Distribution	pmf/pdf	Mean	Variance	Moment Generating Function: $\psi(t) = E[e^{tX}]$; $E[X^k] = \psi^{(k)}(0)$	Etc.
$X \sim Bernoulli(p)$	$f_X(x) = p^x(1-p)^{1-x} * \mathbb{1}_{\{x=0,1\}}$	$E[X] = p$	$Var(X) = p(1-p)$	$\psi(t) = 1 - p + pe^t$	X =binary event w/ $P(success) = p$ $E[X] = E[X^2] = \cdots = E[X^k] = p$
$X \sim Binomial(n,p)$	$f_X(x) = \binom{n}{x} p^x(1-p)^{n-x} * \mathbb{1}_{\{x=0,1,\dots,n\}}$	$E[X] = np$	$Var(X) = np(1-p)$	$\psi(t) = (1-p+pe^t)^n$	X = # of successes in n trials. $X_i \sim Bin(n_i, p) \rightarrow \sum_{i=1}^n X_i \sim Bin(\sum n_i, p)$
$X \sim Poisson(\lambda)$ $\lambda > 0$	$f_X(x) = e^{-\lambda}\frac{\lambda^x}{x!} * \mathbb{1}_{\{x=0,1,\dots\}}$	$E[X] = \lambda$	$Var(X) = \lambda$	$\psi(t) = \exp\{\lambda(e^t - 1)\}$	X = # of events occurring in a fixed time-period with a known rate λ .
$X \sim Uniform(a,b)$	$f_X(x) = \frac{1}{b-a} * \mathbb{1}_{\{a \le x \le b\}}$	$E[X] = \frac{a+b}{2}$	$Var(X) = \frac{(b-a)^2}{12}$	$\psi(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$	$Unif(0,1) \equiv Beta(1,1)$
$X \sim Exponential(\lambda)$ $\lambda > 0$	$f_X(x) = \lambda e^{-\lambda x} * \mathbb{1}_{\{x \ge 0\}}$	$E[X] = \frac{1}{\lambda}$	$Var(X) = \frac{1}{\lambda^2}$	$\psi(t) = \frac{\lambda}{\lambda - t} * \mathbb{1}_{\{t<\lambda\}}$	Memoryless Distribution: $P(X > s + t \mid X > s) = P(X > t)$
$X \sim Gamma(\alpha,\beta)$ $\alpha, \beta > 0$	$f_X(x) = \frac{\beta^\alpha e^{-\beta x} x^{\alpha-1}}{\Gamma(\alpha)} * \mathbb{1}_{\{x \ge 0\}}$	$E[X] = \frac{\alpha}{\beta}$	$Var(X) = \frac{\alpha}{\beta^2}$	$\psi(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha} * \mathbb{1}_{\{t<\beta\}}$	$X \sim Gamma(1,\beta) \equiv Exponential(\beta)$ $\Gamma(\alpha) = (\alpha - 1)!$; $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
$X \sim Beta(\alpha,\beta)$ $\alpha, \beta > 0$	$f_X(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} * \mathbb{1}_{\{0 \le x \le 1\}}$	$E[X] = \frac{\alpha}{\alpha+\beta}$	$Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	–	$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
$X \sim Normal(\mu,\sigma^2)$ $\sigma^2 < 0$	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} * \mathbb{1}_{\{x \in \mathbb{R}\}}$	$E[X] = \mu$	$Var(X) = \sigma^2$	$\psi(t) = \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$	–
$X \sim \chi_m^2$ CHI-SQUARE DIST. m degrees of freedom	$f_X(x) = \frac{1}{2^{m/2}\Gamma(m/2)} x^{(m/2)-1} e^{-(x/2)} * \mathbb{1}_{\{x>0\}}$	$E[X] = m$	$Var(X) = 2m$	$\psi(t) = (1-2t)^{-m/2} * \mathbb{1}_{\{t<1/2\}}$	$Gamma\left(\alpha = \frac{m}{2}, \beta = \frac{1}{2}\right) \equiv \chi_m^2$ & $Exponential\left(\beta = \frac{1}{2}\right) \equiv \chi_2^2$ $X_i \sim iid\ \chi_{m_i}^2\ \forall i = 1, \dots, k \Longrightarrow X_1 + \cdots + X_k \sim \chi^2_{(\sum_{i=1}^k m_i)}$. If $X_1, \dots, X_n \sim iid\ N(\mu, \sigma^2)$, then $\sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2 \sim \chi_{n-1}^2$ and $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_n^2$.
$X \sim t_m$ t-DISTRIBUTION	$f_X(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{(m\pi)^{1/2}\Gamma(m/2)}\left(1 + \frac{x^2}{m}\right)^{-(m+1)/2} * \mathbb{1}_{\{x \in \mathbb{R}\}}$	–	$Var(X) = \frac{m}{m-2}$, $m > 2$	–	$Y \sim \chi_m^2$ and $Z \sim N(0,1)$: $X = \frac{Z}{\sqrt{Y/m}} \sim t_m$

Prior : $\xi(\theta)$	Likelihood : $f_n(\mathbb{x} \mid \theta)$	Posterior : $\xi(\theta \mid \mathbb{X} = \mathbb{x})$
$\theta \sim Beta(\alpha,\beta)$	$X_1, \dots, X_n \sim iid\ Bernoulli(\theta)$	$Beta\left(\alpha + \sum x_i, \beta + n - \sum x_i\right)$
$\theta \sim Gamma(\alpha,\beta)$	$X_1, \dots, X_n \sim iid\ Poisson(\theta)$	$Gamma\left(\alpha + \sum x_i, \beta + n\right)$
$\theta \sim Gamma(\alpha,\beta)$	$X_1, \dots, X_n \sim iid\ Exponential(\theta)$	$Gamma\left(\alpha + n, \beta + \sum x_i\right)$
$\theta \sim N(\mu_0, v_0^2)$	$X_1, \dots, X_n \sim iid\ N(\theta, \sigma^2)$ $\sigma^2 > 0$ is known.	$N(\mu_1, v_1^2)$ $\mu_1 = \frac{\sigma^2\mu_0 + nv_0^2\bar{x}_n}{\sigma^2 + nv_0^2}$; $v_1^2 = \frac{\sigma^2 v_0^2}{\sigma^2 + nv_0^2}$

