8/22/23:

In this class, we focus on univariate iid data.
A random variable, X, is a numerical measure of the outcome of an experiment.
Ex: Toss coin, roll die, # of visitors to website in 1hr
Statistical inference asks: Given some data, what can we infer about the process that generated the data?
Ex: If we take a random sample of n voters, can we get an estimate and error bounds for percent of voters that support candidate Q.
Notation: $\bar{X}$ estimates μ, S^2 estimates σ^2. $\bar{x}$ and s^2 are estimates of $\bar{X}$ and S^2 respectively.

Parametric inference chooses a family of probability distribution and tries to generate an estimator of the family's parameters based on measured data. Types of estimators include point estimators (single number), an interval estimator (a range of estimates), or a Bayesian estimator (a distribution on the parameters).
Nonparametric inference, also called distribution free methods, attempts to make as few assumptions as possible.
Bayesian inference performs statistics where the parameters are also random. Then probabilities represent degrees of belief rather than expected number over several trials. You choose a prior distribution to represent parameters, and generate a posterior distribution which updates the prior based on the data

8/24/23:

Recall Bayes Theorem: Given events A and B with $P(B) \neq 0$,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Suppose in town i, we survey for support of party A or B.

Given a probabilistic experiment with outcomes Ω, and a set function $\mathbb{P} : A \subseteq \Omega \to [0,1]$, we must satisfy the axioms

1) $\mathbb{P}(A) \in [0,1]$ for each $A \subseteq \Omega$
2) $\mathbb{P}(\Omega) = 1$
3) For disjoint $A_1, A_2, \dots \subseteq \Omega$, $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

All random variables have a range of possible outcomes

Its distribution or law describes the probability of the variable taking on certain values or ranges

Common Examples of Discrete Random Variables

1) Bernoulli: One trial, probability p of success
2) Binomial: N trials, count # of binary successes
3) Geometric: Trial at which first success occurs in iid trials
4) Negative Binomial: Trial at which r^{th} success occurs in iid trials
5) Poisson: Counts number of events occuring in time period
6) Multinomial: Probabilities on distributing n objects among k bins
7) Hypergeometric: Probability of k successes in n draws without replacement
8) Discrete Uniform: Every event equally likely

Common types of continuous random variables

1) Uniform :
2) Exponential: Waiting time until event: $f(x) = \lambda e^{-\lambda x}$
3) Gamma: Waiting time until event occurs multiple times: $f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \beta^{\alpha}$
4) Normal: Sum of iid variables: $f(x) = (2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/(2\sigma^2)}$
5) Beta: Various random variables on $[0,1]$, $f(x) = C(\alpha,\beta) \cdot x^{\alpha-1}(1-x)^{\beta-1}$
6) Dirichlet: Random split interval to k pieces: $f(x_1, \dots, x_n) = \frac{1}{B(\alpha)} \prod x_i^{\alpha_i - 1}$
7) χ^2, t, F : From various statistical tests

Discrete distributions described by a probability mass function, pmf, with $p : \Omega \to [0,1]$ with $\sum_{\omega \in \Omega} p(\omega) = 1$. We call

Ω, or the set of events of nonzero probability, the support.
Continuous distributions described by a probability density function, pdf, with $f: \mathbb{R} \to [0, \infty)$ s.t. $\int_{-\infty}^{\infty} f(x)\,dx = 1$ and we have $\mathbb{P}(a \le X \le b) = \int_a^b f(x)\,dx$.

In addition, we define the cumulative density function, cdf, $F: \mathbb{R} \to [0,1]$, s.t.

$$F_X(s) = \int_{-\infty}^{s} f(x)\,dx$$

for continuous random variables, and

$$F_X(s) = \sum_{x=-\infty}^{s} p(x)$$

for discrete random variables.

8/29/23:

CDF method for transformations of random variables:
Suppose given a continuous random variable X with pdf $f_X(x)$. Now, suppose we have an invertible transformation giving a new continuous random variable $Y = \phi(X)$. We can compute the cdf of Y, $F_Y(y)$, by

$$F_Y(y) = \mathbb{P}[Y = \phi(X) < y] = \begin{cases} \mathbb{P}[X < \phi^{-1}(y)] & \text{if } \phi^{-1} \text{ monotone increasing} \\ \mathbb{P}[X > \phi^{-1}(y)] & \text{if } \phi^{-1} \text{ monotone decreasing} \end{cases}$$

$$= \begin{cases} \int_{-\infty}^{\phi^{-1}(y)} f(x)\,dx & \text{if } \phi^{-1} \text{ mono. inc.} \\ \int_{\phi^{-1}(y)}^{\infty} f(x)\,dx & \text{if } \phi^{-1} \text{ mono. dec.} \end{cases}$$

We can then compute the pdf, $f_Y(y)$, by

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} f(\phi^{-1}(y))\, \phi^{-1\prime}(y) & \text{if } \phi^{-1} \text{ mono. inc.} \\ -f(\phi^{-1}(y))\, \phi^{-1\prime}(y) & \text{if } \phi^{-1} \text{ mono. dec.} \end{cases}$$

$$= f(\phi^{-1}(y)) \left| \phi^{-1\prime}(y) \right|.$$

How about if X discrete with pmf $p_X(x)$? Then

$$p_Y(y) = \mathbb{P}[Y = \phi(X) = y] = \mathbb{P}[X \in \phi^{-1}(y)] = \sum_{x \in \phi^{-1}(y)} p_X(x).$$

A much simpler case, don't require ϕ invertible. Back to the continuous case.

What if ϕ not one-to-one? Well, we may need to split-up the domain.

Ex: Let $X \sim \text{Unif}(-1,2)$, $f_X(x) = \frac{1}{3} \cdot \mathbb{1}\{-1 < x < 2\}$, and let $Y = \phi(X) = X^2$. Split $\phi(X)$ into domains $(-\infty, 0)$, $(0,\infty)$.

$$f_Y(y) = \frac{d}{dy} \mathbb{P}[Y < y] = \frac{d}{dy}\left[\mathbb{P}[-\sqrt{y} < X < 0] + \mathbb{P}[0 < X < \sqrt{y}]\right]$$

$$= \frac{d}{dy}\left[\int_{-\sqrt{y}}^{0} f_X(x)\,dx + \int_0^{\sqrt{y}} f_X(x)\,dx\right]$$

$$= \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(\sqrt{y})$$

$$= \frac{1}{6\sqrt{y}} \cdot \mathbb{1}\{0 < y < 1\} + \frac{1}{6\sqrt{y}} \cdot \mathbb{1}\{0 < y < 4\}.$$

$$= \frac{1}{3\sqrt{y}} \cdot \mathbb{1}\{0 < y < 1\} + \frac{1}{6\sqrt{y}} \cdot \mathbb{1}\{1 < y < 4\}.$$

8/31/23:

More generally, for many-to-one functions, the cdf is given by
$$F_Y(y) = \mathbb{P}[Y \in (-\infty, y)] = \mathbb{P}[X \in \phi^{-1}(y)].$$

And we can find the pdf by $f_Y(y) = \frac{d}{dy} F_Y(y)$.

Ex: $X \sim \text{Exp}(\lambda = 1)$, $Y = \phi(X) = (X-1)^2$. We have that $f_X(x) = e^{-x}$. Then,

$$f_Y(y) = \frac{d}{dy} \mathbb{P}[Y \in (-\infty, y)] = \frac{d}{dy} \mathbb{P}[X \in \phi^{-1}(-\infty, y)]$$

$$= \frac{d}{dy} \mathbb{P}[X \in (-\sqrt{y}+1, \sqrt{y}+1)] = \frac{d}{dy} \int_{-\sqrt{y}+1}^{\sqrt{y}+1} f_X(x)\,dx$$

$$= f_X(\sqrt{y}+1)\frac{1}{2\sqrt{y}} + \begin{cases} f_X(-\sqrt{y}+1)\frac{1}{2\sqrt{y}} & \text{if } 0 < \sqrt{y} < 1 \Rightarrow 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$= \frac{1}{2\sqrt{y}}\left(e^{-\sqrt{y}-1} + e^{\sqrt{y}-1} \cdot I\{0 \le y \le 1\}\right) \text{ for } y \ge 0.$$

Now, suppose $X \sim N(-1, 2)$, and $Y = \phi(X) = (X-2)(X-1)(X+1)(X+2)$.
We know that $f_X(x) = \frac{1}{\sqrt{4\pi}} e^{-(x+1)^2/4}$. We can find critical points of ϕ by $\phi'(x) = 4x^3 - 10x = 2x(2x^2-5) = 0 \Rightarrow x = 0, \pm\sqrt{\frac{5}{2}}$. And $\phi(0) = 4$, $\phi(\pm\sqrt{5/2}) = -\frac{9}{4}$.

$$f_Y(y) = \frac{d}{dy} \mathbb{P}[X \in \phi^{-1}(-\infty, y)] = \begin{cases} \frac{d}{dy} \mathbb{P}[X \in (\phi^{-1}(y)[1], \phi^{-1}(y)[2]) \cup (\phi^{-1}(y)[3], \phi^{-1}(y)[4])] & -9/4 < y < 4 \\ + \frac{d}{dy} \mathbb{P}[X \in (\phi^{-1}(y)[1], \phi^{-1}(4)[1]) \cup (\phi^{-1}(4)[2], \phi^{-1}(y)[2])] & 4 < y \end{cases}$$

For a discrete random variable X w/ pmf $p(x)$, its <u>expected value</u> is given by
$$E[X] = \sum_{x \in supp(x)} x\, p(x).$$

And for continuous random variables,
$$E[X] = \int_{supp(x)} x\, f_x(x)\, dx.$$

These behave nicely under transformations.

Properties:

- Monotonicity: If $X \leq Y$, then $E[X] \leq E[Y]$.
- Linearity: $E[aX + bY] = aE[X] + bE[Y]$ <u>even if</u> X, Y dependent
- If c a constant, $E[c] = c$.

<u>Special Cases:</u> Let X be a random variable

- Mean: $\mu_X = E[X]$
- Variance: $Var(X) = E[(X - \mu_X)^2] = E[X^2] - E[X]^2$
- Moment: $E[X^k]$ for any k
- Moment generating function: $M_X(t) = E[e^{tX}]$, $t \in \mathbb{R}$.

Given data $x_1, x_2, \ldots, x_n \in \mathbb{R}$ and a function $\phi: \mathbb{R} \to \mathbb{R}$, the sample expected value of ϕ is the average, $\frac{1}{n}\sum_{i=1}^{n} \phi(x_i)$.

Examples:

- <u>Sample mean:</u> $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

> **Sample variance:** $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ (only $n-1$ independent summands, from $x_1, \dots, x_{n-1}, \bar{x}$, can find x_n)

> **Sample moments:** $\frac{1}{n}\sum_{i=1}^{n} x_i^k$

> **Sample mgf:** $\frac{1}{n}\sum_{i=1}^{n} e^{tx_i}$, $t \in \mathbb{R}$.

Note: x_i denotes a sample, X_i denotes a random variable.

Properties of mgfs:

> $M_X(0) = 1$ and $M_X(t) \geq 0$ $\forall t$, may be infinite.

