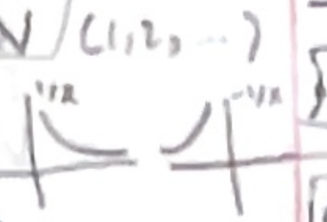
Q^c | Q — Z $(-2,-1,0,1,2)$ — N $(1,2,\dots)$



Recall

- a set can be (in)finite
→ infinite set is (un)countable
→ finite set is finite or countably infinite

$\boxed{\mathbb{R} = Q \cup Q^c}$

Indicator Fx: $\mathbb{1}(A) = \begin{cases} 1, & A \text{ true} \\ 0, & \text{otherwise} \end{cases}$

$\boxed{A \subseteq B \Rightarrow P(A \cap B) = P(A)}$

Geometric series

$\boxed{\forall r \neq 1 \in \mathbb{R}, \sum_{k=0}^{n-1} r^k = \frac{1-r^n}{1-r}}$

$\boxed{\forall |r|<1 \Leftrightarrow r \in (-1,1), \sum_{k=0}^{\infty} r^k = \frac{1}{1-r}}$ otherwise diverge

$\sum_{i=1}^{\infty}(1-p)^{i-1}p = p\sum_{k=0}^{x-1}(1-p)^k$, $k = i-1$ (change-of-var)

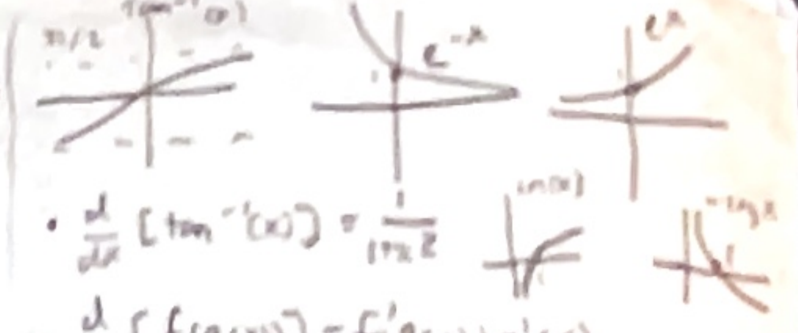
$S: \mathcal{B} := (-\infty, a], a \in Q$

$(-\infty, b] = \cap_{n=1}^{\infty} (-\infty, a_n] \in \mathcal{B}$, $a_n \downarrow b$

$(-\infty, a) = \bigcup_{q<a, q \in Q} (-\infty, q] \in \mathcal{B}$

$\{a\} = (-\infty, a] \setminus (-\infty, a)$
$= (-\infty, a] \cap ((-\infty, a))^c \in \mathcal{B}$

$(a,b] = (-\infty, b] \setminus (-\infty, a)$
$= (-\infty, b] \cap ((-\infty, a))^c \in \mathcal{B}$



- $\frac{d}{dx}[\tan^{-1}(x)] = \frac{1}{1+x^2}$
- $\frac{d}{dx}[f(g(x))] = f'(g(x))g'(x)$
- $e^A e^B = e^{A+B}$
- $(e^A)^B = e^{AB}$

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

1.1: Set Theory

sample space (S): the set of all possible outcomes of an experiment
→ can be (un)countable; dictates prob. assignments
countable: $s \in S$ can be put into a 1-1 correspondence w/ subset of integers

Event: any collection of possible outcomes of an experiment $\boxed{A \subset B \Leftrightarrow x \in A \Rightarrow x \in B}$
→ any subset of S, including S, i.e. $E \subseteq S$. $\boxed{A = B \Leftrightarrow A \subset B \text{ and } B \subset A}$

Elementary set ops.
union, intersection, complementation

add A_i in practice

$\boxed{A \setminus B = \{x \mid x \in A, x \notin B\}}$ $A \cap B^c$

$\boxed{A \cup B = \{x : x \in A \text{ or } x \in B\}}$

Disjoint (mutually exclusive): $= \boxed{A \cap B = \emptyset}$ Pairwise Disjoint $:= \boxed{A_i \cap A_j = \emptyset \ \forall i \neq j}$

multiply A_i in practice — $\boxed{A \cap B = \{x : x \in A \text{ and } x \in B\}}$

Partition (of S): $A_1, A_2, \dots$ pairwise disjoint and $\bigcup_{i=1}^{\infty} A_i = S$.

$\boxed{A^c = \{x : x \notin A\}}$ $\emptyset$ - empty set

Properties

Commutativity	$A \cup B = B \cup A$ $A \cap B = B \cap A$
Associativity	$A \cup (B \cup C) = (A \cup B) \cup C$ $A \cap (B \cap C) = (A \cap B) \cap C$
Distributivity	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
DeMorgan's (law)	$(A \cup B)^c = A^c \cap B^c$ $(A \cap B)^c = A^c \cup B^c$

$\boxed{\bigcup_{i=1}^{\infty} A_i = \{x : x \in A_i \text{ for some } i < \infty\}}$

countable union (existence: ∃) w at least one A_i: $= \{x \in S : x \in A_i \text{ for some } i\}$ $= \lim_{n\to\infty} \bigcup_{i=1}^{n} A_i$

$\boxed{\bigcap_{i=1}^{\infty} A_i = \{x : x \in A_i \text{ for all } i = 1,2,\dots\}}$

countable intersection (∀ - for all): $= \{x \in S : x \in A_i \text{ for all } i\}$ $= \lim_{n\to\infty} \bigcap_{i=1}^{n} A_i$

$\boxed{(\bigcup_{i=1}^{\infty} A_i)^c = \bigcap_{i=1}^{\infty} A_i^c}$

1.2: Probability Theory Basics | 1.2.1: Axiomatic Foundations

Def: a collection on subsets of S is a sigma algebra (Borel field), $\mathcal{B}$ if it satisfies:

i) $\boxed{\emptyset \in \mathcal{B}}$ ii) $\boxed{A \in \mathcal{B} \Rightarrow A^c \in \mathcal{B}}$ iii) $\boxed{A_1, A_2, \dots \in \mathcal{B} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{B}; \bigcap_{i=1}^{\infty} A_i \in \mathcal{B}}$

- $S \in \mathcal{B}$; $\{\emptyset, S\}$:= trivial sigma algebra

Q: How to show that sets are $\in \mathcal{B}$?
A: verify the set in question can be obtained from known elements of $\mathcal{B}$ via countably many set operations
note: use fact that any real number ($\mathbb{R}$) can be constructed as the limit of a sequence of rational #'s (Q)

Prob. Function satisfies AoP

$\boxed{1) P(A) \geq 0 \ \forall A \in \mathcal{B}}$ $\boxed{2) P(S) = 1}$
$\boxed{3) A_i \cap A_j = \emptyset \ \forall i \neq j}$ $\boxed{\Rightarrow P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} A_i}$

Assume finite additivity: $A, B \in \mathcal{B}$
$\boxed{A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)}$

Thm. $\forall A \in \mathcal{B}$, $\boxed{P(A) := \sum_{\{i : s_i \in A\}} p_i}$
$S = \{s_1, \dots, s_n\}$ - finite
or $S = \{s_1, s_2, \dots\}$ - countable
- physical reality (context) of the experiment motivates prob. assignment

1.2.2: Calculus of Prob.'s

use AoP to build properties $A, B \in \mathcal{B}$, P: prob function

$\boxed{P(\emptyset) = 0}$ $\boxed{P(A) \leq 1}$ $\boxed{P(A^c) = 1 - P(A)}$ $\boxed{P(A \cap B) \geq P(A) + P(B) - 1}$

$\boxed{P(B \cap A^c) = P(B) - P(A \cap B)}$ Bonferroni's inequality: bound simultaneous event (∩) in terms of individual events w.r.t. prob; only useful if ind. prob. events are sufficiently large

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

$\boxed{A \subset B \Rightarrow P(A) \leq P(B)}$

∀ partition $C_1, C_2, \dots$
$\boxed{P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)}$

(★) $\boxed{P(\bigcap_{i=1}^{\infty} A_i) \geq \sum_{i=1}^{\infty} P(A_i) - (n-1)}$

$\boxed{P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)}$ (#)

(#) Boole's inequality is a more general version of (★) - Bonferroni's

1.2.3: Counting

FTOC: k separate tasks, i^{th} of which done in n_i ways; $i = 1, \dots, k$ then the entire job can be done in $\overline{n_1 \times \dots \times n_k} = n^k$ ways

$\underline{n!} := n \times (n-1) \times (n-2) \times \dots \times 3 \times 2 \times 1$, $0! = 1$
- # distinct ways the original n objects can be shuffled

Number of possible arrangements of size r from n objects

	w/o replacement	w/ replacement
ordered	$P_r^n = \frac{n!}{(n-r)!}$	n^r (FTOC)
unordered	$C_r^n = \frac{n!}{r!(n-r)!}$	$\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!}$

$P_r^n = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$
- ordered arrangement of r distinct objects out of n total

$C_r^n = \binom{n}{r} = \frac{P_r^n}{r!} = \frac{n!}{r!(n-r)!}$
- # unordered combos of n objects chosen r at a time

Notes: - $\binom{n}{r}$ divides by redundant orderings
- $C_r^n \leq P_r^n$ since $r! \geq 1$

1.2.4: Enumerating options

use (1.2.3) techniques for finite S and if all outcomes equally likely
↳ then P(events) found via counting # outcomes in the event (classical model)

Prop. $S = \{s_1, \dots, s_n\}$. $P(\{s_i\}) = \frac{1}{N} \ \forall s_i \in S$.
$\Rightarrow \boxed{P(A) = \sum_{s_i \in A} P(\{s_i\}) = \sum_{s_i \in A} \frac{1}{N} = \frac{\# A \text{ elements}}{\# N \text{ elements}}}$

Inclusion/Exclusion: For n events $A_1, A_2, \dots, A_n$ (★)
$\boxed{P(\bigcup_{i=1}^{n} A_i) = \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) \right)}$

$\boxed{P(\text{at least one } A_i) = P(\bigcup_{i=1}^{n} A_i) = 1 - P(\bigcap_{i=1}^{n} A_i^c)}$

1.3: Conditional Prob & Independence

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

(⇒) Joint = marginal × cond. ⇒ $\boxed{P(A \cap B) = P(B)P(A|B)}$

$\boxed{P(A_1 \cap \dots \cap A_n) = \prod_{i=1}^{n} P(A_i | A_1 \cap \dots \cap A_{i-1})}$

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

Bayes' Formula $\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$

ind $A \perp\!\!\!\perp B$: $\boxed{\text{Ind} := P(A \cap B) = P(A)P(B)}$

$\boxed{A \perp\!\!\!\perp B \Leftrightarrow A \perp\!\!\!\perp B^c, A^c \perp\!\!\!\perp B, A^c \perp\!\!\!\perp B^c}$

Bayes' Thm. $\boxed{S = \bigcup_{i=1}^{\infty} A_i, \ P(A_i | B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)} (*)}$

turning around condit. prob's; (*) Law of Total Prob

$\boxed{\text{mutual Ind}: P(A_1 \cap \dots \cap A_k) = \prod_{i=1}^{k} P(A_i)}$

- mutual ind ⇔ ind

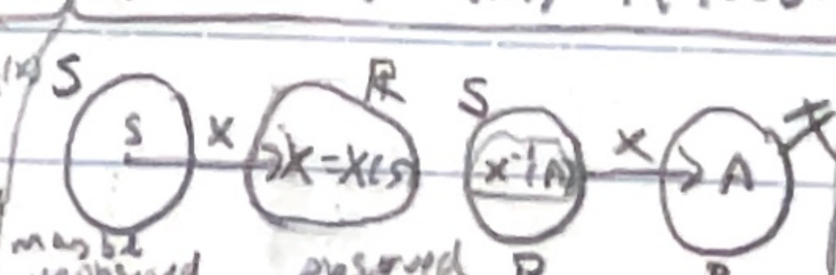
$\boxed{\forall i \ P(A_i|B) \propto P(B|A_i)P(A_i)}$
$P(A_i|B)$: Bayesian Posterior Dist.; $P(B|A_i)$: obs data likelihood; $P(A_i)$: Prior Dist.

