Sample space:
- Mutually exclusive
- Collectively exhaustive
- Right Granularity

Probability axioms:
- P(A)≥0
- P(Ω)=1
- P(AUB)=P(A)+P(B) if A and B are disjoint

Countable additivity axiom:
$P(A_1 \cup A_2 \cup \dots)=P(A_1)+P(A_2)+\dots$
If $A_1,A_2,\dots$ Is an infinite sequence of disjoint events(holds only for countable sets)

Union Bound : $P(A\cup B)\le P(A)+P(B)$

Bonferroni inequality:
$P(A_1\cap A_2)\ge P(A_1)+P(A_2)-1$
General form:
$P(A_1\cap A_2\cap\dots A_n)\ge P(A_1)+P(A_2)+\dots+P(A_n)-(n-1)$

De Morgan's Law:
1. $\left(\bigcup_n S_n\right)^C=\bigcap_n S_n{}^C$ 2. $\left(\bigcap_n S_n\right)^C=\bigcup_n S_n{}^C$

Combination:
$\binom{n}{k}$ = no of k − elements subsets from a given n − *element set*/Binomial coefficient
$$\sum_{k=0}^{n}\binom{n}{k}=2^n$$
$\frac{n!}{n_1!*n_2!*\dots*n_r!}$ = Dividing a n − element set into r partition each with n_i distinct elements/multinomial coefficient

Reliability:
Given n components and probability of components working is p

Series: P(System is up)=p^n

Parallel: P(System is up)=1-$(1-p)^n$

Inclusion-exclusion formula:
$P(\bigcup_{k=1}^{n}A_k)=\sum_i P(A_i)-\sum_{i_1<i_2}P(A_{i_1}\cap A_{i_2})+\sum_{i_1<i_2<i_3}P(A_{i_1}\cap A_{i_2}\cap A_{i_3})-\dots+(-1)^{n-1}*P(\bigcap_{k=1}^{n}A_k)$

Conditional Probability for EVENTS:
$P(A|B)=\frac{P(A\cap B)}{P(B)}$ where $P(B)>0$
Conditional probabilities share same properties like ordinary probabilities (follow probability axioms)
Multiplicative rule:
$P(A\cap B)=P(A)*P(B|A)=P(B)*P(A|B)$
General form:
$P(A_1\cap A_2\cap\dots A_n)=P(A_1)*\prod_{i=2}^{n}P(A_i|A_1\cap A_2\cap\dots\cap A_{i-1})$
Total Probability theorem:
$P(B)=\sum_i P(A_i)*P(B|A_i)$
Independence:
P(B|A)=P(B) occurrence of A provides no new information about B
$P(A\cap B)=P(A)*P(B)$ then A and B are independent
- If A and B are independent then A and B^C are independent
- If A and B are independent then A^C and B^C are independent
- If A and B are independent then B and A^C are independent

Pairwise independence :
A,B and C are pairwise independent means:
$P(A\cap B)=P(A)*P(B)$
$P(B\cap C)=P(B)*P(C)$
$P(A\cap C)=P(A)*P(C)$
It does not imply A,B and C are independent i.e.:
$P(A\cap B\cap C)=P(A)*P(B)*P(C)$
General form:
$P(A_1\cap A_2\cap\dots A_n)=P(A_1)*P(A_2)*\dots*P(A_n)$ then $A_1,A_2,\dots A_n$ are Independent

Conditional Independence:
$P(A\cap B|C)=P(A|C)*P(B|C)$
If A and B are independent it doesn't imply conditional independence

BAYES THEOREM:
$$P(A_i|B)=\frac{P(A_i)*P(B|A_i)}{\sum_j P(A_j)P(B|A_j)}$$

Conditional PMF, given an event:
$p_{(X|A)}(x)=P(X=x|A)$
where $P(A)>0$
$\sum_x p_{(X|A)}(x)=1$
Total Probability theorem:
$P(X)=\sum_i P(A_i)*P(X|A_i)$
Expectation:
$\sum_x g(x)*p_{(X|A)}(x)$
$E[X]=\sum_i P(A_i)*E(X|A_i)$
Independence:
$p_{X|A}(x)=p_{(X)}(x)$ for all x
$P(X=x \text{ and } A)=P(X=x)*P(A)$ for all x

Conditional PDF, given an event:
$$f_{(X|A)}(x)=\begin{cases}0 & if\ x\notin A\\ \frac{f_X(x)}{P(A)} & if\ x\in A\end{cases}$$
For small interval δ:
$f_{(X|A)}(x)*\delta\approx P(x\le X\le x+\delta|A)$where $P(A)>0$
$\int f_{(X|A)}(x)dx=1$
Total Probability theorem:
$f_X(x)=\sum_i P(A_i)*f_{X|A_i}(x)$
Expectation:
$E(X|A)=\int x*f_{(X|A)}(x)dx$
$E[g(X)|A]=\int g(x)*f_{(X|A)}(x)dx$
$E[X]=\sum_i P(A_i)*E(X|A_i)$

Probability Mass Function(PMF)(discrete r.v.)	**Probability Density function(PDF)(continuous r.v.)**
$P_X(x)=P(X=x)$ $=P(\{\omega\epsilon\Omega$ such that $X(\omega)=x\})$	$P(a\le X\le b)=\int_a^b f_X(x)dx$ For small interval δ: $P(a\le X\le a+\delta)\approx f_X(a)*\delta$ PDF are not probabilities but probabilities per unit length i.e. densities
Properties: • $P_X(x)\ge0$ • $\sum_x P_X(x)=1$	Properties: • $f_X(x)\ge0$ • $\int_x f_X(x)dx=1$ • P(X=x)=0 since area at point is zero • P(a≤X≤b)=P(a<X<b)
Cumulative distribution function: $F_X(x)=P(X\le x)=\sum_{k\le x}P_X(k)$ The graph of $F_X(x)$ is a step(staircase) function $P[X=x]=F_X(x+1)-F_X(x)$(Size of the jump or increase in height of the stair)	Cumulative distribution function: $F_X(x)=P(X\le x)=\int_{-\infty}^{x}f_X(x)\,dx$ $\frac{dF_X(x)}{dx}=f_X(x)$ this derivative might be a bit difficult to evaluate at edge points where $f_X(x)$ is not differentiable
Expectation/mean: $E[X]=\sum_x x*P_X(x)$ *exists only when* $\sum_x\lvert x\rvert*P_X(x)<\infty$	Expectation/mean: $E[X]=\int_x x*f_X(x)$ *exists only when* $\int_x\lvert x\rvert*f_X(x)<\infty$
Variance : $Var(X)=V(X)=E[(X-\mu)^2]$ $Var(X)=E[X^2]-[E[X]]^2$ Standard deviation=$\sigma_x=\sqrt{Var(X)}$	Variance: $Var(X)=V(X)=E[(X-\mu)^2]=E[X^2]-[E[X]]^2$ Standard deviation=$\sigma_x=\sqrt{Var(X)}$
Joint PMF: $p_{X,Y}(x,y)=P(X=x\ and\ Y=y)$	**Joint PDF:** $P(a\le X\le a+\delta,c\le Y\le C+\delta)\approx f_{X,Y}(x,y)*\delta^2$ Joint PDF are not probabilities but probabilities per unit area
$\sum_x\sum_y p_{X,Y}(x,y)=1$ $p_{X,Y}(x,y)\ge0$	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{X,Y}(x,y)dxdy=1$ $f_{X,Y}(x,y)\ge0$
$p_X(x)=\sum_y p_{X,Y}(x,y)$ $p_Y(y)=\sum_x p_{X,Y}(x,y)$	$f_X(x)=\int_y f_{X,Y}(x,y)dy$ $f_Y(y)=\int_x f_{X,Y}(x,y)dx$
Expected value rule: $E[g(X,Y)]=\sum_x\sum_y g(x,y)*p_{X,Y}(x,y)$	Expected value rule: $E[g(X,Y)]=\int\int g(x,y)*f_{X,Y}(x,y)dxdy$

Properties of expectation which is true for both continuous and discrete r.v.:
- If X≥0,then E[X]≥0
- If a≤X≤b then a≤E[X]≤b
- E(a)=a where a is a constant. E[a*X+b]=a*E[X]+b where a and b are constants
- E[X+Y]=E[X]+E[Y] more generally E[X1+X2+..+Xn]=E[X1]+E[X2]+..+E[Xn] (this is applicable even if the variables are not independent)
- If X and Y are independent E[XY]=E[X] * E[Y]
- Discrete r.v.:$E[g(X)]=\sum_x g(x)*P_X(x)$
- Continuous r.v: $E[g(X)]=\int_x g(x)*f_X(x)$
- E[g(X)]≠g(E[X]) But if g(x) is linear function(g(X)=a*X+b) then only E[g(X)]=g(E[X])
- When the PMF is symmetric about a pint then that point is the expected value of the PMF
- When the PMF is not symmetric the expected value is interpreted as center of gravity of the PMF
- Discrete r.v:If X takes non negative integer value $E[X]=\sum_{x=1}^{\infty}P(X\ge x)$
- Continuous r.v:If X takes non negative integer value $E[X]=\int_0^{\infty}1-F_X(x)\,dx=\int_0^{\infty}P(X\ge x)\,dx$

Properties of variance which is true for both continuous and discrete r.v.:
- $V(aX+b)=a^2*V(X)$
- V(X)≥0 (ALWAYS)
- If X and Y are independent then V(X+Y)=V(X)+V(Y) or else it is V(X+Y)=V(X)+V(Y)+2*Cov(X,Y)
General form:$V(X_1+X_2+\dots X_n)=\sum_{i=1}^{n}V(X_i)+\sum_{i\ne j}Cov(X_i,X_j)$ (there are n^2-n where i≠j