9/5/23:

> They generate the moments: $E[X^k] = \frac{d^k}{dt^k}M(t)\Big|_{t=0}$

$$\frac{d^k}{dt^k}M(t) = \frac{d^k}{dt^k}\int_{-\infty}^{\infty} f_x(x)e^{tx}\,dx = \int_{-\infty}^{\infty} f_x(x)\cdot x^k e^{tk}\,dx$$

$$\Rightarrow \frac{d^k}{dt^k}M(t)\Big|_{t=0} = \int_{-\infty}^{\infty} f_x(x)\,x^k\,dx = E[X^k].$$

> Given $Y = aX + b$, $M_Y(t) = e^{tb}M_X(at)$.

$$M_Y(t) = E[e^{atX+tb}] = E[e^{atX}e^{tb}] = e^{tb}E[e^{atX}] = e^{tb}M_x(at).$$

Ex: $X \sim \text{Unif}(-1, 1)$.

$$M_x(t) = \int_{-1}^{1} e^{tx}\cdot\frac{1}{2}\,dx = \begin{cases} 1 & \text{if } t = 0 \\ \frac{e^{tx}}{2t}\Big|_{-1}^{1} & \text{otherwise} \end{cases} = \begin{cases} 1 & \text{if } t = 0 \\ \frac{e^{t}-e^{-t}}{2t} & \text{otherwise} \end{cases}$$

Ex: $X \sim \text{Poisson}(\lambda)$.

$$M_x(t) = \sum_{k=0}^{\infty} e^{tk}\cdot\frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(e^t\lambda)^k}{k!} = e^{-\lambda}\cdot e^{e^t\lambda} = \exp(\lambda(e^t-1))$$

which exists and is positive $\forall t \in \mathbb{R}$.

> If $X_1, \dots, X_n$ are independent, $Y = X_1 + \dots + X_n$, then $M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$

$$M_Y(t) = E[e^{tX_1}e^{tX_2}\cdots e^{tX_n}] \overset{\text{(by indep.)}}{=} E[e^{tX_1}]E[e^{tX_2}]\cdots E[e^{tX_n}] = \prod_{i=1}^{n} M_{X_i}(t)$$

> MGFs uniquely define a random variable, $M_X(t) = M_Y(t) \Rightarrow X = Y$.

The **Strong Law of Large Numbers** says that given iid variables $\{X_n\}_{n=1}^{\infty}$, and letting $Y_n = \frac{1}{N}\sum_{n=1}^{N} X_n$, that assuming $\mu = E[X_i]$ is finite, then Y_n converges to μ, i.e., $P[\lim_{n\to\infty} Y_n = \mu] = 1$

Ex: Exponential, $\lambda = 1$.

$$M_x(t) = \int_0^\infty e^{-x} \cdot e^{tx} \, dx = \int_0^\infty e^{x(t-1)} \, dx \overset{t \neq 1}{=} \frac{e^{x(t-1)}}{t-1} \Big|_0^\infty \overset{t<1}{=} \frac{1}{1-t}, \quad t < 1.$$

And

$$E(X) = \frac{d}{dt} M_x(t) \Big|_{t=0} = \frac{1}{(1-t)^2} \Big|_{t=0} = 1,$$

$$E(X^2) = \frac{d^2}{dt^2} M_x(t) \Big|_{t=0} = \frac{2}{(1-t)^3} \Big|_{t=0} = 2,$$

$$E(X^n) = \frac{d^n}{dt^n} M_x(t) \Big|_{t=0} = \frac{n!}{(1-t)^{n+1}} \Big|_{t=0} = n!.$$

So the n^{th} moment is $n!$.

9/7/23:

Now, let $X_1, X_2, \ldots, X_n$ be random variables, and let $\underline{X}$ be the random vector $\underline{X} = (X_1, \ldots, X_n)^T$. We can describe it with a joint pdf $f_{\underline{x}} : \mathbb{R}^n \to \mathbb{R}^{\geq 0}$, which satisfies that $\int_{\mathbb{R}^n} f_{\underline{x}}(\underline{x}) \, d\underline{x} = 1$.

The multivariate normal distribution has joint pdf

$$f(\underline{x}) = \frac{1}{(2\pi)^{n/2} (\det \underline{\underline{\Sigma}})^{1/2}} \exp\left(-\frac{1}{2} (\underline{x} - \underline{\mu})^T \underline{\underline{\Sigma}}^{-1} (\underline{x} - \underline{\mu})\right)$$

$$\sim N(\underline{\mu}, \underline{\underline{\Sigma}}), \quad \underline{x} \in \mathbb{R}^n.$$

Note that $\Sigma_{ii} = \sigma_{ii} = \sigma_i^2 = Var(X_i)$, and that for $i \neq j$, $\Sigma_{ij} = \sigma_{ij} = Cov(X_i, X_j)$.

A marginal is the distribution of a subset of random variables after having accounted for all others.

Ex: $n = 4$, $\underline{X} = (X_1, X_2, X_3, X_4)$. The marginals for X_2, X_3 are given by

$$f_{X_2, X_3}(x_2, x_3) = \int_{-\infty}^\infty \int_{-\infty}^\infty f_{\underline{x}}(x_1, x_2, x_3, x_4) \, dx_1 \, dx_4.$$

Think of "integrating-out" the other variables.

The distribution of a subset of the variables given some other variables have taken on, is called the conditional distribution

Ex: $n=3$ $$f_{X_3|X_1,X_2}(x_3|x_1,x_2) = \frac{f_{\underline{X}}(x_1,x_2,x_3)}{f_{X_2,X_3}(x_2,x_3)}.$$
The relationship is joint = conditional × marginal.

Ex: Suppose (X,Y) have joint pdf $f(x,y) = 2(x+y)\cdot I\{0<x<y<1\}$

The marginal of X is
$$f_X(x) = \int_x^1 2(x+y)\,dy = [2xy+y^2]_x^1 = 2x+1-3x^2$$
for $0<x<1$.

The conditional of Y given X is then
$$f_{Y|X}(y|x) = \frac{2(x+y)\cdot I\{0<x<y<1\}}{(2x+1-3x^2)\cdot I\{0<x<1\}} = \frac{2(x+y)}{2x+1-3x^2} I\{0<x<y<1\}.$$

Marginals & conditionals are useful for sampling from joint pdfs. First sample from marginal, and then sample remaining variables from corresponding conditionals.

9/12/23:

Let $X_1,\dots,X_n$ be random variables and $f(\underline{x})$ be the joint pdf of $\underline{X} = (X_1,\dots,X_n)^T$. Let $f_{X_i}(x)$ be the marginal pdf of each X_i. Then $X_1,\dots,X_n$ are <u>independent</u> if and only if
$$f_{\underline{X}}(\underline{x}) = \prod_{i=1}^n f_{X_i}(x_i).$$

If X and Y are <u>independent</u>, then the product of their expected values is the expected value of the products
$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy\,dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x) f_Y(y)\,dy\,dx$$
$$= \int_{-\infty}^{\infty} f_X(x)\,dx \int_{-\infty}^{\infty} f_Y(y)\,dy = E[X]E[Y].$$

And this extends to $E[\prod_{i=1}^{n} X_i] = \prod_{i=1}^{n} E[X_i]$ if independent.

Notation: Sometimes use $X \perp Y$ to notate independence of X & Y.

Let $X_1, \dots, X_n$ be independent and $h_1, \dots, h_n : \mathbb{R} \to \mathbb{R}$. Then

$$E\left[\prod_{i=1}^{n} h_i(X_i)\right] = \prod_{i=1}^{n} E[h_i(X_i)].$$

Now, suppose we have random $\underline{X} = [X_1, \dots, X_n]^T \in \mathbb{R}^n$ with joint pdf $f_{\underline{x}}(\underline{x})$. Let $\underline{\Phi} : \mathbb{R}^n \to \mathbb{R}^n$ be an n-dimensional transformation. How do we get the joint pdf of $\underline{\Phi}(\underline{X})$?
We will focus on $\underline{\Phi}$ being one-to-one, but if not, can consider various regions where $\underline{\Phi}$ is one-to-one as before.
This will depend on the determinant of the Jacobian of $\underline{\Phi}^{-1}$.

Ex: Let X_1, X_2 be independent standard normal, $Y_1 = X_1 + X_2$, $Y_2 = X_1 - X_2$. What is the joint pdf of (Y_1, Y_2)?
We know

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1) f_{X_2}(x_2) = \frac{1}{2\pi} e^{-(x_1^2 + x_2^2)/2}.$$

We can also solve for X_1, X_2 as

$$\begin{aligned} Y_1 &= X_1 + X_2 \\ Y_2 &= X_1 - X_2 \end{aligned} \implies \begin{aligned} X_1 &= (Y_1 + Y_2)/2 \\ X_2 &= (Y_1 - Y_2)/2. \end{aligned}$$

This is the inverse transform. So,

$$|\det(\nabla \Phi^{-1})| = \left|\det \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}\right| = \left|-\frac{1}{4} - \frac{1}{4}\right| = \left|-\frac{1}{2}\right| = \frac{1}{2}.$$

Finally, we find the joint distribution of Y_1, Y_2 by

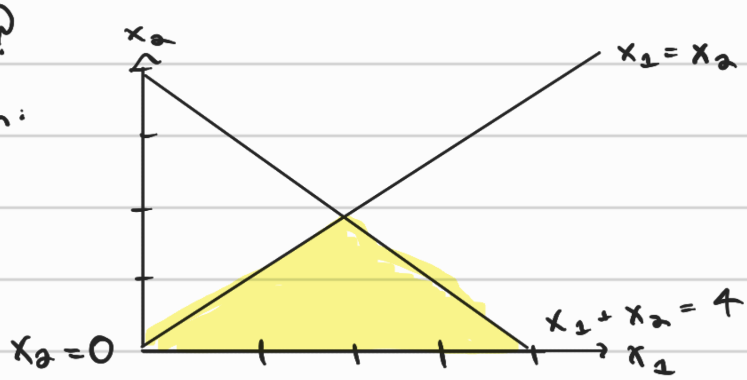
$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1,X_2}(\Phi_1^{-1}, \Phi_2^{-1}) \, |\det \nabla \Phi^{-1}|$$

$$= \frac{1}{4\pi} e^{-((y_1+y_2)^2 + (y_1-y_2)^2)/8}$$

$$= \frac{1}{4\pi} e^{-(y_1^2 + y_2^2)/4}.$$

Ex: Let (X_1, X_2) be uniform random variables on the area bounded by $X_2 > 0$, $X_1 + X_2 < 4$, $X_2 < X_1$. Find the joint pdf of $(X_1, X_1 + X_2)$. And what is the marginal density of $X_1 + X_2$?

