Random Variables

Continuous Random Variables

	Uniform Distribution	Exponential Distribution	Gamma Distribution	Normal Distribution
Notations	$U(a,b)$	$Exp(\lambda)$	$Gamma(\alpha,\lambda)$	$N(\mu,\sigma^2)$
Variables	***a*** – Lower bound of the range ***b*** – Upper bound of the range	***λ*** – Expected number of occurrences per unit time	***α*** – Shape parameter $(\alpha>0)$ ***λ*** – Scale parameter $(\lambda>0)$	***μ*** – Mean ***σ*** – Standard Deviation
PDF	$f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & otherwise\end{cases}$ $f(x)=\frac{1}{b-a+1}$	$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases}$	$g(t)=\begin{cases}\frac{\lambda^a}{\Gamma(\alpha)}t^{\alpha-1}e^{-\lambda t}, & t\ge 0\\ 0, & t<0\end{cases}$	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty<x<\infty$
CDF	$F(x)=\begin{cases}0, & x<a\\ \frac{x-a}{b-a}, & a\le x<b\\ 1, & b\le x\end{cases}$	$F(x)=\begin{cases}1-e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}$	$G(t)=\frac{1}{\Gamma(\alpha)}\gamma(\alpha,\beta X)$	$F(x)=\int_{-\infty}^{\infty}f(x)dx$, $-\infty<x<\infty$
E(X)	$E(X)=\frac{a+b}{2}$	$E(X)=\frac{1}{\lambda}E(X^2)=\frac{2}{\lambda}$	$E(X)=\frac{\alpha}{\lambda}$	$E(X)=\mu$
Var(X)	$Var(X)=\frac{(b-a)^2}{12}$	$Var(X)=\frac{1}{\lambda^2}$	$Var(X)=\frac{\alpha}{\lambda^2}$	$Var(X)=\sigma^2$
Note	Standard Uniform Distribution: $U(0,1)$ $U(0,1)\equiv Beta(1,1)$ PDF: $f(x)=1,\ for\ 0\le x\le 1$ $CDF: F(x)=x,\ for\ 0\le x\le 1$	Memoryless Property: $P(T>t+s\mid T>s)=P(T>t)$ Connection with Gamma Distribution: $Exp(\lambda)=Gamma(1,\lambda)$ If we have *n* independent Exponential Random Variables, $X_i\sim Exp(\lambda)$ $X_1+X_2+\cdots+X_n\sim Gamma(n,\lambda)$	Gamma Function: $\Gamma(x)=\int_0^\infty u^{x-1}e^{-u}\,du,\ x>0$ $\Gamma(x)=(x-1)!,\ \Gamma\left(\frac12\right)=\sqrt{\pi}$ $\Gamma\left(\frac{n}{2}\right)=\frac{(n-1)!\sqrt{\pi}}{2^{n-1}\left(\frac{n-1}{2}\right)!}$ Incomplete Gamma Function: Upper Incomplete: $\Gamma(s,x)=\int_x^\infty t^{s-1}e^{-t}\,dt$ Lower Incomplete: $\gamma(s,x)=\int_0^x t^{s-1}e^{-t}\,dt$ $\gamma(s,x)+\Gamma(s,x)=\Gamma(s)$ Suppose $X_1,\cdots,X_n$ are i.i.d $\sim Gamma(\alpha,\lambda)$ $\sum_{i=1}^n X_i\sim Gamma(n\alpha,\lambda)$ $cX_i\sim Gamma\left(\alpha,\frac{\lambda}{c}\right)$	σ rule: $P(\lvert X-\mu\rvert\le 3\sigma)$ $=99.74\%P(\lvert X-\mu\rvert\le 2\sigma)$ $=95.6\%P(\lvert X-\mu\rvert\le\sigma)=68\%$
MGF	$M_x(t)=\frac{e^{\beta t}-e^{\alpha t}}{(\beta-\alpha)t}$	$M_x(t)=\frac{\lambda}{\lambda-t},\ for\ t<\lambda$	$M_x(t)=\left(\frac{\lambda}{\lambda-t}\right)^\alpha$	$M_x(t)=e^{\mu t+\frac{\sigma^2t^2}{2}}$ $X\sim N(\mu_1,\sigma_1^2)$ $Y\sim N(\mu_2,\sigma_2^2)$ $M_{X+Y}(t)=e^{(\mu_1+\mu_2)t+\frac12(\sigma_1^2+\sigma_2^2)t^2}$

	Standard Normal Distribution	Beta Distribution	Chi-Squared Distribution	Cauchy Distribution
Notations	$N(0,1)$	$Beta(\alpha,\beta)$	$\chi^2(k)\ or\ \chi_k^2$	
Variables	-	***α*** – Shape Parameter ***β*** – Shape Parameter	***k*** – Number of independent standard normal random variables	
Probability Mass Function	$\phi(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$	$f(u)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}u^{\alpha-1}(1-u)^{b-1},\ 0\le u\le 1$	$f(x)=\begin{cases}\frac{x^{\frac{k}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)}, & x<0\\ x, & x\ge 0\end{cases}$	$f(x)=\frac{1}{\pi}\left(\frac{1}{1+x^2}\right),\ -\infty<x<\infty$
Continuous Distribution Function	$\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-\frac{t^2}{2}}\,dt$	Regularized Incomplete Beta Function: $I_x(\alpha,\beta)$	$F(x)=\frac{1}{\Gamma\left(\frac{k}{2}\right)}\gamma\left(\frac{k}{2},\frac{x}{2}\right)$	
E(X)	$E(X)=0$	$E(X)=\frac{a}{a+b}$	$E(X)=k$	
Var(X)	$Var(Z)$	$Var(X)=\frac{ab}{(a+b)^2(a+b+1)}$	$Var(X)=2k$	
Note	We can make use of the Z table to compute the values of probability	Beta Function: $B(a,b)=\int_0^1 t^{a-1}(1-t)^{b-1}\,dt=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ Incomplete Beta Function: $B(x;a,b)=\int_0^x t^{a-1}(1-t)^{b-1}\,dt$ Regularized Incomplete Beta Function: $I_x(a,b)=\frac{B(x;a,b)}{B(a,b)}$	Changes the shape parameter: $\chi^2(n)\equiv Gamma\left(\frac{n}{2},\frac{1}{2}\right)$	
MGF	$M_x(t)=e^{\frac{t^2}{2}}$		$M_y(t)=(1-2t)^{-\frac{n}{2}}$	