1.4: Random variables

RV: function mapping from $X: S \to \mathbb{R}$
- easier to deal w/ summary var than original prob. structure

$S = \{s_1, \dots, s_n\}$, $\mathcal{X} = \{x_1, \dots, x_m\}$, $\boxed{P_X(X = x_i) = P(\{s_j \in S : X(s_j) = x_i\})}$

$\boxed{P_X(X \in A) = P(X^{-1}(A)) = P(\{s \in S : X(s) \in A\})}$ $\mathcal{X}$: uncountable range



$\boxed{f_X(x) = P(X = x) \ \forall x}$

$\boxed{F_X(x) = \int_{-\infty}^{x} f_X(t)dt \ \forall x}$

CDF Properties (a) $\lim_{x\to -\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$ (b) nondecreasing function of x: $F(x) \geq 0$ increasing
(c) right continuous, i.e. $\lim_{x \downarrow x_0} F(x) = F(x_0)$; $\lim_{\epsilon \downarrow 0} F_X(x + \epsilon) = F_X(x)$

- can construct new RV from known constituent RV's

1.5: Distribution Functions

CDF: $= F_X(x) = P(X \leq x) \ \forall x \in \mathcal{X}$
cont RV ⇔ continuous F_X
Discrete RV ⇔ Discrete F_X

$\boxed{X \stackrel{d}{=} Y \Leftrightarrow F_X(a) = F_Y(a) \ \forall a, b \in \mathbb{R}}$

$\boxed{X \stackrel{d}{=} Y := P(X \in A) = P(X \in B) \ \forall A, B \in \mathcal{B}^1}$

$\mathcal{B}^1$: smallest σ-algebra containing $(a,b), [a,b), (a,b], [a,b]$

1.6: Density & Mass Functions

- $X \sim F_X$ or $X \sim f_X$ contain all info abt dist. of RV X

$\boxed{f_X(x) = \frac{d}{dx} F_X(x)}$

$\boxed{P(X \leq x) = F_X(x)}$ $\boxed{X\text{-cont.}: P(a < X < b) = \dots = P(a \leq X \leq b)}$

$\boxed{X\text{-cont.} \Rightarrow P(X = x) = 0}$ PDF/PMF satisfies

$\boxed{a) f_X(x) \geq 0 \ \forall x}$ $\boxed{b) \sum_x f_X(x) = 1, \int_{-\infty}^{\infty} f_X(x)dx = 1}$

$X\sim \text{Binomial}(n,p)$; $f(x)=\binom{n}{x}p^x(1-p)^{n-x}$, $x=1,\dots,n$; $E(X)=np$, $V(X)=np(1-p)$, $M_X(t)=[(1-p)+pe^t]^n$

$X\sim \text{Normal}(\mu,\sigma^2)$; $f=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$; $\sigma>0$; $E(X)=\mu$, $V(X)=\sigma^2$, $M_X(t)=\exp(\mu t+\frac{1}{2}(\sigma^2t^2))$

$X\sim N(0,1)$; $f=\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})$; $\chi^2_{df}=\sum_{i=1}^{df} Z_i^2$, $Z_i\sim N(0,1)$; $X\sim\text{Bernoulli}(p)$; $f=p^x(1-p)^{1-x}$, $x=0,1$, $p\in(0,1)$; $E(X)=p$, $V(X)=p(1-p)$, $M_X(t)=(1-p)+pe^t$

$X\sim\text{unif}(a,b)$; $f=\frac{1}{b-a}$, $a\le x\le b$

$X\sim\text{Exp}(\theta)$, $f=\frac{1}{\theta}\exp(-\frac{x}{\theta})$, $x\ge0$, $\theta>0$; $E(X)=\theta$, $V(X)=\theta^2$; $M_X(t)=(1-\theta t)^{-1}$, $t<1/\theta$

$X\sim\text{Gamma}(\alpha,\beta)$: $\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}\exp(-\frac{x}{\beta})$, $x>0$, $\alpha,\beta>0$; $E(X)=\alpha\beta$, $V(X)=\alpha\beta^2$, $M_X(t)=\left(\frac{1}{1-\beta t}\right)^\alpha$, $t<1/\beta$; $X/\beta\sim\text{Gamma}(\alpha,1)$

$X\sim\text{Cauchy}(\theta,\sigma)$: $\frac{1}{\pi\sigma}\left(1+\left(\frac{x-\theta}{\sigma}\right)^2\right)^{-1}$; $\sigma>0$; $x\in\mathbb{R}$; $E(X)$, $V(X)$, $M_X(t)$ DNE; $X\sim\text{NegBin}(r,p)=\binom{r+x-1}{x}p^r(1-p)^x$; $p\in(0,1)$

$X\sim\text{Cauchy}(0,1)$: $\frac{1}{\pi}(1+x^2)^{-1}$

IBP: $\int u\,dv=uv-\int v\,du$; choose u: LIATE; dv s.t. integration is easy

$\frac{d}{d\theta}\tan(\theta)=\sec^2(\theta)=\frac{1}{\cos^2(\theta)}$

u-sub: $\int\frac{x}{1+x^2}dx$, $u=1+x^2$, $du=2x\,dx$; $=\int\frac{1}{u}\frac{1}{2}du=\frac{1}{2}\log(1+x^2)$

$|x|\le y$: $\begin{cases}x\le y\\ x\ge -y\end{cases}$; $|x|\ge y$: $\begin{cases}x\ge y\\ x\le -y\end{cases}$; $\ln(AB)=\ln(A)+\ln(B)$

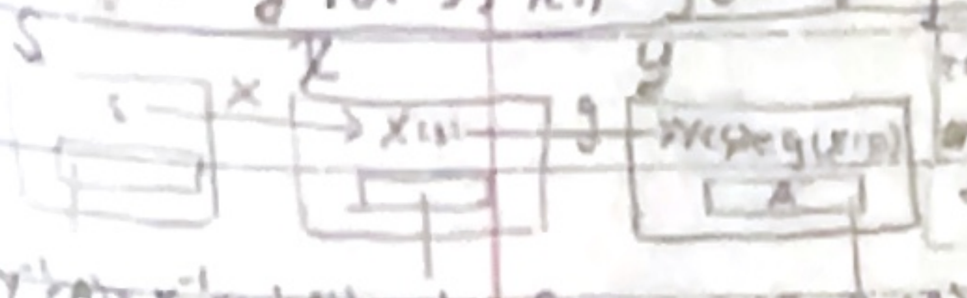
$\int_{-\infty}^{\infty}f(|x-a|)dx=\int_{-\infty}^{a}f(a-x)dx+\int_{a}^{\infty}f(x-a)dx$

$tx-\frac{x^2}{2}=-\frac{1}{2}(x^2-2tx)=-\frac{1}{2}((x-t)^2-t^2)=-\frac{(x-t)^2}{2}+\frac{t^2}{2}$

2.1: Dist. of RV Functions

$Y=g(X)$ new RV

Let $g:\mathcal{X}\to\mathcal{Y}$ for s; $Y(s)=g(X(s))=g\circ X(s)$



$Y^{-1}(A)=X^{-1}(g^{-1}(A))$, $g^{-1}(A)=\{x: g(x)\in A\}$, $P_Y(A)$

$g^{-1}(y)=\{x: g(x)=y\}$ is singleton set

Discrete X; $Y=g(X)$; $\mathcal{Y}=\{y: y=g(x), x\in\mathcal{X}\}$

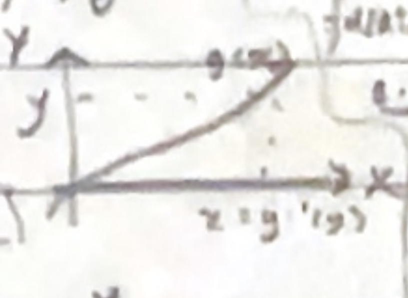
$f_Y(y)=P(Y=y)=\sum_{x\in g^{-1}(y)}P(X=x)=\begin{cases}\sum_{x\in g^{-1}(y)}f_X(x), & y\in\mathcal{Y}\\ f(y)=0, & y\notin\mathcal{Y}\end{cases}$

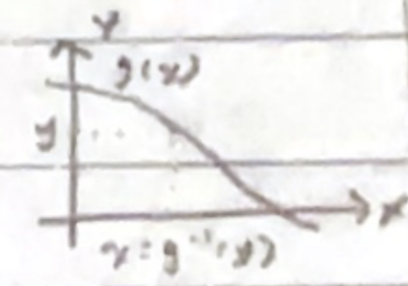
e.g. (Binom): $Y=g(X)=n-X$; $X\sim\text{Bin}(n,p)$; $\binom{n}{y}=\binom{n}{n-y}$ $\to Y\sim\text{Bin}(n,1-p)$

Cont. X; $Y=g(X)$; $F_Y(y)=P(Y\le y)=P(g(X)\le y)=\int_{\{x: g(x)\le y\}}f_X(x)\,dx$

e.g. (Square): $F_Y(y)=P(Y\le y)$; $Y=g(X)=X^2\sim\chi^2_1$

Thm. X has $F_X(x)$; $Y=g(X)$; $g:\mathcal{X}\to\mathcal{Y}$





$f_Y(y)=f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|$

e.g. $X\sim\text{unif}(0,1)$, $Y=g(X)=-\log(X)$; $g^{-1}(y)=x=e^{-y}$; $g\downarrow$ since $g'(x)=-1/x<0$ for $x\in(0,1)$

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

Generating RVs from unif RV

2.2: Expectation

Prop. X RV $\sim P_X$ for $X_1,\dots,X_n$ n-i.i.d. of P_X; then $Y\sim P_n$ and $P(Y=x_i)=\frac{1}{n}$ for $i=1,\dots,n$ $\Rightarrow E(Y)=\frac{1}{n}(X_1+\dots+X_n)\ne E(X)$; avg. of data sample

$E[g(X)]=\begin{cases}\int_{-\infty}^{\infty}g(x)f_X(x), & X\text{ cont}\\ \sum_{x\in\mathcal{X}}g(x)f_X(x), & X\text{ disc.}\end{cases}$

- If $E(X)\notin\mathbb{R}$, then $E(X)$ may be undefined, where $f(x)$ continuous but defined
- Proper exp. $\exists$ s.t. $\left(\int|x|f_X(x)dx<\infty\right)$ i.e. integral finite
- $E(X)$ DNE if $\int x f_X(x)dx$ not abs convergent

$E(X)=\begin{cases}\int x f_X(x)dx, & X\text{ cont}\\ \sum x f_X(x), & X\text{ disc}\end{cases}$

e.g. (Exp): $X\sim\text{Exp}(\lambda)$, $\int f(x)dx=1$; $E(X)=\int x f(x)dx=\lambda$ (IBP)

e.g. (Bin): $X\sim\text{Bin}(n,p)$; $\sum f(x)=1$; binom. thm: $(x+a)^n=\sum_{k=0}^{n}\binom{n}{k}x^ka^{n-k}$; $=(p+(1-p))^n=1$