Area sketch:



The area of the given region is $\frac{1}{2} \cdot 4 \cdot 2 = 4$. We then know $f_{X_1,X_2}(x_1, x_2) = \frac{1}{4} \cdot I\{(x_1, x_2) \text{ in shaded region}\}$. From def of Y_1, Y_2, $\text{supp}(Y_1) = \text{supp}(X_1) = [0, 4]$. And $\text{supp}(Y_2) = \text{supp}(X_1 + X_2) = [0, 4]$. The inverse transform is

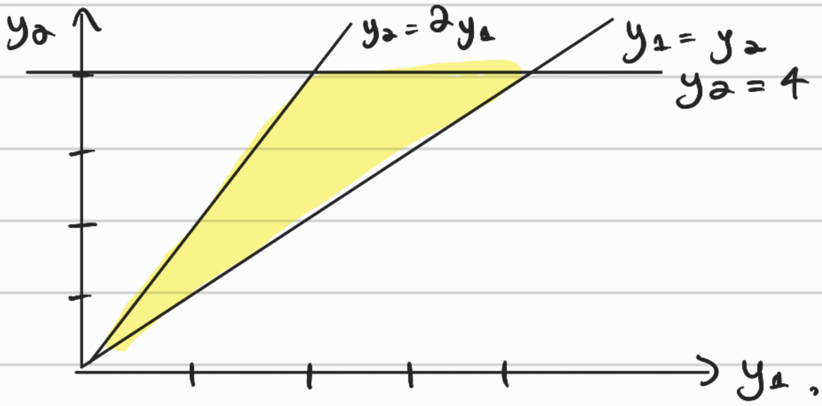
$$\phi^{-1}(\underline{y}) = \begin{bmatrix} y_1 \\ y_2 - y_1 \end{bmatrix} \Rightarrow \nabla \phi^{-1}(\underline{y}) = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$$

$$\Rightarrow \left| \det \nabla \phi^{-1}(\underline{y}) \right| = 1.$$

Hence,
$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(\underline{\phi}^{-1}(\underline{y})) \left| \det \nabla \phi^{-1}(\underline{y}) \right|$$

$$= \frac{1}{4} \cdot I\{y_2 - y_1 > 0,\ y_2 < 4,\ y_2 < 2y_1\}$$

Area sketch:



Finally, the marginal density of $Y_2 = X_1 + X_2$ is

$$f_{Y_2}(y_2) = \int_0^4 f_{Y_1,Y_2}(y_1, y_2)\, dy_1 = \int_{y_2/2}^{y_2} \frac{1}{4} \cdot I\{0 < y_2 < 4\}\, dy_1$$

$$= \frac{y_2}{8} \{0 < y_2 < 4\}.$$

Similarly, the marginal of $Y_1 = X_1$ is

$$f_{Y_1}(y_1) = \int_{y_1}^{2y_1} \frac{1}{4} I\{0<y_1<2\}\, dy_2 + \int_{y_1}^{4} \frac{1}{4} \cdot I\{2<y_1<4\}\, dy_2$$

$$= \frac{y_1}{4} \cdot I\{0<y_1<2\} + \left(1 - \frac{y_1}{4}\right) \cdot I\{2<y_1<4\}.$$

Ex: $U_1, U_2 \sim \text{Unif}(0,1)$. What is joint pdf of (U_1, U_1U_2)?
We know $f_{U_1,U_2}(u_1,u_2) = 1 \cdot I\{0<u_1,u_2<1\}$. And the inverse transform is $U_1 = Y_1$, $U_2 = Y_2/Y_1$. Hence,

$$f_{Y_1,Y_2}(y_1,y_2) = f_{U_1,U_2}\left(y_1, \frac{y_2}{y_1}\right) \cdot \left| \det \begin{bmatrix} 1 & 0 \\ -y_2/y_1^2 & 1/y_1 \end{bmatrix} \right|$$

$$= 1 \cdot I\left\{0<y_1<1,\ 0<\frac{y_2}{y_1}<1\right\} \cdot \frac{1}{y_1}$$

$$= \frac{1}{y_1} \cdot I\{0<y_2<y_1<1\}.$$

9/14/23:

Given iid random variables $X_1, \dots, X_n$ with common pdf $f_X(x)$, we will study the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. We can study the distribution of $\bar{X}$ through moment generating functions. By independence,

$$M_{\bar{X}}(t) = \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{n}\right) = M_X\left(\frac{t}{n}\right)^n.$$

From this, we can obtain

$$E[\bar{X}] = \frac{d}{dt} M_{\bar{X}}(t)\Big|_{t=0} = n\, M_X\left(\frac{t}{n}\right)^{n-1} \cdot M_X'\left(\frac{t}{n}\right) \cdot \frac{1}{n}\Big|_{t=0}$$

$$= E[X].$$

Ex: Assume $X_1, \dots, X_n$ are iid $N(\mu, \sigma^2)$. Then from above,

$$M_{\bar{X}}(t) = \left(M_X\left(\frac{t}{n}\right)\right)^n = \left(\exp\left(\mu\frac{t}{n} + \frac{\sigma^2 \frac{t^2}{n^2}}{2}\right)\right)^n = \exp\left(\mu t + \frac{\sigma^2 t^2}{2n}\right)$$

implying that $\bar{X} \sim N(\mu, \sigma^2/n)$.

Ex: Assume $X_1, \dots, X_n$ iid Gamma(k, θ). What is the dist. of $\bar{X}$. Note, $M_x(t) = (1 - \theta t)^{-k}$ for $t < \frac{1}{\theta}$.

$$M_{\bar{x}}(t) = \left[\left(1 - \theta \frac{t}{n}\right)^{-k}\right]^n = \left(1 - \frac{\theta}{n} t\right)^{-kn} \Rightarrow \bar{X} \sim \text{Gamma}\left(nk, \frac{\theta}{n}\right)$$

$$(t/n < 1/\theta \Rightarrow t < n/\theta)$$

Note that since $E[\bar{x}] = E[x]$, we say that $\bar{X}$ is an <u>unbiased estimator</u> for $\mu = E[X_i]$. By the strong law of large numbers, as $n \to \infty$, $E[\bar{x}] \to \mu$.

How about the variance of $\bar{X}$? Recall $\text{var}(\bar{x}) = E[(\bar{x} - E[\bar{x}])^2]$ $= E[\bar{x}^2] - E[\bar{x}]^2$. So we need the second moment.

$$E[\bar{x}^2] = \frac{d^2}{dt^2} M_{\bar{x}}(t)\Big|_{t=0} = \left[\frac{d}{dt}\left(M_x\left(\tfrac{t}{n}\right)^{n-1} M_x'\left(\tfrac{t}{n}\right)\right)\right]_{t=0}$$

$$= \left[\frac{n-1}{n} M_x\left(\tfrac{t}{n}\right)^{n-2}\left(M_x'\left(\tfrac{t}{n}\right)\right)^2 + \frac{1}{n} M_x\left(\tfrac{t}{n}\right)^{n-1} M_x''\left(\tfrac{t}{n}\right)\right]_{t=0}$$

$$= \frac{n-1}{n} E[X]^2 + \frac{1}{n} E[X^2]$$

$$= E[X]^2 + \frac{1}{n}\left(E[x^2] - E[x]^2\right)$$

$$= \mu^2 + \frac{1}{n} \text{var}(X)$$

$$\Rightarrow \text{var}(\bar{x}) = E[\bar{x}]^2 - \mu^2 = \frac{1}{n} \text{var}(X) =: \frac{\sigma^2}{n}$$

So, in the asymptotic behavior, as $n \to \infty$, $\text{var}(\bar{x}) \to 0$. So as n increases, the variability of $\bar{X}$ decreases.

<u>Chebyshev Inequality</u>: For $0 < \varepsilon << 1$, as $n \to \infty$, we have

$$P\left[|\bar{x} - E[\bar{x}]| > \varepsilon\right] \leq \frac{\text{var}(\bar{x})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

We define the <u>sample variance</u> as

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

The distribution of S^2 is difficult to compute in general. This is because it is a sum of terms like

$$(X_i - \bar{X})^2 = \left(X_i - \frac{\sum_{j=1}^n X_j}{n}\right)^2 = \left(\frac{n-1}{n} X_i - \frac{1}{n} \sum_{j=1, j\neq i}^n X_j\right)^2.$$

If $X_1, \dots, X_n$ iid normal, $N(\mu, \sigma^2)$, then we can show $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$.

Claim: S^2 is an unbiased estimator of σ^2. (Any dist)

$$E[S^2] = \frac{1}{n-1} E\left[\sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2)\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^n E[X_i^2] - nE[\bar{X}^2]\right]$$

$$= \frac{1}{n-1}\left[nE[X_i^2] - nE[\bar{X}^2]\right].$$

Now, recall $E[\bar{X}^2] = \mu^2 + \frac{1}{n}\sigma^2$, and we know that $E[X_i^2] - \mu^2 = \sigma^2 \Rightarrow E[X_i^2] = \mu^2 + \sigma^2$. Hence,

$$E[S^2] = \frac{1}{n-1}\left[n(\mu^2 + \sigma^2) - n\left(\mu^2 + \frac{1}{n}\sigma^2\right)\right]$$

$$= \frac{1}{n-1}\left[\sigma^2(n-1)\right]$$

$$= \sigma^2 \quad \checkmark$$

9/19/23:

Can we compute $\mathrm{var}(S^2)$? Takes more involved algebra, but can show

$$\mathrm{var}(S^2) = E[S^4] - E[S^2]^2 = \frac{(n-1)\left((n-1)E[X^4] + (n^2-2n+3)E[X^2]^2\right)}{n^3} - (\sigma^2)^2$$

$$= \frac{n-1}{n^3}\left((n-1)E[X^4] - (n-3)\sigma^4\right)$$

$\to 0$ as $n \to \infty$

And if we input S^2 into Chebyshev's inequality, we have

$$P[\,|S^2 - \sigma^2| > \varepsilon\,] \le \frac{1}{\varepsilon^2} \operatorname{var}(S^2).$$

Let $X_1, X_2, \dots, X_n$ be independent $N(\mu_i, \sigma_i^2)$ random variables. Let $a_1, a_2, \dots, a_n \in \mathbb{R}$, then we know that

$$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).$$

Also, recall the special case

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N\left(\frac{1}{n} \sum_{i=1}^{n} \mu_i, \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2\right).$$

Now, suppose $X_1, \dots, X_n$ iid $N(\mu, \sigma^2)$. Then we can show that

1) $\bar{X}$ and S are independent
2) $\bar{X}$ and $(X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X})$ are independent
3) $\frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1}$

(Will be proven below)

Suppose we model a dataset generating iid $N(\mu, \sigma^2)$. Suppose given sample, $n = 25$, $s^2 = 14.5$. What is the probability that σ^2 is within 10% of 16?

$$P[14.4 \le S^2 \le 17.6] = P\left[\frac{14.4 \cdot 24}{\sigma^2} \le \chi^2_{24} \le \frac{17.6 \cdot 24}{\sigma^2}\right]$$

supp $\sigma^2 = 16$ $\qquad \approx 33\%$

<u>Proof of statements above 1-3</u>

Define $Y_1 = \bar{X}$, then $Y_i = X_i - \bar{X}$ for $i = 2, 3, \dots, n$. We can see from this definition that the inverse is given by $X_i = Y_1 + Y_i$ for $i = 2, \dots, n$, and that $X_1 = nY_1 - \sum_{i=2}^{n} X_i = Y_1 - \sum_{i=2}^{n} Y_i$. Hence, we can compute $f_{\underline{Y}}$ by

$$f_{\underline{Y}}(\underline{y}) = f_{\underline{X}}\left(y_1 - \sum_{i=2}^{n} y_i, \; y_1 + y_2, \dots, y_1 + y_n\right) \cdot \begin{vmatrix} 1 & -1 & -1 & \cdots & -1 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ 1 & 0 & 0 & \cdots & 1 \end{vmatrix}$$

Can show det = n

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_1 - \sum_{i=2}^{n} y_i - \mu)^2/(2\sigma^2)} \cdot \prod_{i=2}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_1 + y_i - \mu)^2/(2\sigma^2)} \cdot n$$

$$= \frac{n}{(2\pi\sigma^2)^{n/2}} \exp\left(\frac{-1}{(2\sigma^2)}\left(\sum_{i=1}^{n} y_i^2 + 2\sum_{i=2}^{n}\sum_{j=i}^{n} y_i y_j - 2y_1\mu - 2\sum_{i=2}^{n} y_1 y_i + 2\sum_{i=2}^{n} y_i \mu + \mu^2 \right.\right.$$

$$\left.\left. + (n-1)y_1^2 + \sum_{i=2}^{n} y_i^2 + 2\sum_{i=1}^{n} y_1 y_i - 2(n-1)y_1\mu - 2\sum_{i=2}^{n} y_i\mu + (n-1)\mu^2\right)\right)$$

$$= \frac{n}{(2\pi\sigma^2)^{n/2}} \exp\left(\frac{-1}{(2\sigma^2)}\left(ny_1^2 + 2\sum_{i=2}^{n}\sum_{j=i}^{n} y_i y_j - 2ny_1\mu + n\mu^2 + 2\sum_{i=2}^{n} y_i^2\right)\right)$$

from which we see we can factor out y_1, i.e. $Y_1 = \bar{X}$ is independent from all other $Y_i = (X_i - \bar{X})$. To show independence to $X_1 - \bar{X}$, modify defn of $Y_2 = X_1 - \bar{X}$. This proves statement ② above. For statement ①, we note that since $\bar{X} \perp X_i - \bar{X} \ \forall i$, this implies $\bar{X} \perp \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S^2$.