Discrete Random Variables

	Bernoulli Random Variable	Poisson Distribution	Geometric Distribution
Description	The random variable only has 2 possible outcomes. Probability of one of them is ***p***	The number of events occurring in a fixed time interval or region of opportunity Number of events per single unit of time	Number of *Bernoulli* trials to obtain the **first** success
Notations	$Be(p)$	$Poisson(\lambda)$	$Geom(p)$
Variables	***p*** – Probability of success	***λ*** – Expected number of occurrence Note: $\lambda>0$	***p*** – Probability of success
PMF	$P(X=k)=\begin{cases}p, & k=1\\ 1-p, & k=0\end{cases}$	$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$	$P(X=k)=p(1-p)^{k-1}$
E(X)	$E(X)=p$	$E(X)=\lambda$	$E(X)=\frac{1}{p}$
Var(X)	$Var(X)=p(1-p)$	$Var(X)=\lambda$	$Var(X)=\frac{1-p}{p^2}$
MGF	$M_x(t)=1-p+pe^t$	$M_x(t)=e^{\lambda(e^t-1)}$ $X\sim Poisson(\lambda_1)$ $Y\sim Poisson(\lambda_2)$ $M_{X+Y}(t)=e^{(\lambda_1+\lambda_2)(e^t-1)}$	$M_x(t)=\frac{p\,e^t}{1-(1-p)e^t}$

	Binomial Distribution	Negative Binomial Distribution	Hypergeometric Distribution
Description	Number of successes in ***n*** *Bernoulli* trials	Number of *Bernoulli* trials required to obtain ***r*** successes	Number of successes in a sample size ***m*** from a population of size ***n*** where we have ***r*** items that will equate to success
Notations	$Bin(n,p)$	$NB(r,p)$	$H(r,n,m)$
Variables	***n*** – Number of trials ***p*** – Probability of success	***r*** – Number of Successes ***p*** – Probability of success	***r*** – Number of items selected that will equate to success ***n*** – Population size ***m*** – Sample size
PMF	$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$	$P(X=k)=\binom{k-1}{r-1}p^r(1-p)^{k-r}$	$P(X=k)=\frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$
E(X)	$E(X)=np$	$E(X)=\frac{pr}{1-p}$	$E(X)=m\frac{r}{n}$
Var(X)	$Var(X)=np(1-p)$	$Var(X)=\frac{pr}{(1-p)^2}$	$Var(X)=m\left(\frac{r}{n}\right)\left(\frac{n-r}{n}\right)\left(\frac{n-m}{n-1}\right)$
Note	Normal Approximation: When $n\to\infty,\ q=1-p$ $Bin(n,p)\approx N(np,npq)$ Poisson Approximation: When ***n*** is large and ***p*** is small, ***np*** is moderate $Bin(n,p)\to Poisson\ (np)$	$Geom(p)=NB(1,p)$	For a fixed, ***n*** , E(X) is large means either ***m*** is large, ***r*** is large or both are large
MGF	$M_x(t)=(1-p+pe^t)^n$ X ~ Bin(n, p) Y ~ Bin(m, p) $M_{X+Y}(t)=[1-p+pe^t]^{n+m}$	$M_x(t)=\left(\frac{1-p}{1-pe^t}\right)^r$ $t<-\log p$	

General Properties of Functions of a Random Variable:

1) **Probability Integral Transform**: Distribution of a CDF is uniform on [0, 1]
   $Let\ F(X)\ be\ the\ CDF\ of\ X, Let\ Z=F(X)\to Z\sim U(0,1)$
2) **Inverse Transform Sampling**: To get a random sample from a distribution
   $Let\ U\sim U(0,1),$
   $Let\ X=F^{-1}(U),\Rightarrow CDF\ of\ X\ is\ F$

   Steps:
   a) Get the cdf, F(X) of the random variable X
   b) Get the inverse of the cdf
   c) Deliver the inverse as a random variable and generate out random samples
3) **Manpulating Functions of Distributions**: Suppose that $Y=aX+b$. If we know the distribution of X, we can manipulate to be in terms of X
   CDF of Y: $F_y(y)=P(Y\le y)=P(aX+b\le y)=P\left(X\le\frac{y-b}{a}\right)=F_x\left(\frac{y-b}{a}\right)$
   PMF or PDF: $f_y(y)=\frac{d}{dy}F_y(y)=\frac{d}{dy}F_x(x)=\frac{d}{dy}F_x\left(\frac{y-b}{a}\right)=\frac{1}{a}f_x\left(\frac{y-b}{a}\right)$
4) **Functions of Normal Distribution**: $If\ X\sim N(\mu,\sigma^2)\ and\ Y=aX+b\Rightarrow Y\sim N(a\mu+b,a^2\sigma^2)$
5) **Finding PDF of Function that is differentiate and strictly monotonic**:
   X -Continuous Random Variables, $PDF\ -f(x), Y=g(X)$. Where g is differentiatble, strictly monotonic function
   $$f_Y(y)=f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|$$
6) **Binomial Distribution**: $(a+b)^n=\sum_{k=0}^n x^ky^{n-k}$

Joint Distributions:

Note that Discrete Case is just Sum instead of Integral

Continuous Random Variables:

Joint Density $P((X, Y) \in C) = \int\int_{X,Y \in C} f_{X,Y}(x, y)\, dx\, dy$

Marginal PDF: $f_x(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$ & $f_y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$

Joint Distribution Function: $F_{X,Y}(x, y) = P(X \le x, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)\, dv du \;\; for\ x, y \in \mathbb{R}$

Conversion from PDF to CDF: $f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$

Marginal Distribution Function: $F_x(X) = \lim_{y \to \infty} F_{X,Y}(x, y)$ & $F_y(Y) = \lim_{x \to \infty} F_{X,Y}(x, y)$

Multiple Random Variables:

Suppose $f(x, y, z) = P(X = x, Y = y, Z = z)$:

Marinal Frequency: $f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y, z)\, dy\, dz$

2D Marginal: $f_{XY}(x, y) = \int_{-\infty}^{\infty} f(x, y, z)\, dz$

Useful Formulas:

1) $P(X \in A, Y \in B) = \int_A \int_B f_{X,Y}(x, y)\, dy\, dx$
2) $P(a_1 < X \le a_2, b_1 < Y \le b_2) = \int_{a_1}^{a_2}\int_{b_1}^{b_2} f_{X,Y}(x, y)\, dy\, dx$

Independence:

PDF/PMF: $f_{x,y}(x, y) = f_x(x) f_y(y)$

CDF: $F_{XY}(X, Y) = F_x(X) F_y(Y)$

They are independent if we can factor them into functions of X and Y individually

$F_{XY}(x, y) = g(x)h(y)$ & $f_{XY}(x, y) = g(x)h(y)$

Conditional Distribution:

Continuous Case:

PDF: $f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$ $for\ all\ Y\ such\ that\ f_y(y) > 0 \Leftrightarrow f_{X,Y}(x, y) = f_{X|Y}(x|y) f_Y(y)$

Marginal Density: $f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)\, dy$

CDF: $F_{X|Y}(x|y) = P(X \le x \,|Y = y) = \int_{-\infty}^{x} f_{X|Y}(t|y)\, dt$

Independence: For all Y such that $p_y(y) > 0$: $f_{X|Y}(x|y) = f_x(x)$

Theorems:

1) **Farlie-Morgenstern Family**:

If F(x) and G(y) are one-dimensional CDF, we can choose any $|\alpha| \le 1$

Bivariate Cumulative Distribution Function:

$$H(x, y) = F(x)G(y)\{1 + \alpha[1 - F(x)][1 - G(y)]\}$$

Marginal Distributions:

$$H(x, \infty) = F(x), \qquad H(\infty, y) = G(y)$$

Note:

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

Because they are cdfs so when it is infinity, it is all of the values

Using this Bivariate distribution, we can generate marginals that are of the form that we want

2) **Copula**:

Characteristics:

Have uniform marginal distributions

Nondecreaseing in each variable because it is a cdf

Bivariate Cumulative Distribution Function:

$$C(u, v)$$

Joint Probability Density Function:

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v) \ge 0$$

Marginal Distributions:

$$C(u, 1) = P(U \le u) = uC(1, v) = P(V \le v) = v$$

If $U = F_x(X),\ V = F_y(Y)$

(By Probability Integral Transform, they are uniform random variables)

$$F_{XY}(x, y) = C(F_X(x), F_Y(y))$$

The joint distribution will be a copula since each of the marginals are the cdfs

Any joint distribution will be a copula if we take the marginal cdfs as the marginal distributions for the copula

Joint Density:

$$f_{XY}(x, y) = c\left(F_X(x), F_y(y)\right) f_X(x) f_Y(y)$$

3) **Unform area over some region of space**:

In a plane, the random point (X, Y) is uniform over a region R, if any $A \subset R$. Note that | | denotes area

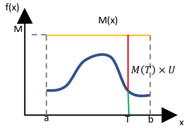
$$P((X, Y) \in A) = \frac{|A|}{|R|}$$

4) **Rejection Method**:

Suppose that we have a density function $f(x)$

Let $M(x)$ be a function such that $M(x) \ge f(x)\ on\ [a, b]$:

$$m(x) = \frac{M(x)}{\int_a^b M(x)\, dx}$$



Algorithm:

1. We will want to choose $M(x) = M\ (constant)$ such that $m(x)$ uniform on [a, b]
2. Generate a random value T with density m (This will be a random point on the line M(x))
3. Generate U, uniform on [0, 1] and independent of T. This will generate out which point it is that we are on the vertical line of M(T). If $M(T) \times U \le f(T)$, then let X = T (accept T). Else, go back to step 2 and reject T.

Functions of Joint Distributions:

Suppose U and V are functions of X and Y, $u = g_1(x, y) v = g_2(x, y), J(x, y) \ne 0$

Joint Density of U and V:

$$f_{UV}(u, v) = f_{XY}\left(h_1(u, v), h_2(u, v)\right)\left|\frac{1}{J\left(h_1(u, v), h_2(u, v)\right)}\right|$$

Note that $h_1(u, v)$ is x represented by u, v only. $h_2(u, v)$ is y represented by u, v only.

Method to find Joint Density:

1) Find the derivative of u and v with respect to the variables
2) Find the Jacobian and the columns will be the variables and the rows will be the functions
3) Represent the Jacobian in terms of u and v only by substituting x and y away
4) Use the formula above where we will sub in only u and v into the joint distribution for X and Y and the Jacobian as well

Jacobian: $J(x, y) = det\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} = \left(\frac{\partial u}{\partial x}\right)\left(\frac{\partial v}{\partial y}\right) - \left(\frac{\partial v}{\partial x}\right)\left(\frac{\partial u}{\partial y}\right)$, Rows – Functions, Columns, Variables

---

Extrema and Order Statistics:

$X_1, X_2, \cdots, X_n$ are independent random variables, have a common CDF F, and density f. (I.I.D)

U – Maximum of X_i

V – Minimum of X_i

Maximum:

CDF: $F_U(u) = P(U \le u) = [F_x(u)]^n$

PDF: $f_U(u) = nf_x(u)[F_X(u)]^{n-1}$ (Differentiate from CDF)

Minimum:

CDF: $F_V(v) = P(V \le v) = 1 - P(V \ge v) = 1 - [1 - F_x(v)]^n$

PDF: $f_V(v) = nf_x(v)[1 - F_x(v)]^{n-1}$ (Differentiate from CDF)

$X_1, X_2, \cdots, X_n$ are independent random variables, have a common CDF F, and density f. (I.I.D)

Order Statistics: $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$

Note: X_1 not necessarily $X_{(1)}$

$X_{(1)}$ – Minimum, $X_{(n)}$ – Maximum

If n is odd, n = 2m+1 , $X_{(m+1)}$ is the median

PDF kth Order Statistics: $f_k(x) = \frac{n!}{(k-1)!(n-k)!} f_x(x)[F_X(x)]^{k-1}[1 - F_X(x)]^{n-k}$

CDF kth Order Statistics: $F(X_{(k)}) \sim Beta(k, n - k + 1)$

Statistical Measures:

Note that for Discrete Case it is Sum instead of Integral

Expected Values:

Continuous Case:

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

Functions of Random Variables: Suppose that $Y = g(x)$

$$E(Y) = \int_{-\infty}^{\infty} g(x)f(x)\, dx, \qquad Y = g(X_1, \cdots, X_n)$$

$$E(Y) = \int\int \cdots \int g(x_1, \cdots, x_n)\, dx_1 \cdots dx_n$$

Independence: If X and Y are independent random variables, g and h are fixed functions

$$E(XY) = E(X)E(Y)$$

$$E(g(X)h(Y)) = E(g(X))E(g(Y))$$

Linear Combinations of Random Variable:

If $X_1, \cdots, X_n$ are jointly distributed random variables with expectations $E(X_i)$, Y = a $+\sum_{i=1}^{n} b_i X_i$

$$E(Y) = a + \sum_{i=1}^{n} b_i E(X_i)$$

Note: $E[g(x)] \ne g[E(X)]$

Condtional Expectation:

$$E(Y|X = x) = \int y f_{Y|x}(y|x)\, dy$$

Functions of Random Variable: Suppose $h(Y)$ is a function of Y

$$E(h(Y)|X = x) = \int h(y) f_{Y|x}(y|x)\, dy$$

Properties:

1) $E(Y) = E[E(Y|X)]$
2) $Var(Y) = Var[(E(Y|X)] + E[Var(Y|X)]$

Sum of Conditional Probabilities:

$$E\left[\sum_{k=1}^{n} X_k | Y = y\right] = \sum_{k=1}^{n} E[X_k | Y = y]$$

If they are independent: $E[\sum_{k=1}^{n} X_k | Y = y] = \sum_{k=1}^{n} X_k$

Expectation of Random Sum:

Suppose that $X_1, \cdots, X_n$ denote the expenditure of the i-th customer, with a common mean E(X)

N is the number of customers entering the store

T is the total revenue from all the customers

$$E(T) = E[E(T|N)], E(T|N = n) = nE(X)$$

$$E(T) = E(N)E(X)$$

Variance:

$$Var(X) = E\left[\left(X - E(X)\right)^2\right] = E(X^2) - [E(X)]^2$$

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

$Var(X) = (\sigma)^2$ where σ is the standard deviation

Properties:

1) Y = aX +b $\Rightarrow Var(Y) = b^2 Var(X)$
2) $Var(X) = 0 \Rightarrow P(X = \mu) = 1$

Covariance:

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y)$$

Independence: If X and Y are independent

$$E(XY) = E(X)E(Y)\ and\ Cov(X, Y) = 0$$

Independence $\Rightarrow Cov(X, Y) = 0$

$Cov(X, Y) = 0 \nRightarrow Independence$

Properties:

1) $Cov(a + X, Y) = Cov(X, Y)$
2) $Cov(aX, bY) = abCov(X, Y)$
3) $Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$ (Cross Product Terms)
4) $Cov(aW + bX, cY + dZ) = ac\, Cov(W, Y) + bc\, Cov(X, Y) + ad\, Cov(W, Z) + bd\, Cov(X, Z)$
5) Suppose that $U = a + \sum_{i=1}^{n} b_i X_i$ and $Y = c + \sum_{i=1}^{n} b_j d_j Cov(X_i, Y_k)$

$$Cov(U, V) = \sum_{i=1}^{n}\sum_{i=1}^{m} b_i d_j Cov(X_i, Y_j)$$

6) $Var(a + \sum_{i=1}^{n} b_i X_i) = \sum_{i=1}^{n}\sum_{i=1}^{n} b_i b_j Cov(X_i, X_j)$
7) $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$, provided X_i are independent
8) $Cov(X, X) = Var(X)$
9) $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
10) $Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$
11) Cov(X, a) = 0, where a is a constant

Correlation Coefficient:

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$\sigma_{XY} = \rho \sigma_x \sigma_y$$

Interpretation:

Range: $-1 \le \rho(X, Y) \le 1$

Results:

1 - Strong Positive Correlation

0 - No Correlation

-1 - Strong Negative Correlation

Note: Independence implies uncorrelated but uncorrelated does not imply independence

---

Prediction:

Mean Squared Error: Predicting the value of Y as a function of X

$MSE = E[(Y - h(X))]^2 = E(E\{[Y - h(X)]^2 \,|\, X\}$

Minimising function: $h(X) = E(Y|X)$

Moment Generating Functions:

The moment generating function (MGF) of a Random Varialbe X is $M(t) = E(e^{tX})$ if the expectation is defined

Discrete Case: $M(t) = E(e^{tX}) = \sum_x e^{tx} p(x)$

Continuous Case: $M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$

Uniqueness Property: If the MGF exists for t in an open interval containing 0, it uniquely determins the probability distribution

Generating Moments: Differentiate *r* times, set *t = 0* and we will get the r[th] moment

$M^{(r)}(0) = E(X^r)$

Linear Combination of Random Variable:

If X has MGF $M_X(t)$ and $Y = a + bX \Rightarrow$ Y has the MGF $M_Y(t) = e^{at} M_X(bt)$

Independence:

If X and Y are independent random variables, $Z = X + Y$. On the common interval where both MGF exists:

$M_Z(t) = M_X(t) M_Y(t)$

Recovering Marginal MGF from Joint MGF:

Suppose $Y = X_1, X_2, \cdots, X_n$ has a joint MGF, $M_Y(s_1, \cdots, s_n) = E(e^{s_1 x_1, \cdots, s_n x_n})$

$M_{X_i}(s_i) = M_Y(0, \cdots, s_i, \cdots, 0)$ (Substitute everything to be 0 instead)

Approximate Methods: **δ Method (Method of Propagation)**

When we have $Y = g(x)$ which means Y is a function of X. We can approximate the expectation and variance of Y by:

- This is useful when we know the mean of X and variance of X. We can use it to find an approximate for Y

$E(Y) \approx g(\mu_X) + \frac{1}{2}\sigma_X^2 g''(\mu_X)$

$Var(Y) = \sigma_X^2 [g'(\mu_X)]^2$

Limit Theorems:

Markov's Inequality: If X is a random variable with $P(X \geq 0) = 1$ and for which $E(X)$ exists

$$P(X \geq t) \leq \frac{E(X)}{t}$$

Chebyshev's Inequality: Gives an upper bound if we know what is the mean and variance