$E(X)=\sum x f(x)$; $x\binom{n}{x}=n\binom{n-1}{x-1}$; let $y=x-1$ ... $=np(1+1-p)^{n-1}=np$

e.g. (Cauchy): $X\sim\text{Cauchy}(0,1)$, $E|X|=\int_{-\infty}^{\infty}|x|f(x)=2\int_0^\infty\frac{x}{\pi(1+x^2)}=\lim_{M\to\infty}\frac{1}{\pi}\log(1+M^2)=\infty$

$\int f(x)dx$: change of var $\tan(\theta)$ $=\frac{1}{\pi}\left(\frac{\pi}{2}+\frac{\pi}{2}\right)=1$; $E(X)$ DNE

Linearity: $E(aX+bY+c)$ for $a,b,c\in\mathbb{R}$ $=aE(X)+bE(Y)+c$; $x=g_1(z)$, $y=g_2(z)$

Pf. X,Y RV $\in\mathbb{R}$; $E(aX+bY+c)=\int(ag_1(z)+bg_2(z)+c)f(z)=aE(g_1(z))+bE(g_2(z))+c$

e.g. Let $X\sim\text{Bin}(n,p)$; $Y_i\sim\text{Bernoulli}(p)$; $E(Y_i)=\sum y f(y)=1(p)+0(1-p)=p$; $E(X)=E(Y_1)+\dots+E(Y_n)=np$

Prop. 1) $X\ge0$ a.s. $(P(X\ge0)=1)\Rightarrow E(X)\ge0$

2) $X\ge Y$ a.s. $(P(X\ge Y)=1)\Rightarrow E(X)\ge E(Y)$

3) $P(X\in[a,b])=1\Rightarrow E(X)\in[a,b]$

Prop. Let $X_1,\dots,X_n$ n-i.i.d., $X=X_1+\dots+X_n$; $V(X)=V(X_1)+\dots+V(X_n)=n\cdot Var(X_1)$

MGF: $M_X(t)=E(e^{tX})$ provided $t\in$ some $(-h,h)$ for $h>0$
- can characterize dist.
- uniquely determines all moments

$\frac{d}{dt}E(Y)=\frac{d}{dt}\int e^{tx}f(x)dx$; $\frac{d}{dt}$ at $t=0$: $\int xe^{tx}f(x)|_{t=0}=\int xf(x)dx$; $\forall n\in\mathbb{N}$, $E(X^n)=\frac{d^n}{dt^n}M_X(t)\Big|_{t=0}=M_X^{(n)}(0)$
- gives meaning of moment generating

e.g. (Gamma): $\Gamma(\alpha)\beta^\alpha=\int x^{\alpha-1}e^{-x/\beta}dx$; letting $y=x/\beta$, $x=\beta y\Rightarrow dx=\beta\,dy$ $=\int y^{\alpha-1}e^{-y}dy$

$\int f(x)dx=\frac{\Gamma(\alpha)}{\Gamma(\alpha)}=1$, $M_X(t)=\int e^{tx}f(x)$

$=\frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha-1}\exp\left(-\frac{(1-\beta t)x}{\beta}\right)=\left(\frac{1}{1-\beta t}\right)^\alpha$, $t<1/\beta$

$t\ge\frac{1}{\beta}\Rightarrow M_X(t)=+\infty$

$E(X)=\alpha(1-\beta t)^{-\alpha-1}\beta|_{t=0}=\alpha\beta$

e.g. (Binomial): $X\sim\text{Bin}(n,p)$

$M_X(t)=E(e^{tX})=\sum e^{tx}f(x)$; binom. thm: $=(e^tp+(1-p))^n$

$E(X^k)=M_X^{(k)}(0)|_{t=0}$.

Thm. $\forall a,b\in\mathbb{R}$,

$M_{aX+b}=E(e^{t(aX+b)})=e^{bt}E(e^{taX})=e^{bt}M_X(at)$

e.g. $M_X(t)=p^r/(1-(1-p)e^t)^r$. Let $Y=2pX$; $M_Y(t)=M_{2pX}(t)=M_X(2pt)=p^r/(1-(1-p)e^{2pt})^r$

2.3: Moments

n-th moment: $\mu'_n:=E(X^n)$; n-th central mom.: $\mu_n:=E(X-E(X))^n$, $\forall n\in\mathbb{N}$

$V(X)=E(X-E(X))^2=E(X^2)-2E^2(X)+E^2(X)=E(X^2)-[E(X)]^2$; $SD(X):=\sqrt{Var(X)}$

$V(aX+b)=E[(aX+b)^2-(aE(X)+b)^2]=\dots=a^2(E(X^2)-E^2(X))=a^2Var(X)$

e.g. (Exp): $X\sim\text{Exp}(\lambda)$, $V(X)=E(X^2)-E^2(X)$ where $E(X^2)=\int x^2f(x)$ (IBP) $\Rightarrow V(X)=2\lambda^2-\lambda^2=\lambda^2$

e.g. (Bin): $X\sim\text{Bin}(n,p)$; $E(X^2)=\sum x^2 f(x)$, $x\binom{n}{x}=n\binom{n-1}{x-1}$; $=\sum xn\binom{n-1}{x-1}p^x(1-p)^{n-x}$

$y=x-1$ ... expand; $n\sum(x+a)^n=\sum_{k=0}^{n}\binom{n}{k}x^ka^{n-k}$

$=np[(n-1)p(p+(1-p))^{n-2}+(p+(1-p))^{n-1}]=n(n-1)p^2+np$

$\Rightarrow V(X)=E(X^2)-E^2(X)=n(n-1)p^2+np-(np)^2=np-np^2=np(1-p)$

$\Leftarrow$ $Y_i\sim\text{Bernoulli}(p)$, $V(X)=V(Y_1)+\dots+V(Y_n)=n\,Var(Y_1)=np(1-p)$

Thm. Let $F_X(x)$, $F_Y(y)$ valid cdfs:

i) X and Y have bounded support $\{x: f(x)>0\}$; $F_X(u)=F_Y(u)\ \forall u\iff E(X^r)=E(Y^r)\ \forall r=0,1,\dots$

ii) $M_X(t)$ and $M_Y(t)$ exist: $M_X(t)=M_Y(t)$, $t\in(-h,h)\Rightarrow F_X(u)=F_Y(u)\ \forall u$

- moments do not capture all info. about dist. in unbounded setting, e.g. $X\sim f_1$, $Y\sim f_2$ unbounded

$Z\sim N(0,1)$. Let $X=e^Z\Rightarrow Z=g^{-1}(X)=\ln(X)$

$f(x)=f_Z(g^{-1}(x))\left|\frac{d}{dx}g^{-1}(x)\right|\sim\text{LogNorm}(0,1)$

$f_2=f_1(x)(1+\sin(2\pi\log(x)))$ where $E(X^r)=E(Y^r)$, $r=0,1,\dots$ but $f(x)\ne f(y)$.

2.3: Moments (cont.)

Def: $X_n\xrightarrow{d}X$ i.f.f. $F_{X_n}(x)\to F_X(x)$ at all continuity points of F_X

Thm. Let $X_1,\dots,X_n$ RVs w/ $M_{X_i}(t)$; $\lim_{i\to\infty}M_{X_i}(t)\to M_X(t)\Rightarrow X_i\xrightarrow{d}X$; $\exists h>0$ s.t. $\forall t\in(-h,h)$, $M_{X_i}(t)=M_X(t)$

e.g. (Poisson Approx): $X_n\sim\text{Bin}(n,p)$, $Y\sim\text{Pois}(\lambda)$; assume $\begin{cases}np=\lambda\text{ (constant)}\\ n\to\infty,\ p\to0\end{cases}$

$M_{X_n}(t)=(pe^t+(1-p))^n$; using $p=\frac{1}{n}\lambda$, $np=\lambda$

$\Rightarrow\lim_{n\to\infty}(1+\frac{\lambda}{n}(e^t-1))^n=\exp(\lambda(e^t-1))=M_Y(t)\Rightarrow F_{X_n}(x)\to F_Y(x)$ as $n\to\infty$ $\forall x$ where F_Y is cont., i.e. $x\in\mathbb{R}$

2.4: Tools

- $f(x,\theta)$, $a(\theta)$, $b(\theta)$ differentiable wrt θ

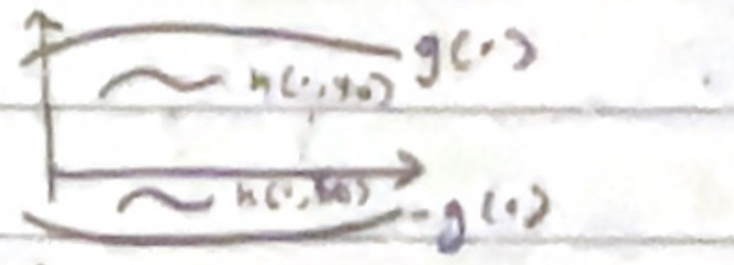
Thm (Leibniz's): $\frac{\partial f}{\partial\theta}(x,\theta)$ cont. on $x\in[x_0,x_1]$

$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)}f(x,\theta)dx=f(b(\theta),\theta)\frac{d}{d\theta}b(\theta)-f(a(\theta),\theta)\frac{d}{d\theta}a(\theta)+\int_{a(\theta)}^{b(\theta)}\frac{\partial}{\partial\theta}f(x,\theta)dx$

Let $a(\theta)=a$, $b(\theta)=b$; $\frac{d}{d\theta}\int_a^b f(x,\theta)dx=\int_a^b\frac{\partial f}{\partial\theta}(x,\theta)dx$

Thm. - $h(x,y)$ cont. at $y=y_0$ for each fixed x - there is a function $g(x)$ s.t. $|h(x,y)|\le g(x)\ \forall x,y$, $\int_{-\infty}^{\infty}g(x)dx<\infty$ (converge cond.)

$\lim_{y\to y_0}\int_{-\infty}^{\infty}h(x,y)dx=\int_{-\infty}^{\infty}\lim_{y\to y_0}h(x,y)dx$



Thm. - For each x, $f(x,\theta)$ is differentiable wrt θ, $\theta\in\Theta$
- For each $\theta\in\Theta$, there is a function $g(\cdot,\theta)$ and $\delta>0$ s.t.