Definition $(7.4.3)$: $(Bayes\ Estimator/Estimate)$. Let $L(\theta, a)$ be a loss function. For each possible value $x_1, \dots, x_n$ of $X_1, \dots, X_n$ let $\delta^*(x_1, \dots, x_n)$ be a value of a such that $E[L(\theta,a) \mid x_1, \dots, x_n] = \int_{\theta \in \Omega} L(\theta, a)\xi(\theta \mid x_1, \dots, x_n)d\theta$ is minimized. Then δ^* is called a *Bayes estimator* of θ. Once $X_1 = x_1, \dots, X_n = x_n$ is observed $\delta^*(x_1, \dots, x_n)$ is called a *Bayes estimate* of θ.
Definition $(7.4.4)$: $(Squared\ Error\ Loss\ Function)$. The loss function $L(\theta, a) = (\theta - a)^2$ is called the *squared loss error*.
Corollary $(7.4.1)$: Let θ be a real-valued parameter. Suppose that the squared error loss function $L(\theta, a) = (\theta - a)^2$ is used and the posterior mean of θ, $E[\theta \mid X_1, \dots, X_n]$, is finite. Then the Bayes estimator of θ is $\delta^*(X_1, \dots, X_n) = E[\theta \mid X_1, \dots, X_n]$.
Definition $(7.4.5)$: $(Absolute\ Error\ Loss\ Function)$. The loss function $L(\theta, a) = |\theta - a|$ is called the *absolute error loss*.
For every observed value $x_1, \dots, x_n$, the Bayes estimate $\delta^*(x_1, \dots, x_n)$ will now be the value of a for which the expectation $E[(|\theta - a|) \mid x_1, \dots, x_n]$ is a minimum. This is when a is the median of the posterior distribution.

Proof : Using the property $\sum_{i=1}^n (x_i - \theta)^2 = n(\theta - \bar{x}_n)^2 + \sum_{i=1}^n (x_i - \bar{x}_n)^2$. Note that in the likelihood proportionality we omit terms only dependent on x , i.e., $\sum_{i=1}^n (x_i - \bar{x}_n)^2$. Hence, $f_n(\mathbb{x} \mid \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \theta)^2\right\} \propto \exp\left\{-\frac{1}{2\sigma^2}[n(\theta - \bar{x}_n)^2]\right\} =$. $\xi(\theta) = \frac{1}{\sqrt{2\pi v_0^2}} \exp\left\{-\frac{1}{2v_0^2}(\theta - \mu_0)^2\right\} \propto \exp\left\{-\frac{1}{2v_0^2}(\theta - \mu_0)^2\right\}$. $\xi(\theta \mid \mathbb{X} = \mathbb{x}) = \frac{f_n(\mathbb{x} \mid \theta)\xi(\theta)}{g_n(\mathbb{x})} \propto f_n(\mathbb{x} \mid \theta)\xi(\theta) \propto \exp\left\{-\frac{1}{2\sigma^2}[n(\theta - \bar{x}_n)^2]\right\} \exp\left\{-\frac{1}{2v_0^2}(\theta - \mu_0)^2\right\} = \exp\left\{-\frac{1}{2}\left[\frac{n}{\sigma^2}(\theta - \bar{x}_n)^2 + \frac{1}{v_0^2}(\theta - \mu_0)^2\right]\right\}$. Now we complete the squares again. $\frac{n}{\sigma^2}(\theta - \bar{x}_n)^2 + \frac{1}{v_0^2}(\theta - \mu_0)^2 = \frac{1}{v_1^2}(\theta - \mu_1)^2 + \frac{n}{\sigma^2 + nv_0^2}(\bar{x}_n - \mu_0)^2$. Since the final term no θ relation, drop in proportionality. $\xi(\theta \mid \mathbb{X} = \mathbb{x}) \propto \exp\left\{-\frac{1}{2v_1^2}(\theta - \mu_1)^2\right\}$. $\theta \mid \mathbb{X} = \mathbb{x} \sim N(\mu_1, v_1^2)$ ■

Question 1 : Given a random sample $\{X_i \mid \theta\}_{i \le n} \sim Binomial(m_i, \theta)$. Prior $\theta \sim Beta(\alpha, \beta) \equiv \xi(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$. Let $Y = \sum_{i=1}^n X_i$. Find the PMF of Y. Calculate the posterior distribution and find the Bayes estimator under squared error loss.
Answer 1 : $f_n(\mathbb{x} \mid \theta) = \prod_{i=1}^n \binom{m_i}{x_i}\theta^{x_i}(1-\theta)^{m_i - x_i} = \left[\prod_{i=1}^n \binom{m_i}{x_i}\right] * \theta^{\sum x_i}(1-\theta)^{\sum m_i - \sum x_i}$. $\xi(\theta \mid \mathbb{X} = \mathbb{x}) \propto f_n(\mathbb{x} \mid \theta) * \xi(\theta) \propto \left[\theta^{\sum x_i}(1-\theta)^{\sum m_i - \sum x_i}\right] * \left[\theta^{\alpha-1}(1-\theta)^{\beta-1}\right] = \theta^{\alpha + \sum x_i - 1}(1-\theta)^{\sum m_i - \sum x_i + \beta - 1}$. Therefore, $\theta \mid \mathbb{X} = \mathbb{x} \sim Beta(\alpha + \sum x_i, \beta + \sum m_i - \sum x_i)$. Bayes estimator under square error loss is $E[\theta \mid \mathbb{X}] = (\alpha + \sum x_i)/(\alpha + \beta + \sum m_i)$. Let $M = \sum_{i=1}^n m_i$, the PMF: $X_i \sim Bin(m_i, \theta) \rightarrow Y \mid \theta = \sum_{i=1}^n X_i \sim Bin(M, \theta)$. $f_{Y|\theta}(y \mid \theta) = \binom{M}{y}\theta^y(1-\theta)^{M-y}$. $f_Y(y) = \int_{\theta=0}^1 f_{Y|\theta}(y \mid \theta) * f_\theta(\theta)d\theta = \int_{\theta=0}^1 \left[\binom{M}{y}\theta^y(1-\theta)^{M-y}\right] * \left[\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\right]d\theta = \binom{M}{y}\frac{B(\alpha+y, \beta+M-y)}{B(\alpha,\beta)}\int_{\theta=0}^1 \frac{1}{B(\alpha+y,\beta+M-y)}\theta^{\alpha+y-1}(1-\theta)^{\beta+M-y-1}d\theta = \binom{M}{y}\frac{B(\alpha+y,\beta+M-y)}{B(\alpha,\beta)}$.■
Question 2 : Let $X_1, \dots, X_n \sim iid\ N(\theta, \sigma^2)$ and σ^2 is known. Prior $\theta \sim N(\mu_0, v_0^2)$. Find the bayes estimator for μ under the squared error loss function. **Answer 2** : $\hat{p}_B = E[\mu \mid \mathbb{X} = \mathbb{x}] = (\sigma^2\mu_0 + nv_0^2\bar{x}_n)/(\sigma^2 + nv_0^2) = \mu_0 + \frac{n(\bar{x}_n - \mu_0)}{(\sigma^2/v_0^2) + n}$. Therefore a v_0^2 increases $\hat{p}_B$ increases if $\bar{x}_n > \mu_0$