Properties of CDF which is true for both continuous and discrete r.v.:
- It's always a monotonically increasing function i.e:y≥x then $F_X(y)\ge F_X(x)$
- $F_X(x)$ tends to 1 as $x\to\infty$ i.e: $F_X(\infty)=1$
- $F_X(x)$ tends to 0 as $x\to-\infty$ i.e: $F_X(-\infty)=0$
- F(x) is right-continuous; that is, for every number x_0, $\lim_{x\to x0^+}F(x)=F(x_0)$

Mixed distribution:
$X=\begin{cases}Y \text{ with probability } p\\ Z \text{ with probability } 1-p\end{cases}$
Where Y is discrete r.v. and Z is a continuous r.v.
X is neither continuous nor discrete it is a mixed r.v.
$F_X(x)=p*F_Y(x)+(1-p)*F_Z(x)$
E[X]= p*E[Y] + (1-p)*E[Z]

Indicator r.v:
If X is the indicator of an event A
$X=I_A$ X=1 iff A occur E[IA]=P(A)

Random Variable:
A function from Ω to the real numbers

Sum of a random number of independent r.v.:
$Y=X_1+X_2+\dots+X_N$,where N is a non-negative integer r.v., X_i's are i.i.d and independent of N
E[Y]= E[N]*E[X] Var(Y)= E[N]*Var(X) + $(E[X])^2$*Var(N)

Conditional PMF:	**Conditional PDF:**
$p_{(X\|Y)}(x\|y)=\frac{p_{X,Y}(x,y)}{p_Y(y)}$ where $p_Y(y)>0$ $\sum_x p_{(X\|Y)}(x\|y)=1$	$f_{(X\|Y)}(x\|y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}$ where $f_Y(y)>0$ For small interval δ,Ɛ: $P(a\le X\le a+\delta\|c\le Y\le C+\varepsilon)\approx f_{X,Y}(x,y)*\delta$ The conditional PDF is slice of joint PDF
Total Probability theorem: $p_X(x)=\sum_y p_Y(y)*p_{(X\|Y)}(x\|y)$	Total Probability theorem: $f_X(x)=\int_y f_Y(y)*f_{(X\|Y)}(x\|y)dy$
Expectation: $E(X\|Y)=\sum_x x*p_{(X\|Y)}(x\|y)$ $E[g(X)\|Y]=\sum_x g(x)*p_{(X\|Y)}(x\|y)$ $E[X]=\sum_y p_Y(y)*E(X\|Y=y)$	Expectation: $E(X\|Y=y)=\int_x x*f_{X\|Y}(x\|y)dx$ $E[g(X)\|Y]=\int_x g(x)*f_{X\|Y}(x\|y)dx$ $E[X]=\int_y f_Y(y)E[X\|Y=y]dy$
Independence: $p_{X\|Y}(x)=p_X(x)$ for all x,y $p_{X,Y}(x,y)=p_X(x)*p_Y(y)$for all x,y If X,Y are independent then g(X) and h(y) are also independent and: E[XY]=E[X]*E[Y] E[g(X)*h(Y)]=E[g(X)] * E[h(Y)] If X and Y are independent then V(X+Y)=V(X)+V(Y)	Independence: $f_{X,Y}(x,y)=f_X(x)*f_Y(y)$ for all x,y Also: $f_{(X\|Y)}(x\|y)=f_X(x)$ $F_{X,Y}(x,y)=F_X(x)*F_Y(y)$ If X,Y are independent then g(X) and h(y) are also independent and: E[XY]=E[X]*E[Y] E[g(X)*h(Y)]=E[g(X)] * E[h(Y)] If X and Y are independent then V(X+Y)=V(X)+V(Y)

More than two r.v. joint pmf: The same goes for joint PDF just replace summation by integration
$p_{X,Y,Z}(x,y,z)=P(X=x\ and\ Y=y\ and\ Z=z)$
$\sum_x\sum_y\sum_z p_{X,Y,Z}(x,y,z)=1$ $p_X(x)=\sum_y\sum_z p_{X,Y,Z}(x,y,z)$ $p_{X,Y}(x,y)=\sum_z p_{X,Y,Z}(x,y,z)$

Joint CDF:
$F_{X,Y}(x,y)dxdy=P(X\le x,Y\le y)=\int_y\int_x f_{X,Y}(x,y)dxdy$ $f_{X,Y}(x,y)=\frac{\partial^2F_{X,Y}}{\partial x\partial y}(x,y)$

Covariance:
It tells us if X and Y are moving in same directions or not
Cov(X,Y)=E[(X-E[X])*(Y-E[Y])]= E[XY]-E[X]*E[Y]
If X,Y are independent then Cov(X,Y)=0 converse not true
Properties:
- Cov(X,X)=Var(X)
- Cov(a*X+b,Y)=a*Cov(X,Y)
- Cov(X,Y+Z)=Cov(X,Y)+Cov(X,Z)

Correlation coefficient:
$$\rho(X,Y)=E\left[\left(\frac{X-E[X]}{\sigma_X}\right)*\left(\frac{Y-E[Y]}{\sigma_Y}\right)\right]=\frac{Cov(X,Y)}{\sigma_X*\sigma_Y}$$
Properties:
- $-1\le\rho\le1$ *i.e.* $\lvert\rho\rvert\le1$
- *If X,Y are independent then* ρ *=0* is true converse is not true
- *If |ρ|=1 then (X-E[X])=c*(Y-E[Y]) (X and Y are linearly related)*
- $\rho(a*X+b,Y)=sign(a)*\rho(X,Y)$
- When $\rho=\frac{1}{2}$ this means that X and Y have a underlying, common, hidden factor (e.g:X=V+W Y=V+Z where V is common hidden factor)

Derived distribution-the discrete case:	**Derived distribution-the continuous case :**
Y=g(X) $p_Y(y)=\sum_{x:g(x)=y}p_X(x)$	**Y=g(X)** *Two step procedure:* • *Find CDF of Y=P(Y≤y)=P(g(X)≤y)* • $f_Y(y)=\frac{dF_Y(y)}{dy}$
Y=a*X+b $p_Y(y)=p_X\left(\frac{y-b}{a}\right)$	**Y=a*X+b** $f_Y(y)=\frac{1}{\lvert a\rvert}*f_X\left(\frac{y-b}{a}\right)$

For g(X) being strictly increasing or decreasing we get :
Y=g(X)
$X=g^{-1}(Y)=h(y)$
$f_Y(y)=\lvert\frac{dh(y)}{dy}\rvert*f_X(h(y))$

Derived distribution of multiple r.v.'s:

Z=g(X,Y)=X+Y **Discrete case:** $p_Z(z)=\sum_x p_X(x)*p_Y(z-x)$ Given that X and Y are independent	**Z=g(X,Y)=X+Y** **Continuous case:** Same two step procedure: Find CDF of Z and then differentiate $f_Z(z)=\int_x f_X(x)*f_Y(z-x)\,dx$ Given that X and Y are independent $f_{(Z\|X)}(z\|x)=f_Y(z-x)$ $f_{X,Z}(x,z)=f_X(x)f_Y(z-x)$

Bayes rule one discrete and continuous r.v:

K:Discrete r.v. Y:Continuous r.v.

$p_{K|Y}(k|y) = \frac{p_K(k) * f_{(Y|K)}(y|k)}{f_Y(y)}$

Where $f_Y(y) = \sum_{k'} p_K(k') * f_{(Y|K)}(y|k')$

$f_{(Y|K)}(y|k) = \frac{f_Y(y) * p_{K|Y}(k|y)}{p_K(k)}$

Where $p_K(k) = \int f_{Y'}(y') * p_{K|Y'}(k|y')dy'$

Frequentist	*Bayesian*
Parameter is not a r.v. but a fixed value	*Paremeter is a r.v. with certain prior distribution*
P(a<θ<b)=0 or 1 depending upon if the true parameter lies in the interval	

Bayesian Inference/ Bayesian approach:

1. Unknow θ which is r.v. -Prior distribution (π_θ)
2. Observation X:Model($L_n(X_1, X_2, ..X_n|\theta)$)
3. Posterior distribution: ($\pi(\theta|X_1, X_2, ..X_n)$)

$\pi(\theta|X_1, X_2, ..X_n) \propto \pi_\theta L_n(X_1, X_2, ..X_n|\theta)$

Conditional probability of error:

$P(\hat{\theta} \neq \theta|X = x)$

Overall probability of error: (Depending upon Θ and X been discrete or continous replace the formula with summation or integration)

$P(\hat{\theta} \neq \theta) = \int P(\hat{\Theta} \neq \Theta|X = x)f_X(x)dx$

Or $P(\hat{\theta} \neq \theta) = \sum_\theta P(\hat{\Theta} \neq \theta|X = x)p_\Theta(\theta)$

Maximum a posterior Probability (MAP):

$p_{\theta|X}(\theta^*|X) = \max_\theta p_{\theta|X}(\theta|x)$

$f_{\theta|X}(\theta^*|X) = \max_\theta f_{\theta|X}(\theta|x)$

MAP rule achieves smallest conditional probability of error as well as overall probability or error

Least mean Squares(LMS)

$\hat{\theta}$=E[θ|X=x]

MSE= Var($\theta|X = x$)(Conditional MSE)

MSE=E[Var($\theta|X = x$)] (Overall MSE)

Properties of LMS:

Error:$\tilde{\theta} = \hat{\theta} - \theta$

$E[\tilde{\theta}|X] = 0$

$Cov(\tilde{\theta}, \hat{\theta}) = 0$

Var(θ)=Var($\tilde{\theta}$) + Var($\hat{\theta}$)

Linear Least Mean Square(LLMS):

$\hat{\theta}$=E[θ|X=x]=a*X+b

Where a =$\frac{Cov(\theta,X)}{Var(X)}$ b=E[θ]-a*E[X]

$E(\widehat{\theta_{LLMS}} - \theta) = (1 - \rho^2) * Var(\theta)$

Statistics	*Probability*
Includes estimating paramters deriving C.I. after the experiment has been done (Random)	We already know that the variable has cetain distribution and taking about chances and expectations(deterministic)

Stochastic Process:

Infinite sequence of r.v. where we are interested in $E[X_i]$,$Var[X_i]$,$p(X_1, X_2,....)$,$p(X_i)$.It is also interpreted as set of infinite sequence of 0's and 1's

Bernoulli Process:

A sequence of independent Bernoulli trials (X_i).Each trial has P(Success)=p.It is simplest stochastic process

Assumptions: Independence,Time-homogenity

Fresh-start after a random time N: as long as N is determined causally

The process X_{N+1}, X_{N+2},..is: a bernoulli process,is independent of N, $X_1, X_2, ..X_N$

Geometric : X:Time untill 1[st] success	Pascal distribution Time of the k^{th} success $Y_k = T_1 + T_2 + .. + T_k$
$P[X = x] = (1-p)^{x-1} * p$ x=1,2,.....	$P[Y_k = t] = \binom{t-1}{k-1}(1-p)^{t-k} * p^k$
$E[X] = 1/p$	$E[X] = k/p$
$V[X] = \frac{(1-p)}{p^2}$	$V[X] = \frac{k*(1-p)}{p^2}$

Merging of bernoulli process: when we are merging two independent bernoulli process the resulting process is a bernoulli process and we can compute its P(Success)=p using the two process and each trails are also independent of each other.

Splitting of bernoulli process: Splitting success of a Bernoulli into two streams using independent flips of a coin with bias q then the resulting process gives two bernoulli process which are dependent on each other and one has P(Success)=p*q and the other process had P(Success)=p*(1-q)

Prior Distribution	*Conditional Distribution*	*Posterior Distribution*
$\pi(p) \sim Beta(a,b)$ where $p\epsilon(0,1)$ $\pi(p) = p^{a-1} * (1-p)^{b-1}$	$L_n(X_1,X_2,..X_n\|p) \sim Ber(p)$ $L_n(X_1,X_2,..X_n\|p) = p^{\sum_{i=1}^n X_i} * (1-p)^{n-\sum_{i=1}^n X_i}$	$\pi(p\|X_1,X_2,..X_n) \sim Beta\left(a = a + \sum_{i=1}^n X_i, b = b + n - \sum_{i=1}^n X_i\right)$ $\pi(p\|X_1,X_2,..X_n) \propto p^{a+\sum_{i=1}^n X_i - 1} * (1-p)^{b+n-\sum_{i=1}^n X_i - 1}$
$\pi(p) \sim U(0,1)$ where $p\epsilon(0,1)$ $\pi(p) = 1$	$L_n(X_1,X_2,..X_n\|p) \sim Ber(p)$ $L_n(X_1,X_2,..X_n\|p) = p^{\sum_{i=1}^n X_i} * (1-p)^{n-\sum_{i=1}^n X_i}$	$\pi(p\|X_1,X_2,..X_n) \sim Beta\left(a = 1 + \sum_{i=1}^n X_i, b = 1 + n - \sum_{i=1}^n X_i\right)$ $\pi(p\|X_1,X_2,..X_n) \propto p^{(\sum_{i=1}^n X_i + 1)-1} * (1-p)^{(n-\sum_{i=1}^n X_i+1)-1}$
$\pi(\lambda) \sim Expo(a)$where $\lambda\epsilon(0,\infty)$ $\pi(\lambda) = a * e^{-a*\lambda}$	$L_n(X_1,X_2,..X_n\|\lambda) \sim Poi(\lambda)$ $L_n(X_1,X_2,..X_n\|\lambda) \propto \lambda^{\sum_{i=1}^n X_i} * e^{-n*\lambda}$	$\pi(\lambda\|X_1,X_2,..X_n) \sim Gamma\left(\alpha = 1 + \sum_{i=1}^n X_i, \beta = n + a\right)$ $\pi(\lambda\|X_1,X_2,..X_n) \propto \lambda^{(\sum_{i=1}^n X_i+1)-1} * e^{-\lambda*(n+a)}$
Improper: $\pi(\theta) \propto 1$ where $\theta\epsilon\mathbb{R}$	$L_n(X_1,X_2,..X_n\|\theta) \sim N(\theta,1)$ $L_n(X_1,X_2,..X_n\|\theta) \propto e^{\left\{\frac{-1}{2}\sum_{i=1}^n (X_i-\theta)^2\right\}}$	$\pi(\theta\|X_1,X_2,..X_n) \sim N(\mu = \overline{X}, \sigma^2 = 1/n)$ $\pi(\lambda\|X_1,X_2,..X_n) \propto e^{\left\{\frac{-n}{2}\sum_{i=1}^n (\theta-\overline{X})^2\right\}}$
$\pi(\lambda) \sim Expo(a)$where $\lambda\epsilon(0,\infty)$ $\pi(\lambda) = a * e^{-a*\lambda}$	$L_n(X_1,X_2,..X_n\|\lambda) \sim Expo(\lambda)$ $L_n(X_1,X_2,..X_n\|\lambda) = \lambda^n * e^{-\lambda*\sum_{i=1}^n X_i}$	$\pi(\lambda\|X_1,X_2,..X_n) \sim Gamma\left(k = n + 1, \theta = \frac{1}{\sum_{i=1}^n X_i + a}\right)$ $\pi(\lambda\|X_1,X_2,..X_n) \propto \lambda^{(n+1)-1} * e^{-\lambda*(\sum_{i=1}^n X_i + a)}$
$\pi(\lambda) \sim Gamma(a,b)$where $\lambda\epsilon(0,\infty)$ $\pi(\lambda) \propto \lambda^{a-1} * e^{-\frac{\lambda}{b}}$	$L_n(X_1,X_2,..X_n\|\lambda) \sim Expo(\lambda)$ $L_n(X_1,X_2,..X_n\|\lambda) = \lambda^n * e^{-\lambda*\sum_{i=1}^n X_i}$	$\pi(\lambda\|X_1,X_2,..X_n) \sim Gamma\left(k = n + a, \theta = \frac{1}{\sum_{i=1}^n X_i + \frac{1}{b}}\right)$ $\pi(\lambda\|X_1,X_2,..X_n) \propto \lambda^{(n+a)-1} * e^{-\lambda*(\sum_{i=1}^n X_i + \frac{1}{b})}$
Improper: $\pi(\lambda) \propto 1$where $\lambda\epsilon(0,\infty)$	$L_n(X_1,X_2,..X_n\|\lambda) \sim Expo(\lambda)$ $L_n(X_1,X_2,..X_n\|\lambda) = \lambda^n * e^{-\lambda*\sum_{i=1}^n X_i}$	$\pi(\lambda\|X_1,X_2,..X_n) \sim Gamma\left(k = n + 1, \theta = \frac{1}{\sum_{i=1}^n X_i}\right)$ $\pi(\lambda\|X_1,X_2,..X_n) \propto \lambda^{(n+1)-1} * e^{-\lambda*(\sum_{i=1}^n X_i)}$
Jeffreys Prior: $\pi(\theta) \propto \frac{1}{\theta}$where $\theta\epsilon(0,\infty)$	$L_n(X_1,X_2,..X_n\|\theta) \sim N(0,\theta)$ $L_n(X_1,X_2,..X_n\|\theta) \propto \theta^{\frac{-n}{2}} e^{\left\{\frac{-\sum_{i=1}^n X_i^2}{2*\theta}\right\}}$	$\pi(\|X_1,X_2,..X_n) \sim InverseGamma\left(\alpha = \frac{n}{2}, \beta = \frac{\sum_{i=1}^n X_i^2}{2}\right)$ $\pi(\theta\|X_1,X_2,..X_n) \propto \theta^{\frac{-n}{2}-1} e^{\left\{\frac{-\sum_{i=1}^n X_i^2}{2*\theta}\right\}}$
$\pi(p) = 1$ where $p\epsilon(0,1)$	$L_n(X_1,X_2,..X_n\|p) \sim U(0,p)$ $L_n(X_1,X_2,..X_n\|p) = \frac{1}{p^n} I(X_{(n)} < p)$	$\pi(p\|X_1,X_2,..X_n) \sim Pareto(b = \max(X_i), k = n - 1)$ $E(p\|X_1,X_2,..X_n) = (k*b)/(k-1)$

Improper Prior:

When your prior distribution does not integrate to 1 over its domain then it called as an improper Prior.

