

F_X increasing!

So $P(g(U) \leq x) = P(F_X^{-1}(u) \leq x)$

$= P(u \leq F_X(x))$ $\quad$ because $u \in [0,1]$

$= F_X(x)$

$\boxed{\text{Joint Distribution of R.V.}}$

$$\Omega \xrightarrow{x_1} \mathbb{R}, \quad \Omega \xrightarrow{x_2} \mathbb{R}, \quad \vdots \quad \Omega \xrightarrow{x_n} \mathbb{R}$$

<u>CDF</u>

$$\boxed{F_{\tilde{X}}(x_1, \dots x_n) = P\left(X_1 \leq x_1, \dots, X_n \leq x_n\right)}$$

<u>or</u> $\tilde{X} : \Omega \longmapsto \mathbb{R}^n$

$\boxed{\text{Case 1}}$ $\{x_i\}$ are all discrete

<u>Joint Probability mass function</u>: $p(k_1, k_2 \dots k_n) = P(X_1 = k_1 \dots X_n = k_n)$

So $$\boxed{\sum_{(k_1 \dots k_n)} p(k_1 \dots k_n) = 1}$$

↳ no Expected value here. <u>But</u> if we have $g : \mathbb{R}^n \to \mathbb{R}$

$$\boxed{E\left[g(x_1, \dots x_n)\right] = \sum_{k_1 \dots k_n} g(k_1, \dots k_n)\, P(X_1 = k_1, \dots X_n = k_n)}$$

<u>Recover</u> pmf for X_1 from $P_{\tilde{X}}(x_1, \dots x_n)$

<u>Note</u> $\{x_1 = k_1\} = \bigcup$ of $\{x_1 = k_1, x_2 = k_2 \dots x_n = k_n\}$ over all choices of $k_2 \dots k_n \dots$

<u>So</u> $$P(x_1 = k_1) = \sum_{(k_2 \dots k_n)} P_{\tilde{x}}(x_1 = k_1, \dots x_n = k_n)$$

<u>Eg</u>

X_2 \ X_1	1	2	3
0	1/6	1/6	1/6
5	1/3	0	1/6

$P_{\tilde{x}}(k_1, k_2) = ?$

$P_{X_1}(2) = \frac{1}{6}, \quad P_{X_1}(1) = \frac{1}{6} + \frac{1}{3}$

So $\boxed{p_{X_j}(k_j) = \sum_{k_1, k_2 \ldots k_{j-1}, k_{j+1} \ldots k_n} p_{\vec{X}}(k_1 \ldots k_{j-1}, \boxed{k_j}, k_{j+1} \ldots k_n)}$

marginal

Ex Roll 2 dice. X_1: roll of #1, X_2: roll of #2

and $Y_1 = \min(X_1, X_2)$

$Y_2 = \text{abs}(X_1 - X_2) = |X_2 - X_1|$

Legal values (i, j); $1 \leq i, j \leq 6$ and $p_{X_1 X_2}(i, j) = 1/36$

Y_1 : possible values $1 \ldots 6$

Y_2 : possible values $0 \ldots 5$

Y_1 \ Y_2	0	1	2	3	4	5
1	1/36	1/18	1/18	1/18	1/18	1/18
2	1/36	1/18	1/18	1/18	1/18	0
3	1/36	1/18	1/18	1/18	0	0
4	1/36	1/18	1/18	0	0	0
5	1/36	1/18	0	0	0	0
6	1/36	0	0	0	0	0

$p_{Y_1 Y_2}(i, 0) = p_{X_1 X_2}(i, i) = \frac{1}{36}$

$p_{Y_2}(0) = \frac{1}{6}$

$p_{Y_1 Y_2}(1, 5) = \frac{2}{36}$

Multinomial Distribution

n independent trials. r possible outcomes $1, 2, \ldots, r$ with probabilities $p_1, p_2, \ldots, p_r$

So $p_1 + p_2 + \ldots + p_r = 1$

$X_1, X_2, \ldots, X_r$ where X_i is # of occurrences of outcome i

Possible values are $(k_1, k_2, \ldots, k_r)$ with $k_1 + k_2 + \ldots + k_r = n$

So # of outcomes, each with the same probability $p_1^{k_1} \cdot p_2^{k_2} \cdot p \ldots p_r^{k_r}$. # of terms is $\binom{n}{k_1, k_2 \ldots k_r}$

<u>Multinomial coeff</u> $\binom{n}{k_1, k_2, \dots, k_r} = \dfrac{n!}{k_1!\, k_2! \dots k_r!}$

pmf for Multinomial →

$$\text{Multinom}(n, r, p_1, p_2, \dots p_r)$$

$$P(k_1 \dots k_r) = \binom{n}{k_1 \dots k_r} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}$$

Normalized? Yes

$$(x_1 + \dots + x_r)^n = \sum_{k_1 + k_2 + \dots + k_r = n} \binom{n}{k_1 \dots k_r} x_1^{k_1} x_2^{k_2} \dots x_r^{k_r}$$

Sampling with replacement (no order)

↳ $p_1 = p_2 = p_3 = \dots = p_r = \frac{1}{r}$

<u>Note</u> Marginals are binomial

$$P_{X_1}(k_1) = \binom{n}{k_1} p_1^{k_1} (1-p_1)^{n-k_1} \quad \text{(binomial)}$$

<u>Ex</u> Roll a die 100 times P (10 sixes & 7 fives)

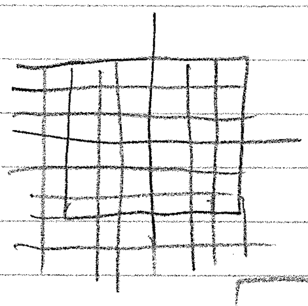
P (10 sixes and 7 fives) → $\text{Multinom}\left(100, 3, \frac{1}{6}, \frac{1}{6}, \frac{2}{3}\right)$

$\boxed{\text{Joint Dist, Cont}}$

Nov 16, 2018

CDF $F(x_1, \dots x_n) = P(X_1 \le x_1, X_2 \le x_2 \dots X_n \le x_n)$

Ex $F_{XY}(x,y) = P(X \le x, Y \le y)$

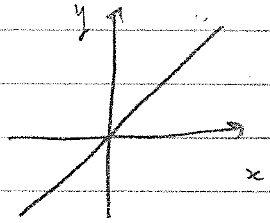


But what if X & Y continuous RVs?

$\boxed{\text{There might not be a joint density}}$

We would like $\boxed{P(X \le x, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\, dt\, ds}$

Suppose $X = Y$, then the entire P is on the line



$f(x,y) = 0$ if $x \ne y$

so $\int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\, dt\, ds = P(X \le x, Y \le y)$ $\boxed{\text{can't be true}}$

$\boxed{\text{Definition}}$

X & Y are jointly continuously distributed if $\exists\, f(x,y)$ s.t. $\int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\, dt\, ds = P(X \le x, Y \le y)$

It follows that

$$P((X,Y) \in B) = \iint_B f(x,y)\, dA$$

$$\boxed{\begin{aligned}&\underline{\text{Note}}\ \textbf{(1)}\ f(x,y) \geq 0\\ &(2)\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s,t)\,dt\,ds = 1\end{aligned}}$$

Expectation value... $g : \mathbb{R}^2 \to \mathbb{R}$

$$\to \boxed{E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\,f(x,y)\,dx\,dy}$$

How to find $P(X \leq u)$?

$$\hookrightarrow \boxed{P(X \leq u) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(x,y)\,dy\,dx}$$

so

$$\boxed{\begin{aligned}&\frac{d}{dx} P(X \leq u) = f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy\\ &\frac{d}{dx}(Y \leq y) = f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx\end{aligned}}$$

