Distribution	pmf/pdf	Mean	Variance	Moment Generating Function: $\psi(t) = E[e^{tX}]$	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/ $P(success) = p$ $E[X] = E[X^2] = \cdots = E[X^k] = p$
$X \sim Binomial(n,p)$	$f_X(x) = \binom{n}{x} p^x(1-p)^{n-x} * \mathbb{1}_{\{x=0,1,\ldots,n\}}$	$E[X] = np$	$Var(X) = np(1-p)$	$\psi(t) = (1 - p + pe^t)^n$	X = # of successes in n trials.
$X \sim Poisson(\lambda)$ $\lambda > 0$	$f_X(x) = e^{-\lambda} \frac{\lambda^x}{x!} * \mathbb{1}_{\{x=0,1,\ldots\}}$	$E[X] = \lambda$	$Var(X) = \lambda$	$\psi(t) = \exp\{\lambda(e^t - 1)\}$	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: $P(X > s + t \mid X > s) = P(X > t)$
$X \sim Gamma(\alpha,\beta)$ $\alpha, \beta > 0$	$f_X(x) = \frac{\beta^\alpha e^{-\beta x} x^{\alpha-1}}{\Gamma(\alpha)} * \mathbb{1}_{\{x \ge 0\}}$	$E[X] = \frac{\alpha}{\beta}$	$Var(X) = \frac{\alpha}{\beta^2}$	$\psi(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha} * \mathbb{1}_{\{t<\beta\}}$	$X \sim Gamma(1,\beta) \equiv Exponential(\beta)$ $\Gamma(\alpha) = (\alpha-1)!$; $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
$X \sim Beta(\alpha,\beta)$ $\alpha, \beta > 0$	$f_X(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} * \mathbb{1}_{\{0 \le x \le 1\}}$	$E[X] = \frac{\alpha}{\alpha+\beta}$	$Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$-$	$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
$X \sim Normal(\mu,\sigma^2)$ $\sigma^2 < 0$	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} * \mathbb{1}_{\{x \in \mathbb{R}\}}$	$E[X] = \mu$	$Var(X) = \sigma^2$	$\psi(t) = \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$	$-$
$X \sim \chi^2_m$ CHI-SQUARE DIST. m degrees of freedom.	$f_X(x) = \frac{1}{2^{m/2}\Gamma(m/2)} x^{(m/2)-1} e^{-(x/2)} * \mathbb{1}_{\{x>0\}}$	$E[X] = m$	$Var(X) = 2m$	$\psi(t) = (1-2t)^{-m/2} * \mathbb{1}_{\{t<1/2\}}$	$Gamma\left(\alpha = \frac{m}{2}, \beta = \frac{1}{2}\right) \equiv \chi^2_m$ $Exponential\left(\beta = \frac{1}{2}\right) \equiv \chi^2_2$ $X_i \sim iid\ \chi^2_{m_i}\ \forall i = 1,\ldots,k \Longrightarrow X_1 + \cdots + X_k \sim \chi^2_{(\sum_{i=1}^k m_i)}$
$X \sim t_m$ t-DISTRIBUTION	$f_X(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{(m\pi)^{1/2}\Gamma(m/2)} \left(1 + \frac{x^2}{m}\right)^{-(m+1)/2} * \mathbb{1}_{\{x \in \mathbb{R}\}}$	$-$	$Var(X) = \frac{m}{m-2}$, $m > 2$	$-$	$Y \sim \chi^2_m$ and $Z \sim N(0,1)$: $X = \frac{Z}{\sqrt{Y/m}} \sim t_m$

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 4.7.4: $Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$.