$\left|\frac{\partial}{\partial\theta}f(x,\theta)|_{\theta=\theta'}\right|\le g(x,\theta)\ \forall\theta'\in(\theta-\delta_0,\theta+\delta_0)$; $\int_{-\infty}^{\infty}g(x,\theta)dx<\infty$ for each $\theta\in\Theta$

then $\frac{d}{d\theta}\int_{-\infty}^{\infty}f(x,\theta)dx=\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}f(x,\theta)dx\ \forall\theta\in\Theta$

Pointwise conv. $\lim_{n\to\infty}\sum_{x=0}^{n}h(x,\theta)=:S_n(\theta)=\sum_{x=0}^{\infty}h(x,\theta)=:S(\theta)$

Sup: smallest $s\in\mathbb{R}$ s.t. $s\ge a\ \forall a\in A$, e.g. $\sup(u,v)=v$.

uniform conv. $\sup_{\theta\in(a,b)}|S_n(\theta)-S(\theta)|\to0$ as $n\to\infty$

Thm. - $\sum_{x=0}^{\infty}h(x,\theta)$ exists (converges pointwise) $\forall\theta\in(a,b)$

1) $\frac{\partial}{\partial\theta}h(x,\theta)$ is cont. in θ for each x 2) $\sum_{x=0}^{\infty}\frac{\partial}{\partial\theta}h(x,\theta)$ converges uniformly $\forall\theta$ in closed sub-int of (a,b)

$\frac{d}{d\theta}\sum_{x=0}^{\infty}h(x,\theta)=\sum_{x=0}^{\infty}\frac{\partial}{\partial\theta}h(x,\theta)$; $X\sim\text{Geo}(\theta)$, $f=\theta(1-\theta)^x$, $x=1,2,\dots$, $\theta\in(0,1)$

e.g. $X\sim\text{Geo}(\theta)$; $\frac{d}{d\theta}\sum f=\sum\frac{\partial}{\partial\theta}f\Rightarrow h(x,\theta)=\theta(1-\theta)^x$

$\frac{\partial}{\partial\theta}h(x,\theta)=(1-\theta)^x-\theta x(1-\theta)^{x-1}$ cont. in θ

$S_k(\theta)=\sum_{x=0}^{k}\frac{\partial}{\partial\theta}h(x,\theta)=(k+1)(1-\theta)^k\to0$ as $k\to\infty$ (PC)

$\sup_{\theta\in[c,d]}|S_k(\theta)-0|\to0$ as $k\to\infty$ $\forall[c,d]\in(0,1)$ (UC)

e.g. (1): $X\sim\text{Exp}(\lambda)$, $d/d\lambda\,E(X^n)=\frac{d}{d\lambda}\int_0^\infty=\int_0^\infty\frac{\partial}{\partial\lambda}$; $\left|\frac{\partial}{\partial\theta}f(x,\theta)\right|=x^ne^{-x/\theta}\frac{1}{\theta^2}\left|\frac{x}{\theta}-1\right|\le x^ne^{-x/\theta}\frac{1}{\theta^2}\left(\frac{x}{\theta}+1\right)\le x^ne^{-x/(\theta+\delta_0)}\frac{1}{(\theta-\delta_0)^2}\left(\frac{x}{\theta-\delta_0}+1\right)=:g(x,\theta)$, $x>0$, $\theta>\delta_0$

$X\sim\text{LogNormal}(\mu,\sigma^2)$; $f=\frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x}\exp\left(-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right)$, $x>0$, $\sigma>0$; $E(X)=\exp(\mu+\frac{\sigma^2}{2})$, $V(X)=\exp(2(\mu+\sigma^2))-\exp(2\mu+\sigma^2)$, $E(X^r)=\exp(r\mu+\frac{r^2\sigma^2}{2})$

$X\sim\text{Poisson}(\lambda)$; $f=(e^{-\lambda}\lambda^x)/x!$, $\lambda\ge0$, $x\ge0$; $E(X)=\lambda=V(X)$, $M_X(t)=e^{\lambda(e^t-1)}$

$\lim_{x\to\infty}(1+\frac{a}{x})^x=e^a$; L'Hopital: $\lim_{x\to a}\frac{f(x)}{g(x)}$ ind $=\lim_{x\to a}\frac{f'(x)}{g'(x)}$

$\sum_{k=0}^{\infty}r^k=\frac{1}{1-r}$, $|r|<1$

Thm. suppose $\sum_{x=0}^{\infty}h(x,\theta)$ converges pointwise $\forall\theta\in[a,b]$

1) $h(x,\theta)$ cont. in θ $\forall$ fixed x 2) $\sum_{x=0}^{\infty}h(x,\theta)$ converges uniformly on $[a,b]$. Then $\int_a^b\sum_{x=0}^{\infty}h(x,\theta)d\theta=\sum_{x=0}^{\infty}\int_a^b h(x,\theta)d\theta$

Pre-Test 1

$\bigcup_{i=1}^{\infty} A_i = \{x \in S \mid x \in A_i \text{ for some } i\}$

$\bigcap_{i=1}^{\infty} A_i = \{x \in S \mid x \in A_i \text{ for all } i\}$

$\mathcal{B}$: closed under complementation, $\cap$, $\cup$;
Show sets $\in \mathcal{B}$ via obtaining from known elements of $\mathcal{B}$ via countably many set ops.

	w/o replacement	w/ replacement
ordered	$P_r^n = \frac{n!}{(n-r)!}$	n^r
unordered	$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!}$	$\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!}$

$P(\text{at least one } A_i) = P(\bigcup_{i=1} A_i) = 1 - P(\bigcap_{i=1} A_i^c)$

I.E. generalizes: $P(A \cup B \cup C) = P(A) + P(B) + P(C) - [P(A\cap B) + P(A \cap C) + P(B \cap C)] + P(A \cap B \cap C)$

Bayes: $P(A|B) = P(A\cap B)/P(B) \Leftrightarrow P(A \cap B) = P(A|B)P(B)$

chain: $P(A\cap B \cap C) = P(A)P(B \cap C|A) = P(A)P(B|A)P(C|B,A)$

Bayes: let $S = \bigcup_{i=1}^{\infty} A_i$, $P(B) = \sum_i P(B \cap A_i) = \sum_i P(B|A_i)P(A_i)$ → law of tot. prob.

$P(A_i|B) = [P(B|A_i)P(A_i)] / [\sum_i P(B|A_i)P(A_i)]$

RV: $P_X(X \in A) = P(X^{-1}(A)) = P(\{s \in S \mid X(s) \in A\})$

CDF: $F_X(x) = P(X \le x)\ \forall x \in \mathcal{X}$ $\begin{cases} \lim_{x\downarrow -\infty} = 0, \lim_{x \uparrow \infty} = 1 \\ \text{nondecreasing}, \lim_{x \downarrow x_0} F(x) = F(x_0) \end{cases}$

PDF: $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt\ \forall x \in \mathcal{X} \Leftrightarrow f_X(x) = \frac{d}{dx}F_X(x)$

PMF: $f_X(x) = P(X = x)\ \forall x$; $\begin{cases} f(x) \ge 0\ \forall x \in \mathcal{X} \\ \sum_x f(x) = 1,\ \int_{-\infty}^{\infty} f(x) = 1 \end{cases}$

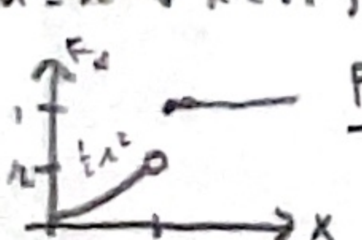
Pre-Test 2 $Y(s) = g(X(s))$; $Y^{-1}(A) = X^{-1}(g^{-1}(A))$

Prop. Let X-cont, $Y = g(X)$; then $F(y) = P(Y \le y) = P(g(x) \le y) = P(x \in \mathcal{X} : g(x) \le y) = \int_{\{x : g(x) \le y\}} f(x)\,dx$.

Thm. $X : F_X(x)$, $\mathcal{X} = \{x \mid f(x) > 0\}$, $Y = g(X)$, then
i) $g : \mathcal{X} \to \mathcal{Y} \uparrow$, then $F_Y(y) = F_X(g^{-1}(y))\ \forall y \in \mathcal{Y}$
ii) $g : \mathcal{X} \to \mathcal{Y} \downarrow$, then $F_Y(y) = 1 - F_X(g^{-1}(y))\ \forall y \in \mathcal{Y}$

cov: $f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$

(min. set equivalent)
Inf: Let $A \subset \mathbb{R}$. $\inf A :=$ largest $a \in \mathbb{R} \cup \{-\infty\}$ s.t. $a \le x\ \forall x \in A$, e.g. $\inf((a,b]) = a \ne \min(a,b)$ DNE

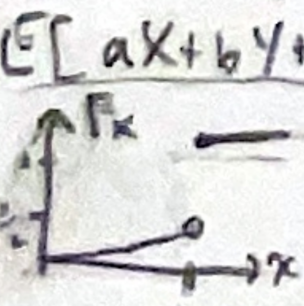


$F^{-1}(p) = \inf\{x \in \mathbb{R} : F(x) \ge p\} = \begin{cases} -\infty, & p < 0 \\ \sqrt{2p}, & 0 \le p < 1/2 \\ 1, & 1/2 \le p \le 1 \\ \infty, & p > 1 \end{cases}$

$F_X^{-1}(1) = +\infty$, $F_X^{-1}(0) = -\infty$ $\forall p \in (0,1)$

$E[g(X)] = \begin{cases} \int_{\mathbb{R}} g(x) f(x)\,dx \\ \sum_{x \in \mathcal{X}} g(x) f(x)\,dx \end{cases}$ e.g. $g(x) = x \Rightarrow \begin{cases} \int x f(x)\,dx \\ \sum x f(x) \end{cases}$

$E[aX + bY + c] = aE(X) + bE(Y) + c$, $a, b, c \in \mathbb{R}$: const.



$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}x^2, & x \in [0,1) \\ 1, & x \ge 1 \end{cases}$ $f(x) = x, x \in [0,1)$
$P(X=1) = \Delta$: $F(1) - F(1^-) = \frac{1}{2}$
$E(X) = \int_0^1 x f(x)\,dx + x_0 \cdot \Delta$

$x_0 = 1$

assign Δ mass @ discontinuity point x_0.

$\mu'_n := E(X^n)$; $\mu_n := E(X - E(X))^n$; $V(X) = E(X - E(X))^2$

$V(X) = E(X^2) - [E(X)]^2$; $V(aX+b) = a^2 \text{var}(X)$

Thm. Let $X_i \overset{iid}{\sim}$, $X = X_1 + \dots + X_n$; $\text{var}(X) = n \cdot \text{var}(X_i)$

mgf: $M_X(t) = E(e^{tX})$; Provided $E(e^{tX})$ $\exists$ in some nghbd of $t = 0$, i.e. $t \in (-h, h)$ for $h > 0$
(characterize RV dist.)

• Justify $\int_{\mathbb{R}} e^{tx} f(x)\,dx$ via $|e^{tx} f(x)|$ dom. conv. thm.
(or/or finite $M_X(t)\ \forall x \in \mathcal{X}$)

Thm. $\forall n \in \mathbb{N}$, $EX^n = \frac{d^n}{dt^n} M_X(t) \big|_{t=0}$

Thm. $\forall a, b \in \mathbb{R}$: $M_{aX+b} = e^{bt} M_X(at)$

• Justify $M_X'(t) = \frac{d}{dt} \int_{\mathbb{R}} e^{tx} f(x)\,dx \overset{(*)}{=} \int_{\mathbb{R}} x e^{tx} f(x)\,dx$ via $|x e^{tx} f(x)|$ dom. conv thm.

Thm $X_i \xrightarrow{d} X$ means $F_{X_i}(x) \to F_X(x)\ \forall$ cont. point of F_X.