Similarly

- $$E[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)\,f(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,f(x,y)\,dy\,dx + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\,f(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x,y)\,dy\,dx + \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f(x,y)\,dx\,dy$$

$=$

$$= \int_{-\infty}^{\infty} x f_X(x)\,dx + \int_{-\infty}^{\infty} y f_Y(y)\,dy$$

$$= E(X) + E(Y) \quad \& \quad \boxed{E(X+Y) = E(X) + E(Y)}$$

$\boxed{Ex}$

$$f(x,y) = \begin{cases} \frac{3}{2}\left(x^2y + y\right) & \text{if } (x,y) \in [0,1]\times[0,1] \\ 0 & \text{if not} \end{cases}$$

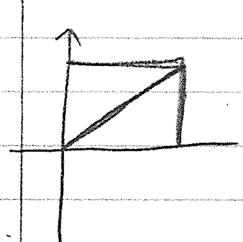
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x^2y + y)\,dx\,dy = \int_0^1\int_0^1 x^2y + y\,dx\,dy$$

$$= \int_0^1 \frac{1}{3}x^3y + xy\Big|_0^1 dy$$

$$= \int_0^1 \frac{4}{3}y\,dy = \frac{1}{6} + \frac{1}{2} = \frac{2}{3}$$

$$\rightarrow \quad f(x,y) = \frac{3}{2}\left(x^2y + y\right)$$

$$P(X<Y) = ? = \int_0^1\int_0^y \frac{3}{2}\left(x^2y + y\right)dx\,dy$$



$$= \frac{3}{2}\int_0^1 \left(\frac{y^4}{3} + y^3\right)dy$$

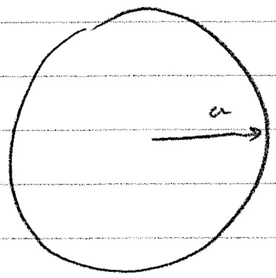
$$= \frac{3}{2}\left(\frac{1}{15} + \frac{1}{3}\right) = \frac{3}{5}$$

Uniform Dist on Bounded set R

$$\boxed{f(x,y) = \begin{cases} \dfrac{1}{\text{area}(R)} & \text{if } (x,y) \in R \\ 0 & \text{if } (x,y) \notin R \end{cases} \qquad \mathbb{P}\big((x,y) \in B\big) = \frac{\text{area}(B)}{\text{area}(R)}}$$

(unif)

Ex Region $x^2 + y^2 \le a^2$



$$f(x,y) = \begin{cases} \dfrac{1}{\pi a^2} & \text{if } (x,y) \in R \\ 0 & \text{if not} \end{cases}$$

$$f_X(x) = \int_{-\sqrt{a^2-x^2}}^{\sqrt{a^2-x^2}} \frac{1}{\pi a^2}\, dy = \cancel{\frac{2a}{\pi a^2}} \cancel{\ell \frac{2}{\pi a}} = \frac{2\sqrt{a^2-x^2}}{\pi a^2}$$

So $\boxed{f_X(x) = \dfrac{2\sqrt{a^2-x^2}}{\pi a^2}, \quad f_Y(y) = \dfrac{2\sqrt{a^2-y^2}}{\pi a^2}}$ (not unif)

What does $f(x,y)$ mean?

Note $\boxed{f(x,y) = \dfrac{\partial^2 F(x,y)}{\partial x\, \partial y} = \dfrac{\partial^2}{\partial y\, \partial x} F(x,y)}$

$$\boxed{P\big(x-\varepsilon < X < x+\varepsilon,\ y-\delta < Y < y+\delta\big) \simeq (4\varepsilon\delta)\, f(x,y)}$$

Next ... Independence ...

Joint Distribution & Independence

Nov 19, 2018

if $P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$

then $X_1 \ldots X_n$ independent

Equivalent to $F_{\vec{X}}(x_1, \ldots x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \ldots F_{X_n}(x_n)$

$\forall\ x_1 \ldots x_n$

Discrete case

$p_{\vec{x}}(k_1 \ldots k_n) = \prod_{i}^{n} p_{X_i}(k_i)$

Theorem

(a) If X, Y are jointly continuously distributed, and $f_{XY}(x,y) = f_X(x$
then X, Y are independent

(b) If X, Y are independent, then $f_X(x) f_Y(y) = f_{XY}(x,y)$
i.e, is their joint pdf. ,i.e

Independent $\Leftrightarrow$ Jointly continuously distributed

Proof (a) $P(X \in A, Y \in B) = \iint_{A \times B} f_{XY}(x,y)\,dx\,dy$

$$= \iint_{A \times B} f_X(x) f_Y(y)\,dx\,dy = \int_A f_X(x)\,dx \int_B f_Y(y)\,dy$$

$$= P(X \in A) \cdot P(Y \in B)$$

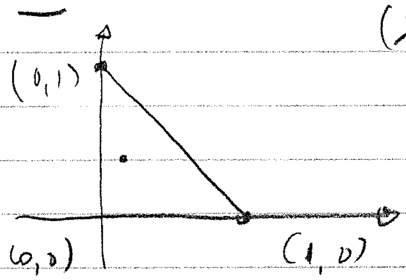
(b) Proof by reversing the direction of (a).

Theorem

Suppose $X_1 \dots X_\ell, X_{\ell+1} \dots X_n$ are indep r.v., and

$Y_1 = g_1(X_1 \dots X_\ell)$
$Y_2 = g_2(X_{\ell+1} \dots X_n)$, then Y_1, Y_2 independent

Ex



(X,Y) $\quad f_{XY}(x,y) = \begin{cases} 2 & \text{if } (x,y) \in \text{triangle} \\ 0 & \text{if not} \end{cases}$

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = \int_0^{1-x} 2\,dy = 2(1-x)$$

if $(x,y) \in \Delta$

$$f_Y(y) = 2(1-y)$$

So $f_X(x) f_Y(y) = 4(1-x)(1-y) \neq 2$

Several Independent Geometric RVs

$X_1 \dots X_n$ indep. $X_i \sim \text{Geom}(p_i)$. $Y = \min(X_1 \dots X_n)$

$P(Y > k) = P(X_1 > k \dots X_n > k)$

where $P(X_i > k) = (1-p_i)^k$

$$P(Y > k) = \prod_i^n (1-p_i)^k = \left[\prod_i (1-p_i)\right]^k = (1-r)^k$$

Let $r = 1 - \prod_i^n (1-p_i) \Rightarrow P(Y>k) = (1-r)^k \Rightarrow \boxed{Y \sim \text{Geom}(r)}$

$\boxed{\text{Let } N = \text{which of } X_i \text{ is minimum}}$

↳ **Exercise**: Show that Y & N are independent

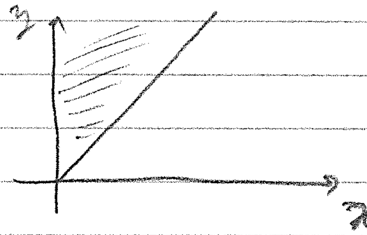
Since true for geometric → expect true for exp as well..