Finally, we wish to show ③, $\frac{n-1}{\sigma} S^2 \sim \chi^2_{n-1}$. Well,

$$(n-1)S^2 = \sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2$$

$$= \sum_{i=1}^{n}\left[(X_i - \mu)^2 + 2(X_i - \mu)(\mu - \bar{X})\right] + n(\mu - \bar{X})^2$$

$$= \sum_{i=1}^{n}(X_i - \mu)^2 - 2(\bar{X} - \mu)(n\bar{X} - n\mu) + n(\mu - \bar{X})^2$$

$$= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2$$

$$\Rightarrow \frac{n-1}{\sigma^2} S^2 = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} - \frac{(\bar{X} - \mu)^2}{\sigma^2/n}$$

$$\sim \chi^2_n - \chi^2_1 \quad \left(\text{Since } \begin{array}{l} X_i \sim N(\mu, \sigma^2) \\ \bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \end{array}\right)$$

Finally, need to show $\chi^2_n - \chi^2_1 \sim \chi^2_{n-1}$ with first terms dependent. By S^2, $\bar{X}$ independent,

$$M_{\frac{(X_i-\mu)^2}{\sigma^2}}(t) = M_{\frac{n-1}{\sigma^2}S^2}(t)\, M_{\frac{\bar{X}-\mu}{\sigma^2/n}}(t)$$

$$\Rightarrow M_{\frac{n-1}{\sigma^2}S^2}(t) = \frac{M_{\frac{(X_i-\mu)}{\sigma^2}}(t)}{M_{\frac{\bar{X}-\mu}{\sigma^2/n}}(t)} = \frac{M_{\chi^2_n}(t)}{M_{\chi^2_1}(t)} \overset{\text{(can show)}}{\sim} \chi^2_{n-1}.$$

9/26/23:

Let $X_1, X_2, \ldots$ be iid random variables with mean μ, variance σ^2.

Weak Law of Large Numbers: For every $\varepsilon > 0$, then

$$\lim_{n\to\infty} P[\,|\bar{X}_n - \mu| < \varepsilon\,] = 1$$

So we say $\bar{X}_n$ converges in probability to μ, or $\bar{X}_n \xrightarrow{P} \mu$.

General Statement: Let $X_1, X_2, \ldots$ be a sequence of random variables. We say that (X_n) converges to X in probability if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P[\,|X_n - X| < \varepsilon\,] = 1.$$

Or, reversing the inequality:

$$\lim_{n\to\infty} P[\,|X_n - X| > \varepsilon\,] = 0.$$

Now, recall Chebyshev's inequality,

$$P[\,|X - E[X]| > \varepsilon\,] \leq \frac{Var(X)}{\varepsilon^2}.$$

So plugging in $\bar{X}$ or S^2, then we can prove that $\bar{X} \to \mu$ and $S^2 \to \sigma^2$ as long as $Var(\bar{X})$, $Var(S^2) \to 0$ as $n \to \infty$, which we've already shown.

Strong Law of Large Numbers: Let $X_1, \ldots, X_n$ be iid random variables with finite expectation $E[X_i] = \mu$. Then, for all $\varepsilon > 0$,

$$P[\lim_{n\to\infty} |\bar{X}_n - \mu| < \varepsilon\,] = 1$$

also referred to as pointwise convergence or almost-sure convergence.

General Statement: Given $(X_n)_{n=1}^{\infty}$, this converges almost-surely to a random variable X if for every $\varepsilon > 0$,

$$P\left[\lim_{n\to\infty} |X_n - X| < \varepsilon\right] = 1$$

which is equivalent to

$$P\left[\lim_{n\to\infty} X_n = X\right] = 1.$$

The strong law is more difficult to prove.

We say that $(X_n)_{n=1}^{\infty}$ converges in distribution to X if

$$\lim_{n\to\infty} P[X_n \le a] = P[X \le a]$$

for every $a \in \mathbb{R}$. I.e., they converge in cdf. Note that this is the weakest form of convergence. We also call this convergence in law. Write $X_n \xrightarrow{D} X$.

So, almost-sure convergence (pointwise) implies convergence in probability (consistency) implies probability in distribution. And convergence in distribution implies convergence in probability only if the limit is a constant.

Central Limit Theorem (CLT): Let $X_1, \ldots, X_n$ be iid random variables. The approximate form states that $\sum_{i=1}^{n} X_i$ is approximately normal with mean $n \cdot E[X_i]$, and variance $n \cdot \mathrm{var}(X_i)$. The formal statement says, letting $\mu = E[X_i]$, $\sigma^2 = \mathrm{var}(X_i)$, that

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} \xrightarrow{d} N(0,1) \quad \text{as} \quad n \to \infty$$

where $\xrightarrow{d}$ denotes convergence in distribution (cdfs converge). We could also write this as, for any $a \in \mathbb{R}$,

$$P\left[\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} \leq a\right] \longrightarrow \int_{-\infty}^{a} \phi(s)\,ds \quad \text{as} \quad n \to \infty.$$

Overview, from strongest to weakest:

1) <u>Almost-sure convergence</u> (Strong law)

$$P\left[\lim_{n\to\infty} |X_n - X| < \varepsilon\right] = 1$$

2) <u>Convergence in probability</u> (Weak law)

$$\lim_{n\to\infty} P\left[|X_n - X| < \varepsilon\right] = 1$$

3) <u>Convergence in distribution / law:</u>

$$\forall a \in \mathbb{R}, \quad \lim_{n\to\infty} P[X_n < a] = P[X < a]$$

<u>Why?</u> Convergence in distribution happens for any iid variables as their cdfs are the same, but not for convergence in probability or almost-sure convergence. Almost-sure convergence more complicated.

<u>9/28/23:</u>

Recall the central limit theorem. Given iid X_i with $E[X_i] = \mu$, $\text{var}(X_i) = \sigma^2$, then

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty.$$

<u>Proof of CLT:</u> Let $Y_n = \left(\sum_{i=1}^{n} X_i - n\mu\right)/(\sqrt{n}\,\sigma)$. We wish to show

$$\lim_{n\to\infty} m_{Y_n}(t) = m_{N(0,1)}(t) = e^{t^2/2}.$$

First, we'll calculate

$$m_{Y_n}(t) = E[e^{tY_n}] = E\left[\prod_{i=1}^{n} \exp\left(\frac{t}{\sqrt{n}\,\sigma}(X_i - \mu)\right)\right]$$

$$= E\left[\exp\left(\frac{t}{\sqrt{n}\,\sigma}(X_i - \mu)\right)\right]^n$$

$$= \left[m_{X_i-\mu}\left(\frac{t}{\sqrt{n}\,\sigma}\right)\right]^n.$$

Then,

$$\lim_{n\to\infty} m_Y(t) = \lim_{n\to\infty}\left[m_{X_i-\mu}\left(\frac{t}{\sqrt{n}\sigma}\right)\right]^n.$$

Taylor expanding,

$$m_{X_i-\mu}\left(\frac{t}{\sqrt{n}\sigma}\right) = 1 + \frac{t}{\sqrt{n}\sigma}\overset{(0)}{E[X_i-\mu]} + \frac{t^2}{2n\sigma^2}\overset{(\sigma^2)}{E[(X_i-\mu)^2]} + O\left(\frac{t^3}{n^{3/2}\sigma^6}\right)$$

$$= 1 + \frac{t^2}{2n} + O\left(n^{-3/2}\right).$$

And, it can be shown that *

$$\lim_{n\to\infty}\left[1+\frac{t}{n}+a_n\right]^n = e^t \quad \text{if} \quad \lim_{n\to\infty} a_n\cdot n = 0, \text{ i.e. } a_n = o(n).$$

Hence,

$$\lim_{n\to\infty}\left[m_{X_i-\mu}\left(\frac{t}{\sqrt{n}\sigma}\right)\right)^n = \lim_{n\to\infty}\left[1+\frac{t^2/2}{n}+O\left(n^{-3/2}\right)\right]^n = e^{t^2/2}.$$

Hence, we have that $\left(\sum_{i=1}^n X_i-\mu\right)/(\sqrt{n}\sigma) \sim N(0,1)$.

$$* \log\left[\lim_{n\to\infty}\left(1+\frac{t}{n}+a_n\right)^n\right] = \lim_{n\to\infty} n\cdot\log\left(1+\frac{t}{n}+a_n\right)$$

$$\overset{(T.S.)}{=} \lim_{n\to\infty} n\cdot\sum_{k=1}^\infty \frac{(-1)^{k+1}}{k}\left(\frac{t_n}{n}+a_n\right)^k \quad \left(\log(1+x) = \overset{0}{\log(1)} + \frac{x}{(1+y)|_{y=0}} - \frac{x^2/2}{(1+y)^2|_{y=0}} + \cdots\right)$$

$$= \lim_{n\to\infty}\left(t_n + a_n\cdot n\right) + \lim_{n\to\infty} n\cdot O\left(n^{-2}\right)$$

$$= t_n.$$ And since $\log\left(\lim_{n\to\infty}\left(1+\frac{t_n}{n}+a_n\right)^n\right) = t$, then $\lim_{n\to\infty}\left(1+\frac{t}{n}+a_n\right)^n = e^t$. *

10/3/23:

The $\chi^2(r)$ or χ^2_r is a one-parameter family of distributions, indexed by $r\in\mathbb{Z}^{>0}$. It has the pdf

$$f(x;r) = \frac{x^{r/2-1}e^{-x/2}}{2^{r/2}\Gamma(r/2)}\cdot I[x\geq 0].$$

It has mean r and variance $2r$. Given iid $Z_1, \ldots, Z_r$, which are $N(0,1)$, we have that

$$Z_1^2 + \cdots + Z_r^2 = \sum_{i=1}^{r} Z_i^2 \sim \chi_r^2.$$

It is a special case of the Gamma distribution, χ_r^2 is the same as $\text{Gamma}(\frac{r}{2}, \frac{1}{2})$. Recall,

$$f_{\text{Gamma}(\alpha,\beta)}(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \cdot I[x \geq 0].$$

Let $X \sim \text{Gamma}(\alpha, \beta)$, and $Y = 2X\beta$. Then $Y \sim \chi^2(r)$ where $r = 2\alpha$.