Question 1 : Show that the $Beta(\alpha, \beta)$ distribution is a conjugate prior for iid data $X_1, \dots, X_n$ from a $Bernoulli(p)$ distribution with unknown parameter p. Explicitly state the parameters of the posterior distribution. **Answer 1** : $f_n(\mathbb{x} \mid \theta) = p^{\sum x_i}(1-p)^{n - \sum x_i}$. $\xi(p) = \frac{1}{B(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1} * \mathbb{1}_{\{0 \le p \le 1\}}$. $\xi(p \mid \mathbb{X} = \mathbb{x}) \propto f_n(\mathbb{x} \mid p) * \xi(p) \propto p^{\alpha + \sum x_i - 1}(1-p)^{\beta + n - \sum x_i - 1}$. Therefore, $p \mid \mathbb{X} = \mathbb{x} \sim Beta(\alpha + \sum x_i, \beta + n - \sum x_i)$. **Question 2** : With the above setup from question 1, find the Bayes estimator $\hat{p}_B$ of p under squared error loss. Next, express $\hat{p}_B$ as a weighted average of the sample mean and the prior mean, i.e., $\hat{p}_B = W_n\bar{X}_n + (1 - W_n)\mu$ where μ is the mean of the $Beta(\alpha, \beta)$ prior. What happens to $\hat{p}_B$ as $n \rightarrow \infty$? . **Answer 2** : $\hat{p}_B = E[p \mid \mathbb{X} = \mathbb{x}] = \frac{\alpha + \sum x_i}{\alpha + \beta + n} = \frac{\alpha}{\alpha + \beta + n} + \frac{n}{\alpha + \beta + n}\bar{X}_n$. Therefore, for $W_n = \frac{n}{\alpha + \beta + n}$ and $\mu = \frac{\alpha}{\alpha + \beta}$, $\hat{p}_B = W_n\bar{X}_n + (1 - W_n)\mu = \frac{n}{\alpha + \beta + n}\bar{X}_n + \left(1 - \frac{n}{\alpha + \beta + n}\right)\mu$. Thus, $\lim_{n \to \infty} \frac{n}{\alpha + \beta + n}\bar{X}_n + \left(1 - \frac{n}{\alpha + \beta + n}\right)\mu = \lim_{n \to \infty} \bar{X}_n = \mu$ by LLN.

Theorem $(7.2.1)$: Suppose that the n random variables $X_1, \dots, X_n$ form a random sample from a distribution for which the pdf is $f(x \mid \theta)$. Suppose also that the value of the parameter θ is unknown and the prior pdf of θ is $\xi(\theta)$. Then the posterior pdf is given by: $\xi(\theta \mid x_1, \dots, x_n) = \frac{\xi(\theta)f(x_1, \dots, x_n \mid \theta)}{g_n(x_1, \dots, x_n)} \overset{iid}{\cong} \frac{\xi(\theta)\prod_{i=1}^n f(x_i \mid \theta)}{g_n(x_1, \dots, x_n)}$ for $\theta \in \Omega$. We can depict this equivalently by: $\xi(\theta \mid x_1, \dots, x_n) \propto f_n(x_1, \dots, x_n \mid \theta) * \xi(\theta)$ .Where the proportionality symbol $\propto$ is used to convey that the left-hand side is equal to the right-hand side except possibly up to a constant. The appropriate constant can be determined by using: $\int_{\theta \in \Omega} \xi(\theta \mid x_1, \dots, x_n)d\theta = 1$ or $g_n(x_1, \dots, x_n) = \int_\Omega f_n(x_1, \dots, x_n \mid \theta)\xi(\theta)d\theta$.

Confidence Intervals : meaning = range of values that you expect your estimate to fall between a certain percentage of the time if you run your experiment again.
$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$. If σ^2 is known and we want a $100(1-\alpha)\%$ CI for μ : $P\left(-z_{\alpha/2} < \frac{\bar{X} - E[\bar{X}]}{\sqrt{Var(\bar{X})}} < z_{\alpha/2}\right) = P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$. ■ $\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim T_{n-1}$ as n increases $s^2 \xrightarrow{p} \sigma^2$ thus $T_{n-1} \rightarrow N(0,1)$. If σ^2 is unknown use s^2 : $100(1-\alpha)\%$ CI for μ : $P\left(-t_{\alpha/2,n-1} < \frac{\bar{X} - E[\bar{X}]}{\sqrt{Var(\bar{X})}} < t_{\alpha/2,n-1}\right) = P\left(\bar{X} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right) = 1 - \alpha$. ■ $X_1, \dots, X_n \sim iid\ N(\mu, \sigma^2)$ where μ, σ^2 are unknown. Construct CI for σ^2. $\sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2 \sim \chi_{n-1}^2 \rightarrow P\left(\chi^2_{1-\frac{\alpha}{2}n-1} < \sum_{i=1}^n \frac{(X_i - \bar{X}_n)^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}n-1}\right) = 1 - \alpha \rightarrow P\left(\frac{1}{\chi^2_{\frac{\alpha}{2}n-1}} < \sum_{i=1}^n \frac{\sigma^2}{(X_i - \bar{X}_n)^2} < \frac{1}{\chi^2_{1-\frac{\alpha}{2}n-1}}\right) = 1 - \alpha$. Since $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. After observing $\mathbb{X} = \mathbb{x}$, this yields a $(1-\alpha)100\%$ CI for σ^2: $P\left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}n-1}}\right) = 1 - \alpha$.