$\boxed{\text{Same deal for Exp}}$

X = time until Mom calls
Y = time until Grandma calls

Assume X & Y indep. $X \sim Exp(\lambda)$, $Y \sim Exp(\mu)$

$P(X<Y) = ?$ $\qquad f(x,y) = \lambda\mu e^{-\lambda x} e^{-\mu y}$



$$P(X<Y) = \int_{-\infty}^{\infty}\int_{x}^{\infty} \lambda\mu e^{-\lambda x} e^{-\mu y}\, dy\, dx$$

$$= \lambda\mu \int_{-\infty}^{\infty}\int_{x}^{\infty} e^{-\lambda x} e^{-\mu y}\, dy\, dx$$

$$= \lambda \int_{-\infty}^{\infty} e^{-\lambda x} \int_{+x}^{\infty} \mu e^{-\mu y}\, dy\, dx = \lambda \int_{-\infty}^{\infty} e^{-\lambda x}\left[e^{-\mu x}\right] dx$$

$$= \lambda \int_{-\infty}^{\infty} e^{(-\lambda-\mu)x}\, dx = \frac{\lambda}{\lambda+\mu}\int_{-\infty}^{\infty} (\lambda+\mu) e^{-(\lambda+\mu)x}\, dx - \cancel{\int_{-\infty}^{\infty} \mu e^{-(\lambda+\mu)x}\, dx}$$

$$= \frac{\lambda}{\lambda+\mu} \rightarrow \boxed{P(X<Y) = \frac{\lambda}{\lambda+\mu}}$$

76

who calls first

$$\boxed{\text{Can make } I = \left\{ \begin{array}{ll} 1 & \text{if } X < Y \text{ w/ } p = \frac{\lambda}{\lambda+\mu} \\ 0 & \text{if } X > Y \ \ldots \end{array} \right\} \quad \underline{I, T \text{ independent}} \\ \text{and } T = \min(X, Y)}$$

when the first call is

$$P(T>t) = P(X>t, Y>t) = P(X>t)P(Y>t)$$

$$= e^{-\lambda t} e^{-t\mu} = e^{-(\lambda+\mu)t}$$

<u>So</u> $\boxed{T \sim \text{Exp}(\lambda+\mu)}$

26, 2012

Standard Bivariate Normal Distribution

<u>Ideas</u>

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \quad \ldots \text{Choose } \rho, \ -1 < \rho < 1$$

$$\boxed{f_{xy}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[- \frac{x^2+y^2-2\rho xy}{2(1-\rho^2)} \right]}$$

$\boxed{\text{If } \rho = 0, \ X, Y \text{ independent } \underline{\text{and}} \ X \sim N(0,1), \ Y \sim N(0,1)}$

Marginal density of X

$$f_X = \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[- \frac{x^2+y^2-2\rho xy}{2(1-\rho^2)} \right] dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[- \frac{x^2+(y-\rho x)^2-\rho^2x^2}{2(1-\rho^2)} \right] dy = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \int_{-\infty}^{\infty} \text{normal density}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \Rightarrow X \sim \mathcal{N}(0,1) \to f_{XY} \text{ is a prob. dist.}$$

$\boxed{\rho \text{ is the correlation coefficient}}$

Sums of R.V.

Let X, Y R.Vs, and $Z = X+Y$. X & Y discrete

$$P_Z(n) = P(X+Y=n) \overset{?}{=} \sum_k P(X=k)\,P(Y=n-k)$$

$$= \sum_k P(X=k, Y=n-k) \quad \text{only if } X, Y \text{ indep}$$

If X, Y indep, then $P_Z(n) = \sum_k P(X=k)\,P(Y=n-k)$

$$= \sum_k P_X(k)\,P_Y(n-k) \quad \text{converges}$$

$$= \sum_\ell P_X(n+\ell)\,P_Y(\ell)$$

Convolution

iscrete

P_X, P_Y two functions that are zero outside a countable set

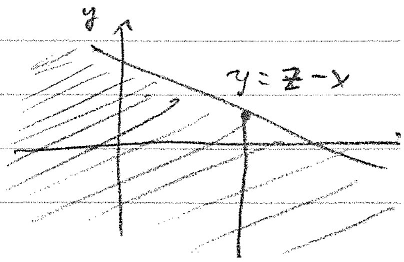
Let $(P_X * P_Y)(n) = \sum_k P_X(k)\,P_Y(n-k)$

If X, Y are indep discrete R.V., then $P_{X+Y} = P_X * P_Y$

ontinuous

Let X, Y jointly continuous. $Z = X+Y$. What is $F_Z(z)$?

$$F_Z(z) = P(X+Y \le z) = \iint_{x+y \le z} f_{XY}(x,y)\,dy\,dx$$

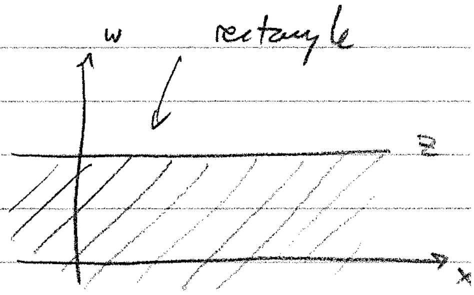


$$= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f_{XY}(x,y)\,dy\,dx = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-x} f_{XY}(x,y)\,dy\right)dx$$

let $w = y + x \Rightarrow dw = dy$

So $$F_Z(z) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{z} f_{XY}(x, w-x)\,dw\right]dx$$

rectangle



swap order of integration

$$= \int_{-\infty}^{z}\left(\int_{-\infty}^{\infty} f_{XY}(x, w-x)\,dx\right)dw$$

$$\Rightarrow \boxed{f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z-x)\,dx}$$

this integral converges

o.k. but what if X, Y <u>independent</u>?

$$\rightarrow f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z-x)\,dx = \int_{-\infty}^{\infty} f_X(x)\,f_Y(z-x)\,dx$$

continuous convolution

So, if X, Y are independent, continuous R.V., then (jointly)

$$f_{X+Y} = f_X * f_Y$$

#

$\boxed{Ex}$ if $X \sim Poisson(\lambda)$ $Y \sim Poisson(\mu)$, X, Y indep

then $\boxed{(X+Y) \sim Poisson(\lambda+\mu)}$

7.4 $\boxed{Ex}$ $X \sim Bin(n,p)$, $Y \sim Bin(m,p)$, X, Y indep

$\boxed{X+Y \sim Bin(n+m, p)}$

7.5 $\boxed{Ex}$ $X \sim Geom(p)$ $Y \sim Geom(p)$ X, Y indep.

n−1 ways

$X+Y = \#$ of tries until 2^{nd} success $= \boxed{p^2(1-p)^{n-2}(n-1)}$

$\boxed{\text{Generalize}}$ $X_i \sim Geom(p)$, ~~X_i~~ all indep

$$n = X_1 + X_2 + X_{\dots} + X_k$$

k^{th} success on the n^{th} trial

$$\boxed{P(X_1 + X_2 + \dots + X_k) = p^k (1-p)^{n-k} \binom{n-1}{k-1}}$$

→ negative binomial distribution (k,p)

$\mu = \frac{k}{p}$ $\quad E(X_i) = 1/p$

Nov 28, 2018

Ex 7.8 $X \sim N(\mu_1, \sigma_1^2)$ $Y \sim N(\mu_2, \sigma_2^2)$

$Z = X + Y \Rightarrow Z \sim N(\mu_1 + \mu_2, \sigma_2^2 + \sigma_1^2)$

Proof Complete the square in $e^{-\text{stuff}\dots}$

$\sum a_i^2 \sigma_i^2$

Corollary

$X_i \sim N(\mu_i, \sigma_i^2)$, independent

let $Z = \sum_i a_i X_i + b$, then $Z \sim N(\sum \mu_i a_i + b, \sum a_i^2 \sigma_i^2)$