Proof: This is an invertible map with inverse $X = \frac{1}{2\beta} Y =: \phi^{-1}(Y)$. And $\frac{d}{dy}\phi^{-1} = \frac{1}{2\beta}$. Hence,

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

$$= \frac{\beta^\alpha \left(\frac{y}{2\beta}\right)^{\alpha-1} e^{-\beta y/(2\beta)}}{\Gamma(\alpha)} \cdot \frac{1}{2\beta} \cdot I\left[\frac{y}{2\beta} \geq 0\right]$$

$$= \frac{y^{\alpha-1} e^{-y/2}}{2^\alpha \Gamma(\alpha)} \cdot I[y \geq 0]$$

$$= f_{\chi_r^2}(y) \text{ where } \alpha = \frac{r}{2} \checkmark$$

The mgf of a $\text{Gamma}(\alpha, \beta)$ is given by

$$m_{\text{Gamma}(\alpha,\beta)}(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha}, \quad t < \beta$$

so, the mgf of a χ_r^2 is given by

$$m_{\chi_r^2}(t) = (1 - 2t)^{-r/2}, \quad t < \frac{1}{2}.$$

<u>10/17/23:</u>

From the MGF,

$$E[\chi_r^2] = r(1 - 2(0))^{-r/2-1} = r,$$

$$\mathrm{var}(\chi^2_r) = -2r\left(\frac{-r}{2}-1\right) - r^2 = 2r.$$

<u>Claim:</u> If $Z \sim N(0,1)$, $Z^2 \sim \chi^2_1$.

$$M_{Z^2}(t) = E[e^{tZ^2}]$$

$$tx^2 - x^2/2$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx^2} e^{-x^2/2}\, dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(1-2t)x^2/2}\, dx \quad \left(<\infty \text{ for } t - \tfrac{1}{2} < 0\right)$$

$$= \int_{-\infty}^{\infty} \frac{1/\sqrt{1-2t}}{\sqrt{2\pi/(1-2t)}} e^{-\frac{1}{2}(x/\sqrt{1-2t})^2}\, dx$$

$$= (1-2t)^{-1/2} = M_{\chi^2_1}(t).$$

Can use similar argument to show for iid std normals Z_i,

$$M_{\sum_{i=1}^n Z_i^2}(t) = (1-2t)^{-n/2} = M_{\chi^2_n}(t).$$

Now, recall that if $X_1, \dots, X_n$ iid $N(\mu, \sigma^2)$, that $\bar{X} \sim N(\mu, \sigma^2/n)$,

so
$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

What if instead use $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$? The <u>Student's t-distribution</u> can be written as

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1).$$

Or, more specifically, given $Z \sim N(0,1)$, $V \sim \chi^2_r$, independent,

$$\frac{Z}{\sqrt{V/r}} \sim t(r).$$

Former example, we showed previously that $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, and we know $(\bar{X}-\mu)/\sqrt{\sigma^2/n} \sim N(0,1)$, so

$$\frac{(\bar{X}-\mu)/\sqrt{\sigma^2/n}}{\sqrt{(n-1)S^2/\sigma^2}/\sqrt{n-1}} = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1) \checkmark$$

It has a pdf of

$$f(t,r) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)} \cdot \frac{1}{\sqrt{r\pi}} \left(1 + \frac{t^2}{r}\right)^{-(r+1)/2}$$

<u>Note:</u> When plotted, looks very similar to standard normal as r grows large. As $r \to \infty$ $t \to N(0,1)$.

<u>10/19/23:</u>

The t-distribution is symmetric, mean 0, variance $\frac{r}{r-2}$ when $r > 2$, undefined otherwise.

<u>Proof of pdf:</u> Let $Z \sim N(0,1)$, $V \sim \chi^2_r$ independent. By independence, the joint pdf is given by the product of the marginals:

$$f_{Z,V}(z,x) = f_Z(z) f_V(x) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \cdot \frac{x^{r/2-1} e^{-x/2}}{\Gamma(r/2)\, 2^{r/2}} \cdot I[x \geq 0].$$

Now, consider the transformation $T = Z/\sqrt{V/r} = \phi_1(z,v)$ and $Y = V$. The inverse is given by $V = Y$, and $Z = T\sqrt{V/r} = T\sqrt{Y/r}$. The Jacobian of this transform is given by

$$\begin{vmatrix} \sqrt{y/r} & \frac{1}{2r} t / \sqrt{y/r} \\ 0 & 1 \end{vmatrix} = \sqrt{y/r}.$$

Hence,

$$f_{T,Y}(t,y) = f_{Z,V}\left(t\sqrt{y/r}, y\right) \cdot \sqrt{y/r}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-t^2 y/(2r)} \cdot \frac{y^{r/2-1} e^{-y/2}}{\Gamma\left(\frac{r}{2}\right) 2^{r/2}} \cdot \sqrt{\frac{y}{r}} \cdot I[y \geq 0]$$

We can then get the marginal of T by

$$f_T(t) = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2 y/(2r)} \cdot \frac{y^{r/2} e^{-y/2}}{\Gamma\left(\frac{r}{2}\right) 2^{r/2}} \sqrt{\frac{y}{r}}\, dy$$

$$= \frac{1}{\sqrt{2\pi r}\, \Gamma\left(\frac{r}{2}\right) 2^{r/2}} \int_0^\infty y^{r/2+1} e^{-\frac{y}{2}(1+t^2/r)}\, dy.$$

Now, the above has the form of a Gamma(α, β) with $\alpha = (r+1)/2$, $\beta = (t^2+r)/(2r)$. All we are missing is $\frac{\beta^\alpha}{\Gamma(\alpha)}$.

Hence,

$$f_T(t) = \frac{\Gamma\left(\frac{r+1}{2}\right) / \left(\frac{t^2+r}{2r}\right)^{\frac{r+1}{2}}}{\sqrt{2\pi r}\,\Gamma\left(\frac{r}{2}\right) 2^{r/2}}$$

$$= \frac{\Gamma\left(\frac{r+1}{2}\right)\cdot(2r)^{\frac{r+1}{2}}}{(t^2+r)^{\frac{r+1}{2}}\sqrt{2\pi r}\,\Gamma\left(\frac{r}{2}\right) 2^{r/2}}$$

$$= \frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)} \cdot \frac{1}{\sqrt{r\pi}} \left(1 + \frac{t^2}{r}\right)^{-(r+1)/2} \quad \checkmark$$

It does not have an mgf, and it only as up to moment $r-1$. So $E[T] = 0$ only for $r \geq 2$, and $\operatorname{var}(T)$ only exists for $r \geq 3$.

Let $X_1, \dots, X_n$ be iid $N(\mu, \sigma^2)$. We showed that both

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{and} \quad \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1).$$

So it makes a lot of sense that since $S \to \sigma$, that $t(n) \to N(0,1)$ as $n \to \infty$.

Ex: Suppose $X_1, \dots, X_n$ iid $N(\mu, \sigma^2)$, σ^2 unknown. We wish to find $P[|\bar{x} - \mu| < 2]$. We know $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$. So $\bar{X} - \mu \sim N(0, \frac{\sigma^2}{n})$. Since σ^2 unknown, calculate S^2. We know that $(\bar{x}-\mu)/(S/\sqrt{n}) \sim t(n-1)$. So, letting $T \sim t(n-1)$,

$$P[|\bar{x}-\mu| < 2] \approx P\left[\left|\frac{\bar{X}-\mu}{S/\sqrt{n}}\right| < \frac{2\sqrt{n}}{s}\right]$$

* Ask why can div by S and s at once
(S: R.V., s: sample st.dev.)

$$= P\left[\frac{-2\sqrt{n}}{s} < T < \frac{2\sqrt{n}}{s}\right],$$

and then we can use a table to give us proper values.

Suppose we want $c > 0$ s.t. (a confidence interval)

$$P[\bar{X} - c < \mu < \bar{X} + c] = p \quad (\text{say } 0.95).$$

We can do this by equating to $|\bar{x} - \mu| < c$:

$$p = P[|\bar{X} - \mu| < c]$$

$$\overset{\text{former work}}{=} P\left[\frac{-c\sqrt{n}}{S} < T < \frac{c\sqrt{n}}{S}\right]$$

and choose c properly.

The <u>F-distribution</u> appears when analyzing ratios of sample variances. Let $V_1 \sim \chi^2(v_1)$, $V_2 \sim \chi^2(v_2)$ be independent. Then

$$F = \frac{V_1 / v_1}{V_2 / v_2} \sim F(v_1, v_2).$$

The pdf is

$$f_F(x; v_1, v_2) = \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} \left(\frac{v_1}{v_2}\right)^{v_1/2} x^{v_1/2 - 1} \left(1 + \frac{v_1}{v_2}x\right)^{-(v_1+v_2)/2}$$

for $x > 0$.

PDF derivation same as before. First, $f_{V_1,V_2} = f_{V_1} \cdot f_{V_2}$ by independence. Then, transform $F = \frac{V_1/v_1}{V_2/v_2}$, $Y = V_2/v_2$. Clearly this is invertible, compute $f_{F,Y}(x,y)$, and then integrate out Y from $y=0$ to ∞ to get the marginal $f_F(x)$.

<u>10/24/23:</u>

Suppose given $X_1, \ldots, X_n$ iid $N(\mu_1, \sigma_1^2)$ and $Y_1, \ldots, Y_m$ iid $N(\mu_2, \sigma_2^2)$. We don't know μ_1, σ_1, μ_2, or σ_2, want to test whether or not $\sigma_1 = \sigma_2$. Recall, $\frac{(n-1)S_1^2}{\sigma_1^2}$ and $\frac{(m-1)S_2^2}{\sigma_2^2}$ are χ^2 with $n-1$, $m-1$ degrees of freedom. Then, using independence, $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n-1, m-1)$. Additionally, if we assume $\sigma_1 = \sigma_2$ (a null hypothesis), then we'd have that $\frac{S_1^2}{S_2^2} \sim F(n-1, m-1)$. Then, we can obtain a rejection region under some p-value, and see if $\frac{S_1^2}{S_2^2}$ lies in that range or not.