Thm. $F_X(x)$, $F_Y(y)$ - moments $\exists$. then
$F_X(u) = F_Y(u)\ \forall u \Leftrightarrow E(X^r) = E(Y^r)\ \forall r = 0, 1, \dots$
$M_X(t) = M_Y(t)\ \forall t \in (-h, h)$ for $h > 0 \Leftrightarrow F_X(u) = F_Y(u)\ \forall u$

PT2 under conditions

• $\frac{d}{d\theta} \int_a^b f(x,\theta)\,dx = \int_a^b \frac{\partial f}{\partial \theta}(x,\theta)\,d\theta$

• $\lim_{y \to y_0} \int_{\mathbb{R}} h(x,y)\,dx = \int_{\mathbb{R}} \lim_{y \to y_0} h(x,y)\,dx$

• $\frac{d}{d\theta} \int_{\mathbb{R}} f(x,\theta)\,dx = \int_{\mathbb{R}} \frac{\partial}{\partial \theta} f(x,\theta)\,d\theta$

• $\frac{d}{d\theta} \sum_{x=0}^{\infty} h(x,\theta) = \sum_{x=0}^{\infty} \frac{\partial}{\partial \theta} h(x,\theta)$

• $\int_a^b \sum_{x=0}^{\infty} h(x,\theta)\,d\theta = \sum_{x=0}^{\infty} \int_a^b h(x,\theta)\,d\theta$

Thm. Suppose $h(x,y)$ cont. @ $y = y_0$ for all x, then, there is a dominating (envelope) fx g s.t.
$\begin{cases} |h(x,y)| \le g(x)\ \forall x, y \\ \int_{\mathbb{R}} g(x)\,dx < \infty \end{cases}$

Pointwise convergence: $\lim_{n \to \infty} \sum_{x=0}^{n} h(x,\theta) = \sum_{x=0}^{\infty} h(x,\theta)$ ($S_n(\theta)$, $S(\theta)$)

Sup. Let $A \subset \mathbb{R}$; $\sup A :=$ smallest $s \in \mathbb{R}$ s.t. $s \ge a\ \forall a \in A$, e.g. $\sup(A \cap (u,v)) = v \ne \max(u,v)$

unif. conv: $\sup_{\theta \in (a,b)} |S_n(\theta) - S(\theta)| \to 0$ as $n \to \infty$

P2(0)

• odd: $f(-x) = -f(x)$, e.g. $x, x^3, \sin^3(x)$

• even: $f(-x) = f(x)$, e.g. $x^2, |x|, \cos(x)$

Thm. $\int_{\mathbb{R}} g(x)\,dx = 0$ if g is odd and $g(x)$ dist. symmetric about zero.

Prop. odd×odd = even; odd×even = odd; even×even = even

• $\int_{\mathbb{R}} f(x - a)\,dx = \int_{-\infty}^{a} f(a - x)\,dx + \int_a^{\infty} f(x-a)\,dx$ ($x < a$, $a < x$)

• computing the square: $tx - \frac{a}{2}x^2 = -\frac{a}{2}(x^2 - \frac{2t}{a}x)$
$= -\frac{a}{2}(x - \frac{t}{a})^2 + \frac{t^2}{a}$

$\int_{\mathbb{R}} \exp(-\frac{a}{2}(x-\mu)^2)\,dx = \sqrt{\frac{2\pi}{a}}$ for $a > 0$

• $\int_{\mathbb{R}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \sqrt{2\pi\sigma^2}$

Ch 3: Dist Fam's

• $X \sim \text{unif}(1, N)$; $p(x) = \frac{1}{N}$, $n = 1, \dots, N$; $E(X) = \frac{N+1}{2}$,
$V(X) = [(N+1)(N-1)]/12 = (N-1)^2/12$

• $X \sim \text{HypGeom}(N, M, K)$: N balls, M-red, N-M G; Pick K balls w/o replac.; X: # red
$p(x) = \frac{1}{\binom{N}{K}} \cdot \binom{M}{x}\binom{N-M}{K-x}$, $x = 0, \dots, K$; $E(X) = \frac{KM}{N}$,
$V(X) = \frac{KM}{N} \cdot \frac{(N-M)(N-K)}{N(N-1)}$; $\sum_{x=0}^{K} \binom{M}{x}\binom{N-M}{K-x} = \binom{N}{K}$

• $X \sim \text{Bernoulli}(p)$; $p(x) = \begin{cases} P(X=1|p) = p \\ P(X=0|p) = 1-p \end{cases}$ $EX = p$, $VX = p(1-p)$

• Suppose $X_i \overset{iid}{\sim} \text{Bernoulli}(p)$ $i = 1, \dots, n$; Y: # successes

$Y \sim \text{Binomial}(n, p)$; $p(y) = \binom{n}{y} p^y (1-p)^{n-y}$; $EY = np$; $VY = np(1-p)$; $M_Y(t) = (pe^t + 1 - p)^n$

$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$, $x\binom{n}{x} = n\binom{n-1}{x-1}$

• $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$; $p(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$, $x = 0, 1, \dots$; $E(X) = \lambda = V(X)$

$e^\lambda = 1 + \lambda + \frac{\lambda^2}{2!} + \dots + \frac{\lambda^x}{x!} + \dots$; $M_X(t) = \exp(\lambda(e^t - 1))$

approx: $X_n \sim \text{Bin}(n, p_n)$, $np_n \to \lambda$ as $n \to \infty$, then $X_n \xrightarrow{d} Y \sim \text{Pois}(\lambda)$

• $X \sim \text{NegBinom}(r, p)$ $p \in [0,1]$, $r \in \mathbb{N}^+$ | X: # ind. Bern(p) trials to obtain r-th success
$p(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$, $x = r, r+1, \dots$

$Y = X - r$; # failures before r-th success (X counts success/failures)
$p(y) = (-1)^y \binom{-r}{y} p^r (1-p)^y$, $y = 0, 1, \dots$; $EY = \frac{r(1-p)}{p}$, $VY = \frac{r(1-p)}{p^2}$

Prop. $\mu = EY$; $\text{var}Y = \frac{1}{r}\mu^2 + \mu$; approx: $r \to \infty$, $p \to 1$, $r(1-p) \to \lambda$ $\Rightarrow$ $EY = \lambda$, $VY = \lambda$ $\Rightarrow Y \xrightarrow{d} \text{Pois}(\lambda)$

• $X \sim \text{Geo}(p) = \text{NegBinom}(p, r=1)$: # trials before first success
$p(x) = p(1-p)^{x-1}$, $x = 1, 2, \dots$; $EX = (1-p)/p$; $VX = (1-p)/p^2$; Prop. if $s > t$, (memoryless) $P(X > s | X > t) = P(X > s - t)$

• $X \sim \text{unif}([a,b])$, $a < b$: $f(x) = \frac{1}{b-a}$, $x \in [a,b]$ | $EX = (b+a)/2$; $VX = (b-a)^2/12$

• $X \sim \text{Gamma}(\alpha, \beta)$ | $\alpha > 0$: shape; $\beta > 0$: scale
$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$, $x > 0$
$EX = \alpha\beta$, $VX = \alpha\beta^2$; $\frac{X}{\beta} \sim \text{Gam}(\alpha, 1)$
$M_X(t) = (\frac{1}{1-\beta t})^\alpha$, $t < 1/\beta$
| $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\,dt$, $\alpha > 0$
| $\Gamma(\alpha)\beta^\alpha = \int_0^\infty x^{\alpha-1} e^{-x/\beta}\,dx$
| $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$
| $\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)!$

Prop. $\forall x > 0$ $P(X \le x) = P(Y \ge \alpha)$ for $X \sim \text{Gam}(\alpha, \beta)$, $Y \sim \text{Pois}(x/\beta)$

• $\chi_p^2 = X \sim \text{Gam}(p/2, 2)$; $f(x) = \frac{1}{\Gamma(p/2) 2^{p/2}} x^{p/2 - 1} e^{-x/2}$ | $x > 0$, $p \in \mathbb{Z}$

• $X \sim \text{Exp}(\beta) = \text{Gam}(1, \beta > 0)$; $f(x) = \frac{1}{\beta} e^{-x/\beta}$, $x > 0$ | maintains memoryless prop

Thm. $X \sim \text{Exp}(\beta > 0)$, $Y = X^{1/\gamma}$, $\gamma > 0$. Then $Y \sim \text{Weibull}(\gamma, \beta)$
$F_Y(y) = 1 - e^{-y^\gamma/\beta}$; $f(y) = \frac{\gamma}{\beta} y^{\gamma-1} e^{-y^\gamma/\beta}$, $0 < y < \infty$

• $X \sim \text{Normal}(\mu, \sigma^2)$ | $\mu \in \mathbb{R}$, $\sigma > 0$ $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$, $x \in \mathbb{R}$; $M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$

$\int f(x)\,dx = 1$, $\int x f(x)\,dx = \mu$, $\int x^2 f(x)\,dx = \mu^2 + \sigma^2$

$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$; $f(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2)$, $z \in \mathbb{R}$ | $EZ = 0$, $VZ = 1$

$\int_{\mathbb{R}} e^{-1/2 \cdot y^2}\,dy = \sqrt{2\pi} \Rightarrow \frac{1}{\sqrt{2}} \int_{\mathbb{R}} \exp(-\frac{1}{2}x^2)\,dx = \sqrt{\pi}$; $\Gamma(1/2) = \sqrt{\pi}$; $Y = Z^2 \sim \chi_1^2$

approx to Bin: $X \sim \text{Bin}(n, p)$, $\begin{cases} np \to \infty \\ np(1-p) \to \infty \end{cases}$ Then $Y \approx N(np, np(1-p))$

cont correction: $P(X = k) \approx P(k - 0.5 < Y < k + 0.5)$; $P(X \le k) \approx P(Y \le k + 0.5)$

$X \sim \text{Beta}(\alpha, \beta)$ | $\alpha > 0$, $\beta > 0$; $f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, $x \in (0,1)$

$EX^n = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+n)}{\Gamma(\alpha)\Gamma(\alpha+\beta+n)}$; $EX = \frac{\alpha}{\alpha+\beta}$, $VX = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$; $X \sim \text{Cauchy}(\theta)$, $\theta \in \mathbb{R}$
$f(x) = \frac{1}{\pi} \frac{1}{1 + (x-\theta)^2}$, $x \in \mathbb{R}$
$EX = \infty$ (DNE); bell-shaped.
Prop. $X, Y \overset{iid}{\sim} N(0,1) \Rightarrow \frac{X}{Y} \sim \text{Cauchy}(0)$

$Y \sim N(\mu, \sigma^2)$, $X = e^Y \sim \text{LogN}(\mu_Y, \sigma_Y)$
$\Leftrightarrow Y = \log X \sim N(\mu_Y, \sigma_Y)$
$f(x|\sigma_Y, \mu_Y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \frac{1}{x} \exp\left(\frac{-1}{2\sigma_Y^2}(\log x - \mu_Y)^2\right)$

Exp. Family: $f(x|\theta) = c(\theta) h(x) \exp\left\{\sum_{i=1}^{k} w_i(\theta) t_i(x)\right\} = c(\theta) h(x) \exp\langle w(\theta), t(x) \rangle$

$\int_{\mathbb{R}} f(x|\theta) = 1 \Rightarrow \frac{1}{c(w)} = \int_{\mathbb{R}} h(x) \exp\langle w(\theta), t(x) \rangle\,dx$; $\exp(x \ln(\frac{A}{B})) = (\frac{A}{B})^x$

$E\left[\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)\right] = \frac{\partial}{\partial \theta_j} \log\left(\frac{1}{c(\theta)}\right)$, $j = 1, \dots, k$

differentiating log of normalizing const. wrt a parameter results in simple expectations of sufficient stats; $\log(a^k) = k\log(a)$, natural

closed under multiplication and reparametrization; not addition; mixtures arise addition at density level (no lumping in algebra to achieve)

Satisfies:
• finite-dimensional param.
• fixed support (not depending on unknown params)
• single exp form (no sums of exponentials)

3 (cont.) Thm. Let $f(x)$: valid pdf. Then define

l-s Family of dist. | μ: location param, σ: scale param.