Ex $\boxed{\text{Sum of 2 independent Gamma R.V.s}}$

$X \sim \text{Gamma}(a, \lambda)$ indep.
$Y \sim \text{Gamma}(b, \lambda)$

$Z = X + Y$. Claim $Z \sim \text{Gamma}(a+b, \lambda)$

Recall $f_X(x) = \dfrac{\lambda^a x^{a-1}}{\Gamma(a)} e^{-\lambda x}$ if $x \geq 0$, $f_Y(y) = \dfrac{\lambda^b x^{b-1}}{\Gamma(b)} e^{-\lambda y}$ if $y \geq 0$

Want $f_Z(z) = f_Z(x+y) = \dfrac{\lambda^{a+b} z^{a+b-1}}{\Gamma(a+b)} e^{-\lambda z}$

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, d\cancel{x} \quad = 0 \text{ if } x < 0 \text{ or } x > z \quad \to 0 \leq x \leq z$$

$$= \int_0^z \frac{\lambda^a x^{a-1}}{\Gamma(a)} \frac{\lambda^b (z-x)^{(b-1)}}{\Gamma(b)} e^{-\lambda z} dx$$

$$= \frac{\lambda^{(a+b)}}{\Gamma(a)\Gamma(b)} e^{-\lambda z} \int_0^z x^{a-1} (z-x)^{b-1} dx \qquad \text{let } x = zt, \quad dx = z\,dt$$

$$= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} e^{-\lambda z} \int_0^1 (zt)^{a-1} z^{b-1} (1-t)^{b-1} z\,dt$$

$$= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} e^{-\lambda z} \int_0^1 z^{a+b-1} t^{a-1} (1-t)^{b-1} dt$$

$$\boxed{f_Z(z) = \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} e^{-\lambda z} z^{a+b-1} \int_0^1 t^{a-1} (1-t)^{b-1} dt}$$

both density functions

We know that $\int A f(z)\,dz = 1 = \int B f(z)\,dz \Rightarrow A = B$

So $$\boxed{\int_0^1 t^{a-1}(1-t)^{b-1}\,dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}}$$

and $\boxed{Z \sim \text{Gamma}(a+b, \lambda)}$

Note $$\boxed{\beta(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\,dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}}$$

Beta function (Euler)

Note Poisson Process (section 7.3)

T_1 T_2 T_3

$\text{Exp}(\lambda)$ $\Gamma(2,\lambda)$ $\Gamma(3,\lambda)$

So, $T_2 - T_1 \sim \text{Exp}(\lambda)$

because $T_2 - T_1 + \underset{\text{Exp}(\lambda)}{T_1} \sim \text{Gamma}(2,$

$\text{Exp}(\lambda) + \text{Exp}(\lambda) \longrightarrow \text{Gamma}(2,\lambda)$

Note

Suppose $Z \sim N(0,1)$, we know that $Z^2 \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right.$

then $Z_i \sim N(0,1)$ independent

then $\boxed{\sum_i^n Z_i^2 \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)}$

In fact, $\boxed{\sum_i^n Z_i^2 \sim \chi^2_n}$ → chi-square

Also

$$\boxed{\sqrt{Z_1^2 + Z_2^2 + \dots + Z_n^2} = \sqrt{\sum_i^n Z_i^2} \sim \chi_n} \longrightarrow \text{chi dist}$$

Note $$\boxed{f(x) = \frac{2^{1-n/2} x^{n-1}}{\Gamma\left(\frac{n}{2}\right)} e^{-x^2/2}} \longleftarrow \text{chi dist}$$

Exchangability

Note Equal in distribution

$$\tilde{X} = (X_1, X_2, \dots, X_n)$$

$$\tilde{Y} = (Y_1, Y_2, \dots, Y_n)$$

$\boxed{X \text{ \& } Y \text{ are equal in dist iff } P(X_1 \in B_1, \dots X_n \in B_n) = P(Y_1 \in B_1, \dots Y_n \in B_n)}$

no matter which set B_i you choose, get same values.

or $\boxed{F_{\vec{x}}(x_1, x_2, x_3, \dots) = F_{\vec{y}}(y_1, y_2, y_3, \dots)}$ (sufficient) for $\vec{X} \overset{d}{=} \vec{Y}$

Note $\boxed{\text{Permutation}}$ $k: \{1, 2, \dots, n\} \xrightarrow[\text{bijection}]{} \{1, 2, \dots n\}$

$$(k_1, k_2, \dots, k_n) = \mathrm{Im}(k)$$

o.k...

Def $(x_1, x_2, \ldots, x_n)$ is exchangeable if for any permutation k, we have

$$(X_{k_1}, X_{k_2}, \ldots, X_{k_n}) \overset{d}{=} (X_1, X_2, \ldots, X_n)$$

very few of these things...

It's easy to see that this is equivalent to saying that the pmf or pdf are invariant under permutation of the variables

Ex of Symmetric functions

- products $\prod_i^n x_i$
- Sums $\sum_i^n x_i^k$
- Constants k
- Products of two $\sum \prod_{i,j} x_i x_j = x_1 x_2 + x_1 x_3 + \ldots x_{n-1} \ldots$

Back to probability

Suppose $X_1 \ldots X_n$ all have the same prob. dist $(f(x))$, and that they are all independent. → iid

$$f_{\vec{x}}(x_1, x_2, \ldots x_n) = \prod_i^n f_{x_i} \quad \text{symmetric}$$

So: $\boxed{\text{iid}}$ → $\boxed{\text{iid R.V.'s are exchangeable}}$

Check that all exchangeable R.V.s are ~~iid~~ equal in distribution

$P(X_1 \le x, X_2 < +\infty, \ldots, X_n < \infty) = P(X_k \le x, X_1 \le \infty \ldots)$

$\Rightarrow P(X_1 \le x) = P(X_k \le x) \rightarrow \boxed{X_1 \overset{d}{=} X_k}$

Nov 30 2018

Exchangeability $\left\{ \begin{array}{l} X_1 \dots X_k \text{ iid} \Rightarrow \text{exchangeable} \\ \text{exchangeability} \Rightarrow \text{identically distributed} \end{array} \right\}$

(not iid) → not independence

Ex <u>Sampling without replacement</u>

n things $\{1, 2, \dots, n\}$ sample k times $X_1 \dots X_k$

$$p_{\tilde{X}}(x_1, \dots x_k) = P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{1}{n!/(n-k)!}$$

$$= \frac{(n-k)!}{n!} \text{, which is a constant}$$

So this is a symmetric function (independent of X)

So $X_1 \dots X_k$ are exchangeable. So X_i are identically dist.

Observation → if $X_1 \dots X_k$ exchangeable & $g: \mathbb{R} \mapsto \mathbb{R}$

then $g(X_1), \dots g(X_k)$ exchangeable.