PROOF : Let $E[Y] = \mu_Y$. $Var(Y) = E[(Y - \mu_Y)^2] = E\left\{\left((Y - E[Y|X]) + (E[Y|X] - \mu_Y)\right)^2\right\} = E[(Y - E[Y|X])^2] + 2E[(Y - E[Y|X])(E[Y|X] - \mu_Y)] + E[(E[Y|X] - \mu_Y)^2]$. We know that $E[(Y - E[Y|X])^2] = E\{E[(Y - E[Y|X])^2|X]\} = E[Var(Y|X)]$ and $E[(Y - E[Y|X])(E[Y|X] - \mu_Y)] = E[E[(Y - E[Y|X])(E[Y|X] - \mu_Y)] \mid X] = E[(E[Y|X] - \mu_Y) * E(Y - E[Y|X] \mid X)] = E[(E[Y|X] - \mu_Y) * (E[Y|X] - E\{E[Y|X]|X\})] = E[(E[Y|X] - \mu_Y) * 0] = 0$. Finally, we know $E[(E[Y|X] - \mu_Y)^2] = Var(E[Y|X])$. Therefore, $Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$ ■
RESULTS – RB ; The following results show that for any unbiased estimator of a parameter (θ), we can find another unbiased estimator (a function of a sufficient statistic only) with a smaller variance. In summary, let T_2 be an unbiased, $E[T_2] = \theta$, let T_1 be a sufficient statistic for θ. Then the random variable $\varphi(T_1) = E[T_2|T_1]$ has $E[\varphi(T_1)] = E[E[T_2|T_1]] = E[T_2] = \theta$ and a smaller variance.
RAO BLACKWELL : Let $X_1, \ldots, X_n \sim iid\ f_X(x;\theta), \theta \in \Omega$. Let $T_1 = t(x_1, \ldots, x_n)$ be a sufficient statistic for θ and let $T_2 = t_2(x_1, \ldots, x_n)$ (not a function of T_1 alone) be another unbiased estimator of θ. Then, the random variable $\varphi(T_1) = E[T_2|T_1]$ is an unbiased estimator of θ and $Var(T_2) \ge Var(\varphi(T_1)) = Var(E[T_2|T_1])$.
PROOF : Theorem 4.7.4 $\to Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$. Let $Y = T_2$ and $X = T_1 \to Var(T_2) = E[Var(T_2|T_1)] + Var(E[T_2|T_1])$. As Variance > 0 , $Var(T_2) > Var(E[T_2|T_1]) = Var(\varphi(T_1))$.
Therefore, in our search for "best" estimators of θ we can and should restrict our attention to functions of the sufficient statistic.
COMPLETE FAMILY: The family of pdf's $\{f_X(x;\theta);\ \theta \in \Omega\}$ is said to be complete if $E[u(x)] = 0$ for every $\theta \in \Omega$ requires $u(x)$ is zero except on a set of points with probability zero.
CRLB : (UNBIASED ESTIMATOR) Cramer-Rao Lower Bound: $Var(\widehat{\Theta}) \ge \frac{1}{n * \boldsymbol{I}(\theta)}$. where $\boldsymbol{I}(\theta) = E\left[\left(\frac{\partial}{\partial\theta} \ln f_X(x)\right)^2\right]$. (BIASED ESTIMATOR) $Var(\widehat{\Theta}) \ge \frac{\left(\frac{\partial}{\partial\theta} E[\widehat{\Theta}]\right)^2}{n * \boldsymbol{I}(\theta)}$.
If $\widehat{\Theta}$ is an unbiased estimator of θ and $Var(\widehat{\Theta})$ attains the CRLB, then $\widehat{\Theta}$ is a **MVUE** of θ .
MINIMUM VARIANCE UNBIASED ESTIMATOR (MVUE): Let $f_X(x;\theta);\ \theta \in \Omega$ be an exponential family. Let Y_1 be a sufficient statistic for θ. Then $f_X(x;\theta)$ is complete and Y_1 is said to be a complete sufficient statistic for θ.

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 $T_{n-1} \to N(0,1)$. If σ^2 is unknown use s^2 : $100(1-\alpha)\%$ CI for μ : $P\left(-t_{\alpha/2,n-1} < \frac{\bar{X} - E[\bar{X}]}{\sqrt{Var(\bar{X})}} < t_{\alpha/2,n-1}\right) = P\left(\bar{X} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right) = 1 - \alpha$.