Suppose X is a random variable with mean μ and variance σ^2, for any t > 0

$$P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$$

Law of Large Numbers:

Let $X_1, X_2, \cdots, X_n$ be a seqence of independent random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, $\overline{X_n} = n^{-1}\sum_{i=1}^n X_i$

For any $\epsilon > 0$:

- **Weak Law of Large Numbers**: $P(|\overline{X_n} - \mu| > \epsilon) \to 0$ as $n \to \infty$
  - Converges in Probability to μ
- **Strong Law of Large Numbers**: $P\left(\lim_{n\to\infty} \overline{X_n} = \mu\right) = 1$
  - Converges almost surely to μ

Convergence in Distribution:

1) **Using the Cumulative Distribution**:
   Let $X_1, X_2, \cdots$ be a sequence of Random Variables with CDF $F_1, F_2, \cdots$. Let X be a random variable with CDF F
   $X_n \to X$ (Converges in Distribution) $\Leftrightarrow \lim_{n\to\infty} F_n(x) = F(x)$
2) **Using the Moment Generating Function (Continuity Theorem)**:
   Let F_n be a sequence of CDF with corresponding MGF M_n. Let F be a CDF with MGF M
   $M_n(t) \to M(t)$ for all t in an open interal containing 0 $\Leftrightarrow F_n(x) \to F(x) \Leftrightarrow X_n \to X$ (Converges in Distribution)

Central Limit Theorem:

Let $X_1, X_2, \cdots,$ be a sequence of iid random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. $\overline{X_n} = \frac{\sum_{i=1}^n X_i}{n}$

$$\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} \to N(0,1)$$

It converges in distribution to a standard normal distribution

Note that it is only applicable if $E(X_i)$ is a constant

Distributions from Normal Distribution:

Chi-Squared (χ^2) Distribution:

If Z is a standard normal random variable, the distribution of $U = Z^2$ is called the *chi-squared distribution* with *1* degree of freedom

Characteristics:

1) χ_n^2 where we have *n* degrees of freedom, is a special case of Gamma Distribution: $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$
2) If $X \sim N(\mu, \sigma^2) \Rightarrow [(X-\mu)/\sigma^2] \sim \chi_1^2$
3) If $U_1, U_2, \cdots, U_n$ are independent χ_1^2 random variables, the distribution $V = U_1 + U_2 + \cdots + U_n$ is a χ_n^2 random variable (n degrees of freedom)

t-Distribution:

If $Z \sim N(0,1)$ and $U \sim \chi_n^2$, Z and U are independent, then the distribution $Z/\sqrt{U/n}$ is called the *t distribution* with *n* degrees of freedom

PDF of t_n: $f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)}\left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$

Characteristics:

1) t_1 is the Cauchy Distribution
2) The t-distribution is symmetric about 0. As the degrees of freedom increases, it tends towards a standard normal (~ 20-30 and the distributions will be relatively similar)
3) $E(t) = 0, n > 1$ & $Var(t) = \frac{n}{n-2}, n > 2$

F-Distribution:

Let U and V be independent χ^2 random variables with *m* and *n* degrees of freedom respectively.

The distribution $W = \frac{U/m}{V/n}$ is a F distribution with *m* and *n* degrees of freedom, $F_{m,n}$

PDF of $F_{m,n}$: $f(w) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2} w^{(m/2)-1}\left(1 + \frac{m}{n}w\right)^{-(m+n)/2}, w \geq 0$

Characteristics:

1) Let $T \sim t_n \to T^2 \sim F_{1,n}$

Samples:

Let $X_1, \cdots, X_n$ be independent normal random varaibles

Sample Mean: $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$

Sample Variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$

Characteristics:

1) Random variable $\bar{X}$ and the vector of $(X_1 - \bar{X}, X_2 - \bar{X}, \cdots, X_n - \bar{X})$ are independent
2) $\bar{X}$ and S^2 are independently distributed
3) The distribution $(n-1)S^2/\sigma^2$ is a χ^2 distribution with $n-1$ degrees of freedom
4) $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

Sampling Distribution: Distribution of the Samples

Mean of Sampling Distribution: $E(\bar{X}) = \mu$

Variance of Sampling Distribution: $Var(\bar{X}) = \frac{\sigma^2}{n}$

Standard Error of Sampling Distribution: $\hat{\sigma} = \frac{\sigma}{\sqrt{n}}$

Parameter Estimation and Fitting of Probability Distribution:

Method of Moments:

Steps:

1) Calculate the k[th] moments, $\mu_k = E(X^k)$
2) Find expressions for moments in terms of the parameters that we want to estimate. The number of order of moments that is required is normally the same as the number of parameters $\bar{X} = E(X)$ (Sample Mean)
3) Make the parameters the subject of the equation
4) Subtitute the sample moments inside and this gives us a parameter estimate in terms of the sample moments

Method of Moments Estimate for Distributions:

Poisson Distribution: $\hat{\lambda} = \bar{X}$

Normal Distribution: $\hat{\mu} = \bar{X}$ & $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$

Gamma Distribution: $\hat{\lambda} = \frac{\bar{X}}{\hat{\sigma}^2}$ & $\hat{\alpha} = \frac{\bar{X}^2}{\hat{\sigma}^2}$

Consistent Estimators:

Let $\widehat{\theta_n}$ be an estimate for a parameter based on a sample size n. $\widehat{\theta_n}$ is consistent in probability if $\widehat{\theta_n}$ converges in probability to θ as n approaches ∞ (Makes use of WLLN)

$$\text{For any } \epsilon > 0, \quad P(|\widehat{\theta_n} - \theta| > \epsilon) \to 0, \text{as } n \to \infty$$

Bootstrap: Used to approximate the sampling distribution

Steps:

1) Assume the distribution with parameters, α, β gives a good fit to the data
2) Simulate *N* (i.e. 1000) random samples of size *n* from the distribution with parameters, $\hat{\alpha}, \hat{\beta}$
3) For each random sample, compute the MOM/ MLE estimates for α and β
4) Use the N values of $\hat{\alpha}$ and N values of $\hat{\beta}$ to approximate the sampling distribution of $\hat{\alpha}$ & $\hat{\beta}$

Maximum Likelihood Function:

Suppose we have random variables $X_1, \cdots, X_n$

Steps to find MLE:

1) Find likelihood function:
   $lik(\theta) = f(x_1, x_2, \cdots, x_n | \theta)$
   $= \prod_{i=1}^n f(X_i|\theta)$ (If X_i are independent)
2) Find the log likelihood function
   $$l(\theta) = \sum_{i=1}^n \log(f(X_i|\theta))$$
3) Find the $\frac{\partial l(\theta)}{\partial \theta} = 0$
4) Solve for $\hat{\theta}$ under the above equation

Invariance Property of MLE:

Let $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_k)$ be a MLE of $\theta = (\theta_1, \cdots, \theta_k)$ in the density function $f(x|\theta_1, \cdots, \theta_k)$

If $\tau(\theta) = (\tau_1(\theta), \cdots, \tau_r(\theta)), 1 \leq r \leq k$ is transformation of the parameter space $\underline{\theta}$, then

MLE of $\tau(\theta)$ is $\tau(\hat{\theta}) = \left(\tau_1(\hat{\theta}), \cdots, \tau_r(\hat{\theta})\right)$

Maximum Liklihood Estimate of Multinomial Cell Probabilities:

Suppose $X_1, \cdots, X_m$, the counts in cells $1, \cdots, m$ follows a multinominal distribution with a total count of n and cell probabilties $p_1, \cdots, p_m$

Joint Likelihood Function: $f(x_1, \cdots, x_m | p_1, \cdots, p_m) = \frac{n!}{\prod_{i=1}^m x_i!}\prod_{i=1}^m p_i^{x_i}$

Log Likelihood Function: $l(p_1, \cdots, p_m) = \log n! - \sum_{i=1}^m \log(x_i)! + \sum_{i=1}^m x_i \log(p_i)$

MLE of $\hat{p}_j = \frac{x_j}{n}$

Large Sample Theory for MLE:

Consistency of MLE:

Under appropriate smoothness conditions on f, the MLE from an i.i.d sample is consistent

1) The pdf have common support for all θ (**Uniform Distribution is a good counterexample**)

Information Function:

$I(\theta) = E\left[\frac{\partial}{\partial\theta}\log f(X|\theta)\right]^2 = -E\left[\frac{\partial^2}{\partial^2\theta}\log f(X|\theta)\right]$

Note that we use the marginal likelihood here as compared to the joint likelihood for MLE

Second Form is usually easier to compute but it is only valid under the same smoothness conditions as consistency

Large Sample Distribution of MLE:

Under smoothness conditions (stated above) on f, the MLE with mean θ_0 and variance $\frac{1}{[nI(\theta_0)]}$ is approximately normal

When *n (sample size)* tends to ∞: $\hat{\theta} \sim N\left(\theta_0, \frac{1}{[nI(\theta_0)]}\right) \Rightarrow \frac{\hat{\theta}-\theta_0}{\sqrt{1/n(I(\theta_0))}} \sim N(0,1)$

Asymptotic Variance of MLE: $\frac{1}{nI(\theta_0)}$

Confidence Intervals for MLE:

1) **Exact Method**: When we already know the distribution of the MLE and we can find the confidence interval using the specific distribution
   a. Standard Deviation and Chi Squared: $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-1}^2$
   b. Sample Mean and t distribution: $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$
2) **Approximations**:
   a. Large Sample Theory: $\sqrt{nI(\hat{\theta})}(\hat{\theta} - \theta_0) \sim N(0,1) \Rightarrow \hat{\theta} \pm z(\alpha/2)\frac{1}{\sqrt{nI(\hat{\theta})}}$
3) **Bootstrap**:
   Suppose that $\hat{\theta}$ is an estimate of θ_0
   a. Assume $\hat{\theta} - \theta_0$ is known:
      i. We find $\underline{\delta}$ and $\bar{\delta}$ from the $\hat{\theta} - \theta_0$ distribution such that $P(\underline{\delta} \leq \hat{\theta} - \theta_0 \leq \bar{\delta}) = 1 - \alpha$
      ii. $P(\hat{\theta} - \bar{\delta} \leq \theta_0 \leq \hat{\theta} - \underline{\delta}) = 1 - \alpha$
      iii. Confidence Interval for θ_0: $(\hat{\theta} - \bar{\delta}, \hat{\theta} - \underline{\delta})$
   b. Make use of $\theta^* - \hat{\theta}$ to approximate $\hat{\theta} - \theta_0$

c. Find the $\frac{\alpha}{2}(\underline{\delta}^*)$ and $1-\frac{\alpha}{2}(\overline{\delta}^*)$ quantiles for θ^* (Use Boostrap Estimate to get this distribution)
   i. Find the $\frac{\alpha}{2}(\underline{\delta})$ and $1-\frac{\alpha}{2}(\overline{\delta})$ quantiles for $\theta^*-\hat{\theta}$ by taking:
   ii. $\underline{\delta} = \underline{\delta}^* - \hat{\theta}$ & $\overline{\delta} = \overline{\delta}^* - \hat{\theta}$ (Note that this is just taking a shift downwards from the θ^* distribution)

d. We make use of $\underline{\delta}$ & $\overline{\delta}$ from this distribution into the actual value of $\theta_0: (\hat{\theta}-\overline{\delta}, \hat{\theta}-\underline{\delta})$

Bayesian Approach to Parameter Estimation:

Posterior Distribution: $f_{(\Theta|X)}(\theta|X) = \frac{f_{(X|\theta)}(x|\theta)\, f_\Theta(\theta)}{\int f_{(X|\theta)}(x|\theta)\, f_\Theta(\theta)\, d\theta}$

(Note that the denominator is constant since we are integrating it) Useful fact to find distribution of Posterior with respect to Likelihood and Prior Density

Posterior Density ∝ Likelihood × Prior Density
$f_{(\theta|X)}(\theta|X) \propto f_{(X|\theta)}(X|\theta) \times f_\theta(\theta)$

Large Sample Normal Approximation to Posterior:
When *n (sample size)* tends to ∞, the posterior distribution is **approximately normal** with Mean $= \hat{\theta}\ (MLE\ Estimate)$ & Variance $= -1/[l''(\hat{\theta})]$ (Where $l''(\hat{\theta})$ is the log likelihood function)