Expectation & Variance

(a) $\boxed{\text{Expectation is linear } E(aX + bY) = aE(X) + bE(Y) \text{ true } \forall \text{ r.v. } X, Y}$

Ex If $S \sim Bin(n, p)$, $X_i \sim Ber(p)$

$$S = X_1 + \dots + X_n \rightarrow E(S) = np$$

Can use this whenever $X = I_1 + I_2 + \dots + I_n$

Ex $I_i = \begin{cases} 1 & p_i \\ 0 & (1 - p_i) \end{cases}$ $\boxed{E(X) = \sum_{i=1}^{n} p_i}$

(b) X, Y independent, $\boxed{E(XY) = E(X)E(Y)}$

(c) Variance

Theorem: if X, Y independent, then $Var(X+Y) = Var(X) + Var(Y)$

analogy: "Pythagorean theorem"

Proof
$E(X_1) = \mu_1$ $\quad Var(X_1) = \sigma_1^2$
$E(X_2) = \mu_2$ $\quad Var(X_2) = \sigma_2^2$

$$Var(X_1+X_2) = E\left((X_1+X_2-\mu_1-\mu_2)^2\right)$$
$$= E\left(\left((X_1-\mu_1)+(X_2-\mu_2)\right)^2\right)$$
$$= E\left((X_1-\mu_1)^2 + 2(X_1-\mu_1)(X_2-\mu_2) + (X_2-\mu_2)^2\right)$$
$$= E\left((X_1-\mu_1)^2\right) + E\left((X_2-\mu_2)^2\right) + E\left(2(X_1-\mu_1)(X_2-\mu_2)\right)$$

use indep $\longrightarrow$ $= Var(X_1) + Var(X_2) + 2\underbrace{E(X_1-\mu_1)}_{0}\underbrace{E(X_2-\mu_2)}_{0}$ if indep

$$Var(X_1+X_2) = Var(X_1) + Var(X_2)$$

Ex Negative Binomial

$X_i \sim Geom(p)$, independent.

$$X = X_1 + X_2 + \dots + X_k \,. \quad E(X) = kE(X_1) = \frac{k}{p}$$

$$Var(X) = k \cdot Var(X_i) = \frac{k(1-p)}{p^2}$$

Ex <u>Statistics</u>

$X_1, X_2, \dots X_n$ iid $\sim$ $E(X_i) = \mu$, $Var(X_i) = \sigma^2$

$\overline{X}_n = \frac{1}{n}\sum_i^n X_i$ is an r.v. <u>Want</u> $E(\overline{X}_n) = \mu$

" $\overline{X}_n$ is an unbiased estimator "

<u>Proof</u> $E(\overline{X}_n) = \frac{1}{n}(n\mu) = \mu$

$$Var(\overline{X}_n) = Var\left(\frac{1}{n}\sum X_i\right)$$

$$= \frac{1}{n^2} Var(X_1 + \dots + X_n)$$

$$= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$\boxed{\text{But how to estimate } \sigma?}$

$$E\left[(X_i - \overline{X}_n)^2\right] = E\left(\left[(X_i - \mu) - (\overline{X}_n - \mu)\right]^2\right)$$

$$= E\left[(X_i-\mu)^2 - 2(X_i-\mu)(\overline{X}_n-\mu) + (\overline{X}_n-\mu)^2\right]$$

$$= E\left[(X_i-\mu)^2\right] + E\left[(\overline{X}_n-\mu)^2\right] - 2E\left[(X_i-\mu)(\overline{X}_n-\mu)\right]$$

$$= \sigma^2 + \frac{\sigma^2}{n} - 2E\left[(X_i-\mu)(\overline{X}_n-\mu)\right]$$

<u>Add up</u> So $E\left(\sum_{i=1}^n (X_i - \overline{X}_n)^2\right) = E\left(\sigma^2_{\text{sample}}\right)$

$$= n\sigma^2 + \sigma^2 - 2E\left[(\overline{X}_n-\mu)\sum_{i=1}^n (X_i-\mu)\right]$$

$$= \sigma^2(n+1) - 2E\left[(\overline{X}_n-\mu)(n\overline{X}_n - n\mu)\right]$$

$$= \sigma^2(n+1) - n2E\left((\bar{X}_n-\mu)^2\right)$$

$$= \sigma^2(n+1) - \frac{2n\sigma^2}{n}$$

$$= (n-1)\sigma^2$$

So, the unbiased estimator for σ^2 is $\boxed{S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - }$ ← sample

$E(S_n^2) = \sigma^2$

Population $\boxed{\sigma_n^2 = \frac{1}{n}\sum (x_i-\mu)^2}$

Dec 3, 2018 Covariance!

Recall $Var(X+Y) = Var(X) + Var(Y) + 2E((X-\mu_x)(Y-\mu_y))$

If X & Y indep, then final term vanishes...

Def $\boxed{Cov(X,Y) = E((X-\mu_x)(Y-\mu_y)) = E(XY) - \mu_x\mu_y}$ → "like a dot product"

$= 0$ if X, Y indep.

"Proof" $E((X-\mu_x)(Y-\mu_y)) = E(XY - \mu_x Y - \mu_y X + \mu_x\mu_y)$

$$= E(XY) - \underbrace{E(\mu_x Y) - E(\mu_y X)} + \mu_y\mu_x$$

$$= E(XY) - 2\mu_x\mu_y + \mu_y\mu_x$$

$$= E(XY) - \mu_x\mu_y$$

Note $\boxed{Cov(X,X) = Var(X)}$

Observation

1. Sign of covariance is significant.
2. A, B, with indicator var I_A, I_B

$E(I_A) = p_A$, $E(I_B) = p_A$

$I_A I_B = I_{A \cap B}$

$$Cov(I_A, I_B) = E[I_{A\cap B}] - E(I_A)E(I_B)$$

$$= P(AB) - P(A)P(B)$$

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

$$Cov(I_A, I_B) = P(B)\left[P(A|B) - P(A)\right]$$

Properties

1. $Cov(X, Y) = Cov(Y, X)$
2. $Cov(X+b, Y) = Cov(X, Y)$
3. $Cov(a_1X_1 + a_2X_2, Y) = a_1 Cov(X_1, Y) + a_2 Cov(X_2, Y)$

1, 3 Cov is symmetric & bilinear. (true for dot product)

4. ~~Var(X+Y)~~

$$\boxed{\begin{aligned} Var\left(\sum_i^n X_i\right) &= \sum_i^n Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j) \\ &= \sum_{i,j} Cov(X_i, X_j) \end{aligned}}$$

Def X & Y are uncorrelated if $Cov(X,Y)=0$

Know if X, Y independent, then uncorrelated

$Cov(X,Y)=0 \not\Rightarrow$ independence

Property ⑤ $\boxed{[Cov(X,Y)]^2 \leq Var(X)\,Var(Y)}$ Cauchy-Schwarz inequality

So $\left| \dfrac{Cov(X,Y)}{\underbrace{\sqrt{Var(X)}\sqrt{Var(Y)}}} \right| \leq 1$ → like $a(\cos\theta)$

$\rho =$ "Correlation coefficient" of X & Y

So $\rho = 0 \to X$ & Y uncorrelated
$\rho = \pm 1 \Rightarrow Y = aX + b$

Note

$$\begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(Y,X) & Var(Y) \end{pmatrix}$$

Dec 5, 2018

(Observation) ① W - r.v. and $P(W \geq 0) = 1$, then $E(W) \geq 0$

② If X r.v. $Var(X) \geq 0$ (if defined)

Theorem $\boxed{Cov(X,Y)^2 \leq Var(X)\,Var(Y)}$

Proof Choose $t \in \mathbb{R}$ $Cov(tX+Y, tX+Y) \geq 0$

$$\text{Cov}(tX+Y,\ tX+Y) = \text{Var}(X)t^2 + 2\text{Cov}(X,Y)t + \text{Var}(Y) \geq 0$$

So $\quad 4\,\text{Cov}(X,Y)^2 - 4\,\text{Var}(X)\,\text{Var}(Y) \geq 0$

or $\quad \text{Cov}(X,Y)^2 \geq \text{Var}(X)\,\text{Var}(Y) \quad$ (proved)

$\boxed{\text{Corollary}}$

$$-1 \leq \underbrace{\frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}} \leq 1$$

$\text{Corr}(X,Y) = \rho =$ "correlation coeff. of X, Y"

$\boxed{\text{Consequence}}$ $\quad \boxed{\text{Corr}(aX+b,\ Y) = \text{sgn}(a) \cdot \text{Corr}(X,Y)}$

$\boxed{\text{Neat Little Trick}}$ $\quad$ Let $\tilde{X} = \dfrac{X-\mu_X}{\sigma_X}, \quad \tilde{Y} = \dfrac{Y-\mu_Y}{\sigma_Y}$

So
$$\left.\begin{array}{l} E(\tilde{X}) = 0 = E(\tilde{Y}) \\ \text{Var}(\tilde{X}) = 1 = \text{Var}(\tilde{Y}) \end{array}\right\} \Rightarrow E(\tilde{X}^2) = 1 = E(\tilde{Y}^2)$$

So
$$\text{Cov}(\tilde{X},\tilde{Y}) = E(\tilde{X}\tilde{Y})$$
$$= E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right)\right]$$
$$= \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} = \rho_{X,Y}$$

So $\quad \boxed{\text{Cov}(\tilde{X},\tilde{Y}) = \rho_{X,Y}}$

What happens when $\rho = 1$ or -1 ?