$X_1, \ldots, X_n \sim iid\ N(\mu, \sigma^2)$ where μ, σ^2 are unknown. Construct CI for σ^2. $\sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2 \sim \chi^2_{n-1} \to P\left(\chi^2_{1-\frac{\alpha}{2}n-1} < \sum_{i=1}^n \frac{(X_i - \bar{X}_n)^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}n-1}\right) = 1 - \alpha \to P\left(\frac{1}{\chi^2_{\frac{\alpha}{2}n-1}} < \sum_{i=1}^n \frac{\sigma^2}{(X_i - \bar{X}_n)^2} < \frac{1}{\chi^2_{1-\frac{\alpha}{2}n-1}}\right) = 1 - \alpha$. 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: $P\left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}n-1}}\right) = 1 - \alpha$.

If $X \sim N(0,1)$; $Y = g(X) = X^2 \to Y \sim Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$. **PROOF** : $F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \Longrightarrow f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}\{F_X(\sqrt{y}) - F_X(-\sqrt{y})\} = f_X(\sqrt{y}) * \frac{d}{dy}\{\sqrt{y}\} - f_X(-\sqrt{y}) * \frac{d}{dy}\{-\sqrt{y}\} \Longrightarrow f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})]$. As $f_X(x) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x^2\right\} \Longrightarrow f_Y(y) = \frac{1}{2\sqrt{y}}\left[2 * \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}y\right\}\right]$. Thus, $Y \sim Gamma\left(\frac{1}{2}, \frac{1}{2}\right) \equiv \chi^2_1 \equiv f_Y(y) = \frac{1}{\sqrt{2\pi y}}\exp\left\{-\frac{1}{2}y\right\}$ ■

We are given from above $X^2 = Y \sim Gamma\left(\frac{1}{2}, \frac{1}{2}\right) \equiv \chi^2_1$. We know chi-square distribution is a sum of iid standard normal random variables squared: $\chi^2_n = X_1^2 + \cdots + X_n^2 = Y_1 + \cdots + Y_n$; (Y_i iid) . Therefore, we can use MGF as follows to derive distribution of χ^2_n : $\psi_{Y_j}(t) = \left(1 - \frac{t}{1/2}\right)^{-1/2} * \mathbb{1}_{\{t<1/2\}}$ for $j = 1, \ldots, n$. Therefore, let $W = Y_1 + \cdots + Y_n \to \psi_W(t) = E[e^{tW}] = E\left[e^{t(Y_1 + \cdots + Y_n)}\right] = \prod_{j=1}^n E[e^{tY_j}] = \prod_{j=1}^n \psi_{Y_j}(t) = \left(1 - \frac{t}{1/2}\right)^{-n/2}$. Since MGF uniquely describes a distribution. $\psi_W(t) = \left(1 - \frac{t}{1/2}\right)^{-n/2}$ is the MGF given by $Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$. Therefore, $W = \chi^2_n \sim Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$. ■

THEOREM : Suppose that $X_1, \ldots, X_n \sim iid\ N(\mu, \sigma^2)$. Then the sample mean $\bar{X}_n$ and sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ are independent random variables where $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ and $S^2 \sim \chi^2_{n-1}$.