Fisher's Information: $I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\ln f_X(x;\theta)\right)^2\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(x;\theta)\right] = Var\left(\frac{\partial}{\partial\theta}\ln f_X(x;\theta)\right)$ **Proof** : $\int f_X(x;\theta)dx = 1 \rightarrow \int \frac{\partial}{\partial\theta} f_X(x;\theta)dx = \frac{\partial}{\partial\theta}\int f_X(x;\theta)dx = \frac{\partial}{\partial\theta}(1) = 0$. Since $\frac{\partial}{\partial\theta}\ln f_X(x;\theta) = \frac{1}{f_X(x;\theta)} * \frac{\partial}{\partial\theta} f_X(x;\theta)$, then $E\left[\frac{\partial}{\partial\theta}\ln f_X(x;\theta)\right] = \int\left[\frac{\partial}{\partial\theta}\ln f_X(x;\theta)\right] * f_X(x;\theta)dx = \int\left[\frac{1}{f_X(x;\theta)} * \frac{\partial}{\partial\theta}f_X(x;\theta)\right] * f_X(x;\theta)dx = \int \frac{\partial}{\partial\theta}f_X(x;\theta)dx = 0$. Therefore, $E\left[\frac{\partial}{\partial\theta}\ln f_X(x;\theta)\right] = 0$ so $Var\left(\frac{\partial}{\partial\theta}\ln f_X(x;\theta)\right) = E\left[\left(\frac{\partial}{\partial\theta}\ln f_X(x;\theta)\right)^2\right] - \left(E\left[\frac{\partial}{\partial\theta}\ln f_X(x;\theta)\right]\right)^2 = E\left[\left(\frac{\partial}{\partial\theta}\ln f_X(x;\theta)\right)^2\right]$. Next, note that $\frac{\partial^2}{\partial\theta^2}\ln f_X(x;\theta) = \frac{\partial}{\partial\theta}\left\{\frac{1}{f_X(x;\theta)}f_X'(x;\theta)\right\} = -\frac{1}{f_X(x;\theta)^2}f_X'(x;\theta)^2 + \frac{1}{f_X(x;\theta)}f_X''(x;\theta) = -\left(\frac{f_X'(x;\theta)}{f_X(x;\theta)}\right)^2 + \frac{f_X''(x;\theta)}{f_X(x;\theta)}$. So, $-E\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(x;\theta)\right] = -E\left[-\left(\frac{f_X'(x;\theta)}{f_X(x;\theta)}\right)^2 + \frac{f_X''(x;\theta)}{f_X(x;\theta)}\right] = I(\theta) - E\left[\frac{f_X''(x;\theta)}{f_X(x;\theta)}\right] = I(\theta) - E\left[\frac{f_X''(x;\theta)}{f_X(x;\theta)}\right] = I(\theta)$. As $E\left[\frac{f_X''(x;\theta)}{f_X(x;\theta)}\right] = \int f_X''(x;\theta)dx = 0$. ■ More information when $I(\theta)$ is larger…
Theorem : $I_n(\theta) = n * I(\theta)$, i.e., fishers' information for a random sample is simple n times the fishers' information in a single observation. $I_n(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\ln f_n(\mathbb{x};\theta)\right)^2\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\ln f_n(\mathbb{x};\theta)\right]$.■

Theorem (CRLB) : (iid sample size n) If $\widehat{\Theta}$ is an unbiased estimator of θ and $Var(\widehat{\Theta}) \ge \frac{1}{nI(\theta)}$ then $\widehat{\Theta}$ is a minimum variance unbiased estimator. (BIASED ESTIMATOR) $Var(\widehat{\Theta}) \ge \frac{\left(\frac{\partial}{\partial\theta}E[\widehat{\Theta}]\right)^2}{n * I(\theta)}$.
Proof : Using Cauchy-Swartz inequality, $\left(Cov(W,Y)\right)^2 \le Var(W)Var(Y)$. Set $W = \widehat{\Theta}$ and $Y = \frac{\partial}{\partial\theta}\ln f_X(x;\theta) = score$. We know that the expectation of the score is 0, i.e., $E[Y] = 0$. Thus, $\left(Cov(W,Y)\right)^2 \le Var(\widehat{\Theta})Var\left(\frac{\partial}{\partial\theta}\ln f_X(x;\theta)\right)$. Where $Var\left(\frac{\partial}{\partial\theta}\ln f_X(x;\theta)\right) = E\left[\left(\frac{\partial}{\partial\theta}\ln f_X(x;\theta)\right)^2\right] = I(\theta)$ (proved above). Therefore, $Var(\widehat{\Theta}) \ge \frac{(Cov(W,Y))^2}{I(\theta)}$. $\left(Cov(W,Y)\right)^2 = E[WY] - E[W]E[Y] = E[WY] = E\left[\widehat{\Theta}\frac{\partial}{\partial\theta}\ln f_X(x;\theta)\right] = \int\left(\widehat{\Theta}\frac{\partial}{\partial\theta}\ln f_X(x;\theta)\right)f_X(x;\theta)dx = \int \widehat{\Theta}\frac{\partial}{\partial\theta}\{f_X(x;\theta)\}dx = \frac{\partial}{\partial\theta}\int \widehat{\Theta}f_X(x;\theta)dx = \frac{\partial}{\partial\theta}E[\widehat{\Theta}] = \frac{\partial}{\partial\theta}\theta = 1$. Hence, we have proved $Var(\widehat{\Theta}) \ge \frac{1}{I(\theta)}$ .WLOG if we have iid sample $Var(\widehat{\Theta}) \ge \frac{1}{nI(\theta)}$. ■ ↓prof notes: $\sqrt{n}(\hat{\theta} - \theta) \sim^{approx} N\left(0, \frac{1}{I(\theta)}\right)$.
Theorem (Asymptotic Distribution of MLE): Suppose the MLE $\hat{\theta}$ exists and is twice differentiable. Then the asymptotic distribution of $\sqrt{nI(\theta)}(\hat{\theta} - \theta) \sim^{approx} N(0,1)$. Equivalently, the distribution of $\hat{\theta} \sim^{approx} N\left(\theta, \frac{1}{nI(\theta)}\right)$.

Factorization Theorem (sufficiency/joint sufficiency) : $f_{\mathbb{X}}(x_1,\dots,x_n|\theta) = u(x_1,\dots,x_n)v[T_1,\dots,T_k,\theta]$ where $T_j = r_j(x_1,\dots,x_n)$ for $j = 1,\dots,k$

k −parameter exponential family : $f_X(x|\theta) = a(\theta)b(x)\,exp\{\sum_{i=1}^{k} c_i(\theta)d_i(x)\}$

Invariance Property of MLE (f is one-to-one) : If $\hat{\theta}$ is the MLE of θ , then for any function $f(\theta)$, the MLE of $f(\theta)$ is $f(\hat{\theta})$, e.g. let $X_1,\dots,X_n \sim Binomial(1,p)$. $\hat{p}_{MLE} = \bar{x}_n$. If $f(p) = Var_p(x) = p(1-p)$, then the MLE for $Var_p(x)$ is $\hat{p}_{MLE}(1-\hat{p}_{MLE}) = \bar{x}_n(1-\bar{x}_n)$.

Method of Moments: Let $X_1,\dots,X_n$ be a random sample from a population. Method of moment estimation (MOME): Equate sample moments to population moments. If the population has r parameters, the MOME consists of solving the system of r equations $m_k' = \mu_k'$, $k\epsilon\{1,2,\dots,r\}$ for the r parameters, where $m_k' = \frac{1}{n}\sum_{i=1}^{n} x_i^k$, i.e., (kth sample moment) $\mu_k' = E[X_i^k]$ (kth population moment).