$\boxed{\text{Theorem}}$ $\quad \rho_{X,Y} = 1 \Rightarrow Y = aX + b$ with $a > 0$ with probability 1

<u>Proof</u> $0 \le E\left((\tilde{X}-\tilde{Y})^2\right) = E(\tilde{X}^2) - 2E(\tilde{X}\tilde{Y}) + E(\tilde{Y}^2)$

$$= 2 - 2\rho_{X,Y} = 2 - 2(1) = 0$$

So if $\rho_{X,Y} = 1 \Rightarrow E\left((\tilde{X}-\tilde{Y})^2\right) = 0$

with $E(\tilde{X}-\tilde{Y}) = E(\tilde{X}) - E(\tilde{Y}) = 0$

$\Rightarrow Var(\tilde{X}-\tilde{Y}) = 0 \Rightarrow \tilde{X}-\tilde{Y} = 0$ with probability 1

So $\frac{X-\mu_X}{\sigma_X} - \frac{Y-\mu_Y}{\sigma_Y} = 0$ with probability 1

So
$$\boxed{\begin{aligned} Y = \frac{\sigma_Y}{\sigma_X}(X-\mu_X) + \mu_Y &= \frac{\sigma_Y}{\sigma_X}X + \mu_Y - \frac{\sigma_Y}{\sigma_X}\mu_X \\ &= aX + b \end{aligned}}$$

$\hookrightarrow > 0$

<u>Proof</u>
If $\rho_{X,Y} = -1$, look at $E\left[(\tilde{X}+\tilde{Y})^2\right] \ldots$ $Y = aX + b$, $a < 0$

If $\rho_{X,Y} = 0$, X & Y are not correlated, but not necessarily independent

STANDARD BIVARIATE NORMAL

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right]$$

$\begin{cases} X \sim N(0,1) \\ Y \sim N(0,1) \end{cases}$

If $\rho = 0$ then X & Y independent.

If X, Y independent then $\rho = 0$

$$\rightarrow \begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(Y,X) & Var(Y) \end{bmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = S, \quad S^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$$

$\hookrightarrow$ covariance matrix

so $S^{-1} = \begin{pmatrix} 1/1-\rho^2 & -\rho/1-\rho^2 \\ -\rho/1-\rho^2 & 1/1-\rho^2 \end{pmatrix}$

if $(x\ y)\, S^{-1} \begin{pmatrix} x \\ y \end{pmatrix} = (x^2 + y^2 - 2\rho x y) / (1-\rho^2)$

so the bivariate normal can be written like

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[- \frac{x^2+y^2-2\rho xy}{2(1-\rho^2)} \right]$$

$$f_{X,Y}(x,y) = \frac{1}{\sqrt[2]{2\pi}^2 \sqrt{\det S}} e^{-\vec{x}^T S^{-1} \vec{x}/2}$$

① $\boxed{\text{MARKOV'S INEQUALITY}}$

$X \geq 0$ r.v. finite μ_X, then $P(X \geq c) \leq \mu_X / c$ ($c \geq 0$)

Proof $I = I(x \geq c)$ (1 if $x \geq c$, 0 if not)

then $X \geq X \cdot I \geq c \cdot I$ but since $X \geq c$

$\rightarrow E(X) \geq c\, E(I) = c\, P(X \geq c)$

so $\boxed{P(X \geq c) \leq \frac{\mu_X}{c}}$ $E(X)$

② $\boxed{\text{CHEBYSHEV'S INEQUALITY}}$ X. r.v. ($\rightarrow$ no assumption) Assume μ_X, $Var(X)$ finite.

For any $c > 0$, $P(|X - \mu_X| \geq c) \leq \frac{Var(X)}{c^2}$ ($|X-\mu_X|, c > 0$)

Proof $\boxed{|X - \mu_X| \geq c \Leftrightarrow |X - \mu_X|^2 \geq c^2}$. By Markov, $P(|X-\mu_X|^2 \geq c^2)$

By Markov, $\underbrace{P(|X-\mu|^2 \geq c^2)}_{P(|X-\mu_X| \geq c)} \leq \frac{E((X-\mu_X)^2)}{c^2} = \frac{\sigma_X^2}{c^2}$

$P(|X-\mu_X| \geq c) \leq \frac{Var(X)}{c^2}$

③ Weak law of large numbers (bounded)

Let $X_1, X_2, \dots$ iid with μ, σ^2 finite.

$$\lim_{n\to\infty} P\left(\left|\frac{\sum^n X_i}{n} - \mu\right| < \varepsilon\right) = 1 \qquad \forall \varepsilon > 0$$

Proof
$$E\left[\frac{X_1+X_2+\dots}{n}\right] = \mu \quad \text{and} \quad Var\left(\frac{X_1+X_2+\dots}{n}\right) = \frac{\sigma^2}{n}$$

By Chebyshev, $P\left(\left|\frac{\sum^n X_i}{n} - \mu\right| \geq \varepsilon\right) \leq \left(\frac{\sigma^2}{n}\right)\frac{1}{\varepsilon^2} \to 0$ as $n \to \infty$

So $\lim_{n\to\infty} P\left(\left|\frac{\sum^n X_i}{n} - \mu\right| \geq \varepsilon\right) \Rightarrow 0$

So $\lim_{n\to\infty} P\left(\left|\frac{\sum^n X_i}{n} - \mu\right| < \varepsilon\right) = 1 \quad \forall \varepsilon > 0$

④ STRONG LAW OF LARGE NUMBERS (equality)

$X_1\ X_2 \dots$ iid r.v. with finite μ

$$P\left(\lim_{n\to\infty}\left(\frac{X_1 + X \dots X_n}{n}\right) = \mu\right) = 1$$

Proof in "finer points", but by assuming $E(X^4) < \infty$ (ex 7b)

Dec 7, 2018 **CENTRAL LIMIT THEOREM** → universality theorem

$X_1, X_2 \dots$ iid r.v. with finite mean μ, finite variance σ^2 then

$$\lim_{n\to\infty} P\left(\frac{(X_1+X_2+\dots+X_n) - n\mu}{\sqrt{\sigma^2 n}} \leq x\right) = \Phi(x) = P(Z \leq x)$$

$Z \sim N(0,1)$

Equivalently, $\lim_{n\to\infty} P\left(\dfrac{\frac{X_1+X_2+\dots X_n}{n}-\mu}{\sqrt{\sigma^2/n}} \le x\right) = \Phi(x)$

Generalized version $X_1, X_2, \dots$ independent (no need id), and $\mu_i = E(X_i)$, $\sigma_i^2 = Var(X_i)$ finite + a few mild assumptions ...

then $\lim_{n\to\infty} P\left(\dfrac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \le x\right) = \Phi(x)$