PROOF : WLOG (FOR CASE: $n = 2$; $\sigma^2 = 1, \mu = 0$) : Show $S^2 \perp \bar{X}$. Firstly, $\bar{X} = \frac{X_1 + X_2}{2}$ and $S^2 = \frac{1}{2-1}\sum_{i=1}^2 (X_i - \bar{X})^2 = \left(X_1 - \frac{X_1 + X_2}{2}\right)^2 + \left(X_2 - \frac{X_1 + X_2}{2}\right)^2 = 2\left(\frac{X_1 - X_2}{2}\right)^2$. Let $Y_1 = \bar{X} = \frac{X_1 + X_2}{2}$ and $Y_2 = S^2 = 2\left(\frac{X_1 - X_2}{2}\right)^2$. We will apply Jacobian transformation and show that the joint pdf for Y_1, Y_2 factors. Finding the inverse is done with $\begin{cases} X_1 - X_2 = 2\sqrt{\frac{Y_1}{2}} \\ X_1 + X_2 = 2Y_2 \end{cases}$ which yields $\begin{cases} X_1 = \sqrt{\frac{Y_1}{2}} + Y_2 \\ X_2 = Y_2 - \sqrt{\frac{Y_1}{2}} \end{cases}$. Therefore, our inverse functions are given by,
$x_1 = s_1(y_1, y_2) = \sqrt{\frac{y_1}{2}} + y_2$ and $x_2 = s_2(y_1, y_2) = y_2 - \sqrt{\frac{y_1}{2}}$. Since $X_1, X_2 \sim iid\ N(0,1)$ their joint pdf is the product of 2 standard normal pdf given by $f_{X_1,X_2}(x_1, x_2) = \prod_{i=1}^2 \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x_i^2\right\} = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}(x_1^2 + x_2^2)\right\}$.
Therefore, $f(s_1, s_2) = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}(s_1^2 + s_2^2)\right\}$. Note that $s_1^2 + s_2^2 = \left(\sqrt{\frac{y_1}{2}} + y_2\right)^2 + \left(y_2 - \sqrt{\frac{y_1}{2}}\right)^2 = 2y_2^2 + y_1$. Hence this simplifies to $f(s_1, s_2) = \frac{1}{2\pi}\exp\left\{-y_2^2 - \frac{1}{2}y_1\right\}$. Now to compute the Jacobian. : $\mathcal{J} = \det\begin{bmatrix} \frac{\partial s_1}{\partial y_1} & \frac{\partial s_1}{\partial y_2} \\ \frac{\partial s_2}{\partial y_1} & \frac{\partial s_2}{\partial y_2} \end{bmatrix} = \det\begin{bmatrix} \frac{1}{2}\left(\frac{y_1}{2}\right)^{-\frac{1}{2}} & 1 \\ -\frac{1}{2}\left(\frac{y_1}{2}\right)^{-\frac{1}{2}} & 1 \end{bmatrix} = \frac{1}{2}\left(\frac{y_1}{2}\right)^{-\frac{1}{2}} - -\frac{1}{2}\left(\frac{y_1}{2}\right)^{-\frac{1}{2}} = \left(\frac{y_1}{2}\right)^{-\frac{1}{2}} = \sqrt{\frac{2}{y_1}}$. Thus, $g(y_1, y_2) = f(s_1, s_2)|\mathcal{J}| \to g(y_1, y_2) = \frac{1}{2\pi}\exp\left\{-y_2^2 - \frac{1}{2}y_1\right\} * \sqrt{\frac{2}{y_1}}$. We can now factor this into a product of marginals to prove independence between $Y_1 = \bar{X}$ and $Y_2 = S^2$. $g(y_1, y_2) = \left[\frac{1}{\sqrt{2\pi y_1}}\exp\left\{-\frac{1}{2}y_1\right\}\right] * \left[\frac{1}{\sqrt{\pi}}\exp\{-y_2^2\}\right]$. NOTE: $f_{Y_1}(y_1) = \frac{1}{\sqrt{2\pi y_1}}\exp\left\{-\frac{1}{2}y_1\right\} \equiv Gamma\left(\frac{1}{2}, \frac{1}{2}\right) \equiv \chi^2_1$ AND $f_{Y_2}(y_2) = \frac{1}{\sqrt{\pi}}\exp\{-y_2^2\} \equiv N\left(0, \frac{1}{n} \middle| n = 2\right)$. Hence, $\bar{X}_n$ and S^2 are independent random variables where $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ and $S^2 \sim \chi^2_{n-1}$. ■
REMARKS: Using the fact {if V, W are independent, then $g(V)$, $h(W)$ are independent for any functions g , h not depending on V or W}, we also proved $g(Y_1) \perp h(Y_2)$.

Theorem : If $X_1, \dots, X_n \sim iid\ N(\mu, \sigma^2)$, then the statistic $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.