Example:$\pi(\lambda) \propto 1$ where $\lambda\epsilon(0,\infty)$

Jeffreys Prior:

$\pi_J(\theta) \propto \sqrt{\det(I(\theta))}$

Poisson Process:

λ is arrival rate, τ is length of interval, N_τ is no. of arriavls in interval of duration τ

Number of arrivals in disjoint time intervals are independent

P(k,τ)= $P(N_\tau = k)$ =Prob. of k arrivals in interval of duration τ

$$P(k,\tau) = \begin{cases} 1 - \lambda\delta & if\ k = 0 \\ \lambda\delta & if\ k = 1 \\ 0 & if\ k > 1 \end{cases}$$

Assumptions: Independence,Time-homogenity

$P(k,\tau) = P(N_\tau = k) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}$ $k = 0,1, ...$ $E[N_\tau] = \lambda\tau$

$Var[N_\tau] = \lambda\tau$ $\lambda = \frac{E[N_\tau]}{\tau}$

Fresh-start and memoryless :

Even if we start wayching at time t we see Poisson process,independent of history until time t

Time T_1 until the first arrival : Exponential (λ) $f(x) = \lambda e^{-\lambda x}$

Memorylessness: Conditioned on T_1>t the pdf of T_1 -t is again exponential

Time Y_k of the k^{th} arrival: Erlang distribution of order k $f_{Y_k}(y) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}$ $y \geq 0$ $E[Y_k]$=k/ λ $Var[Y_k]$=k/ λ^2

$T_k = Y_k - Y_{k-1}$ follows Exp(λ) for k≥2

Merging of independent Poisson process: If two Poisson process with rates λ_1 and λ_2 are merged then the resulting process will be a Poisson process with rate $\lambda_1 + \lambda_2$

Also, from the merging process if we had to trace back to one of the two orginial process then the probability is $\lambda_1/\lambda_1 + \lambda_2$ if we trace it to first similary it would be $\lambda_2/\lambda_1 + \lambda_2$ for second

Splitting of Poisson process: Splitting arrivals of a Poisson into two streams using independent flips of a coin with bias q then the resulting process gives two Poisson process which are independent on each other and one has λ_1=λ*q and the other process had λ_2=λ*(1-q)

Backward Running Poisson process :Poisson process in reverse time is same as the initial Poisson process

Renewal Process:

The interarrival are independent and identically distributed but they have a general distribution

E[X|Y]	*Var(X|Y)*
• It is a function of Y E[X|Y]=g(Y) • It is a random variable • It has its own distribution,mean,variance • It takes a value(number) when Y=y i.e. E[X|Y=y] is not a random variable but a number/constant	• It is a function of Y Var[X|Y]=g(Y) • It is a random variable • It takes a value(number) when Y=y i.e. Var[X|Y=y] is not a random variable but a number/constant
Law of iterated expectations: E[X]=E[EX|Y]]	Law of total variance: Var(X)=E[Var(X|Y)]+Var(E[X|Y]) Var(X)=(average variability within sections)+(variability within sections)
Properties: • E[g(Y)X|Y]=g(Y) * E[X|Y] • If h is an invertible function then E[X|Y]=E[X|h(Y)]	

E[Y|X=x] is often refered as regression function and when it is of the form a*X+b it called linear regression

Generalized linear model:

Y|X=x follows some distribution

Regression function :$g(\mu(x)) = x^T\beta$ where g is called link function it must be monotone increasing(Strictly) and differentiable and range all over real line

Exponential family:

A distribution is said to belong to exponential family if it can be represented as:

$$f_\theta(y) = \exp\left[\sum_{i=1}^k \eta_i(\theta)T_i(y) - B(\theta)\right]h(y)$$

One-parameterexponential family:

$f_\theta(y) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y,\phi)\right)$ $g(\mu) = (b')^{-1}(\mu)$

ϕ is called dispersion parameter if it is known then it is called one parameter exponential family and θ is canonical parameter or else it is two parameter exponential family

E [Y]=b'(θ) Var(Y)=b''(θ) ϕ

Conjugate:

When the Prior and posterior have same family distribution then the prior

Discrete-time infinite state Markov chain:

X_n:state after n transitions

p_{ij}=P(X_1=j|X_0=i)= P(X_{n+1}=j|X_n=i)

Markov property:

"Given current stae,the past doesn't matter"

p_{ij}=P(X_{n+1}=j|X_n=i)= P(X_{n+1}=j|X_n=i, X_{n-1}.., X_0)

n-step transition probabilites:

r_{ij}(n))=P(going from sate I to j in n steps)= =P(X_n=j|X_0=i)= P(X_{n+s}=j|X_s=i)

- $\sum_{j=1}^n r_{ij}(n) = 1$
- $r_{ij}(0) = \begin{cases} 1 & if\ i = j \\ 0 & if\ i \neq j \end{cases}$
- $r_{ij}(1) = p_{ij}\ \forall i,j$
- $r_{ij}(n) = \sum_{k=1}^m r_{ik}(n-1)p_{kj}$
- $r_{ij}(n) = \sum_{k=1}^m p_{ik}r_{kj}(n-1)$

P(X_n=j)= $\sum_{j=1}^m P(X_0 = i) * r_{ij}(n)$

Recurrent states: state I is recurrent if starting from I and from wherever you can go there is a way of returning to I

Transient sate: If a stater is not recurrent then its transient.

Reccurent Class: Acollection of recurrent sate is called a class if within a class all recurrent states have a way to communicate to each other but there is not communication between recurrent states of different classes

Periodic states in recurrent class: The states in a recurrent class are periodic if they can be grouped into d>1 groups so that all transitions from one group lead to the next group

Golden rule: If we have a self transition then Markov chain is aperiodic

Steady state probabilities:

Golden rule: Recurrent states are all in a single classs and this single recurrent class is not periodic $r_{ij} \xrightarrow{n\to\infty} \pi_j$

$$\pi_j = \sum_k \pi_k p_{kj} \quad \sum_{j=1}^m \pi_j = 1$$

Simple Linear regression:

Y=a*X+b+ where (a,b)=$\underset{(a,b)\in\mathbb{R}^2}{argmin}\, E[(Y - a - bX)^2]$

b=$\frac{Cov(X,Y)}{var(X)} = \frac{\overline{XY} - \bar{X}\bar{Y}}{\overline{X^2} - \bar{X}^2}$ a=E[y]-b*E[X]=$\bar{Y} - b * \bar{X}$

Noise:E[Ɛ]=0 and Cov(X, Ɛ)=0

Properties of LSE

- LSE = MLE
- Distribution of $\hat{\beta}$: $\hat{\beta} \sim \mathcal{N}_p\left(\beta^*, \sigma^2(\mathbb{X}^\top\mathbb{X})^{-1}\right)$.
- Quadratic risk of $\hat{\beta}$: $\mathbb{E}\left[\|\hat{\beta} - \beta^*\|_2^2\right] = \sigma^2 tr\left((\mathbb{X}^\top\mathbb{X})^{-1}\right.$
- Prediction error: $\mathbb{E}\left[\|\mathbf{Y} - \mathbb{X}\hat{\beta}\|_2^2\right] = \sigma^2(n-p)$.
- Unbiased estimator of σ^2: $\hat{\sigma}^2 = \frac{1}{n-p}\|\mathbf{Y} - \mathbb{X}\hat{\beta}\|_2^2$.

Theorem

- $(n-p)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$.
- $\hat{\beta} \perp\!\!\!\perp \hat{\sigma}^2$.

Multivariate regression:

Y|X=x is gaussian with mean $\mu(x) = x^T\beta$(Linear)

$Y_i = X_i^T\beta + Ɛ_i$ where $\beta = \underset{\beta\in\mathbb{R}^p}{argmin} \sum_{i=1}^n (Y_i - X_i^T\beta)^2$

$\beta = (X^TX)^{-1}X^TY$ $X\hat{\beta}$ is orthogonal projection of Y onto the subspace spanned by the columns X

Y is a nx1,X is nxp,Ɛ is nx1, β is px1

Assumptions: Ran(X)=p, $Ɛ_i$ are i.i.d,$\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$ where $\sigma^2 > 0$(Homoscedastic)

Significance test: To test β_j is significant where $\gamma_j$is the j-th diagonal coefficient of X^TX

H_0: $\beta_j = 0$ vs H_1: $\beta_j \neq 0$

$$T_n = \frac{\widehat{\beta_j}}{\sqrt{\sigma^2\gamma_j}} \sim t_{n-p}$$

Inequalities	**Required Conditions:**
Hoeffding's Inequality: for all $\varepsilon > 0$ $P(\|X_n - \mu\| \geq \varepsilon) \leq 2 * e^{\frac{-2n\varepsilon^2}{(b-a)^2}}$	$X_1, X_2, .. X_n$ be i.i.d r.v. such that E[X]= μ and X∈ [a, b] n need not be large
Markov inequality: If X≥0 and $\mu = E[X]$is small, then X is unlikely to be very large P(X≥t)≤$\frac{\mu}{t}$	X≥0 with mean $\mu > 0$ and any number t>0 applicable only to non-negative r.v.
Chebyshev inequality: "If the variance is small, then X is unlikely to be too far from the mean" $P(\|X - \mu\| \geq t) \leq \frac{\sigma^2}{t^2}$	X has finite mean μ and variance $\sigma^2$and for any number t>0 Relation: When we substitute $(X - \mu)^2$ in Markov inequality we obtained Chebyshev PRO: When t is larger Chebyshev bound is smaller than Markov.
Jensen inequality: For any convex function g we have E[g(X)] ≥ g(E[X])	*Concave and convex Functions:* A function twice differentiable function h is said to be concave if: h''(θ)≤0. Its said to be strictly concave if h''(θ)<0

Almost surely convergence (a.s.):

A sequence Yn converges to a r.v. Y almost surely iff $P[\{\omega: Y_n(\omega) \xrightarrow[n\to\infty]{} Y(\omega)\}] = 1$