Proof? → use generating functions: Sketch: Assume $M_{X_i}(t)$ all exist on $(-\varepsilon, \varepsilon)$. $\boxed{\text{Key Lemma}}$ suppose X, Y_i are r.v. such that all M_X, M_Y are defined on $(-\varepsilon,\varepsilon)$, and

If $\lim_{n\to\infty} M_{Y_n}(t) = M_X(t)$ for $t \in (-\varepsilon,\varepsilon)$

Then $Y_n \xrightarrow{d} X$, i.e., $\lim_{n\to\infty} P(Y_n \le x) = P(X \le x)$

...ory CLT

...oivre

↓

...ace {CLT

...nov (good proof of CLT)

Sequence $X_1\ X_2 \dots$ iid. Let $\tilde{X}_i = \dfrac{X_i-\mu_i}{\sigma_i} = \dfrac{X_i-\mu}{\sigma}$ $Var(\tilde{X}_i)=1$

Let $Y_n = (\tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_n)\left(1/\sqrt{n}\right)$

$= \dfrac{\tilde{X}_1}{\sqrt{n}} + \dfrac{\tilde{X}_2}{\sqrt{n}} + \dots + \dfrac{\tilde{X}_n}{\sqrt{n}}$ iid

$M_{Y_n} = M_{\tilde{X}_1/\sqrt{n}}(t)\, M_{\tilde{X}_2/\sqrt{n}}(t) \dots = \prod_{i=1}^n M_{\tilde{X}_i/\sqrt{n}}(t) = \left[M_{\tilde{X}/\sqrt{n}}(t)\right]^n$

$= \left(1 + \underbrace{E\left(\frac{\tilde{X}}{\sqrt{n}}\right)}_{0} t + E\left[\left(\frac{\tilde{X}}{\sqrt{n}}\right)^2\right] t^2/2 + \dots\right)^n$

$= \left(1 + \dfrac{t^2}{2n} + \dots\right)^n \approx \left(1+\dfrac{t^2}{2n}\right)^n$ $= M_Z(t)$

so $\boxed{\lim_{n\to\infty} M_{Y_n}(t) = \lim_{n\to\infty}\left(1 + t^2/2n\right)^n \to e^{t^2/2} = M \text{ of } N(0,1)}$

Probability Exam 2

HUAN BUI

NOV 12, 2018

- $E(XY) = E(X)E(Y)$
- F cdf. $P(X<a) = \lim_{x\to a^-} F(x)$

requires converge $\sum |k| P(X=k) < \infty$, $\int_{-\infty}^{\infty} |x| f(x) dx$

$E(X) = \sum_k k P(X=k)$ or $\int_{-\infty}^{\infty} x f(x) dx$

Theory

- If A & B disjoint $\Rightarrow A \cap B = \emptyset$. Converse not true $\Rightarrow P(A\cup B) = P(A) + P(B)$
- If $B_1 \subset B_2 \subset \dots \subset B_n$, then $P(\bigcup^\infty B_i) = P(B) = \lim_{n\to\infty} P(B_n)$
- If $A \subset B \Rightarrow P(A) \le P(B)$
- In-Ex $P(A_1 \cup A_2 \dots \cup A_n) = \sum_k (-1)^{k+1} \sum_{i_1 < \dots < i_k} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k})$
- $\bigcap_{i=1}^n A_i^c = (\bigcup A_i)^c$

$$\boxed{M_Z(t) = \sum_n^\infty \frac{M^{(n)}(0)}{n!} t^n = \sum_n^\infty \frac{t^n}{n!} E(X^n)}$$

- R.V. A r.v. X is f_n $\Omega \mapsto \mathbb{R}$. X degenerate if $P(X=b) = 1$ for some $b \in \mathbb{R}$
- X discrete if $P(X\in K) = 1 = \sum_k^\infty P(X=k)$, $K = \{k_1 \dots k_n\}$
- Bayes $P(B_j|A) = \frac{P(AB_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(A|B_i)P(B_i)}$
- Ind $\Leftrightarrow P(A)P(B) = P(AB)$. Theorem $\boxed{A_1 \dots A_n \text{ ind} \Rightarrow A_1^*, A_2^* \dots \text{ desire } * = c \text{ or not, are ind}}$
- cdf props (1) $F_X > 0$

(2) $\lim_{x\to\infty} F_X = 1$, (3) $\lim_{x\to-\infty} F_X = 0$, (4) $\lim_{x\to a^+} F_X = F(a)$

F^* s.t. $\lim_{x\to-\infty} F = 0$, $\lim_{x\to\infty} F = 1$, non decreasing f_n

- Cont r.v. $\exists f(x)$ s.t. $F(x) = \int_{-\infty}^x f(\tilde{x}) d\tilde{x} = P(X \le x)$, $f(x) \equiv$ pdf
- f_X props: (1) $f_X \ge 0$ (2) $\int_{-\infty}^\infty f_X dx = 1$ (3) piecewise continuous

$E(g(X)) = \int_\Omega g(x) = \begin{cases} \sum g(x) P(X=x) & \text{discrete} \\ \int_{-\infty}^\infty g(x) f_X(x) dx & \text{cont.} \end{cases}$

- Moments n^{th} moment $\to E(X^n)$. If $E(X^n) \exists$ for n, then $E(X^l) \exists$ for $l \le n$
- Def Quantile: x quantile if $P(X \le x) \ge p$ & $P(X \ge x) \ge 1-p$ $(0 \le p \le 1)$
- $Var(X) = \int (x-\mu)^2 f_X dx = \sum (x-\mu)^2 P(X=x) = E(X^2) - E(X)^2$
- $E(aX+b) = aE(X) + b$, $Var(aX+b) = a^2 Var(X)$
- $Var(X) = 0 \Leftrightarrow P(X=a) = 1$ for some $a \Leftrightarrow$ X degenerate r.v. (X r.v.)

"finitely many"

- $X \sim \mathcal{N}(\mu, \sigma^2)$ if $\varphi(x) = 1/\sqrt{2\pi\sigma^2} \exp[-(x-\mu)^2/2\sigma^2]$. $\Phi(x) = \int_{-\infty}^x \varphi(x) dx$. $\Phi(-x) = 1 - \Phi$
- $\mu, a, b \in \mathbb{R}$, $a \ne 0$ if $X \sim \mathcal{N}(\mu, \sigma^2)$ & $Y = aX + b$, then $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$
- CLT $0 < p < 1$, p fixed, $S_n \sim Bin(n,p)$. $-\infty \le a \le b \le \infty$,

$$\lim_{n\to\infty} P\left(a \le (S_n - np)/\sqrt{np(1-p)} \le b\right) = \int_a^b 1/\sqrt{2\pi}\, e^{-x^2/2} dx = \Phi(b) - \Phi(a)$$

iid, μ, $\boxed{\text{finite}}$ ← Approx n large, p not small $\Big\}$ $P\left(a \le \frac{S_n - np}{\sqrt{npq}} \le b\right) = \Phi(b) - \Phi(a)$ $(npq \ge 10)$

Continuity correction $P(k_1 \le S_n \le k_2) = P(k_1 - 1/2 \le S_n \le k_2 + 1/2)$

Stirling $n! \sim n^n e^{-n} \sqrt{2\pi n}$. Chebyshev ineq $P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) \le \frac{Var(X)}{\varepsilon^2}$

$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| \ge \varepsilon\right) = 0 \Leftrightarrow$ WLLN $\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) = 1$ $(\varepsilon > 0)$. Theorem $\boxed{\left|P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \le x\right) - \Phi(x)\right| \le \frac{3}{\sqrt{npq}}}$

$\le$ or $<$ ok. $<$ or $\le$ o.k.

discr. S_n/n

if $S_n = \sum_{i=1}^n X_i$ indep, identically dist (iid) $E(X_i) = \mu$, $Var(X_i) = \sigma^2 < \infty$

(ME)

$\boxed{\text{C.I.}}$ $P(|\hat{p} - p| < \varepsilon) \cong 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \ge 2\Phi(2\varepsilon\sqrt{n}) - 1 \ge p' = \text{C.I.}\%$

np^2 small → poisson

- Poisson $X \sim Poisson(\lambda)$ $(\lambda > 0)$ if $X \in \mathbb{Z}^+$, $P(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}$

$\lim_{n\to\infty} P(S_n = k) = \frac{e^{-\lambda}\lambda^k}{k!}$ if $\lambda > 0$, $\frac{\lambda}{n} < 1$, $S_n \sim Bin(n, \frac{\lambda}{n})$

There can be r.v. that's neither discrete nor cont.