Proof : First, note that $\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n [(X_i - \bar{X}) + (\bar{X} - \mu)]^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + 2\sum_{i=1}^n (X_i - \bar{X})(\bar{X} - \mu) + \sum_{i=1}^n (\bar{X} - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$ since $2\sum_{i=1}^n (X_i - \bar{X})(\bar{X} - \mu) = 2(\bar{X} - \mu)\sum_{i=1}^n (X_i - \bar{X}) = 0$. Therefore, $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)^2 = \frac{1}{\sigma^2}[\sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2] = \frac{(n-1)S^2}{\sigma^2} + \frac{(\bar{X} - \mu)^2}{\sigma^2/n}$. Denoting these random variables as $V = \frac{(n-1)S^2}{\sigma^2} \sim ???$; $U = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n$; $W = \frac{(\bar{X} - \mu)^2}{\sigma^2/n} \sim \chi^2_1$ and using the fact that $S^2 \perp \bar{X}_n$ and thus $g(S^2) \perp h(\bar{X}_n)$ we know $V \perp W$. Now to get a relationship between U, V, W we see: $U = V + W$. Since $V \perp W$ the moment generating function is factorable, i.e., $\psi_U(t) = E[e^{Ut}] = E\left[e^{(V+W)t}\right] = E[e^{Vt}] * E[e^{Wt}] = \psi_V(t)\psi_W(t)$. Therefore, since we know $U \sim \chi^2_n$ and $W \sim \chi^2_1$, $\psi_V(t) = \frac{\psi_U(t)}{\psi_W(t)} = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}} = (1-2t)^{-(n-1)/2}$. This is the MGF of χ^2_{n-1} . Therefore, $V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. ■

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 $V_{n-1} \sim \chi^2_{n-1}$ 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 $W \sim N(0,1)$; $V \sim \chi^2_r$ 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 : $T = \frac{W}{\sqrt{V/r}} \to \begin{cases} Y_1 = \frac{W}{\sqrt{V/r}} \\ Y_2 = V \end{cases} \to \begin{cases} W = X_1 = Y_1\sqrt{\frac{Y_2}{r}} \\ V = X_2 = Y_2 \end{cases} \to \begin{cases} x_1 = s_1(y_1, y_2) = y_1\sqrt{\frac{y_2}{r}} \\ x_2 = s_2(y_1, y_2) = y_2 \end{cases}$. 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 ■ .

Theorem (sufficiency w/o factorization criteria): Let $X_1, \dots, X_n \sim iid\ f_X(x; \theta)$ for some $\theta \in \Omega$, then $T_1 = T_1(X_1, \dots, X_n)$ is a sufficient statistic for θ if $f_{X_1,\dots,X_n}(x_1, \dots, x_n; \theta | T_1)$ does not depend on θ .
Example (1): Given $X_1, \dots, X_n \sim f_X(x) = \theta^x(1-\theta)^{1-x} * \mathbb{1}_{\{x=0,1 \cap 0<\theta<1\}}$; $T_1 = \sum_{i=1}^n X_i = Y \sim Bin(n, \theta)$. What is $P(\mathbb{X} = \mathbb{x} | T_1 = t_1)$? $P(\mathbb{X} = \mathbb{x} | T_1 = t_1) = \frac{P(\mathbb{X} = \mathbb{x} \cap T_1 = t_1)}{P(T_1 = t_1)}$. Suppose if $t_1 \neq \sum_{i=1}^n x_i$, then $P(\mathbb{X} = \mathbb{x} | T_1 = t_1) = 0$, therefore, if $t_1 = \sum_{i=1}^n x_i$, $P(\mathbb{X} = \mathbb{x} | T_1 = t_1) = \frac{P(\mathbb{X} = \mathbb{x} \cap \sum_{i=1}^n X_i = \sum_{i=1}^n x_i)}{P(\sum_{i=1}^n X_i = \sum_{i=1}^n x_i)} = \frac{P(X_1 = x_1, \dots, X_n = x_n)}{P(\sum_{i=1}^n X_i = \sum_{i=1}^n x_i)} = \frac{P(X_1 = x_1) * \dots * P(X_n = x_n)}{P(\sum_{i=1}^n X_i = \sum_{i=1}^n x_i)} = \frac{\prod_{i=1}^n \theta^{x_i}(1-\theta)^{n-x_i}}{\binom{n}{\sum_{i=1}^n x_i}\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}} = \frac{\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}}{\binom{n}{\sum_{i=1}^n x_i}\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}} = \frac{1}{\binom{n}{\sum_{i=1}^n x_i}}$. Therefore, $P(\mathbb{X} = \mathbb{x} | T_1 = t_1) = \frac{1}{\binom{n}{\sum_{i=1}^n x_i}}$ does not depend on θ . Therefore, T_1 is a sufficient statistic.