- If $T_n \xrightarrow{(a.s)} T$ then $T_n \xrightarrow{(\mathbb{P})} T$
- $X_n \xrightarrow{(a.s)} a$ $Y_n \xrightarrow{(a.s)} b$ then: $X_n + Y_n \xrightarrow{(a.s)} a+b$
- $X_n \xrightarrow{(a.s)} a$ $Y_n \xrightarrow{(a.s)} b$ then: $X_n * Y_n \xrightarrow{(a.s)} a*b$
- If b≠0 and $X_n \xrightarrow{(a.s)} a$ $Y_n \xrightarrow{(a.s)} b$ then: $X_n/Y_n \xrightarrow{(a.s)} a/b$

Convergence in probability ($\mathbb{P}$):

A sequence Yn converges in probability to a r.v. if for any $\varepsilon > 0$ $\lim_{n\to\infty} P(|Y_n - Y| \geq \varepsilon) = 0$

Properties:

- $X_n \xrightarrow{(\mathbb{P})} a$ $Y_n \xrightarrow{(\mathbb{P})} b$ then: $X_n + Y_n \xrightarrow{(\mathbb{P})} a+b$
- $X_n \xrightarrow{(\mathbb{P})} a$ $Y_n \xrightarrow{(\mathbb{P})} b$ then: $X_n * Y_n \xrightarrow{(\mathbb{P})} a*b$
- If b≠0 and $X_n \xrightarrow{(\mathbb{P})} a$ $Y_n \xrightarrow{(\mathbb{P})} b$ then: $X_n/Y_n \xrightarrow{(\mathbb{P})} a/b$
- If g is continous then $g(X_n) \xrightarrow{(\mathbb{P})} g(a)$
- If $X_n \xrightarrow{(\mathbb{P})} a$ does not imply $E[X_n] \to a$
  Convergence in probablity does not imply convergence of expectation
- Convergence in mean square implies convergence in probability
- Convergence in probability implies convergence in distribution

Convergence in distribution(d):

A sequence Yn converges to a number a in distribution iff $E[f(Y_n)] \xrightarrow[n\to\infty]{} E[f(a)]$ where f is continous and boundedProperties:

- Convergence in distribution implies convergence of probabilites if the limit has a density
- $X_n \xrightarrow{(d)} X$ $Y_n \xrightarrow{(\mathbb{P})} y$ then: $X_n + Y_n \xrightarrow{(d)} X+y$ (Slutsky 's theorem)
- $X_n \xrightarrow{(d)} X$ $Y_n \xrightarrow{(\mathbb{P})} y$ then: $X_n * Y_n \xrightarrow{(d)} X*y$(Slutsky 's theorem)
- If y≠0 and $X_n \xrightarrow{(d)} X$ $Y_n \xrightarrow{(\mathbb{P})} y$ then: $X_n/Y_n \xrightarrow{(d)} X/y$ (Slutsky 's theorem)

Continous Mapping theorem:

If f is a continous function then: $X_n \xrightarrow{(a.s.)/(\mathbb{P})/(d)} X$ implies $f(X_n) \xrightarrow{(a.s.)/(\mathbb{P})/(d)} f(X)$

Central Limit Theorem (CLT): n≥30

$X_1, X_2, .. X_n$ be i.i.d r.v. having finite mean μ and variance σ^2

$\sqrt{n}\frac{\overline{X_n} - \mu}{\sigma} \xrightarrow{(d)} \mathcal{N}(0,1)$ or $\sqrt{n}(\overline{X_n} - \mu) \xrightarrow{(d)} \mathcal{N}(0, \sigma^2)$

When using CLT to approximate probabilites of binomial use the ½ correction rule:

$P(S_n=x)=P(x-1/2<S_n<x+1/2)$ $P(S_n<x)=P(S_n<x-/2)$

Delta Method:

$\sqrt{n}(\overline{Z_n} - \theta) \xrightarrow{(d)} \mathcal{N}(0, \sigma^2)$ for some $\theta \in \mathbb{R}$ and $\sigma^2 > 0$ let g be continously differentiable at point θ then

$\sqrt{n}(g(\overline{Z_n}) - g(\theta)) \xrightarrow{(d)} \mathcal{N}(0, (g'(\theta))^2 * \sigma^2)$

Weak Law of large numbers	*Strong law of large numbers*
$X_1, X_2, .., X_n$be i.i.d. r.v. $\mu = E[X]$ and $\sigma^2 = Var[X]$ as n→∞ $\overline{X_n} \xrightarrow{(\mathbb{P})} \mu$	$X_1, X_2, .., X_n$be i.i.d. r.v. $\mu = E[X]$ and $\sigma^2 = Var[X]$ as n→∞ $\overline{X_n} \xrightarrow{(a.s.)} \mu$

Statistical Model:

$(E,(\mathbb{P}_\theta)_{\theta\in\Theta})$ where E is sample space which does not depend upon the parameter θ, $(\mathbb{P}_\theta)_{\theta\in\Theta}$ is a family of probability measures on E and θ is any set called parameter set

Parametric	Nonparametric
When $\theta \subseteq \mathbb{R}^d$ for some d≥1 and finite dimensions	When θ has infinite dimensions orbelongs to multiple family of distributions

Semiparametric: Mixture of finite dimensions parameters and infinite dimension parameters

Identifiability:

The parameter θ is called identifiable iff map$\theta \in \Theta \mapsto \mathbb{P}_\theta$ is injective:$\theta \neq \theta' \Rightarrow \mathbb{P}_\theta \neq \mathbb{P}_{\theta'}$ or $\mathbb{P}_\theta = \mathbb{P}_{\theta'} \Rightarrow \theta = \theta'$

Estimation:

Estimate	Estimator
It is a number	It is a r.v. . Any statistic whose expression does not depend upon parameter
$\hat{\theta} = g(x)$	$\hat{\theta} = g(X)$

Weakly Consistent estimator: (n indiactes parameter depends on sample)

An estimator $\hat{\theta}$ of θ is weakly consistent if:

$\hat{\theta}_n \xrightarrow{(\mathbb{P})} \theta$(w.r.t $\mathbb{P}_\theta$)

Strongly Consistent estimator:

An estimator $\hat{\theta}$ of θ is weakly consistent if:

$\hat{\theta}_n \xrightarrow{(a.s.)} \theta$(w.r.t $\mathbb{P}_\theta$)

Bias of an estimator:

Bias($\hat{\theta}_n$)=E [$\hat{\theta}_n$]-θ if bais =0 then we say $\hat{\theta}_n$ is an unbiased estimator

Variance of an estimator:

Var($\hat{\theta}_n$)=$E\left[\hat{\theta}_n^{\,2}\right] - (E[\hat{\theta}_n])^2$

Quadratic risk:

R($\hat{\theta}_n$)= $E\left[\left|\hat{\theta}_n - \theta\right|\right]^2$=Var($\hat{\theta}_n$)+ Bias($\hat{\theta}_n$)

Low quadratic risk means both bias and variance are small

Methods of estimation:

Total variation distance	Kullback-Leibler(KL) divergence
Distance between two probability distribution $\mathbb{P}_\theta$and $\mathbb{P}_{\theta'}$.Finding estimator$\hat{\theta}$ such that $\mathbb{P}_\theta$is close to $\mathbb{P}_{\theta^*}$for θ*(true parameter) $\mathrm{TV}(\mathbb{P}_\theta,\mathbb{P}_{\theta'}) = \max_{A\subset E}\lvert\mathbb{P}_\theta(A) - \mathbb{P}_{\theta'}(A)\rvert$	Divergence between two probablility distribution
Discrete: $\mathrm{TV}(\mathbb{P}_\theta,\mathbb{P}_{\theta'}) = \frac{1}{2}\sum_{x\in E}\lvert p_\theta(x) - p_{\theta'}(x)\rvert$ for continous replace submission by integration	Discrete: $\mathrm{KL}(\mathbb{P}_\theta,\mathbb{P}_{\theta'}) = \sum_{x\in E} p_\theta(x)\log\left(\frac{p_\theta(x)}{p_{\theta'}(x)}\right) = E_\theta[\log\left(\frac{p_\theta(x)}{p_{\theta'}(x)}\right)]$ for continous replace submission by integration
Properties: • Symmetric $\mathrm{TV}(\mathbb{P}_\theta,\mathbb{P}_{\theta'}) = \mathrm{TV}(\mathbb{P}_{\theta'},\mathbb{P}_\theta)$ • Positive $0 \leq \mathrm{TV}(\mathbb{P}_\theta,\mathbb{P}_{\theta'}) \leq 1$ • Definite $\mathrm{TV}(\mathbb{P}_\theta,\mathbb{P}_{\theta'}) = 0$ then $\mathbb{P}_\theta = \mathbb{P}_{\theta'}$ • Triangle inequality $\mathrm{TV}(\mathbb{P}_\theta,\mathbb{P}_{\theta'}) \leq \mathrm{TV}(\mathbb{P}_\theta,\mathbb{P}_{\theta''}) + \mathrm{TV}(\mathbb{P}_{\theta''},\mathbb{P}_{\theta'})$	Properties: • Not Symmetric • 0≤ $\mathrm{KL}(\mathbb{P}_\theta,\mathbb{P}_{\theta'})$ • Definite • No Triangle inequality

Statistic:

Any measurable function of the sample e,g: sample mean,sample variance

Robust statistic: When estimators change due to change in sample data it is know as Robust statistic

Steps of calculating MLE and Fishers Information:

- Write the pdf/pmf
- Mulitply it n-times(L_n)
- Take log of this multiplicationand differentiation w.r.t to parameters($\log L_n$)
- Equate first derivative to zero to get the MLE estimate
- For fishers information directly take log of pdf/pmf and then use the formula