$\alpha = P(type\ I\ error) = P(reject\ H_0|H_0 true)$. $\beta = P(type\ II\ error) = P(fail\ to\ reject\ H_0|H_0\ false)$. $\pi = 1 - \beta = P(reject\ H_0|H_0\ false)$. It is customary to refer to the rejection region for H_0 as the critical region of a test. The probability of obtaining a value of the test statistic inside the critical region when H_0 is true is called the **size** of the critical region. Thus, the size of the critical region is just the probability α of committing a type I error.
The only way we can reduce the probabilities of both types of errors is to increase the size of the sample, but as long as n is held fixed, this inverse relationship between the probabilities of type I and type II errors is typical of statistical decision procedures. In other words, if the probability of one type of error is reduced, that of the other type of error is increased.
Note: p-value is area in direction of extremeness of test statistic in direction of alternative calculated assuming H_0 is true. If it is two sides we take sum of both directions, i.e., $P(T \geq a) + P(T \leq -a)$ assuming symmetry here.

Theorem ($Neyman\ Pearson\ Lemma$) : $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$ (simple v simple) & iid sample, $L_0 = f_n(x_1,\dots,x_n;\theta_0) = \prod_{i=1}^{n} f_{X_i}(x_i;\theta_0)$. Reject H_0 if $\frac{L_0}{L_1} \leq k$ is the most powerful test of size α. Note, NP Lemma can be used to test from two different distributions given they are completely specified, i.e., $H_0: f$ vs $H_1: g$.

Example (Discrete case) : If $X \sim Bin(2,\theta)$ is used to test $H_0: \theta = \frac{1}{2}$ vs $H_1: \theta = \frac{3}{4}$ of size α . Using NP lemma, $L_0 = \binom{n}{x}\theta_0^x(1-\theta_0)^{n-x} = \begin{cases} 1/4\ , x = 0 \\ 1/2\ , x = 1 \\ 1/4\ , x = 2 \end{cases}$ and $L_1 = \binom{n}{x}\theta_1^x(1-\theta_1)^{n-x} = \begin{cases} 1/16\ , x = 0 \\ 3/8\ , x = 1 \\ 9/16\ , x = 2 \end{cases}$ thus $\frac{L_0}{L_1} = \begin{cases} 4\ , x = 0 \\ 4/3\ , x = 1 \\ 4/9\ , x = 2 \end{cases}$. To determine a test of size $\alpha = P(reject\ H_0|H_0\ true)$ in discrete case, $k \geq 4 \rightarrow P\left(\frac{L_0}{L_1} \leq k|H_0\ true\right) = P(X \in \{0,1,2\}) = 1$ & $\frac{4}{3} \leq k < 4 \rightarrow P\left(\frac{L_0}{L_1} \leq k|H_0\ true\right) = P(X \in \{1,2\}) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$ & $\frac{4}{9} \leq k < \frac{4}{3} \rightarrow P\left(\frac{L_0}{L_1} \leq k|H_0\ true\right) = P(X = 2) = \frac{1}{4}$ & $k < \frac{4}{9} \rightarrow P\left(\frac{L_0}{L_1} \leq k|H_0\ true\right) = P(X \in \emptyset) = 0$. Therefore, if we want some test of maximum $\alpha \leq \alpha_0$ we select k such that this exists.

Theorem (LRT) : Suppose we wish to test $H_0: \theta \in \omega_0$ vs $H_1: \theta \in \omega_1$. $\omega_0 \cup \omega_1 = \Omega$; $\omega_0 \cap \omega_1 = \emptyset$. The statistic $\Lambda(\mathbb{x})$ is called the likelihood ratio statistic. A likelihood ratio test is to reject $H_0: \theta \in \omega_0$ if $\Lambda(\mathbb{x}) \leq k$ for some constant k.
$\Lambda(\mathbb{x}) = \frac{\sup_{\theta\in\Omega_0} f_n(\mathbb{x}|\theta)}{\sup_{\theta\in\Omega} f_n(\mathbb{x}|\theta)} = \frac{f_n(\mathbb{x}|\hat{\theta})}{f_n(\mathbb{x}|\hat{\hat{\theta}})}$ where $\hat{\theta} = \underset{\theta\in\omega_0}{\operatorname{argmin}} f_n(\mathbb{x}|\theta)$ and $\hat{\hat{\theta}} = \underset{\theta\in\Omega}{\operatorname{argmin}} f_n(\mathbb{x}|\theta)$. **Prof Notes** : (i) Test $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0 \rightarrow$ reject if $\left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right| \geq c$. (ii) $X_{1i} \sim iid\ N(\mu_1,\sigma^2)$, $X_{2i} \sim iid\ N(\mu_2,\sigma^2) \rightarrow T = \frac{\bar{X}_1-\bar{X}_2}{s_p\left[\frac{1}{n_1}+\frac{1}{n_2}\right]^{1/2}} \sim t_{n_1+n_2-2}$ where $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$

Example : Given a random sample $X_1,\dots,X_n \sim N(\mu,\sigma^2)$ and σ^2 is known. Determine a test of size α for $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$. **Answer** : $\hat{\theta} = \mu_0$ and $\hat{\hat{\theta}} = \hat{\theta}_{MLE} = \bar{x}_n$. Some algebra yields $\Lambda(\mathbb{x}) = \exp\left\{-\frac{n}{2\sigma^2}|\bar{x}_n - \mu_0|^2\right\}$. Reject H_0 if $\Lambda(\mathbb{x}) \leq k$ gives the test reject H_0 if $|\bar{x}_n - \mu_0| \geq K$. $\alpha = P(|\bar{x}_n - \mu_0| \geq K|\mu = \mu_0) = P\left(|Z| \geq \frac{K}{\sigma/\sqrt{n}}\right) = P\left(Z \geq \frac{K}{\sigma/\sqrt{n}}\right) + P\left(Z \leq -\frac{K}{\sigma/\sqrt{n}}\right)$. Therefore, $z_{\alpha/2} = \frac{K}{\sigma/\sqrt{n}} \rightarrow K = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. The test is to reject H_0 if $|\bar{x}_n - \mu_0| \geq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$;CI from the compliment, i.e. acceptance region.