Example (2): Given $X_1, \dots, X_n \sim iid\ Gamma(2, \theta)$ and $T_1 = \sum_{i=1}^n X_i \sim Gamma(2n, \theta)$ (← find this using iid → product of MGF). $f_{T_1}(t_1; \theta) = \frac{\theta^{2n}}{\Gamma(2n)}t_1^{2n-1}e^{-\theta t_1} * \mathbb{1}_{\{t_1 \geq 0\}}$. Finding $f(x_1, \dots, x_n | T_1 = t_1) = \frac{f_{\mathbb{X}}(x_1, \dots, x_n \cap \sum_{i=1}^n X_i = \sum_{i=1}^n x_i)}{f_{T_1}(t_1)} = \frac{\prod_{i=1}^n f_{X_i}(x_i)}{f_{T_1}(t_1)} = \frac{\prod_{i=1}^n \theta^2 x_i e^{-\theta x_i} * \mathbb{1}_{\{x_i \geq 0\}}}{\frac{\theta^{2n}}{\Gamma(2n)}t_1^{2n-1}e^{-\theta t_1} * \mathbb{1}_{\{t_1 \geq 0\}}} = \frac{\theta^{2n}(\prod_{i=1}^n x_i)\exp\{-\theta\sum_{i=1}^n x_i\} * \mathbb{1}_{\{\min[x_1, \dots, x_n] \geq 0\}}}{\frac{\theta^{2n}}{\Gamma(2n)}(\sum_{i=1}^n x_i)^{2n-1}\exp\{-\theta\sum_{i=1}^n x_i\} * \mathbb{1}_{\{\sum_{i=1}^n x_i \geq 0\}}} = \frac{\Gamma(2n)(\prod_{i=1}^n x_i)}{(\sum_{i=1}^n x_i)^{2n-1}} * \mathbb{1}_{\{\min[x_1, \dots, x_n] \geq 0\}}$. Therefore, $T_1 = \sum_{i=1}^n X_i$ does not depend on θ and is sufficient.
Theorem ($\mathbf{7.8.1}$) (Factorization Criterion-Jointly Sufficient): Let $r_1, \dots, r_k$ be functions of n real variables. The statistics $T_i = r_i(X_1, \dots, X_n)$ for $i = 1, \dots, k$, are jointly sufficient statistics for θ if and only if the joint pdf $f_n(x_1, \dots, x_n | \theta)$ can be factored as follows for all values of $(x_1, \dots, x_n) \in \mathbb{R}^n$ and all values of $\theta \in \Omega$: ####### $\boldsymbol{f_n(x_1, \dots, x_n | \theta) = u(x_1, \dots, x_n) * v[r_1(x_1, \dots, x_n), \dots, r_k(x_1, \dots, x_n), \theta]}$ ############# .Here the functions u and v are nonnegative, the function u may depend on $(x_1, \dots, x_n)$ but does not depend on θ, and the function v will depend on θ but depends on $(x_1, \dots, x_n)$ only through the k functions $r_1(x_1, \dots, x_n), \dots, r_k(x_1, \dots, x_n)$.