Maximum likelihood: Best method out of the three

Maximum likehood is derived from KL where we maximize $\max_{\theta\in\Theta}\prod_{i=1}^n p_\theta(X_i) = L(x_1,x_2,..,x_n,\theta)$ and minimizing KL-divergence

$L_n(x_1,x_2,..,x_n,\theta) = P_\theta[X_1 = x_1, X_2 = x_2,..,X_n = x_n]$

$\hat{\theta}_n^{MLE} = \underset{\theta\in\Theta}{\operatorname{argmax}}\log L_n(x_1,x_2,..,x_n,\theta)$

Properties: Under the conditions:

1. Parameter is idetifiable
2. E(sample space) does not depend upon θ
3. θ^* is not on the boundary of Θ
4. $I(\theta)$ is invertible in neighbourhood of θ^*

And few more conditions MlE estimators satisfy: as n→∞

- $\hat{\theta}_n^{MLE} \xrightarrow{\mathbb{P}} \theta^*$ (Consistent estimator)
- $\sqrt{n}(\hat{\theta}_n^{MLE} - \theta^*) \xrightarrow{(d)} \mathcal{N}(0, I(\theta)^{-1})$

Fishers Information:

It tells you ,on average how curved the function θ→ln[L(x_1, θ)] is

$l(\theta) = \log L_1(X,\theta)$ $I(\theta) = var[\,l'(\theta)\,] = -E[\,l''(\theta)\,]$

Method of moments:

Compare the moments of original distribution which are function of paramters with computed moments from sample and solve for parameter $m_k(\theta) = \widehat{m_k}$ where $m_k(\theta)$ is population moment

M-estimator: No statistical model needs to be assumed

Making a function $\rho: E\times\mathcal{M} \to \mathbb{R}$ where E is our sample space $\mathcal{M}$ is set of all possible values of the unknown parameters such that:$\mathcal{Q}(\theta) = E[\,\rho(X_1,\theta)]$ we are interested in minimum of this function

- $\rho(x,\mu) = (x-\mu)^2$ is minimum at $\mu = E[X]$
- $\rho(x,\mu) = |x-\mu|$is minimum at $\mu = median\ of\ X$
- $\rho(x,\mu) = C_\alpha(x-\mu)$ is minimum at $\mu = \alpha -$ quantile of $\mathbb{P}$ where $C_\alpha(x) = \begin{cases} -(1-\alpha)x & if\ x < 0 \\ \alpha x & if\ x \geq 0 \end{cases}$
- $\rho(x,\mu) = -\log L_1(x,\theta)$ is minimum at $\theta = \hat{\theta}_n^{MLE}$ MLE is special case of m-eestimators

Hypothesis testing:

H_0:null hypothesis H1:alternative hypothesis	• Both should be subset of parameter space and non overlapping • Asymmetrical
Simple:θ=k(k=constant)	Composite:θ>k(k=constant)
One sided:θ<k or θ>k	Two sided:θ≠k
One-sample: When we have only one parameter to test for	Two-sample: When we have two parameter to test in hypothesis

Test: $\psi = \mathbb{I}\{T_n > c\}$ ($T_n$is some statistic and threshold c ∈ ℝ)

A test is a statistics $\psi \in \{0,1\}$ that does not depend on any unknown quantities(parameters).

$\psi = 0$ $H_0$is rejected

$\psi = 1$ $H_0$is accepted

Rejection Region: $\mathcal{R}_\psi = \{T_n > c\}$ ($T_n$is some statistic)={x∈ E^n: ψ(x) = 1}

$\psi = \mathbb{I}\{\mathcal{R}\}$ where $\mathcal{R}$ is an event called rejection region.It is subset of E(sample space)

Error:

Type1:P(Rejecting H_0 | H_0 is true)=α_ψ=$P_{\theta\in\Theta_0}[\,\psi = 1\,]$

Type2: P(Failing to Reject H_0 | H_1 is true)= β_ψ=$P_{\theta\in\Theta_1}[\psi = 0]$

Power of test: $\boldsymbol{\pi_\psi = \inf_{\theta\in\Theta_1}(1-\beta_\psi(\theta))}$=P(Reject H_0 | H_1 is true)

Level of test:

A test ψhas level α if $\alpha_\psi(\theta) \leq \alpha\forall\theta \in \Theta_0$

A test ψhas asymptotic level α if $\lim_{n\to\infty}\alpha_\psi(\theta) \leq \alpha\,\forall\theta \in \Theta_0$

p-value: The (asymptotic) p-value of a test ψ_α is the smallest (asymptotic) level α at which we ψ_α reject H_0.It is random,it depends on the sample.

p-value≤α ⟺ H_0rejected by ψ_α at the (asymptotic) level α

The smaller the p-value ,the more confident one can reject H_0

C.I. for hypothesis testing:

$P_\theta(\theta \in [A,B]) \geq 1-\alpha$ *where [A,B] is C.I.*

Test: $\psi = \mathbb{I}\{\theta_0 \notin [A,B]\}$

Level of test: $P_{\theta\in\Theta_0}[\,\psi = 1\,]$= α

Confidence Invterval (C.I.) of level 1-α : Any random interval I_n(depending only on sample)whose boundaries do not depend upon θ such that:

$$\mathbb{P}_\theta[\,I_n \ni \theta\,] \geq 1-\alpha \quad \forall\theta\in\Theta$$

Asymptotic level 1- α : $\lim_{n\to\infty}\mathbb{P}_\theta[\,I_n \ni \theta\,] \geq 1-\alpha\ \forall\theta\in\Theta$

Remember C.I. are stil random varaibles as they depend on parameters and sample in order to get a interval with numeric values we use:

- **Conservative bound :** Here we plug in those value of estimator which maximizes the width of the interval
- **Solving** we solve the C.I. for the parameter then plug in the roots of quadratic equation
- **Plug-in:** Directly substitute the estimator for parameter in the variane of parameter and compute the interval

Goodness of fit-test:

H_0:p=p_0 H_0:p≠p_0	Test statistic: $$T_n = n\sum_{j=0}^{K}\frac{(\frac{N_j}{n} - p_j^0)^2}{p_j^0} \sim \chi^2_{K-1}$$

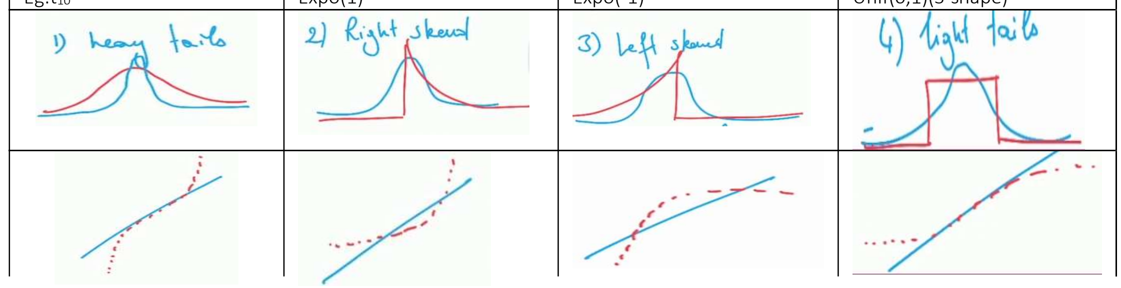
Kolmogorov-Smirnovtest:

Reorder the sample $X_{(1)} \leq X_{(2)} \leq .. X_{(n)}$

H_0:F=F_0 H_0:F≠F_0	Test statistic: $$T_n = \max_{i=1,2,..,n}\left\{\max\left(\left|\frac{i-1}{n} - F^0(X_{(i)})\right|, \left|\frac{i}{n} - F^0(X_{(i)})\right|\right)\right\}$$ $$\sqrt{n}T_n \xrightarrow{(d)} Z$$

Q-Qplot: Compare the plots of data with normal *distribution*

Eg:t_{10}	Expo(1)	Expo(-1)	Unif(0,1)(S-shape)



Mixture of Gaussian:

$X_1 \sim \mathcal{N}(\mu_1,\sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2,\sigma_2^2)$ independent of Z,where Z~Bernoulli(π)

X=$ZX_1 + (1-Z)X_2$

$$f(x) = \pi.\frac{1}{\sigma_1\sqrt{2\pi}}e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}} + (1-\pi).\frac{1}{\sigma_2\sqrt{2\pi}}e^{\frac{-(x-\mu_2)^2}{2\sigma_2^2}}$$

Maximum likelihood:

$$L_n(x_1,\dots,x_n;\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\pi^*) = \prod_{i=1}^n\frac{1}{\sqrt{2\pi}}\left[\frac{\pi^*}{\sigma_1}\exp\left(-\frac{(x_i-\mu_1)^2}{2\sigma_1^2}\right) + \frac{1-\pi^*}{\sigma_2}\exp\left(-\frac{(x_i-\mu_2)^2}{2\sigma_2^2}\right)\right]$$

Moment generating function:

$M_X(t) = \mathrm{Z}M_{X_1}(t) + (1-\mathrm{Z})M_{X_2}(t)$

EM algorithm:

- Initialize $\widehat{\mu_1}, \widehat{\mu_2}$
- Compute weights(E-step):

$$w_i \leftarrow \frac{e^{-\frac{(X_i-\mu_1)^2}{2}}}{e^{-\frac{(X_i-\mu_1)^2}{2}} + e^{-\frac{(X_i-\mu_2)^2}{2}}}, \quad i = 1,\dots,n$$