Suppose we wish to know about the minimum, maximum, or median of some iid random variables $X_1, \ldots, X_n$. The order statistics of $(X_1, \dots, X_n)$ are the quantities $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ such that $X_{(k)}$ provides a distribution for the k'th smallest value of $(X_1, \dots, X_n)$. Let $f(x)$, $F(x)$ denote the pdf & cdf of each X_i respectively. We can compute the cdf by

$$F_{X_{(k)}}(x) = P[X_{(k)} \leq x]$$

$$= P[\text{"at least } k \text{ of } \underline{X} \leq x, \text{ rest} > x\text{"}]$$

$$= \sum_{j=k}^{n} \binom{n}{j} F(x)^j (1 - F(x))^{n-j}.$$

The joint pdf of $X_{(1)}, \dots, X_{(n)}$, is given by

$$f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) \underset{\text{use } \varepsilon\text{-limit}}{\text{"="}} P(X_{(1)} = x_1, \dots, X_{(n)} = x_n)$$

$$= n! \prod_{i=1}^{n} f(x_i) \cdot I[x_1 \leq x_2 \leq \dots \leq x_n].$$

Now, the pdf will be given by

$$f_{X_{(k)}}(x) = \frac{d}{dx} F_{X_{(k)}}(x)$$

$$= \sum_{j=k}^{n} \binom{n}{j} f(x) \left[j F(x)^{j-1} (1 - F(x))^{n-j} - (n-j) F(x)^j (1 - F(x))^{n-j-1} \right]$$

$$= \dots \quad \text{(use induction)}$$

$$= \frac{n!}{(k-1)!(n-k)!} f(x) F(x)^{k-1} (1 - F(x))^{n-k}.$$

Or, can show it by integrating the joint cdf.

$$f_{X_{(2)}, \dots, X_{(n)}}(x_2, \dots, x_n) = \int_{-\infty}^{x_2} f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) \, dx_1$$

$$= n! \cdot F(x_2) \cdot \prod_{i=2}^{n} f(x_i),$$

$$f_{X_{(3)}, \dots, X_{(n)}}(x_3, \dots, x_n) = \frac{n!}{2} \cdot F(x_3)^2 \prod_{i=3}^{n} f(x_i)$$

...

$$f_{X_{(k)},\dots,X_{(n)}}(x_k,\dots,x_n) = \frac{n!}{(k-1)!} F(x_k)^{k-1} \prod_{i=k}^{n} f(x_i),$$

$$f_{X_{(k)},\dots,X_{(n-1)}}(x_k,\dots,x_{n-1}) = \frac{n!}{(k-1)!\cdot 1} F(x_k)^{k-1}(1-F(x_{n-1})) \prod_{i=k}^{n-1} f(x_i)$$

...

$$f_{X_{(k)}}(x_k) = \frac{n!}{(k-1)!\,(n-k)!} f(x)\, F(x)^{k-1}(1-F(x))^{n-k}.$$

Ex: Suppose $X_1,\dots,X_n$ iid Unif(0, 1). Recall then that $F(x) = x$, and $f(x) = 1$. Then,

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!\,(n-k)!} x^{k-1} (1-x)^{n-k} \cdot I[0 \leq x \leq 1]$$

$$= \frac{x^{k-1}(1-x)^{n-k}}{\Gamma(k)\Gamma(n-k+1)/\Gamma(n+1)} \cdot I[0 \leq x \leq 1]$$

$$\Rightarrow X_{(k)} \sim \text{Beta}(k, n-k+1).$$

From this: $$E[X_{(k)}] = \frac{k}{n+1}.$$

In <u>point estimation</u>, we do the following:

1) Assume to generate n iid samples from the same population
2) Model the data as having a common distribution $f(x;\theta)$
3) Create an estimator $\hat{\theta}(\underline{x})$

We wish to understand properties of the estimator

10/26/23:

In <u>estimation theory</u>, some terminology is used:

- An <u>estimate</u> is the observed value of an estimator
- A <u>statistic</u> is any function of some sample

> An <u>estimator</u> is a statistic that estimates some unknown parameter

<u>10/31/23:</u>

The <u>maximum liklihood estimator</u> estimates a parameter based on what is most likely to occur. We create a <u>liklihood function</u>

$$L(\theta ; \underline{x}) = f_{\underline{x}}(\underline{x} ; \theta)$$

and choose θ to maximize it. For iid X_i,

$$\text{iid}: \quad L(\theta ; \underline{x}) = \prod_{i=1}^{n} f(x_i ; \theta).$$

It is commonly easier to maximize the <u>log liklihood function</u>

$$\text{iid}: \ \ell(\theta ; \underline{x}) = \log(L(\underline{x} ; \theta)) = \sum_{i=1}^{n} \log(f(x_i ; \theta)).$$

<u>Ex:</u> $X_1, \dots, X_n$ iid, $f(x) = \theta^2 x e^{-\theta x}, \quad x > 0.$

$$L(\theta ; \underline{x}) = \prod_{i=1}^{n} f(x_i) = \theta^{2n} \left[\prod_{i=1}^{n} x_i \right] e^{-\theta \sum_{i=1}^{n} x_i} \cdot I[\text{each } x_i > 0]$$

$$\Rightarrow \ell(\theta ; \underline{x}) = 2n \log(\theta) + \sum_{i=1}^{n} \log(x_i) - \theta \sum_{i=1}^{n} x_i \cdot I[\text{each } x_i > 0]$$

$$\Rightarrow \frac{d\ell}{d\theta} = \frac{2n}{\theta} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \hat{\theta}_{MLE} = \frac{2}{\bar{X}}.$$

<u>Distinction:</u> MLE written in terms of RV: $\bar{X}$, but then given data, our MLE estimate is $\bar{x}$.

<u>Ex:</u> $X_1, \dots, X_n$ iid Bernoulli(p).

$$L(p ; \underline{x}) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$

$$\Rightarrow \ell(p ; \underline{x}) = \sum_{i=1}^{n} \left(x_i \log(p) + (1-x_i) \log(1-p) \right)$$

$$\Rightarrow \frac{d\ell}{dp} = \sum_{i=1}^{n}\left(\frac{x_i}{p} - \frac{1-x_i}{1-p}\right) = 0 \Rightarrow \frac{n\bar{X}}{p} = \frac{n-n\bar{X}}{1-p}$$

$$\Rightarrow n\bar{X} = p(n-n\bar{X}+n\bar{X}) \Rightarrow p = \bar{X}$$

so we write $\hat{p}_{MLE} = \bar{X}$.

Ex: $X_1, \dots, X_n \sim \text{Unif}(0,\theta)$, $\theta > 0$.

$$L(\theta; \underline{x}) = \prod_{i=1}^{n} \frac{1}{\theta} I[0 \le x_i \le n] = \theta^{-n} \cdot I[0 \le x_1, \dots, x_n \le \theta].$$

Now, clearly $\hat{\theta}_{MLE} \ge x_1, \dots, x_n$. Additionally, L is a decreasing function of θ, so, smallest nonzero optimal value is $\hat{\theta}_{MLE} = \max(\underline{X}) = X_{(n)}$.

11/2/23:

MLE estimators have the <u>invariance property</u> that if $\hat{\theta}_{MLE}$ is the MLE for θ, then given one-to-one differentiable $\tau(\theta)$, $\tau(\hat{\theta}_{MLE})$ is the MLE for $\tau(\theta)$. Why? Given $L(\theta; \underline{x})$, we have that

$$L(\theta; \underline{x}) = L(\theta^{-1}(\tau); \underline{x}) \Rightarrow \frac{\partial L}{\partial \tau} = \frac{dL}{d\theta}\frac{d\theta^{-1}}{d\tau} = 0 \Leftrightarrow \left.\frac{dL}{d\theta}\right|_{\theta = \tau^{-1}(\tau)} = 0$$

since from $\tau(\theta)$ one-to-one, $\frac{d\tau^{-1}(\tau)}{d\tau} \ne 0$. So, $\hat{\theta}_{MLE} = \tau^{-1}(\hat{\tau}_{MLE}) \Rightarrow$ $\hat{\tau}_{MLE} = \tau(\hat{\theta}_{MLE})$. Still should work if $\tau(\theta)$ one-to-one. Since $\hat{\theta}_{MLE}$ maximizes $L(\theta) \Rightarrow \exists\, \varepsilon > 0$ s.t. $\hat{\theta}_{MLE} \ge \theta \;\; \forall\, \theta \in (\hat{\theta}_{MLE} - \varepsilon, \hat{\theta}_{MLE} + \varepsilon)$, so can replace $\hat{\theta}_{MLE}$ with $\tau^{-1}(\tau(\hat{\theta}_{MLE}))$ to get the same result, and $\hat{\tau}_{MLE} = \tau(\hat{\theta}_{MLE})$ maximizes $L(\tau^{-1}(\hat{\tau}_{MLE})) = L(\tau)$. And it works for any well-defined $\tau(\theta)$, just that τ^{-1} can be a one-to-many "function", so there would be several candidates.

The <u>Method of Moments estimator (MME)</u> equates sample moments to theoretical moments to solve for various unknowns from a system of equations.

Ex: Consider $X_1, \dots, X_n$ iid with pdf $f(x; \theta)$, $\theta \in \mathbb{R}$.

The theoretical mean is $E[X] = \int x f(x)\,dx = M_1(\theta)$. The sample mean is $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. So we choose $\hat{\theta}_{MME}$ to match $M_1(\hat{\theta}_{MME}) = \bar{X}$.

In general, if we have k unknowns, $\underline{\theta} \in \mathbb{R}^k$, we set up the k equations

$$M_i(\underline{\theta}) = E[X^i] = \frac{1}{n}\sum_{j=1}^{n} X_j^{\,i} = m_i, \quad i = 1, \dots, k.$$

However, if the system of equations is undetermined or unidentified, may need more than k equations.

Ex: Let $X_1, \dots, X_n \sim \text{Gamma}(K, \theta)$. Recall $f(x) = \frac{1}{\Gamma(K)\theta^K} x^{K-1} e^{-x/\theta}$.

$$E[X] = \int_0^\infty x f(x)\,dx = K\theta \overset{\text{set}}{=} \bar{X}$$

$$E[X^2] = \int_0^\infty x^2 f(x)\,dx = K\theta^2 + (K\theta)^2 \overset{\text{set}}{=} m_2.$$

Combining the above 2,

$$m_2 = K\theta^2 + (K\theta)^2 = \bar{X}\theta + \bar{X}^2 \Rightarrow \hat{\theta}_{MME} = \frac{m_2 - \bar{X}^2}{\bar{X}}$$

$$= \frac{\frac{n-1}{n} S^2}{\bar{X}} \Rightarrow \hat{K}_{MME} = \frac{\bar{X}}{\hat{\theta}_{MME}} = \frac{\bar{X}^2}{\frac{n-1}{n} S^2}$$

Ex: $X_1, \dots, X_n \sim N(1, \sigma^2)$.

MME: $1 = E[X] = \bar{X}$, no information about σ,

$$\sigma^2 + 1 = E[X^2] = m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\Rightarrow \hat{\sigma}^2_{MME} = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 1).$$ (note, not strictly positive, result is unbiased but can be meaningless)

11/7/23:

$\sigma^2 = \text{var}(X) = E[X^2] - E[X]^2$

$$E[\sigma^2_{MME}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i^2] - 1 = \frac{1}{n}\sum_{i=1}^{n}(\sigma^2 + 1) - 1 = \sigma^2$$

$\Rightarrow$ unbiased.

Let $X_1, \dots, X_n$ be iid with pdf $f(x;\theta)$. Suppose $\hat{\theta}$ is an estimator of θ. $\hat{\theta}$ is an unbiased estimator for θ if $E[\hat{\theta}] = \theta$. Otherwise, the bias of $\hat{\theta}$ is $\text{Bias}(\hat{\theta}) = B(\hat{\theta}) = E[\hat{\theta}] - \theta$. A biased estimator may have smaller variance leading to "better" estimates than an unbiased one.