$g(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, $x \in \mathbb{R}$ valid pdf

l-fam: $= \{g(x|\mu) = f(x - \mu) \mid \mu \in \mathbb{R}\}$

S-fam: $= \{g(x|\sigma) = \frac{1}{\sigma} g(x/\sigma) \mid \sigma > 0\}$

Thm. $X \sim f \Rightarrow X + \mu \sim g(x) = f(x - \mu)$
$X \sim f \Rightarrow \sigma X \sim g(x) = \frac{1}{\sigma} f(x/\sigma)$
$X \sim f \Rightarrow \sigma X + \mu \sim g(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$

Thm. Let $Y \sim f$. Then
$g(x|\mu, \sigma) \Rightarrow EX = \sigma EY + \mu$, $VX = \sigma^2 V(Y)$

$\lim_{n \to \infty} (1 + \frac{a}{n})^n = e^a$

$\lim_{n\to\infty} (1 + \frac{a_n}{n})^n = e^{\lim a_n}$ if $a_n \to a$

$e^a = \sum_{t=0}^{\infty} \frac{a^t}{t!}$

Markov: $P(g(X) \ge r) \le \frac{E g(X)}{r}$ $\forall r > 0$; $P(g(X) \ge r) \le$ ... $E g(X) \ge r \cdot P(g(X) \ge r)$
e.g. $P(|X - \mu| \ge 2) \le 1/4$

Chebyshev: $X \sim EV$, $g(x) \ge 0\ \forall x$:
$Y = g(x) = (X - \mu)^2$, $EY = \sigma^2$
$P(|X - \mu| \ge t) \le \sigma^2/t^2$ $(r = t^2)$; $P(Y \ge r) \le EY/r$
$P(Y \ge r) \le EY/r$

Thm. tighter Norm. tail bnd.
$Z \sim N(0,1)$. For $t > 0$, $P(|Z| \ge t) \le \sqrt{\frac{2}{\pi}} \frac{1}{t} \exp(\frac{-t^2}{2})$

Prop. $\sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t} \ll \frac{1}{t^2}$

Identities: • $X \sim \text{Poisson}(\lambda)$:
$P(X = 0) = \exp(-\lambda)$
$P(X = x+1) = e^{-\lambda} \frac{\lambda^{x+1}}{(x+1)!} = P(X = x) [\lambda/(x+1)]$

• $X \sim \text{Gamma}(\alpha, \beta)$; $f(x|\alpha, \beta)$, $\alpha > 1$, $\beta > 0$ | $\forall a, b$
$P(X_{\alpha,\beta} \in (a,b)) = \beta[f(a|\alpha,\beta) - f(b|\alpha,\beta)] + P(X_{\alpha-1,\beta} \in (a,b))$
• For $\alpha \in \mathbb{N}$, then recurse:
$\text{Gam}(\alpha-1, \beta) \to \text{Gam}(\alpha-2, \beta) \to \dots \to \text{Gam}(1, \beta) = \text{Exp}(\beta)$

Steins: $X \sim N(\theta, \sigma^2)$, g: differentiable s.t.
$E|g'(X)| < \infty \Rightarrow E[g(X)(X - \theta)] = \sigma^2 E[g'(X)]$

Hwang's identities: Let g: function, $|Eg(X)| < \infty$, $|g(-1)| < \infty$

• $X \sim \text{Pois}(\lambda)$. Then $E[\lambda g(X)] = E[X g(X-1)]$

• $X \sim \text{NegBin}(r, p)$. $E[(1-p)g(X)] = E\left[\frac{X}{r + X - 1} g(X-1)\right]$

Ch. 4: Multiple RV's

$s \xrightarrow{X} X(s) \in \mathbb{R}^n$ - e.g. 2 (bivariate)

Discrete: $f(x,y) = P(X = x, Y = y) =: f_{XY}(x,y)$ (joint pmf/pdf)

$P((X,Y) \in A) = \sum_{(x,y) \in A} f(x,y)$; $E[g(X,Y)] = \sum_{(x,y) \in \mathbb{R}^2} g(x,y) f_{XY}(x,y)$

$E[a g_1(X,Y) + b g_2(X,Y)] = aE[g_1(X,Y)] + bE[g_2(X,Y)]$

marginal $f_X(x) = \sum_y f_{XY}(x,y)$; $f_Y(y) = \sum_x f_{XY}(x,y)$

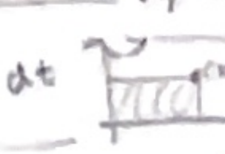
* joint dist. determines marginal dist. but not vice versa

continuous: $P(X \in A) = \iint_{(x,y) \in A} f_{XY}(x,y)\,dx\,dy$

marginal: $f_X(x) = \int_{\mathbb{R}} f_{XY}(x,y)\,dy$
$f_Y(y) = \int_{\mathbb{R}} f_{XY}(x,y)\,dx$

$E[g(X,Y)] = \iint_{\mathbb{R}} g(x,y) f_{XY}(x,y)\,dx\,dy$

$F_{XY}(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(s,t)\,ds\,dt$



$\frac{d^2 F_{XY}(x,y)}{dx\,dy} = f_{XY}(x,y)$ @ cont. points (x,y) of $f_{XY}(x,y)$

Discrete: Let $A = \{x\}$, B: $\{y\}$. $P(Y \in B | X \in A) = \frac{P(X \in A \cap Y \in B)}{P(X \in A)} = \sum_{x \in A} \sum_{y \in B} f_{XY}(x,y) / \sum_{x \in A} f_X(x)$

$f(y|x) = P(Y = y | X = x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_{XY}(x,y)}{\sum_y f_{XY}(x,y)}$; $\sum_y f(y|x) = 1$

Prop. (*) $f_{XY}(x,y) = f_X(x) f_{Y|X}(y|x) = f_Y(y) f_{X|Y}(x|y)$ whenever $f(x), f(y) > 0$

cont.:
$P(Y \in B | X = x) = \lim_{\epsilon \downarrow 0} P(Y \in B | X \in (x - \epsilon, x + \epsilon)) = \int_B \frac{f_{XY}(x,y)}{f_X(x)}\,dy$

$f_{Y|X}(y|x) = f_{XY}(x,y)/f_X(x) = f_{XY}(x,y) / \int_y f_{XY}(x,y)\,dy$

$E[g(Y)|X] = \int g(y) f_{Y|X}(y|x)\,dy$; often easier to identify joint dist. via (*)

Def: X and Y ind. if $f_{XY}(x,y) = f_X(x) f_Y(y)\ \forall x, y \in \mathbb{R} \Rightarrow f(y|x) = f(y)$
• identify (x,y) or (A,B) that violate → for non-independence $\Rightarrow P(Y \in B | X \in A) = P(Y \in B)$

Prop. $X \perp\!\!\!\perp Y$ i.f.f. $f_{XY}(x,y) = g(x) h(y)\ \forall x, y \in \mathbb{R}$

Thm. $E[g(X)h(Y)] = Eg(X) Eg(Y)$; $X \perp\!\!\!\perp Y$, $Z = X + Y$. $M_Z(t) = M_X(t) M_Y(t)$ $\forall t$

Prop. $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, $X \perp\!\!\!\perp Y$; $Z = X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Markov: $P(g(X) \ge r) \le \frac{E[g(X)]}{r}$

$g(X)$: $(X - E(X))^2 \to P(|X - E(X)| \ge t) \le \frac{\text{var}X}{t^2}$

$g(X) = \frac{(X-\mu)^2}{\sigma} \to P(|X - \mu| \ge \sigma\sqrt{r}) \le 1/r$

$\Leftrightarrow P(|X - \mu| \ge t\sigma) \le 1/t^2$, letting $r = t^2$

Ch.4 (cont)

Prop. $\forall A \subset \mathbb{R}^2$, $P((U,V)\in A) = P((X,Y)\in g^{-1}(A))$

$g^{-1}(A) = \{(x,y) \mid g(x,y)\in A\}$; Discrete: $f_{U,V}(u,v) = \sum f_{X,Y}(x,y)$, x,y: $g_1(x,y)=u$, $g_2(x,y)=v$

cont. $g: \mathbb{R}^2 \to \mathbb{R}^2$ 1-1 mapping: $g^{-1}(\{u,v\})$ has at most one element

$g^{-1}(\{u,v\}) = h(u,v) = \begin{pmatrix} h_1(u,v) \\ h_2(u,v) \end{pmatrix}$

$f_{U,V}(u,v) = f_{X,Y}(h_1(u,v), h_2(u,v))\,|J|$

$J = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \begin{vmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{vmatrix}$, $|J| = ad-bc$

$(X,Y) \to g(X,Y)\in\mathbb{R}^2$, many-to-one; partition $A = \{(x,y) \mid f_{X,Y}(x,y)>0\} = A_1 \cup \dots \cup A_k$ s.t. $A_i \to g(A_i)$ 1-1, $i=1,\dots,k$. Then

$f_{U,V}(u,v) = \sum_{i=1}^{k} f_{X,Y}(h_{1i}(u,v), h_{2i}(u,v))\,|J_i|$

$f(x) = f(x_1,\dots,x_n) = P(X_1=x_1,\dots,X_n=x_n)$

$P(X\in A) = \sum_{x\in A} P(X=x) = \sum_{x\in A} f(x)$

$P(X\in A) = \int_A f(x)\,dx = \int\cdots\int_{(x_1,\dots,x_n)\in A} f(x_1,\dots,x_n)\,dx_1\cdots dx_n$

$E g(X) = \begin{cases} \int_{\mathbb{R}^n} g(x) f(x)\,dx \\ \sum_{x\in\mathbb{R}^n} g(x) f(x) \end{cases}$

marginal: $f_{X_1,\dots,X_k}(x_1,\dots,x_k) = \begin{cases} \int_{\mathbb{R}}\cdots\int_{\mathbb{R}} f(x_1,\dots,x_n)\,dx_{k+1}\cdots dx_n \\ \sum_{x_{k+1}}\cdots\sum_{x_n} f(x_1,\dots,x_n) \end{cases}$

e.g. $f(x_1,x_2) = \int_{\mathbb{R}}\int_{\mathbb{R}} f(x_1,x_2,x_3,x_4)\,dx_3\,dx_4$

Conditional: $f(x_{k+1},\dots,x_n \mid x_1,\dots,x_k) = \dfrac{f(x_1,\dots,x_k,x_{k+1},\dots,x_n)}{f(x_1,\dots,x_k)}$

$\Rightarrow$ admits factorization; joint = marginal × cond

Chain: $f(x_1,\dots,x_n) = f(x_1)f(x_2\mid x_1)\cdots f(x_n\mid x_1,\dots,x_{n-1})$

mutual ind. $f(x_1,\dots,x_n) = \prod_{i=1}^n f_{X_i}(x_i)$

$E[g_1(X_1)\cdots g_n(X_n)] = E[g_1(X_1)]\cdots E[g_n(X_n)]$

pdf, cond. pdf, pdf, cond. pmf: $f(x,y,z,t) = f(x,y)f(z,t\mid x,y) = f(z,t)f(x,y\mid z,t)$

$P(X\in A, Y\in B, Z\in C, T\in D) = \sum_{x\in A}\sum_{y\in B}\int_D\int_C f(x,y,z,t)\,dz\,dt$

Multinomial (n trials, p_i's) (n ind. trials, n possible outcomes with prob's $(p_1,\dots,p_n)$, $\sum_{i=1}^n p_i = 1$ | $(X_1,\dots,X_n)$: counts of each of n outcomes s.t. $x_1+\dots+x_n = m$)

multi coefficient: $f(x_1,\dots,x_n) = \binom{m}{x_1,\dots,x_n} p_1^{x_1}\cdots p_n^{x_n} = \frac{m!}{x_1!\cdots x_n!} p_1^{x_1}\cdots p_n^{x_n}$ | $x_1,\dots,x_n\in\mathbb{N}^n$, $\sum_{i=1}^n x_i = m$

$\sum_{x_1+\dots+x_n=m} \binom{m}{x_1,\dots,x_n} p_1^{x_1}\cdots p_n^{x_n} = (p_1+\dots+p_n)^m = 1$ multinomial formula

marginal: $f(x_n) = \sum_{x_1+\dots+x_{n-1}=m-x_n} f(x_1,\dots,x_n) = \frac{m!}{x_n!(m-x_n)!} p_n^{x_n}(1-p_n)^{m-x_n} \sim Bin(m,p_n)$

cond: $f(x_1,\dots,x_{n-1}\mid x_n) = \frac{f(x_1,\dots,x_n)}{f(x_n)} = \frac{(m-x_n)!}{x_1!\cdots x_{n-1}!}\left(\frac{p_1}{1-p_n}\right)^{x_1}\cdots\left(\frac{p_{n-1}}{1-p_n}\right)^{x_{n-1}} \sim Mult\left(m-x_n \text{ trials}, \left(\frac{p_1}{1-p_n},\dots,\frac{p_{n-1}}{1-p_n}\right)\right)$

invariance prop (marg/cond.) holds for many exponential fam dist (e.g. normal)

Thm. X,Y RV, then $EX = E[E(X\mid Y)]$ | iterated expectation

Mixture Dist: (of X) depends on RV Y, e.g.