- Update centers(M-step):

$$\hat{\mu}_1 \leftarrow \frac{\sum_{i=1}^n w_iX_i}{\sum_{i=1}^n w_i}, \qquad \hat{\mu}_2 \leftarrow \frac{\sum_{i=1}^n(1-w_i)X_i}{\sum_{i=1}^n(1-w_i)}$$

ChiSquare: $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n(X_i - \bar{X}_n)^2$ (Expo. Family)

If $Z_1, Z_2, ..Z_d \sim \mathcal{N}(0,1)$ then $V = Z_1^2 + Z_2^2 + .. + Z_d^2 \sim \chi_d^2$

D is degree of freedomE[V]=d Var[V]=2d

Whenχ_2^2~Expo(1/2)

Cochran's theorem:

Cochran's theorem: If $X_1,\dots,X_n \overset{iid}{\sim} \mathcal{N}(\mu,\sigma^2)$, then

- $\frac{(n-1)S_n^2}{\sigma^2} = \sum_{i=1}^n\left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2 \sim \chi^2_{n-1}$.
- $\bar{X}_n$ and S_n^2 are independent r.v.;

Student's T: If $Z \sim \mathcal{N}(0,1)$andV~χ_k^2then $t_k = \frac{Z}{\sqrt{V/k}}$ *is students t with k degrees of freedom*

	Do not reject H_0	Reject H_0
H_0 is true	No discover	***Type 1 error=V***
H_1 is true	***Type 2 error***	***Discovery=D***

FWER(Family wise error rate): The probability of making at least one false discovery or type 1 error.FWER≤α

FDR(False Discovery rate): Number OF false discoveries/Total discoveries=V/V+D

Bonferroni Correction: If you are performing the test m times instead of using α level of significance for each of the test use α/m

Benjamini-Hochberg Correction: Arranged all thep-values of m test in increasing order find k such that p-value of the kth test≤(k*α)/m.Then reject all the H_0till the k-th test

The likehood ratio test is not aplicable on uiform distribution.

The walds test and Likelihood ratio would behave in the same way

The t-test is non asymptotic and works only when parameter of interest will follow normal distribution

Always be careful when computing p-values take in considertion if the hypothesis is two-sided or one

Parametric test	***Walds test***	***T-test***	***Two sample Wald test***	***Two sample t-test***
Test Statistics:	$T_n = \frac{\hat{\theta} - \theta_0}{\sqrt{\widehat{var}(\hat{\theta})}} = \sqrt{nI(\hat{\theta}^{MLE})} * (\hat{\theta}^{MLE} - \theta_0) \xrightarrow{(d)} \mathcal{N}(0,1)$ only guarantees asymptotic level of significance	$T_n = \frac{\sqrt{n}(\overline{X_n} - \mu}{S_n} \sim t_{n-1}$ Applied even when n is small and the parameter for testing is normally distributed	$T_n = \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{\widehat{\sigma_X^2}}{m} + \frac{\widehat{\sigma_Y^2}}{m}}} \xrightarrow{(d)} \mathcal{N}(0,1)$ where $\hat{\theta} = \overline{X_n} - \overline{Y_m}$	$\frac{\bar{X}_n - \bar{Y}_m - (\mu_d - \mu_c)}{\sqrt{\frac{S_d^2}{n} + \frac{S_c^2}{m}}} \sim t_N$ where $N = \frac{(S_d^2/n + S_c^2/m)^2}{\frac{S_d^4}{n^2(n-1)} + \frac{S_c^4}{m^2(m-1)}} \gtrsim \min(n,m)$

Hypothesis:		p-value	Asymptotic p-value:
$H_0: \theta = \theta_0$ $H_1: \theta \neq \theta_0$ $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$\lvert W\rvert > q_{\frac{\alpha}{2}}$ $\lvert T\rvert > q_{\frac{\alpha}{2}}^{t_{n-1}}$	$\mathbb{P}(\lvert W\rvert > \lvert W^{obs}\rvert)$ $\mathbb{P}(\lvert T\rvert > \lvert T^{obs}\rvert)$	$\mathbb{P}(\lvert Z\rvert > \lvert W^{obs}\rvert)$
$H_0: \theta \leq \theta_0$ $H_1: \theta > \theta_1$ $H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$	W>$q_{\frac{\alpha}{2}}$ T>$q_{\frac{\alpha}{2}}^{t_{n-1}}$	$\mathbb{P}(W > W^{obs})$ $\mathbb{P}(T > T^{obs})$	$\mathbb{P}(Z > W^{obs})$
$H_0: \theta \geq \theta_0$ $H_1: \theta < \theta_1$ $H_0: \mu \geq \mu_0$	W<-$q_{\frac{\alpha}{2}}$ T<$-q_{\alpha}^{t_{n-1}}$	$\mathbb{P}(W < W^{obs})$ $\mathbb{P}(T < T^{obs})$	$\mathbb{P}(Z < W^{obs})$

Likelihood Ratio test

Applied only when parameter to test satisfies the asymptotic normality conditions of MLE

$H_0: \theta = \theta_0$

$H_1: \theta = \theta_1$

$: \psi = 1\left\{\frac{L_n(x_1,x_2,..,x_n,\theta_1)}{L_n(x_1,x_2,..,x_n,\theta_0)} > c\right\}$

$$T_n = 2(L_n\left(\widehat{\theta_n}^{MLE}\right) - L_n\left(\widehat{\theta_n}^{C}\right))$$

Where $\widehat{\theta_n}^{C}$ is the constrained MLE

$$L_n\left(\widehat{\theta_n}^{MLE}\right) \geq L_n\left(\widehat{\theta_n}^{C}\right)$$