Ex: Let $X_1, \dots, X_n$ be iid Exponential(θ) (first moment is θ). So $f(x) = \frac{1}{\theta} e^{-x/\theta} \cdot I[x \geq 0]$ for $\theta > 0$. We want to estimate θ. Since $E[X_i] = \theta$, we could generate unbiased estimators X_j, $j \in \{1, \dots, n\}$, or $\bar{X}$, or $2X_1 - X_2$. Also, recall $f_{X_{(1)}}(x) = n \cdot f(x) \cdot (1 - F(x))^{n-1} = \frac{n}{\theta} e^{-x/(\theta/n)}$, so $X_{(1)} \sim \text{Exp}\left(\frac{\theta}{n}\right)$, so another unbiased estimator is $n \cdot X_{(1)}$. Which has least variance?

1) $\text{var}(X_j) = \theta^2$
2) $\text{var}(\bar{X}) = \text{var}(X_i)/n = \frac{\theta^2}{n}$,
3) $\text{var}(2X_1 - X_2) = 2^2 \cdot \text{var}(X_1) + 1^2 \cdot \text{var}(X_2) = 5\theta^2$
4) $\text{var}(n \cdot X_{(1)}) = n^2 \, \text{var}\left(\text{Exp}\left(\frac{\theta}{n}\right)\right) = n^2 \cdot \frac{\theta^2}{n^2} = \theta^2$

So out of these, $\bar{X}$ has the smallest variance, followed by X_j & $n \cdot X_{(1)}$, and lastly $2X_1 - X_2$.

Suppose again that $\hat{\theta}$ is an estimator for θ. The mean squared error (MSE) of $\hat{\theta}$ is given by

$$\text{MSE}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] = E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2$$

$$= E[\hat{\theta}^2] - E[\hat{\theta}]^2 + \underbrace{E[\hat{\theta}]^2 - 2\theta E[\hat{\theta}] + \theta^2}_{(E[\hat{\theta}] - \theta)^2}$$

$$= \text{var}(\hat{\theta}) + \text{Bias}(\hat{\theta})^2$$

which is a way to evaluate variance versus bias.

Ex: Let $Z_1, \dots, Z_n$ be iid $N(\mu, \sigma^2)$ with μ, σ^2 unknown. Recall that $\hat{\sigma}^2_{MLE} = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$ is unbiased for σ^2. But, we also found $\hat{\sigma}^2_{MME} = \frac{1}{n}\sum_{i=1}^{n}(Z_i-\bar{Z})^2 = \frac{n-1}{n}S^2$. So, we have that $Bias(\hat{\sigma}^2_{MME}) = \frac{n-1}{n}\sigma^2 - \sigma^2 = \frac{-\sigma^2}{n}$ while $Bias(\hat{\sigma}_{MLE}) = 0$. But, can show that $MSE(\hat{\sigma}_{MME}) \leq MSE(\hat{\sigma}_{MLE})$ under certain σ^2.

Note, all of above also applies to estimators $\hat{\tau}$ of some function of θ, $\tau(\theta)$.

We say $\hat{\theta}_n = T(X_1, \dots, X_n)$ is <u>exactly unbiased</u> for θ if $E[\hat{\theta}_n] = \theta$. And we say it is <u>asymptotically unbiased</u> for θ if $\lim_{n\to\infty} E[\hat{\theta}_n] = \theta$.

<u>11/9/23:</u>

Let $(T_i)_{i=1}^{\infty}$ be a sequence of estimates for $\tau(\theta)$. We say they are <u>mean squared error consistent</u> for $\tau(\theta)$ if

$$\lim_{n\to\infty} E[(T_n - \tau(\theta))^2] = \lim_{n\to\infty} B(T_n)^2 + \lim_{n\to\infty} var(T_n)$$

$$= 0.$$

Note that MSE consistency implies asymptotic nonbias and consistency.

Let $X_1, \dots, X_n$ be iid with common pdf $f(x; \theta)$. An estimator T^* of $\tau(\theta)$ is a <u>uniformly minimum variance unbiased estimator</u> if $Bias(T^*) = E[T^*] - \tau(\theta) = 0$ and $var(T^*) \leq var(T)$ for any unbiased estimator, T, for $\tau(\theta)$ for every value of θ.

The <u>Cramer-Rao Lower Bound (CRLB)</u> is the smallest possible variance for an unbiased estimator.

<u>CRLB Theorem:</u> Let $X_1, \dots, X_n$ be iid $f(x;\theta)$, and assume T is an unbiased estimator of $\tau(\theta)$. Additionally, suppose that

$\text{var}(T(\underline{X}))$ is finite. Then,

$$\text{var}(T) \geq \frac{(\tau'(\theta))^2}{n \cdot E\left[\left(\frac{\partial}{\partial \theta} \ln(f(x;\theta))\right)^2\right]} > 0$$

where $\frac{\partial}{\partial \theta} \ln(f(x;\theta))$ is called the <u>score function</u>.

<u>Proof:</u> Recall by Cauchy-Schwartz that $\text{Cov}(X,Y)^2 \leq \text{Var}(X)\,\text{Var}(Y)$. So rearranging, and replacing X with $T(\underline{X})$ and Y with the vectorized score function, $s(\underline{x};\theta) = \frac{\partial}{\partial \theta} \log(f_{\underline{x}}(\underline{x};\theta))$,

$$\text{Var}(T(\underline{X})) \geq \frac{\text{Cov}(T(\underline{X}), s(\underline{x};\theta))^2}{\text{var}(s(\underline{x};\theta))}.$$

For the covariance term,

$$\text{cov}(T(\underline{X}), s(\underline{x};\theta)) = E[T(\underline{X}) \cdot s(\underline{x};\theta)] - E[T(\underline{X})]E[s(\underline{x};\theta)].$$

Well, $E[s(\underline{x};\theta)] = E\left[\frac{\partial}{\partial \theta} \ln(f_{\underline{x}}(\underline{x};\theta))\right] = E\left[\frac{\frac{\partial}{\partial \theta} f_{\underline{x}}(\underline{x};\theta)}{f_{\underline{x}}(\underline{x};\theta)}\right]$

$$= \int_{\text{Supp}(\underline{X})} \frac{\partial}{\partial \theta} f_{\underline{x}}(\underline{x};\theta)\, d\underline{x} = \frac{d}{d\theta} \int_{\text{Supp}(\underline{X})} f_{\underline{x}}(\underline{x};\theta)\, d\underline{x} = 0.$$

So the covariance term simply equals $E[T(\underline{X}) \cdot s(\underline{x};\theta)]$. Next,

$$E[T(\underline{X}) \cdot s(\underline{x};\theta)] = \int_{\text{Supp}(\underline{X})} \frac{\partial}{\partial \theta}\left[T(\underline{x}) \cdot f_{\underline{x}}(\underline{x};\theta)\right] d\underline{x}$$

$$= \frac{d}{d\theta} \int_{\text{Supp}(\underline{X})} T(\underline{x}) \cdot f_{\underline{x}}(\underline{x};\theta)\, d\underline{x} = \frac{d}{d\theta} E[T(\underline{X})] = \tau'(\theta).$$

Finally,

$$\text{var}(s(\underline{x};\theta)) = E[s(\underline{x};\theta)^2] - E[s(\underline{x};\theta)]^2$$

$$= E[s(\underline{x};\theta)^2] - 0.$$

Hence,

$$\text{var}(T(\underline{X})) \geq \frac{(\tau'(\theta))^2}{E[s(\underline{x};\theta)^2]}.$$

Finally, if we have that $X_1, \ldots, X_n$ are iid,

$$E[s(\underline{x};\theta)^2] = E\left[\left(\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\log(f(x_i;\theta))\right)^2\right] = \sum_{i=1}^{n} E[s(x_i;\theta)]^2$$

$$+\sum_{i=1}^{n}\sum_{j\neq i} E[s(x_i;\theta)s(x_j;\theta)] \quad n\cdot E[s(x;\theta)^2], \quad \text{qed.}$$

11/14/23:

Let $X_1,\dots,X_n$ be iid Exponential (θ). The CRLB for θ is

$$\frac{1}{n}\cdot E\left[\left(\frac{\partial}{\partial\theta}\ln\left(\frac{1}{\theta}e^{-x/\theta}\cdot I[x>0]\right)\right)^2\right]^{-1}$$

$$= \frac{1}{n}\cdot E\left[\left(\frac{-1}{\theta}+\frac{x}{\theta^2}\right)^2\right]^{-1}$$

$$= \frac{1}{n}\cdot E\left[\frac{x^2}{\theta^4}-\frac{2x}{\theta^3}+\frac{1}{\theta^2}\right]^{-1}$$

$$= \frac{1}{n}\left[\frac{2\theta^2}{\theta^4}-\frac{2\theta}{\theta^3}+\frac{1}{\theta^2}\right]^{-1}$$

$$= \frac{\theta^2}{n}.$$

And, we know that $var(\bar{X}) = var(X_i)/n = \theta^2/n$. So, $\bar{X}$ is a UMVUE of θ (uniformly minimum variance unbiased estimator), as it attains the CRLB.

Ex: $X_1,\dots,X_n$ iid $N(\mu,1)$. We have that

$$f(x;\mu) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2} \Rightarrow \ln(f(x;\mu)) = \frac{-1}{2}\ln(2\pi) - \frac{1}{2}(x-\mu)^2$$

$$\Rightarrow \left(\frac{\partial}{\partial\mu}\ln(f(x;\mu))\right)^2 = (x-\mu)^2 = x^2 - 2x\mu + \mu^2$$

$$\Rightarrow E\left[\left(\frac{\partial}{\partial\mu}\ln(f(x;\mu))\right)^2\right] = E[x^2] - 2E[X]\mu + \mu^2$$

$$= 1+\mu^2 - 2\mu^2 + \mu^2 = 1$$

$$\Rightarrow CRLB = \frac{1}{n}.$$

And, we have that $var(\bar{X}) = \frac{1}{n}$, making $\bar{X}$ a UMVUE for μ.

Ex: $N(\mu, \sigma^2)$, both unknown. The CRLB of σ^2 can be found:

$$\ln(f(x;\mu,\sigma^2)) = \frac{-1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2$$

$$\Rightarrow \left(\frac{\partial}{\partial\sigma^2}\ln(f(x;\mu,\sigma^2))\right)^2 = \left[\frac{-1}{2\sigma^2} + \frac{1}{2\sigma^4}(x-\mu)^2\right]^2$$

$$= \frac{1}{4\sigma^4} - \frac{1}{2\sigma^6}(x-\mu)^2 + \frac{1}{4\sigma^8}(x-\mu)^4$$

$$\Rightarrow E\left[\left(\frac{\partial}{\partial\sigma^2}\ln(f(x;\sigma^2))\right)^2\right] = \frac{1}{4\sigma^4} - \frac{2\sigma^2}{4\sigma^6} + \frac{\overbrace{3\sigma^4}^{3^{rd}\text{ central moment}}}{4\sigma^8} = \frac{1}{2\sigma^4}.$$

$$\Rightarrow CRLB(\sigma^2) = \frac{2\sigma^4}{n}.$$

Is $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ a UMVUE for σ^2 here? We can show that $var(S^2) = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n} = CRLB$, so it is not. However, if μ was known, could show that $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ is a UMVUE for σ.

The quantity

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

is called the <u>Fisher information number</u>.