$X\mid Y \sim Binomial(Y,p)$, $Y\sim Pois(\lambda)$ | mixture of Bin dist's; Pois mixing mechanism

$P(X=x) = \sum_{y\in\mathbb{N}} P(X=x, Y=y) = \sum_{y\in\mathbb{N}} P(Y=y)P(X=x\mid Y=y) = \sum \frac{e^{-\lambda}\lambda^y}{y!}\binom{y}{x}p^x(1-p)^{y-x}$, $x\in\mathbb{N}$

Thm. $X_1,\dots,X_n$ mutually ind., $Z = X_1+\dots+X_n$; $M_Z(t) = M_{X_1}(t)\cdots M_{X_n}(t)$; | if $X_i \overset{iid}{\sim}$, $i=1,\dots,n$ then $M_Z(t) = [M_{X_1}(t)]^n$

Prop. $Y\in\{1,\dots,k\}$, $P(Y=i) = p_i$, $i=1,\dots,k$, $X\mid Y=i \sim N(\mu_i,\sigma_i^2)$

Prop. $X_1,\dots,X_n$ mutually ind. $\Leftrightarrow f(x_1,\dots,x_n) = g_1(x_1)\cdots g_n(x_n)$ for some functions $g_1,\dots,g_n$ $\Rightarrow g_1,\dots,g_n$ are also mutually ind.

$f_X(x) = \sum p_i N(x\mid\mu_i,\sigma_i^2) = \sum p_i \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(\frac{-1}{2\sigma_i^2}(x-\mu_i)^2\right)$

(mixture of k normal components)

Prop. $X_i \sim Gam(\alpha_i,\beta)$, $X_i \overset{ind}{\sim}$, $i=1,\dots,n$. Then $X_1+\dots+X_n \sim Gam(\alpha_1+\dots+\alpha_n,\beta)$

Prop. $X_i \overset{ind}{\sim} N(\mu_i,\sigma_i^2)$ | $i\in\{1,\dots,n\}$, $a_i,b_i\in\mathbb{R}$. Then $Z = \sum_{i=1}^n (a_iX_i+b_i) \sim N\left(\sum_{i=1}^n (a_i\mu_i+b_i), \sum_{i=1}^n a_i^2\sigma_i^2\right)$

$\sum_{n=1}^{\infty}\frac{1}{n} = \infty$ (Harmonic series diverges)

Hierarchical: X dist. obtained w/ many stages of cond. dist's

e.g. $X\mid Y\sim Bin(Y,p)$, $Y\mid\Lambda\sim Pois(\Lambda)$, $\Lambda = Exp(\beta)$

$(\Rightarrow) Y\mid\Lambda=\lambda \sim Pois(\lambda)$

$P(Y=y) = \int_0^\infty f_{Y,\Lambda}(y,\lambda)\,d\lambda = \int_0^\infty P(Y=y\mid\Lambda=\lambda) f_\Lambda(\lambda)\,d\lambda$ (marginalize)

$= \int_0^\infty e^{-\lambda}\frac{\lambda^y}{y!}\frac{1}{\beta}e^{-\lambda/\beta}\,d\lambda \sim NegBin\left(\frac{1}{1+\beta},1\right)$

$(\Rightarrow)$ $X\mid Y\sim Bin(Y,p)$, $Y\sim NegBin\left(\frac{1}{1+\beta},1\right)$

Ch.5: Random Sample Properties

$X_1,\dots,X_n \overset{iid}{\sim} f$. Then f: (n-iid) sample

Statistic: $Y = T(X_1,\dots,X_n)$

e.g. $\bar{X} = \frac{1}{n}(X_1+\dots+X_n)$, $S^2 = \frac{1}{n-1}\sum(X_i-\bar{X})^2$

$X_{(k)} = \min\{X_1,\dots,X_n\}\setminus\{X_{(1)},\dots,X_{(k-1)}\}$

k-th smallest number; order stat: $(X_{(1)},\dots,X_{(n)})$

Thm. Let $(X_1,\dots,X_n)\overset{iid}{\sim}$ with $\mu,\sigma^2<\infty$;

Then $E\bar{X}=\mu$, $Var\bar{X} = \sigma^2/n$, $ES^2 = \sigma^2$

• $\bar{X}$ and S^2 unbiased estimates

Dist. of $\bar{X}$: (1) moments, (2) convolution

(1) $M_{\bar{X}}(t) = Ee^{t\bar{X}} = (M_X(t/n))^n$; easy if we can recognize $\bar{X}$ from known family of dist

(2) $X\sim f_X$, $Y\sim f_Y$, $Z = X+Y$, $X\perp Y$

$f_Z(z) = \int_{\mathbb{R}} f_X(x)f_Y(z-x)\,dx$; useful when $f(x_1,\dots,x_n)$ does not belong to known/well-recognized family

$f(z) = \sum_x f_X(x)f_Y(z-x)$

(2) $Z = X+Y$: $(X,Y)\to(X+Y,X)$

$Z = X-Y$: $(X,Y)\to(X-Y,X)$ and $\int_{\mathbb{R}} f_X(x)f_Y(x-z)\,dx = f(z)$

• $Z = XY$: $(X,Y)\to(XY,X)$, etc.

$\sum_i^n(X_i-\bar{X}) = \sum X_i - n\bar{X} = \sum X_i - \sum X_i = 0$

Thm. Let $X_i \overset{iid}{\sim} N(\mu,\sigma^2)$, $i=1,\dots,n$

• $\bar{X}\perp S^2$ • $\bar{X}\sim N(\mu,\sigma^2/n)$; $E(X_i-E(X))=0$

• $(n-1)\frac{S^2}{\sigma^2}\sim\chi^2_{n-1}$; $Var(X\sim\chi^2_k) = 2k$

$T := (\bar{X}-\mu)/(S/\sqrt{n})$, $t\in\mathbb{R}$

$f(t) = \frac{\Gamma[(p+1)/2]}{\Gamma(p/2)}\frac{1}{(p\pi)^{1/2}}\frac{1}{(1+\frac{t^2}{p})^{(p+1)/2}}$

• T: ratio of 2 ind. RVs: $\bar{X}-\mu\sim N$, $\sqrt{(n-1)S^2}\sim\sqrt{\chi^2_{n-1}}$

$n=2 \Rightarrow T = \frac{\bar{Y}-\mu}{Y\ldots}$ ~ Cauchy

$\int_{\mathbb{R}} e^{-|x|}\,dx = 2$

$\int e^{-x^2}dx = \sqrt{\pi}$, $\int e^{-\frac{1}{2}z^2}dz = \sqrt{2\pi}$, $x = z/\sqrt{2}$

$\int e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx = \sqrt{2\pi\sigma^2}$; $z = \frac{x}{\sigma}$

$F(a+x) = F(a-x)$ $\forall x>0$

$X_n \xrightarrow{d} X := F_n(x)\to F_X(x)$ at all cont. points for F_X | $\Phi(x) = P(Z\le x) = \int_{-\infty}^x\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt$

$X_n\xrightarrow{P} X := \forall\varepsilon>0$, $P(|X_n-X|\ge\varepsilon) \Leftrightarrow P(s: |X_n(s)-X(s)|\ge\varepsilon)\to 0$ as $n\to\infty$

Thm(WLLN): $X_1,\dots,X_n\overset{iid}{\sim}$ with $EX_i=\mu$, $VX_i = \sigma^2<\infty$. Then $\bar{X}_n\xrightarrow{P}\mu$; SLLN: $\bar{X}_n\xrightarrow{a.s.}\mu$.

$X_n\xrightarrow{a.s.}X := P(\lim_{n\to\infty}|X_n-X|=0) = 1 \Leftrightarrow P(\lim_{n\to\infty}X_n = X) = 1 \Leftrightarrow \forall\varepsilon>0$, $P(\lim_{n\to\infty}|X_n-X|<\varepsilon) = 1$

Prop. $X_n\xrightarrow{a.s.}X \Rightarrow X_n\xrightarrow{P}X$; $X_n\xrightarrow{P}X\Rightarrow X_n\xrightarrow{d}X$; Fatou's L: $E[\liminf_{n\to\infty}X_n]\le\liminf_{n\to\infty}E(X_n)$

Thm(CLT): $X_i\overset{iid}{\sim}$, $i=1,\dots,n$. Letting $\mu = EX_i$, $Z\sim N(0,1)$, $\sigma^2 = VX_i>0$ ($\Rightarrow VX_i<\infty$ (mild condition)). Then $\left(\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\right) = \sqrt{n}\left(\frac{\bar{X}_n-\mu}{\sigma}\right)\xrightarrow{d}N(0,1)$.