Theorem ($Wilks'$; $r = 1$) : For large n, the distribution of $-2\ln\Lambda(\mathbb{x})$ converges to the chi-square distribution with 1 degree of freedom, i.e., $-2\ln\Lambda(\mathbb{x}) \sim^{approx} \chi_1^2$. For parameters $r > 1$, $-2\ln\Lambda(\mathbb{x}) \sim^{approx} \chi_r^2$. ■

Example : Check using $\Lambda(\mathbb{X}) = \exp\left\{-\frac{n}{2\sigma^2}|\bar{X}_n - \mu_0|^2\right\}$ for the normal LRT above. $-2\ln\Lambda(\mathbb{X}) = \frac{n}{\sigma^2}(\bar{x}_n - \mu_0)^2 = \left(\frac{\bar{X}_n-\mu_0}{\sigma/\sqrt{n}}\right)^2 = Z^2 \sim \chi_1^2$ ■

Definition (Monotone Likelihood Ratio) : Let $f_n(\mathbb{x}|\theta)$ denote the joint pdf of the observations $\mathbb{X} = \mathbb{x}$. Let $T = r(\mathbb{x})$ be a statistic. It is said that the joint distribution of $\mathbb{X}$ has a *monotone likelihood ratio* (MLR) in the statistic T if the following property is satisfied: For every two values $\theta_1 \in \Omega$ and $\theta_2 \in \Omega$, with $\theta_1 < \theta_2$, the ratio $f_n(\mathbb{x}|\theta_2)/f_n(\mathbb{x}|\theta_1)$ depends on the vector $\mathbb{x}$ only through the function $r(\mathbb{x})$, and this ratio is a monotone function of $r(\mathbb{x})$ over the range of possible values of $r(\mathbb{x})$. Specifically, if the ratio is increasing, we say the distribution of $\mathbb{X}$ has increasing MLR, and if the ratio is decreasing, we say that the distribution has decreasing MLR.

Two sample t-test (equal var): $H_0: \mu_1 - \mu_2 = \lambda$, $H_1: \mu_1 - \mu_2 \neq \lambda$. $T = \frac{(n_1+n_2-2)^{1/2}(\bar{X}_1-\bar{X}_2-\lambda)}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{1/2}(S_{\bar{X}_1}^2+S_{\bar{X}_2}^2)^{1/2}} \sim t_{n_1+n_2-2}$ s.t. $S_{\bar{X}_k}^2 = \sum_{i=1}^{n_k}\left(X_i - \bar{X}_{n_k}\right)^2$ .Reject H_0 if $|T| \geq t_{\frac{\alpha}{2},n_1+n_2-2}$. Alternatively, use two shorthand : $T = \frac{\bar{X}_1-\bar{X}_2}{s_p\left[\frac{1}{n_1}+\frac{1}{n_2}\right]^{1/2}}$ where $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$. ■

Proof : $Z = \frac{\bar{X}_1-\bar{X}_2}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{1/2}\sigma} \sim N(0,1)$ & $W = \frac{S_{\bar{X}_1}^2+S_{\bar{X}_2}^2}{\sigma^2} \sim \chi_{n_1+n_2-2}^2 \rightarrow \frac{Z}{W/(n_1+n_2-2)} \sim t_{n_1+n_2-2}$. ■**F test** : $H_0: \sigma_1^2 = \sigma_2^2$ vs $H_1: \sigma_1^2 \neq \sigma_2^2$ or ($\sigma_1^2 < \sigma_2^2$ or $\sigma_1^2 > \sigma_2^2$) . $F = \frac{larger\ sample\ variance}{smaller\ sample\ variance}$. Reject H_0 if $F > F_{\alpha/2,df_{num},df_{denom}}$ ($F > F > F_{\alpha/2,df_{num},df_{denom}}$ for both)

$\boldsymbol{\chi^2}$ **Test** $(\boldsymbol{r \times c})$ / Use for Test of Independence / Use for testing difference in r proportions between c classes: Let θ_{ij} be the probability of the j^{th} outcome from the i^{th} population. We want to test $H_0: \theta_{1j} = \theta_{2j} = \cdots = \theta_{rj}$ for $j = 1,2,\dots,c$ against $H_1: \theta_{1j}, \theta_{2j}, \dots, \theta_{rj}$ are not equal for at least one value of j . Notation : the observed frequency for the i^{th} row and j^{th} column is f_{ij} , the row total is $f_{i\star}$, the column totals is $f_{\star j}$, and the grand total (sum of all cell frequencies) is f. With this notation, we estimate the probabilities $\theta_{i\star}$ and $\theta_{\star j}$ as $\hat{\theta}_{i\star} = \frac{f_{i\star}}{f}$ and $\hat{\theta}_{\star j} = \frac{f_{\star j}}{f}$ and under the null hypothesis of independence we get $e_{ij} = (\hat{\theta}_{i\star})(\hat{\theta}_{\star j})(f) = \left(\frac{f_{i\star}}{f}\right)\left(\frac{f_{\star j}}{f}\right)f = \frac{f_{i\star}f_{\star j}}{f}$ for the expected frequency for the cell in the i^{th} row and j^{th} column. The test statistic is given by $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$ where we reject H_0 if $\chi^2 > \chi_{\alpha,(r-1)(c-1)}^2$. ■

Goodness of Fit Test : For testing H_0: the population follows the specified distribution vs H_1: the population does not follow the specified distribution. The critical region is given by $\chi^2 \geq \chi_{\alpha,m-t-1}^2$ where $\chi^2 = \sum_{i=1}^{m}\frac{(f_i-e_i)^2}{e_i}$, m is the number of terms in the summation, and t is the number of independent parameters estimated on the basis of the sample data. **Example (1)** : test H_0 that the data follows $N(50,25)$. The categories are $(-\infty,45)$, $(45,50)$, $(50,55)$, $(55,\infty)$ with observed values 21, 38, 27, 14 respectively. To find the expected counts we find probability of each bin times number of samples, i.e., $P(X < 45) * (21 + 38 + 27 + 14) = P\left(Z < \frac{45-50}{5}\right) * 100 = 15.86$ repeat and calculate $\chi^2 = \sum_{i=1}^{m}\frac{(f_i-e_i)^2}{e_i}$, …

Example (2) : Consider Mendel's hypothesis that when crossing two types of peas, the probability of the classifications (a) round and yellow, (b) wrinkled and yellow, (c) round and green, (d) wrinkled and green are 9/16, 3/16, 3/16, 1/16 respectively. If from 160 independent observations, the observed frequemcies of these respective classifications are 86, 35, 26, and 13, are these data consistent with the Mendelian theory? That is, test with $\alpha = 0.01$, the hypothesis that the respective probabilities are 9/16, 3/16, 3/16, 1/16 . **Answer (2)** : For $e_1,\dots,e_4$ are 90, 30, 30, 10 $\rightarrow \chi^2 = \sum_{i=1}^{4}\frac{(f_i-e_i)^2}{e_i} = \frac{22}{9} \approx 2.444$, with $df = 4 - 1 = 3$. p-value $= P(\chi^2 \geq 2.444)) = 0.4854$. Fail to reject H_0, the respective probabilities are 9/16, 3/16, 3/16, 1/16 . ■