Relevant Midterm Predictions Problems: (On sufficiency)
Example (1): *Assuming $f_X(x|\theta)$ comes from an exponential family, i.e., $\boldsymbol{f_X(x|\theta) = a(\theta)b(x)\,exp\{c(\theta)d(x)\}}$. What is a sufficient statistic?* ANSWER: $f_n(\mathbb{x}|\theta) = \prod_{i=1}^n a(\theta)b(x_i)\exp\{c(\theta)d(x_i)\}$. This simplifies to $f_n(\mathbb{x}|\theta) = \{\prod_{i=1}^n b(x_i)\} * \{[a(\theta)]^n\exp\{c(\theta)\sum_{i=1}^n d(x_i)\}\}$. Therefore, $T = \sum_{i=1}^n d(x_i)$ is a sufficient statistic for θ by factorization criterion.
Example (2): *Assuming k −parameter exponential family, $\boldsymbol{f_X(x|\theta) = a(\theta)b(x)\,exp\{\sum_{i=1}^k c_i(\theta)d_i(x)\}}$. Sample $X_1, \dots, X_n \sim iid\ f_{X_j}(x_j|\theta)$. Show $T_i = \sum_{j=1}^n d_i(X_j)$ is jointly sufficient $\forall i = 1, \dots, k$* . ANSWER: $f_n(\mathbb{x}|\theta) = \prod_{j=1}^n a(\theta)b(x_j)\exp\{\sum_{i=1}^k c_i(\theta)d_i(x_j)\}$, this simplifies to $f_n(\mathbb{x}|\theta) = \left[\prod_{j=1}^n b(x_j)\right] * \left[[a(\theta)]^n\exp\{\sum_{i=1}^k c_i(\theta)\sum_{j=1}^n d_i(x_j)\}\right]$. Therefore, T_i is jointly sufficient $\forall i = 1, \dots, k$.
Example (3): *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 $S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n}\left[(\sum_{i=1}^n X_i^2) - n\bar{X}_n\right]$ 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 $f_n(\mathbb{x}) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\right\} \to f_n(\mathbb{x}) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i^2 - 2x_i\mu + \mu^2)\right\} \to f_n(\mathbb{x}) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i^2) + \frac{n}{\sigma^2}\bar{x}_n\mu - \frac{n}{2\sigma^2}\mu^2\right\}$. Jointly sufficient .■

If an unbiased estimator $\widehat{\Theta}$ has the smallest variance among all other unbiased estimators. We call such an estimator $\widehat{\Theta}$ a MVUE.
MVUE and Sufficient Statistics: $f(x; \theta)$. Let $\hat{\theta}$ be an unbiased estimator of θ, $E[\hat{\theta}] = \theta$. It is desirable to find an unbiased estimator of θ with smallest variance $Var(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$ (MVUE) .
Question: How do we find an unbiased estimator of parameter θ with minimum variance? (MVUE) : (1). Find a sufficient statistic of θ . (2). Minimum variance estimators are functions of the sufficient statistic. (3). Assuming the family of distributions is a complete family of distributions, then $\hat{\theta}$ is an unbiased estimator with minimum variance.
Theorem : Let $X_i \sim iid\ f_X(x; \theta)$, $\theta \in \Omega$, a statistic u is called a “best statistic” if $u(x_1, \dots, x_n)$ is an unbiased estimator of θ and $E(u - \theta)^2 \leq E(v - \theta)^2$ for all other estimators v with $E[v] = \theta$.
Example: Given $X_1, \dots, X_n \sim iid\ N(\theta, \sigma^2)$, $\bar{X}_n$ is a better unbiased estimator of θ then X_1 because $E[(\bar{X}_n - \theta)^2] = Var(\bar{X}_n) = \frac{\sigma^2}{n} \leq E[(X_1 - \theta)^2] = Var(X_1) = \sigma^2$

Important Theorems/Properties/Etc .
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 $j = 1, \dots, k$.
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) $\mu'_k = E\left[X_i^k\right]$ (kth population moment).
Result : If $X_1, \dots, X_n \sim iid\ N(\mu, \sigma^2)$, then $\sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2 \sim \chi^2_{n-1}$ and $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n$.