Mean Squared Error: $MSE(\hat{\theta}) = E(\hat{\theta}-\theta_0)^2 = Var(\hat{\theta}) + (E(\hat{\theta})-\theta_0)^2$
where $Var(\hat{\theta})$ – Variance of the Estimate, $(E(\hat{\theta})-\theta_0)^2$ - Bias of the Estimate
Unbiased: $E(\hat{\theta}) = \theta$

Efficiency:
Given 2 estimates $\hat{\theta}$ & $\tilde{\theta}$ of a parameter θ, the efficiency of $\hat{\theta}$ relative to $\tilde{\theta}$:

$$eff(\hat{\theta},\tilde{\theta}) = \frac{Var(\tilde{\theta})}{Var(\hat{\theta})}$$

$eff(\hat{\theta},\tilde{\theta}) < 1 \Rightarrow \hat{\theta}\ larger\ variance\ than\ \tilde{\theta}$ & $eff(\hat{\theta},\tilde{\theta}) > 1 \Rightarrow \tilde{\theta}\ larger\ variance\ than\ \hat{\theta}$
Note that it is most meaningful when both $\hat{\theta}$ & $\tilde{\theta}$ are unbiased or have the same bias

Cramer- Rao Inequality:
Let $X_1,\cdots,X_n$ be i.i.d with density function $f(x|\theta)$. Let $T = t(X_1,\cdots,X_n)$ be an **unbiased estimate** of θ. Under smoothness assumptions on $f(x|\theta)$
$Var(T) \geq \frac{1}{nI(\theta)}$ (Note that $I(\theta)$ here is the information function)

This provides us with the lower bound of the variance of an estimate and whether we have reached the minimum variance for it. It will be an efficient estimator and asymptotically efficient and gives us the asymptotic variance
Efficient Estimator: An unbiased estimator that achieves the CR Lower Bound

Sufficiency:
A statistic $T(X_1,\cdots,X_n)$ is said to be **sufficient** for θ if the conditional distribution of $X_1,\cdots,X_n$ given $T=t$, does not depend on θ for any value of t. We call T a **sufficient statistic**.

MLE Relation: If T is suffiicent for θ, the maximum likelihood estimate is a function of T

Factorization Theorem:
A necessary and sufficient condtion for $T(X_1,\cdots,X_n)$ to be sufficeint for a parameter θ is tha the joint PDF factors into the following form:

$$f(x_1,\cdots,x_n|\theta) = g[T(x_1,\cdots,x_n),\theta]\ h(x_1,\cdots,x_n)$$

Try to find a function g such that it is only in the form of $x_1,\cdots,x_n$ with θ, then if the other part of the function is only in terms of $x_1,\cdots,x_n$. The $x_1,\cdots,x_n$ part of g is a sufficient statistics

Exponential Family of Probability Distributions:

Common Distributions: Normal, Binomial, Poisson, Gamma
One Parameter Members have Density Function: $f(x|\theta) = \exp(c(\theta)T(x) + d(\theta) + S(x)), x \in A$.

Rao-Blackwell Theorem:
Let $\hat{\theta}$ be an estimator of θ with $E(\hat{\theta}^2) < \infty$ for all θ.
Suppose T is sufficient for θ, let $\tilde{\theta} = E(\hat{\theta}|T)$. For all $\theta: E(\tilde{\theta}-\theta)^2 \leq E(\hat{\theta}-\theta^2)$
The relation is strict unless $\tilde{\theta} = \hat{\theta}$

Note that the MSE of the new estimator will be lower than the MSE of the original estimate. Having the expectation based on a sufficient statistics improves the MSE

Testing Hypothesis and Goodness of Fit:

Terminologies:
- Null Hypothesis: Original Status Quo that we want to test against
- **Alternate Hypothesis**: Hypothesis that we want to check against status quo
- **Type I Error**: Rejecting H_0 when H_0 is True
- **Type II Error**: Accepting H_0 when H_0 is False
- **Significance Level**: Probability of Type I Error (denoted by α)
- **Probability of Type II Error**: Probability of accepting H_0 when H_0 is False (denoted by β)
- **Power of Test**: Probability of rejecting H_0 when H_0 is False (denoted by $1-\beta$)
- **Test Statistics**: Used to see whether we reject H_0 under certain criterias, e.g. likelihood ratio test
- **Rejection Region**: The set of values of the test statistics that leads to rejection of the null hypothesis
- **Acceptance Region**: The set of values of the test statistics that leads to acceptance of the null hypothesis
- **Null Distribution**: Probability Distribution of the test statistics when the null hypothesis is True
- **Simple Hypothesis**: When the probability distribution is clearly known (i.e. all the parameters are known)
- **Composite Hypothesis**: If the hypothesis does not specify the probability distribution
- **p-value:** Probability of observing the test statistics as extreme as what has been observed when H_0 is True. The smaller the p-value, the stronger the evidence against H_0

Neyman-Pearson Paradigm:
Suppose H_0 and H_1 are simple hypotheses and that the test that rejects H_0 whenever the likelihood ratio is less than c has significance level α. Then any other test for which the significance level is less than or equal to α has power less than or equal to that of the likelihood ratio test

Likelihood Ratio Test:
Likelihood Ratio of $H_0\ \ against\ H_1$

$$L = \frac{P(H_0|x)}{P(H_1|x)} = \frac{P(H_0)}{P(H_1)}\left(\frac{P(x|H_0)}{P(x|H_1)}\right)$$

Reject H_0 if: $L < 1$ or $\left(\frac{P(x|H_0)}{P(x|H_1)}\right) < c$ (We just consider the prior distribution ratio to be a constant)

Accept H_0 if: $L > 1$ or $\left(\frac{P(x|H_0)}{P(x|H_1)}\right) > c$

Uniformly Most Powerful Test:
H_1: Composite Hypothesis
The test is uniformly most powerful if for every simple alternative in H_1 it is the most powerful

E.g. $H_0: \mu \geq \mu_0$, $H_1: \mu < \mu_0$. If for all values of $\mu < \mu_0$, it is the most powerful then it is uniformly most powerful (We will reject small values of $\mu - \mu_0$

Confidence Interval: Consists Precisely of all those values of μ_0 for which the null hypothesis $H_0: \mu = \mu_0$ is accepted.
It is usually a $100(1-\alpha)\%$ confidence interval where α is the level of signfiicance