■**Simulations** : Simulation is a computational technique that uses models to replicate the behavior of systems, enabling analysis and decision-making when analytical solutions are impractical. It can handle both deterministic systems (fully defined by inputs) and stochastic systems (influenced by randomness). By leveraging the Law of Large Numbers, simulation generates and averages random variables to approximate means, distributions, or probabilities. Key features include its ability to capture dynamic and random behaviors and study processes that are complex or impossible to observe directly. Simulation is particularly useful for estimating uncertainty by analyzing the simulation variance—the variability in repeated simulations—and the simulation standard error, which quantifies the precision of estimates. This makes simulation an essential tool for understanding complex systems and assessing outcomes under uncertainty.

■**ANOVA & GLM** : Analysis of Variance (ANOVA) and Generalized Linear Models (GLM) are statistical tools for analyzing relationships between variables. ANOVA tests for significant differences in means across groups by partitioning total variability into components explained by the groups and residual (unexplained) variation. It is commonly used in experimental designs to evaluate the effects of categorical factors. GLM extends linear regression by accommodating various distributions (e.g., normal, binomial, or Poisson) for the response variable, allowing for a broader range of modeling scenarios. GLM uses a link function to relate the predictors to the expected value of the response, enabling flexibility in handling non-normal data. Both techniques assess the significance of predictors and interactions, making them powerful tools for exploring and interpreting data in diverse fields.

■**Bootstrapping** is a resampling technique in statistics used to estimate the sampling distribution of a statistic by repeatedly drawing samples with replacement from the original data. It is particularly useful when the theoretical distribution of the statistic is complex or unknown. By generating a large number of bootstrap samples, we can compute estimates of standard errors, confidence intervals, and other measures of uncertainty for parameters such as means, medians, or regression coefficients. Bootstrapping is non-parametric, requiring minimal assumptions about the underlying distribution of the data, and is widely applied in cases where sample sizes are small or traditional methods are impractical. This flexibility makes it a robust tool for statistical inference and model validation.

■**The F-test** is a statistical test used to compare the variances of two or more groups or to assess the overall significance of a regression model. In the context of comparing variances, it tests the null hypothesis that the variances of the groups are equal by calculating the ratio of the variances. In the context of regression, the F-test evaluates whether at least one predictor variable significantly contributes to explaining the variability in the response variable, comparing the fit of a model with and without the predictors. The test statistic follows an F-distribution, and a large F-value indicates that the model or group differences are statistically significant. It is commonly used in ANOVA, regression analysis, and comparing nested models.

■ **Non-parametric** tests are statistical methods used when data doesn't meet the assumptions required for parametric tests, such as normality or homogeneity of variance. These tests are distribution-free, meaning they don't assume a specific underlying distribution for the data. Common non-parametric tests include the **Mann-Whitney U test** and **Wilcoxon signed-rank test** (for comparing two independent or paired samples, respectively), the **Kruskal-Wallis test** (for comparing more than two independent groups), and the **Spearman's rank correlation** (for assessing the relationship between two variables). Non-parametric tests are often used with ordinal data or when sample sizes are small or skewed. They are robust and flexible, providing valid results even when data is not normally distributed or when outliers are present.

Example (1): *Suppose* $X_1,\dots,X_n \sim iid\ N(\mu,\sigma^2)$ *where* (μ,σ^2) *are unknown. Prove* $T_1 = \bar{X}_n$ *and* $T_2 = S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ *are jointly sufficient.* ANSWER : Since $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{1}{n}\left[\left(\sum_{i=1}^{n} X_i^2\right) - n\bar{X}_n\right]$ we can equivalently prove $T_1 = \bar{X}_n$ and $T_3 = \sum_{i=1}^{n} X_i^2$ are jointly sufficient as $T_2 = S^2 = \frac{1}{n}T_3 - T_1^2$ is just a function of T_1 and T_3 . Therefore, $f_n(\mathbb{x}) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu)^2\right\}$. Simplifies to $f_n(\mathbb{x}) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\} \rightarrow f_n(\mathbb{x}) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2 - 2x_i\mu + \mu^2)\right\} \rightarrow f_n(\mathbb{x}) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2) + \frac{n}{\sigma^2}\bar{x}_n\mu - \frac{n}{2\sigma^2}\mu^2\right\}$. Jointly sufficient .■

Example (2): Show $E[S^2] = \sigma^2$. $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - n\bar{X}^2$. $E\left[\sum_{i=1}^{n} X_i^2\right] = \sum_{i=1}^{n} E\left[X_i^2\right] = \sum_{i=1}^{n} Var(X_i) + (E[X_i])^2 = n(\sigma^2 + \mu^2)$; $E[n\bar{X}^2] = nE[\bar{X}^2] = n(Var(\bar{X}) + E[\bar{X}]^2) = n\left(\frac{\sigma^2}{n} + \mu^2\right) \rightarrow E[S^2] = n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)$. ■

JACOBIAN TRANSFORMATION : Given X_1, X_2 for which the joint pdf is $f(x_1,x_2)$. Define Y_1, Y_2 as $Y_1 = r_1(X_1,X_2)$ and $Y_2 = r_2(X_1,X_2)$ where we assume the functions r_1 and r_2 are one-to-one. Let the inverse of this transformation be given by $x_1 = s_1(y_1,y_2)$ and $x_2 = s_2(y_1,y_2)$. Then the joint pdf g of Y_1, Y_2 is $g(y_1,y_2) = f(s_1,s_2)|\mathcal{J}|$ where $\mathcal{J}$ is the determinant: $\mathcal{J} = \det\begin{bmatrix} \frac{\partial s_1}{\partial y_1} & \frac{\partial s_1}{\partial y_2} \\ \frac{\partial s_2}{\partial y_1} & \frac{\partial s_2}{\partial y_2} \end{bmatrix}$ and $|\mathcal{J}|$ denotes the absolute value of the determinant $\mathcal{J}$. Thus, the joint pdf $g(y_1,y_2)$ is obtained by starting with the joint pdf $f(x_1,x_2)$ replacing each value x_i by its expression $s_i(y_1,y_2)$ in terms of y_1, y_2 and then multiplying the result by $|\mathcal{J}|$. $\mathcal{J}$ is called the Jacobian of the transformation.