$$T_n \xrightarrow{(d)} \chi^2_{d-r}$$

d-r are the number of know parameters

Discrete uniform	*Uniform*	*Bernoulli(Expo. Family)*	*Binomial(Expo. Family)*	*Poisson(Expo. Family)*	*Exponential(Expo. Family)*	*Normal(Expo. Family)*
Domain/Sample space: $x \in \{a, a+1, \dots, b\}$	Domain/Sample space: $x \in [a, b]$	Domain/Sample space: X:Occurence of success $x \in \{0,1\}$	Domain/Sample space: X:No of success $x \in \{0,1,..n\}$	Domain/Sample space: $x \in \{0,1,\dots\}$ $(natural\ numbers\ starting\ from\ zero)$	Domain/Sample space: X:Interarrival time (Most common example) $x \in [0,\infty)$	Domain/Sample space: $x \in \mathbb{R}$
Parameters: a,b$\in \mathbb{Z}\ with\ b \geq a$ $or\ n = b - a + 1$	Parameters: a,b$\in \mathbb{R},\ b > a$	Parameters: $0 \leq p \leq 1$	Parameters: n$\in \{0,1,..,\}$,p$\in [0,1]$	Parameters: $\lambda \in (0,\infty)$	Parameters: $\lambda \in (0,\infty)$	Parameters: $\mu \in \mathbb{R}, \sigma^2 > 0$
PMF: $P[X = x] = \frac{1}{b-a+1}\ or\ \frac{1}{n}$ $where\ n = b - a + 1$	PDF: $f(x) = \frac{1}{b-a}$	PMF: $P[X = x] = \begin{cases} q = 1 - p & if\ x = 0 \\ p & if\ x = 1 \end{cases}$ DIFFERENT FORMS: $f(x,p) = p^x * (1-p)^{1-x}$ $f(x,p) = px + (1-p)(1-x)$	PMF: $P[X = x] = \binom{n}{x} p^x (1-p)^{n-x}$ or $P[X = x] = \binom{n}{x} p^x\ q^n - x$ $where\ q = 1 - p$	PMF: $P[X = x] = \frac{\lambda^x e^{-\lambda}}{x!}$	PDF: $f(x) = \lambda e^{-\lambda x}$	PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ $f(x) = c * e^{-(\alpha x^2 + \beta x + \gamma)}$ where α>0 Mean=$\frac{-\beta}{2\alpha}$ and variance=$\frac{1}{2\alpha}$
CDF: $F[X] = \begin{cases} 0 & for\ x < a \\ \frac{x-a+1}{b-a+1} & for\ a \leq x < b \\ 1 & for\ x \geq b \end{cases}$	CDF: $F[X] = \begin{cases} 0 & for\ x < a \\ \frac{x-a}{b-a} & for\ a \leq x < b \\ 1 & for\ x \geq b \end{cases}$	CDF: $F[X] = \begin{cases} 0 & if\ x < 0 \\ q = 1 - p & if\ 0 \leq x < 1 \\ 1 & if\ x = 1 \end{cases}$			CDF: $F[X] = \begin{cases} 0 & for\ x < 0 \\ 1 - e^{-\lambda x} & for\ x \in [0,\infty) \end{cases}$	CDF: $\Phi(z) = P[Z \leq z] = P[-Z \geq -z] = 1 - \Phi(-z)$ $\Phi(1.644854) = 0.95$ $\Phi(1.959964) = 0.975$ $\Phi(1.281552) = 0.90$
Mean E[X]: $E[X] = \frac{a+b}{2}$	Mean E[X]: $E[X] = \frac{a+b}{2}$	Mean E[X]: $E[X] = p$	Mean E[X]: $E[X] = np$	Mean E[X]: $E[X] = \lambda$	Mean E[X]: $E[X] = \frac{1}{\lambda}$	Mean E[X]: $E[X] = \mu$
Variance V[X]: $V[X] = \frac{(b-a)(b-a+2)}{12}$	Variance V[X]: $V[X] = \frac{(b-a)^2}{12}$	Variance V[X]: $V[X] = p(1-p)$	Variance V[X]: $V[X] = np(1-p) = npq$	Variance V[X]: $V[X] = \lambda$	Variance V[X]: $V[X] = \frac{1}{\lambda^2}$	Variance V[X]: $V[X] = \sigma^2$
	MLE: $X_i \sim U(0,b)\ then$ $L_n(x_1, x_2, .., x_n, \boldsymbol{b}) = \frac{1}{b^n} \prod_{i=1}^{n} (\max(x_i \leq b))$ $\widehat{b_{MLE}} = X_{(n)}$ $where\ X_{(n)}\ is\ the\ nth\ order\ statistic$	MLE: $L_n(x_1, x_2, .., x_n, p) = p^{\sum_{i=1}^n x_i} * (1-p)^{\sum_{i=1}^n x_i}$ $\widehat{p_{MLE}} = \sum_{i=1}^n X_i / n = \overline{X_n}$ I(p)=1/p(1-p)		MLE: $L_n(x_1, x_2, .., x_n, \lambda) = \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\prod_{i=1}^n x_i!}$ $\widehat{\lambda_{MLE}} = \sum_{i=1}^n X_i / n = \overline{X_n}$ I(λ)=1/ λ	MLE: $L_n(x_1, x_2, .., x_n, \lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}$ $\widehat{\lambda_{MLE}} = n / \sum_{i=1}^n X_i = 1/\overline{X_n}$ I(λ)=1/ λ^2	MLE: $L_n(x_1, x_2, .., x_n, \mu, \sigma^2) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{\frac{\sum_{i=1}^n -(x-\mu)^2}{2\sigma^2}}$ $\widehat{\mu_{MLE}} = \sum_{i=1}^n X_i / n = \overline{X_n}$ $\widehat{\sigma^2_{MLE}} = \sum_{i=1}^{n} (X_i - \bar{X})^2 / n$ I(μ) = $1/\sigma^2$ I(σ^2)= $1/2\sigma^4$
MGF(Moment Generating Function): $M_X(t) = \frac{e^{at} - e^{(b+1)t}}{n(1-e^t)}$	MGF(Moment Generating Function): $M_X(t) = \frac{e^{bt} - e^{at}}{t(b-a)}$ Higher Moments and Cumulants: $E[X^k] = \frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)}\ where\ k \in \mathbb{N}$	MGF(Moment Generating Function): $M_X(t) = q + pe^t$ Higher Moments and Cumulants: $E[X^k] = p$ $E\left[(X - E(X))^k\right] = \mu_k = (1-p)(-p)^k + p(1-p)^k$ $where\ k \in \mathbb{N}$	MGF(Moment Generating Function): $M_X(t) = (q + pe^t)^n$	MGF(Moment Generating Function): $M_X(t) = e^{\lambda(e^t - 1)}$	MGF(Moment Generating Function): $M_X(t) = \frac{\lambda}{\lambda - t}$ Higher Moments and Cumulants: $E[X^k] = \frac{k!}{\lambda^k}\ where\ k \in \mathbb{N}$	MGF(Moment Generating Function): $M_X(t) = e^{\left(\mu t + \frac{\sigma^2 t^2}{2}\right)}$ Higher Moments and Cumulants: $E[(X-\mu)^k] = \begin{cases} 0\ if\ k\ is\ odd \\ \sigma^2(p-1)!!\ in\ k\ is\ even \\ up\ here\ the\ !!\ stands\ for \\ double\ factorial \end{cases}$ $where\ k \in \mathbb{N}$
		CLT: $\sqrt{n}\frac{(\overline{X_n} - p)}{\sqrt{p * (1-p)}} \sim \mathcal{N}(0,1)$		CLT: $\sqrt{n}\frac{(\overline{X_n} - \lambda)}{\sqrt{\lambda}} \sim \mathcal{N}(0,1)$	CLT: Delta method: $\sqrt{n}\frac{\left(\frac{1}{\overline{X_n}} - \lambda\right)}{\lambda} \sim \mathcal{N}(0,1)$ $\sqrt{n}\frac{(\overline{X_n} - \frac{1}{\lambda})}{\lambda} \sim \mathcal{N}(0,1)$	CLT: $n\frac{(\overline{X_n} - \mu)}{\sigma} \sim \mathcal{N}(0,1)$
		C.I.: $\hat{p} - q_{\frac{\alpha}{2}}\sqrt{\frac{p*(1-p)}{n}} < p < \hat{p} + q_{\frac{\alpha}{2}}\sqrt{\frac{p*(1-p)}{n}}$ Conservative:[$\overline{X_n} - \frac{q_{\frac{\alpha}{2}}}{2*\sqrt{n}}$, $\overline{X_n} + \frac{q_{\frac{\alpha}{2}}}{2*\sqrt{n}}$] Quadratic: $\left(1 + \frac{q_{\frac{\alpha}{2}}}{n}\right)p^2 - \left(2\overline{X_n} + \frac{q_{\frac{\alpha}{2}}}{n}\right)p + \overline{X_n}^2 = 0$ Plug-in: [$\overline{X_n} - q_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X_n}*(1-\overline{X_n})}{n}}$, $\overline{X_n} + q_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X_n}*(1-\overline{X_n})}{n}}$]		C.I.: $\hat{\lambda} - q_{\frac{\alpha}{2}}\sqrt{\frac{\lambda}{n}} < \lambda < \hat{\lambda} + q_{\frac{\alpha}{2}}\sqrt{\frac{\lambda}{n}}$ Conservative:[$-\infty, \infty$] Quadratic: $\lambda^2 - \left(2\overline{X_n} + \frac{q_{\frac{\alpha}{2}}}{n}\right)\lambda + \overline{X_n}^2 = 0$ Plug-in: [$\overline{X_n} - q_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X_n}}{n}}$, $\overline{X_n} + q_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X_n}}{n}}$]	C.I.: $\hat{\lambda} - q_{\frac{\alpha}{2}}\frac{\lambda}{\sqrt{n}} < \lambda < \hat{\lambda} + q_{\frac{\alpha}{2}}\frac{\lambda}{\sqrt{n}}$ Conservative:[$-\infty, \infty$] Quadratic: [$\frac{1}{\overline{X_n}}\left(1 + \frac{q_{\frac{\alpha}{2}}}{\sqrt{n}}\right)^{-1}$, $\frac{1}{\overline{X_n}}\left(1 - \frac{q_{\frac{\alpha}{2}}}{\sqrt{n}}\right)^{-1}$] Plug-in: [$\frac{1}{\overline{X_n}}\left(1 - \frac{q_{\frac{\alpha}{2}}}{\sqrt{n}}\right)$, $\frac{1}{\overline{X_n}}\left(1 + \frac{q_{\frac{\alpha}{2}}}{\sqrt{n}}\right)$]	C.I.: $\hat{\mu} - q_{\frac{\alpha}{2}}\frac{\sigma}{n} < \mu < \hat{\mu} + q_{\frac{\alpha}{2}}\frac{\sigma}{n}$ Conservative:[$-\infty, \infty$] Quadratic: same as plug in Plug-in: [$\overline{X_n} - q_{\frac{\alpha}{2}}\frac{\sigma}{n}$, $\overline{X_n} + q_{\frac{\alpha}{2}}\frac{\sigma}{n}$]
		Transformations: If$X_1, X_2, \dots, X_n \sim Bernolli(p) then$, $\sum_{i=1}^{n} X_i \sim B(n,p)$	Transformations: If X~B(n,p) and Y ~B(m,p) are independent binomial variables with same probability p then X+Y~B(n+m,p) $\frac{X - np}{\sqrt{np(1-p)}} \overset{n \to \infty}{\sim} N(0,1)$ X~B(n,p) $\underset{np = \lambda}{\overset{n \to \infty,\ p \to 0}{\sim}} Pois(\lambda)$	Transformations: If forX_i~Poi(λ_i) i=1,2,..n are independent then $\sum_{i=1}^n X_i$~Poi($\sum_{i=1}^n \lambda_i$) X~Pois(λ) $\overset{\lambda \to \infty}{\sim} N(\lambda, \lambda)$	Transformations: Sum of independent exponential r.v. is not exponential r.v.	Transformations: If forX_i~N(μ_i, σ_i^2) i=1,2,..n are independent then $\sum_{i=1}^n X_i$~N($\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2$) If forX_i~N(μ, σ^2) i=1,2,..n are independent then $\sum_{i=1}^n X_i$~N(n$\mu, n\sigma^2$) If forX_i~N(μ, σ^2) i=1,2,..n are independent then $\sum_{i=1}^n \frac{X_i}{n} = \bar{X}$~N($\mu, \sigma^2/n$)
		Link function:Logit:$\boldsymbol{ln}(\frac{\mu}{1-\mu})$ $\mu = \frac{\exp(X\beta)}{1 + \exp(X\beta)} = \frac{1}{1 + \exp(-X\beta)}$	Link function:Logit:$\boldsymbol{ln}(\frac{\mu}{n-\mu})$ $\mu = \frac{\exp(X\beta)}{1 + \exp(X\beta)} = \frac{1}{1 + \exp(-X\beta)}$	Link function:Log:$\boldsymbol{ln(\mu)}$ $\mu = \exp(X\beta)$	Link function:Neagtive Inverse:$\frac{-1}{\mu}$ $\mu = \frac{\boldsymbol{-1}}{X\beta}$	Link function:Identity: $\boldsymbol{\mu}$ $\boldsymbol{\mu} = X\beta$