X, Y can be at different sample spaces... Same pdf/pmf ⟹ Eq in dist ⟹ $E(g(X)) = E(g(Y))$

- $\boxed{\text{Only exp has the memoryless property}}$ (cont case only) | $\Gamma(x)\Gamma(1-x) = \pi/\sin(\pi x)$

There's no "the" uniform dis cont. $0<x<1$

- Markov's ineq $P(X \geq a) \leq \frac{E(X)}{a}$ (proof by int)

- Exp $P(X>t) = e^{-\lambda t}$ | $P(X\leq t) = 1-e^{-\lambda t}$

$(n-1)! = \Gamma(n)$ if n integer. Exp is when $n=0$

- Theorem $X \sim Bin(n,p)$, $Y \sim Poisson(np)$. $A \subseteq \{0,1,2 \dots\} \Rightarrow |P(X\in A) - P(Y \in A)| \leq np^2$ (proof w/ $N(0,1)$)
- $0<\lambda<\infty$. $X \sim Exp(\lambda)$ if $f(x) = (0$ if $x<0)$, $= \lambda e^{-\lambda x}$ if $x \geq 0$ (λ = rate)
- Memoryless if $X \sim Exp(\lambda)$ then $P(X>t+s \mid X>s) = P(X>t)$ $\quad \boxed{\Gamma(1/2) = \sqrt{\pi}}$
- $\lambda>0$, $\frac{\lambda}{n}<1$ if $nT_n \sim Geom(\lambda/n)$, n large $\Rightarrow \lim_{n\to\infty} P(T_n>t) = e^{-\lambda t}$ $\forall t \geq 0$ $\quad \boxed{\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin(\pi x)}}$
- Poisson process = collection of rnd pts on $[0,\infty)$ where (i) points are distinct, (ii) # pts $\in I = N[I]$ then $N[I] \sim Poisson(\lambda(b-a))$ $(I=[a,b])$, (iii) if I_i non overlapping then $N[I_i]$ mutually indep.
- $r,\lambda>0$. $X \sim Gamma(r,\lambda)$ if $x>0$ & $f_X = \frac{\lambda^r x^{r-1}}{\Gamma(r)}$ $x \geq 0$
& $f_X = 0$ if $x<0$. $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx$ $(r>0)$ → $\Gamma(r)$ infinitely diff.
- MGF of r.v. X is $M(t) \stackrel{\circ}{=} E(e^{tX})$. $t \in \mathbb{R}$. $\boxed{M^{(n)}(0) = E[X^n]}$ ($M(t)$ finite in $(-\delta,\delta)$)
- Theorem $\boxed{X, Y \text{ are r.v. } M_X(t) = E(e^{tX}) \text{ \& } M_Y(t) = E(e^{tY}). \text{ If } \exists\, \delta \text{ s.t. } t \in (-\delta,\delta), M_X(t) = M_Y(t) \text{ and } M \text{ finite} \Rightarrow P(X\leq x) = P(Y \leq x) \text{ (equal in dist)}}$

Note $E[X^n] = E[Y^n] \not\Rightarrow$ equality in dist

	Par	pmf / pdf	E(X)	Var(X)	M(t)
Ber(p)	$0\leq p\leq 1$	$P(X=0) = 1-p$, $P(X=1) = p$	p	pq	$1-p+pe^t$
Bin(n,p)	$0 \leq p \leq 1$, $n \geq 1$	$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$, $0 \leq k \leq n$	np	npq	$(1-p+pe^t)^n$
Geom(p)	$0\leq p \leq 1$	$P(X=k) = p(1-p)^{k-1}$	$1/p$	$\frac{1-p}{p^2}$	$pe^t/(1-p+pe^t)$
Poisson(λ)	$\lambda>0$	$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$, $k=0,1,2\dots$	λ	λ	$e^{\lambda(e^t-1)}$
Hypergeom(N, N_A, n)		$P(X=k) = \frac{\binom{N_A}{k}\binom{N-N_A}{n-k}}{\binom{N}{n}}$	$\frac{nN_A}{N}$	$\frac{N-n}{N-1} n \frac{N_A(N-N_A)}{N^2}$	?
Unif[a,b]	$a<b$	$f_X(t) = \frac{1}{b-a}$, $t\in[a,b]$	$\frac{a+b}{2}$	$\frac{1}{12}(a-b)^2$	$(e^{tb}-e^{ta})/t(b-a)$
$N(\mu,\sigma^2)$	$\sigma^2>0$	$f_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{t\mu}e^{t^2\sigma^2/2}$
Exp(λ)	$\lambda>0$	$f_X(t) = \lambda e^{-\lambda t}$	$1/\lambda$	$1/\lambda^2$	$\left(\frac{\lambda}{\lambda-t}\right)$ $t<\lambda$, ∞ $t\geq\lambda$
Gamma(r,λ)	$\boxed{r\geq 1, \lambda>0}$	$f_X(t) = \frac{\lambda^r x^{r-1}}{\Gamma(r)} e^{-\lambda t}$ $(t\geq 0)$	r/λ	r/λ^2	$\left(\frac{\lambda}{\lambda-t}\right)^r$ $t<\lambda$, ∞ $t \geq \lambda$

Wolfram: CDF[Dist[...], x] or InverseCDF[Dist[..], x ...]

$M'(t) = \frac{d}{dt} E(e^{tX}) = E\left[\frac{d}{dt} e^{tX}\right] = E[Xe^{tX}]$

$\boxed{M^{(n)}(t) = E[X^n e^{tX}]}$

$\boxed{Z^2 \sim Gamma(1/2, 1/2)}$

Notes Model T until 1st event, 2nd event (Poisson process)

$P(T_1>t) = P(N[0,t]=0) = e^{-\lambda t} \to P(T_1 \leq t) = 1-e^{-\lambda t} \to T_1 \sim Exp(\lambda)$

$\boxed{\text{Sum of Exp r.v}}$ $P(T_2>t) = P(N[0,t] \leq 1) = e^{-\lambda t}(1+\lambda e^{-\lambda t})$, $P(T_2 \leq t) = \dots$

$(P(T_1 \leq s) = P(N[0,s] \leq 1))$ → diff. to get pdf of T_2 $\boxed{\lambda^2 t e^{-\lambda t}}$

→ n^{th} call ⟹ $\boxed{f_{T_n}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t}}$ → $\boxed{T_n \sim Gamma(n,\lambda) \text{ (n integer)}}$

$n \neq$ integer $f_T(t) = \frac{\lambda^n t^{n-1}}{\Gamma(n)} e^{-\lambda t}$ ↳ Discrete