Theorem : $T = \frac{\bar{X}-\mu}{s/\sqrt{n}}$ has the same pdf as the ratio of: $W \sim N(0,1)$ divided by $\sqrt{\frac{V_{n-1}}{n-1}}$ where $V_{n-1} \sim \chi_{n-1}^2$ and W and V_{n-1} are independent . **Proof** : $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}-\mu)/\sigma}{S/\sigma} = \frac{W}{\sqrt{V_{n-1}/(n-1)}}$.

t-distribution : If $T = \frac{W}{\sqrt{V/r}}$ given $W \sim N(0,1)$; $V \sim \chi_r^2$ and $W \perp V$, then $T \sim t_r$ (t-distribution with r degrees of freedom). This result implies that the pdf is given by (*see chart*) .

Proof : $T = \frac{W}{\sqrt{V/r}} \rightarrow \begin{cases} Y_1 = \frac{W}{\sqrt{V/r}} \\ Y_2 = V \end{cases} \rightarrow \begin{cases} W = X_1 = Y_1\sqrt{\frac{Y_2}{r}} \\ V = X_2 = Y_2 \end{cases} \rightarrow \begin{cases} x_1 = s_1(y_1,y_2) = y_1\sqrt{\frac{y_2}{r}} \\ x_2 = s_2(y_1,y_2) = y_2 \end{cases}$. Therefore, the Jacobian is given by $\mathcal{J} = \det\begin{bmatrix} \frac{\partial s_1}{\partial y_1} & \frac{\partial s_1}{\partial y_2} \\ \frac{\partial s_2}{\partial y_1} & \frac{\partial s_2}{\partial y_2} \end{bmatrix} = \det\begin{bmatrix} \sqrt{\frac{y_2}{r}} & \frac{y_1}{2}\left(\frac{y_2}{r}\right)^{-\frac{1}{2}} \\ 0 & 1 \end{bmatrix} = \sqrt{\frac{y_2}{r}}$. The pdf of $X_1 = W$ and $X_2 = V$ is given by the product of their marginals as $W \perp V$. Therefore, $f_{\mathbb{X}}(x_1,x_2) = \left[\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x_1^2\right\} * \mathbb{1}_{\{x_1\in\mathbb{R}\}}\right] * \left[\frac{1}{2^{r/2}\Gamma(r/2)}x_2^{(r/2)-1}e^{-(x_2/2)} * \mathbb{1}_{\{x_2>0\}}\right] = c_1 * x_2^{(r/2)-1}\exp\left\{-\frac{1}{2}(x_1^2 + x_2)\right\}\mathbb{1}_{\{x_1\in\mathbb{R}\,;\,x_2>0\}}$ where $c_1 = \frac{1}{\sqrt{2\pi}} * \frac{1}{2^{r/2}\Gamma(r/2)}$. Therefore, $f_{\mathbb{X}}(s_1,s_2) = c_1 * s_2^{(r/2)-1}\exp\left\{-\frac{1}{2}(s_1^2 + s_2)\right\}\mathbb{1}_{\{s_1\in\mathbb{R}\,;\,s_2>0\}} = c_1 * y_2^{(r/2)-1}\exp\left\{-\frac{1}{2}\left(\frac{y_1^2}{r} + 1\right)y_2\right\}\mathbb{1}_{\{y_1\in\mathbb{R}\,;\,y_2>0\}}$. As a result, plugging in all values: $g_{Y_1,Y_2}(y_1,y_2) = f_{\mathbb{X}}(s_1,s_2) * |\mathcal{J}|$. Thus, $g_{Y_1,Y_2}(y_1,y_2) = c_1 * y_2^{(r/2)-1}\exp\left\{-\frac{1}{2}\left(\frac{y_1^2}{r} + 1\right)y_2\right\}\mathbb{1}_{\{y_1\in\mathbb{R}\,;\,y_2>0\}} * \frac{\sqrt{y_2}}{\sqrt{r}}$ which simplifies to $g_{Y_1,Y_2}(y_1,y_2) = \frac{c_1}{\sqrt{r}}y_2^{\left[\frac{r+1}{2}-1\right]}\exp\left\{-\left[\frac{1}{2}\left(\frac{y_1^2}{r} + 1\right)\right]y_2\right\}\mathbb{1}_{\{y_1\in\mathbb{R}\,;\,y_2>0\}}$. Now we want to find the pdf of Y_1 so we integrate out the values y_2 . $g_{Y_1}(y_1) = \int_{y_2=0}^{\infty} g_{Y_1,Y_2}(y_1,y_2)dy_2$. Let $\alpha = \frac{r+1}{2}$ and $\beta = \frac{1}{2}\left(\frac{y_1^2}{r} + 1\right)$. $g_{Y_1}(y_1) = \frac{c_1}{\sqrt{r}} * \left\{\frac{\Gamma(\alpha)}{\beta^\alpha}\right\} * \mathbb{1}_{\{y_1\in\mathbb{R}\}}\int_{y_2=0}^{\infty}\left\{\frac{\beta^\alpha}{\Gamma(\alpha)}\right\} * y_2^{\alpha-1}\exp\{-\beta y_2\}\,dy_2$; we see this is the $Gamma(\alpha,\beta)$ which sums to 1. Therefore, $g_{Y_1}(y_1) = \frac{c_1}{\sqrt{r}}\left\{\frac{\Gamma(\alpha)}{\beta^\alpha}\right\}\mathbb{1}_{\{y_1\in\mathbb{R}\}}$. Plugging in c_1, α, β yields: $g_{Y_1}(y_1) = \frac{\frac{1}{\sqrt{2\pi}} * \frac{1}{2^{r/2}\Gamma(r/2)}}{\frac{\sqrt{r}}{1}}\left\{\frac{\Gamma\left(\frac{r+1}{2}\right)}{\left[\frac{1}{2}\left(\frac{y_1^2}{r}+1\right)\right]^{\left(\frac{r+1}{2}\right)}}\right\}\mathbb{1}_{\{y_1\in\mathbb{R}\}} = \frac{\Gamma\left(\frac{r+1}{2}\right)}{(r\pi)^{1/2}\Gamma(r/2)} * 2^{(r+1)/2}\left(\frac{1}{2}\right)^{(r+1)/2}\left(\frac{y_1^2}{r} + 1\right)^{-(r+1)/2}\mathbb{1}_{\{y_1\in\mathbb{R}\}}$. Finally, $g_{Y_1}(y_1) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{(r\pi)^{1/2}\Gamma(r/2)}\left(\frac{y_1^2}{r} + 1\right)^{-(r+1)/2} * \mathbb{1}_{\{y_1\in\mathbb{R}\}}$. Proved ■ .