Cor. Let $Y_i = (X_i-\mu)/\sigma$ s.t. $EY_i=0$, $EY_i^2=1$. Then $\frac{1}{\sqrt{n}}(Y_1+\dots+Y_n)\xrightarrow{d}N(0,1)$

Prop. $X_n\xrightarrow{as}X$, $Y_n\xrightarrow{as}b\in\mathbb{R}\Rightarrow aX_n+Y_n\xrightarrow{as}aX+b$

$X_n\xrightarrow{P}X$, $Y_n\xrightarrow{P}b\in\mathbb{R}\Rightarrow aX_n+Y_n\xrightarrow{P}aX+b$; $X_n\xrightarrow{d}X$, $Y_n\xrightarrow{P}b\in\mathbb{R}\Rightarrow aX_n+Y_n\xrightarrow{d}aX+b$

$X\mid\Theta\sim Bin(n,\theta)$, $\theta\in(0,1)$

$\Theta\sim Beta(\alpha,\beta)$

$P(X=x) = \int P(X=x\mid\Theta=\theta)f_\Theta(\theta)\,d\theta$ (marginalize)

$= \int\binom{n}{x}\theta^x(1-\theta)^{n-x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\,d\theta$

$VarX = E[Var[X\mid Y]] + Var[E[X\mid Y]]$

$P(A) = E(1_A)$ $\forall$ event A where $1_A = \begin{cases}1, & A \text{ occurs}\\ 0, & \text{otherwise}\end{cases}$

$\Phi(t) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-t^2}{2}\right)$; $X\sim N(\mu,\sigma^2)\Rightarrow f(x) = \phi\left(\frac{x-\mu}{\sigma}\right)$

$Cov(X,Y) = E[(X-EX)(Y-EY)]$; | $Cov(X,X) = Var(X)$; $EX=EY=0\Rightarrow Cov(X,Y) = E(XY)$; $Cov(X,cX) = cVar(X)$

$corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{VX}\sqrt{VY}}$ standardized covariance

CSI: $E[XY]\le\sqrt{E(X^2)}\sqrt{E(Y^2)}\Leftrightarrow[E(XY)]^2\le E(X^2)E(Y^2)$

Equality i.f.f. $Y\overset{as}{=}cX\Leftrightarrow P(Y=cX)=1$, c: const $\in\mathbb{R}$

Prop. Let $U = X-EX$, $V = Y-EY$

$E(UV)^2\le E(U^2)E(V^2)\Leftrightarrow|corr(X,Y)|\le 1$

Equality i.f.f. $U\overset{as}{=}cV = P(U=cV)=1\Rightarrow|corr(X,Y)|=1$

Thm. $Cov(X,Y) = E(XY)-E(X)E(Y)$;

$X\perp Y\Rightarrow Cov(X,Y) = corr(X,Y) = 0$ ($\not\Leftarrow$ not necessarily)

Thm. $\forall a,b\in\mathbb{R}$, $Cov(X,A+B) = Cov(X,A)+Cov(X,B)$ (linearity)

$V(aX+bY) = a^2Var(X)+b^2Var(Y)+2abCov(X,Y)$

Thm. $corr(X,Y)=1\Leftrightarrow Y=bX+c$, $b>0$, $c\in\mathbb{R}$

$corr(X,Y)=-1\Leftrightarrow Y=bX+c$, $b<0$, $c\in\mathbb{R}$

Bivariate Normal: $Z = (X,Y)\sim N(\mu,\Sigma)$ | $\mu\in\mathbb{R}^2$, $\Sigma\in\mathbb{R}^{2\times2}$

$f(x,y) = f(z) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(z-\mu)^T\Sigma^{-1}(z-\mu)\right]$

notation: σ_{XY}: $Cov(X,Y)$; $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$: $corr(X,Y)$

$E(Z) = E\left[\begin{pmatrix}X\\Y\end{pmatrix}\right] = \begin{pmatrix}\mu_1\\\mu_2\end{pmatrix}$; $Cov Z = \Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\\Sigma_{21} & \Sigma_{22}\end{pmatrix}$ | $\Sigma_{11} = VX$, $\Sigma_{22} = VY$, $\Sigma_{12} = \Sigma_{21} = Cov(X,Y)$

$= E[(Z-EZ)(Z-EZ)^T]$

marginal: $X\sim N(\mu_1,\Sigma_{11})$, $Y\sim N(\mu_2,\Sigma_{22})$

symmetric abt a:

Identities: $\iint f(x,y)\,dx\,dy = \mu_1$; $\iint f(x,y)\,dx\,dy = 1$; $\iint y f(x,y)\,dx\,dy = \mu_2$; $\iint(z-\mu)(z-\mu)^T f(z)\,dz = \Sigma$

Thm. $Cov(X,Y)=0\Rightarrow X\perp Y$

• $Y\mid X=x\sim N[E(Y\mid X), Var(Y\mid X)]$

$E(Y\mid X) = \mu_2+\Sigma_{21}\Sigma_{11}^{-1}(x-\mu_1)$

$V(Y\mid X) = \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$

Final 3) Exp family: $f(x\mid\theta) = c(\theta)h(x)\exp(\sum w_i(\theta)t_i(x)) = c(\theta)h(x)\exp\langle w(\theta), t(x)\rangle$

NegBin(r,p); $P(X\mid p) = f(x\mid\theta)$ whereas $P(X\mid r,p)\ne f(x\mid\theta)$; cannot factorize $(x-r)!$ (separable)

$f(x\mid k,p,\alpha,\beta) = \sum p_i Gam(x\mid\alpha_i,\beta_i)\ne f(x\mid\theta)$; cannot rewrite $\sum_i p_i e^{a(\theta)}\to e^{ak}$ | Mixtures involve addition of density (not...)

$P(X=x) = p_k$ | $k\in\mathbb{Z}$, $\sum p_k = 1$, co-ly many $\{p_k\}$; $f(x\mid\theta)$ requires finite-dim θ, $p,2$, X-RV's (H: ...)

3) $X\mid p_{Z=1}\sim N(\mu_1,1)$, $X\mid p_{Z=0}\sim N(\mu_0,1)$, $Z\mid p\sim Bernoulli(p)$, $p\sim Beta(\alpha,\beta)$

$P(A) = E(1_A)\Rightarrow P(Z=1) = E[P(Z=1\mid p)]\Rightarrow Z\sim Bern(\alpha/(\alpha+\beta))$; $f(x\mid z) = \phi([x-\mu_z]/\sigma_z)$, $z\in\{0,1\}$.

$EX = E[E[X\mid Z]] = \mu_0P(Z=0)+\mu_1P(Z=1) = \mu_0[\beta/(\alpha+\beta)]+\mu_1[\alpha/(\alpha+\beta)]$ & $Z\mid X\sim Bern(P(Z=1\mid x))$

$P(Z\mid X) = [f(x\mid z)f(z)]/f(x)\mid_{z=1} = [\alpha\phi(x-\mu_1)]/[\beta\phi(x-\mu_0)+\alpha\phi(x-\mu_1)]$ — any linear transformation of a jointly normal vector is jointly normal

1) Roll dice ind. n times; $X_i\in\{1,\dots,6\}$, $i=1,\dots,n$; $X_i\in\{3,6\}$, $Y_i = 1$ ... ; $Y_i = -1$ ...; $i\in\mathbb{N}^n$

$S = Y_1Y_2$; $P(S) = 1/9\,(S=1,-1)$, $2/9\,(S=1,-1)$, $1/3\,(S=0)$ | $P(Y_1Y_2) = P(Y_1)P(Y_2)$ ...

$S_n = \sum_{i=1}^n Y_i$; $E(S_n) = E(\sum Y_i) = 0$; $V(S_n) = n\cdot V(Y_i) = 2n/3$; CLT: $[S_n-ES_n]/\sqrt{V S_n} = S_n/\sqrt{2n/3}\xrightarrow{d}N(0,1)$ | $P(S_n\le t) = \Phi(t/\sqrt{2n/3})$

For $\varepsilon>0$, show $P(|Y_1\cdots Y_n|\ge\varepsilon)\to 0$ as $n\to\infty$;

$|Y_1\cdots Y_n| = 1\Leftrightarrow|Y_1\cdots Y_n|=1$ whenever $Y_i\ne 0$ $\forall i$

$Z_i = \mathbb{1}\{Y_i=0\}$ s.t. $EZ_i = 1/3$, $VZ_i = 1/3(1/3)\ldots = 2/9$

$Z_n = \sum Z_i$ ... $EZ_n = n/3$, $VZ_n = 2n/9$

$P(|Y_1\cdots Y_n|\ge\varepsilon)\le P(Z_n = 0)$ & bound

$Z_n = 0\Rightarrow|Z_n-EZ_n|\ge EZ_n = n/3$

$P(|Z_n-EZ_n|\ge n/3)\le\frac{Var Z_n}{(n/3)^2} = \frac{2}{n}$

$\Rightarrow P(|Y_1\cdots Y_n|\ge\varepsilon)\le\frac{2}{n}\to 0$, n→∞

4) $X\sim N(\mu,\sigma^2)$, $\varepsilon = aX+Z$, $Z\sim N(0,1)\perp X$, observed $Y = X+\varepsilon$

$\begin{pmatrix}X\\Y\end{pmatrix} = \begin{pmatrix}1 & 0\\ 1+a & 1\end{pmatrix}\begin{pmatrix}X\\Z\end{pmatrix}\Rightarrow\begin{pmatrix}X\\Y\end{pmatrix}$ is bivariate normal

$E\left[\begin{pmatrix}X\\Y\end{pmatrix}\right] = \begin{pmatrix}EX\\EY\end{pmatrix} = \begin{pmatrix}\mu\\(1+a)\mu\end{pmatrix}$; $\Sigma = \begin{pmatrix}VX & CovXY\\ CovXY & VY\end{pmatrix} = \begin{pmatrix}\sigma^2 & (1+a)\sigma^2\\ (1+a)\sigma^2 & (1+a)^2\sigma^2+1\end{pmatrix}$

$VY = V[(1+a)X+Z]$; $Cov(X,Y) = Cov(X,(1+a)X)+Cov(X,Z) = (1+a)\sigma^2$; $(1+a)^2\sigma^2+1$

$corr(X,Y) = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}\ldots(1+a)\sigma^2)\ldots(1+a)<0\Leftrightarrow a<-1$

$E(Y\mid X) = \mu_2+\Sigma_{21}\Sigma_{11}^{-1}(x-\mu_1) = \mu_2+\frac{\sigma_{XY}}{\sigma_X^2}(x-\mu_1)$

5) $U,V\sim Uni(0,1)$, $U\perp V$, $f(u)=1$, $f(v) = \begin{cases}1 & 0<u<1\\ 1 & 0<v<1\end{cases}$

$Y = U^2$; $F_Y(y) = P(U^2\le y) = P(U\le\sqrt{y}) = \sqrt{y}$

$F(y) = \frac{d}{dy}F(y) = 1/2\sqrt{y}$, $0<y<1$

$X = U+V$; $F(x) = \int_{\mathbb{R}} f(u)f_V(x-u)\,du = \int_{\mathbb{R}} f(u)1\{0<u<1\}f_V(x-u)1\{0<x-u<1\}$ $x-1<u<x$ $= \int_{\max(0,x-1)\text{ - changes @ }x=1}^{\min(1,x)\text{ - changes @ }x=1} du$

$0<x<1$: $F(x) = \int_0^x du = x$; $1<x<2$: $F(x) = \int_{x-1}^1 du = (2-x)$

$Y=y\Rightarrow U=\sqrt{y}$ $(U\ge 0)\Rightarrow X = \sqrt{y}+V\in[\sqrt{y},\sqrt{y}+1]$: $\{(x,y): 0<y<1, \sqrt{y}<x<\sqrt{y}+1\}$

Choose $y = 1/4$, $x = 0.1$. Then $x = 0.1<\sqrt{1/4} = \frac{1}{2}\Rightarrow P(X=x, Y=y)$ undefined ($=0$) since $x<\frac{1}{2}$, $\notin$ CB

But $f_X(0.1)>0$, $f_Y(1/4)>0$ (outside rectangle on support) when $y=1/4$,

and $f(x,y)\ne f(x)f(y)\Rightarrow X\not\perp Y$

$\begin{cases}X = U+V\\ Y = U^2\end{cases}\to\begin{cases}U = \sqrt{Y}\\ V = X-\sqrt{Y}\end{cases}$, $J = \begin{pmatrix}0 & 1/2\sqrt{y}\\ 1 & -1/2\sqrt{y}\end{pmatrix}$, $|J| = 1/2\sqrt{y}$

$f(x,y) = f_U(\sqrt{y})f_V(x-\sqrt{y})\frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}}$ for $\begin{vmatrix}0<u<1\Rightarrow 0<\sqrt{y}<1\Rightarrow 0<y<1\\ 0<v<1\Rightarrow 0<x-\sqrt{y}<1\Rightarrow\sqrt{y}<x<\sqrt{y}+1\end{vmatrix}$