In comparison with acceptance region: It is the values of the test statistics that will lead to it being accepted. Whereas for confidence interval, we just look at whether the value of μ_0 is within the acceptance region

Duality Theorems for Confidence Intervals and Acceptance Regions:
Suppose that for every value θ_0 in Θ there is a test at level α of the hypothesis test $H_0: \theta = \theta_0$.
We denote the acceptance region of the test by $A(\theta_0)$ – Set of test statistics values, **X**, such that θ_0 is accepted:
$C(X) = \{\theta: X \in A(\theta)\}$ – $100(1-\alpha)\%$ confidence region for θ

Suppose $C(X)$ is a $100(1-\alpha)\%$ confidence region for θ; that is for every θ_0:

$$P[\theta_0 \in C(X)|\theta = \theta_0) = 1-\alpha$$

Then an acceptance region for a test at level α of the hypothesis $H_0: \theta = \theta_0$ is:

$$A(\theta_0) = \{X|\theta_0 \in C(X)\}$$

Generalised Likelihood Ratio Test:
Suppose that the observations $\tilde{X} = (X_1,\cdots,X_n)$ have a joint pdf of $f(\tilde{x}|\theta)$. $H_0: \theta \in \omega_0$ & $H_1: \theta \in \omega_1$
Where $\Omega = \ \omega_0 \cup \omega_1$ and $\omega_0 \cap \omega_1 = \emptyset$

Generalised Likellihood Ratio: $\Lambda^* = \frac{\max_{\theta \in \omega_0}[lik(\theta)]}{\max_{\theta \in \omega_1}[lik(\theta)]}$

Better to use: $\Lambda = \frac{\max_{\theta \in \omega_0}[lik(\theta)]}{\max_{\theta \in \Omega}[lik(\theta)]}$

Formulas

Note that $\Lambda = \min(\Lambda^*, 1)$ and small values of Λ/Λ^* tends to discredit H_0.
$\max_{\theta \in \Omega}[lik(\theta)]$ – Is also the MLE for the region that is specified
Rejection region $\Lambda \leq \lambda_0 \Rightarrow P(\Lambda \leq \lambda_0|H_0) = \alpha$

Note

Large Sample Theory for Generalised Likelihood Ratio Test:
Under smoothness conditions on the frequency function, null distribution of $-2\log\Lambda \sim \chi^2_{\dim\Omega - \dim\omega_0}$ where the $df = \dim\Omega - \dim\omega_0$ as the sample size → ∞

Likelihood Ratio Test for Multinominal Distributions:
x_i is the observed counts in m cells, $\hat{\theta}$ is the MLE of θ, $\hat{p}_i = \frac{x_i}{n}$

$$\Lambda = \prod_{i=1}^{m}\left(\frac{p_i(\hat{\theta})}{\hat{p}_i}\right)^{x_i} \Rightarrow -2log\Lambda = 2\sum_{(i=1)}^{m} O_i \log\left(\frac{O_i}{E_i}\right)$$

$-2\log\Lambda$ will have $df = m-k-1$. m: Number of Cells, k: Number of estimated parameters

Pearson's Chi-Squared Test:
It sees how much the observed values differs from the expected values

$$\chi^2 = \sum_{i=1}^{m}\left(\frac{[x_i - np_i(\hat{\theta})]^2}{np_i(\hat{\theta})}\right) = \sum_{i=1}^{m}\frac{(O_i - E_i)^2}{E_i}$$

$H_0: p_1(\theta),\cdots,p_m(\theta)$ & H_1: otherwise. $O_i = n\hat{p}_i = x_i$ & $E_i = np_i(\hat{\theta})$
Reject H_0 if χ^2 is large
χ^2 will have $df = m-k-1$. m: Number of Cells, k: Number of estimated parameters

Poisson Dispersion Test:
Given counts $x_1,\cdots,x_n$, MLE of $\lambda_i = x_1,\cdots,x_n$. MLE of $\lambda: \hat{\lambda} = \ \bar{X}$
$H_0: Poisson(\tilde{\lambda})$ vs. $H_1: Poisson(\lambda_1),\cdots,Poisson(\lambda_n)$

$$\Lambda = \prod_{i=1}^{n}\left(\frac{\bar{x}}{x_i}\right)^{x_i} e^{x_i - \bar{x}}$$

$$-2\log\Lambda = \ -2\sum_{i=1}^{n}\left[x_i\log\left(\frac{\bar{x}}{x_i}\right) + (x_i - \bar{x})\right]$$

$$\approx \frac{1}{\bar{x}}\sum_{i=1}^{n}(x_i - \bar{x}) \sim \chi^2_{n-1}$$

Since Poisson's variance is λ, we are comparing the dispersion of the variance to see how variable the values are. $df = n-1$ where n: Number of observations.

Variance Stabilisation Transformation:
When we have non-constant variance, we can perform the following transformation
Can be used to find the transformation by finding $f(x)$ or check if the transformation is valid by checking $Var(Y) = constant$

If $Y = f(x), E(X) = \mu,\ Var(X) = \sigma_x^2(\mu)$ (Function of μ),
By Method of Propgation of Error (δ method): $Var(Y) = \sigma_x^2(\mu)[f'(\mu)]^2$

Probability Plots:
Used to check the theoretical against observed values

Checking Non-Linearity:
Given a sample $X_1,\cdots,X_n$ with order statistics $X_{(1)},\cdots,X_{(n)}$ and it assuming it follows a distribution F:
We can plot $F(X_{(k)})\ vs. \frac{k}{n+1} \Leftrightarrow X_{(k)}\ vs. F^{-1}\left(\frac{k}{n+1}\right)$
We know that CDF is a uniform distribution so we can check against a theoretical distribution to check whether the data follows the distribution F

Checking Normality:
Coefficient of Skewness: Checks how symmetric it is. Positive skew means that we have longer right tail and negative skew means we have longer left tail. Uses 3rd Moment

$$b_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^3}{s^3}$$

Where s is the standard deivation. We will reject H_0 that the distribution is Normally Distribution for large values of $|b_1|$

Coefficient of Kurtosis: Fatter tails, higher kurtosis as compared to normal. Smaller tails, lower kurtosis as compared to normal. Uses 4th Moment

$$b_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^4}{s^4}$$