Also, if $\log(f(x;\theta))$ is twice differentiable, and Leibnitz applies, then

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

$$= -E\left[\frac{\frac{\partial^2 f}{\partial\theta^2}(x;\theta)}{f(x;\theta)}\right] + E\left[\left(\frac{\frac{\partial f}{\partial\theta}(x;\theta)}{f(x;\theta)}\right)^2\right]$$

$$= -\frac{d^2}{d\theta^2}\int_{supp(x)} f(x;\theta)\,dx + E\left[\left(\frac{\partial}{\partial\theta}\ln(f(x;\theta))\right)^2\right]$$

$$= 0 + I(\theta) = I(\theta).$$

Now, consider an estimator $T(\underline{X})$ satisfying

$$a(\theta)\left[T(\underline{x})-\tau(\theta)\right] = \frac{\partial}{\partial\theta}\ln(f(x;\theta))$$

for some function $a(\theta)$.

11/16/23:

Why consider this? Well, then

$$T(\underline{X}) = \frac{1}{a(\theta)}\frac{\partial}{\partial\theta}\log(f(x;\theta)) + \tau(\theta)$$

so $T(\underline{X})$ and the score function are perfectly coorelated, so by Cauchy-Schwartz, the CRLB is perfectly held.

11/24/23: Let $X_1,\dots,X_n$ be iid from pdf $f(x;\underline{\theta})$, $\underline{\theta}$ unknown. Loosely speaking, a statistic $S = h(\underline{X})$ is <u>sufficient</u> for $\underline{\theta}$ if it contains as much info about θ as the entire sample does.

Let $X_1,\dots,X_n$ have joint pdf $f(\underline{x};\underline{\theta})$, and let $S=h(\underline{x})$ be a statistic. Then, S is a <u>sufficient statistic</u> for $\underline{\theta}$ if the conditional distribution of $\underline{X}$ given S does not depend on $\underline{\theta}$. Notationally, $f_{\underline{x}|s}(\underline{x},s;\theta)$ is not dependent on $\underline{\theta}$.

<u>Ex</u>: $X_1,\dots,X_n$ iid Exponential(θ).

Recall, $$f_{\underline{x}}(\underline{x};\theta) = \theta^{-n}e^{-\sum_{i=1}^{n}x_i/\theta}\cdot I[x_1,\dots,x_n>0].$$

Now, recall, $$f_{\underline{x}|s}(\underline{x},s;\theta) = \frac{\text{joint}}{\text{marginal}} = \frac{f_{\underline{x},s}(\underline{x},s;\theta)}{f_s(s;\theta)}.$$

Well, from $\sum_{i=1}^{n}x_i = n\bar{x}$, the joint is given by

$$f_{\underline{x},s}(\underline{x},s;\theta) = \theta^{-n}e^{-ns/\theta}\cdot I\left[x_1,\dots,x_n \geq 0, \sum_{i=1}^{n}x_i = ns\right]$$

For the marginal of $S = \overline{X}$, we use that $\sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \theta)$, so $nS \sim \text{Gamma}(n, \theta)$. So, we have that

$$f_S(s;\theta) = f_{\text{Gamma}(n,\theta)}(ns) = \frac{(ns)^{n-1} e^{-ns/\theta}}{\Gamma(n)\,\theta^n} \cdot I[ns \geq 0].$$

Dividing,

$$f_{\underline{X}|S}(\underline{x}, s; \theta) = \frac{\theta^{-n} e^{-ns/\theta} \cdot I[x_1, \dots, x_n \geq 0, \ \sum_{i=1}^{n} x_i = ns]}{(ns)^{n-1} e^{-ns/\theta} / (\Gamma(n)\,\theta^n) \cdot I[ns \geq 0]}$$

$$= \frac{\Gamma(n)}{(ns)^{n-1}} \cdot I[x_1, \dots, x_n \geq 0, \ \sum_{i=1}^{n} x_i = ns]$$

which does not depend on θ. Hence, $\overline{X}$ is a sufficient statistic for θ.

Ex: $X_1, \dots, X_n$ iid Bernoulli(p).

Note, the pmf is $\quad p(x_i; p) = p^{x_i}(1-p)^{1-x_i}, \quad x_i \in \{0, 1\}$.

We know that $\sum_{i=1}^{n} X_i \sim \text{Bernoulli}(n, p)$, so choosing $S = \overline{X}$, we have $nS \sim \text{Bernoulli}(n, p)$. So the marginal is

$$p_n(s; p) = p_{\text{Bernoulli}(n,p)}(ns) = \binom{n}{ns} p^{ns}(1-p)^{n-ns}, \quad ns \in \{0, 1, \dots, n\}.$$

Now, for the joint density, we first calculate

$$p_{\underline{X}}(\underline{x}; p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}, \quad \text{each } x_i \text{ 0 or 1}$$

$$= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \quad \text{———— '' ————}$$

So, $\quad p_{\underline{X},S}(\underline{x}, s; p) = p^{ns}(1-p)^{n-ns} \cdot I[x_1, \dots, x_n \in \{0,1\}, \ \sum_{i=1}^{n} x_i = ns]$

Finally, dividing,

$$p_{\underline{X}|S}(\underline{x}, s; p) = \frac{p_{\underline{X},S}}{p_S} = \binom{n}{ns}^{-1} \cdot I[x_1, \dots, x_n \in \{0,1\}, \ \sum_{i=1}^{n} x_i = ns]$$

which does not depend on p. So, $\overline{X}$ is sufficient for p.

Note that sufficient statistics need not be unbiased. In the previous example, $T = \sum_{i=1}^{n} X_i$ is also sufficient for p, as it works out that $p_{\underline{X}|T}(\underline{x}, t ; p)$ does not depend on T.

<u>Fact:</u> Any invertible function of a sufficient statistic is a sufficient statistic.

<u>11/28/23:</u>

Let $\underline{X}$ have joint pdf $f_{\underline{X}}(\underline{x} ; \underline{\theta})$, and let $\underline{S} = \underline{h}(\underline{X})$ be statistics. The elements of $\underline{S}$ are <u>jointly sufficient statistics</u> for $\underline{\theta}$ if $f_{\underline{X}|\underline{S}}$ does not depend on $\underline{\theta}$.

<u>Ex:</u> Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$. Can show that $\overline{X}$ and S^2 are sufficient statistics for μ and σ^2. Do this in same way as before.

Note, the set of statistics $\underline{S}(\underline{X}) = [X_1, \ldots, X_n]^T$ ~~is~~ are sufficient, but not minimal. A set of statistics can be <u>minimally sufficient</u>. In general, the number of unknowns equals the number of statistics required, but not always. <u>Ex:</u> For $Unif(\theta, \theta + 1)$, you need both $X_{(1)}$ and $X_{(n)}$.

The <u>factorization criteria</u> says that $\underline{S}(\underline{X})$ is jointly sufficient for $\underline{\theta}$ iff we can factorize

$$f_{\underline{X}, \underline{S}}(\underline{x}, \underline{s} ; \underline{\theta}) = g(\underline{s} ; \underline{\theta}) \cdot h(\underline{x}).$$

<u>Ex:</u> $X_1, \ldots, X_n$ iid $Ber(p)$. We can write

$$f_{\underline{X}}(\underline{x} ; p) = p^{\sum_{i=1}^{n} x_i} (1-p)^{1 - \sum_{i=1}^{n} x_i} \cdot I[x_1, \ldots, x_n \in \{0, 1\}]$$

$$= \underbrace{\left(\frac{p}{1-p}\right)^{\sum_{i=1}^{n} x_i} \cdot (1-p)^n}_{h(s;p)} \cdot \underbrace{\text{— '' —}}_{g(\underline{x})}$$

So, we have the statistic being $\sum_{i=1}^{n} X_i$ sufficient for p.

Ex: $X_1, \ldots, X_n$ iid $Exp(\theta)$.

$$f(\underline{x} ; \theta) = \underbrace{\theta^{-n} e^{-\sum_{i=1}^{n} x_i / \theta}}_{g(s;\theta)} \cdot \underbrace{I[x_1, \ldots, x_n \geq 0]}_{h(\underline{x})}$$

So the sufficient statistic is $\sum_{i=1}^{n} X_i$.

Ex: $X_1, \ldots, X_n$ iid Uniform $(0, \theta)$.

$$f(\underline{x} ; \theta) = \theta^{-n} \cdot I[0 \leq x_1, \ldots, x_n \leq \theta]$$

$$= \underbrace{\theta^{-n} \cdot I[0 \leq x_{(1)} \leq x_{(n)} \leq \theta]}_{g(\underline{s};\theta)} \cdot \underbrace{1}_{h(\underline{x})}$$

So the sufficient statistics are $X_{(1)}$ and $X_{(n)}$

From the factorization criteria, and noting that $\hat{\theta}_{MLE}$ is the value of θ when $\partial_\theta f(\underline{x} ; \theta) = 0$, or, $\partial_\theta \log f(\underline{x} ; \theta) = 0$,

$$f(\underline{x} ; \theta) = g(\underline{s} ; \theta) \cdot h(\underline{x})$$

$$\Rightarrow \log(f(\underline{x} ; \theta)) = \log(g(\underline{s} ; \theta)) + \log(h(\underline{x}))$$

$$\Rightarrow \frac{\partial}{\partial \theta} \log(f(\underline{x} ; \theta)) = \frac{\partial}{\partial \theta} \log(g(\underline{s} ; \theta)) + 0$$

$\Rightarrow$ $\hat{\theta}_{MLE}$ is a function of the sufficient statistic $\underline{S}$.

<u>Rao-Blackwell Theorem:</u> Let $X_1, \ldots, X_n$ be jointly $f_{\underline{X}}(\underline{x} ; \underline{\theta})$. Let $\underline{S}$ be a vector of jointly sufficient statistics for $\underline{\theta}$. If T is any unbiased estimator of $\tau(\underline{\theta})$, and letting $T^* = E[T \mid \underline{S}]$, then,

1) T^* is an unbiased estimator of T

2) $Var(T^*) \le Var(T)$

The first statement can be shown by

$$E[T^*] = E[\,E[T \mid S]\,] = \tau(\theta)$$ since T an unbiased estimator, no matter S.

Additionally, we use the law of total variance says that

$$var(T) = var(E[T \mid S]) + \underbrace{E[var(T \mid S)]}_{\ge 0.}$$

$$\ge var(T^*).$$

Note that $E[var(T \mid S)] = 0$ iff $var(T \mid S) = 0$ iff $T = T^*$ as the dependence on S is already accounted for.

11/30/23.

Ex: $X_1, \dots, X_n$ iid $Exp(\theta)$. Suppose we want to estimate $\tau(\theta) = \theta^2$. First, recall that $E[X_i] = \theta$ and $var(X) = \theta^2$. So $\theta^2 = E[X_i^2] - E[X_i]^2 = E[X_i^2] - \theta^2 \Rightarrow E[X_i^2] = 2\theta^2$. So, an unbiased estimator for θ^2 is $\frac{X_i^2}{2}$ for any i. To find a sufficient statistic, we write the joint pdf and factorize

$$f_{\underline{x}}(\underline{x};\theta) = \underbrace{\theta^{-n} e^{-\sum_{i=1}^n x_i/\theta}}_{g(s;\theta)} \cdot \underbrace{I[x_1, \dots, x_n \ge 0]}_{h(\underline{x})}$$

so $S = \sum_{i=1}^n X_i$ is a sufficient statistic

Now, we've shown that $S = \sum_{i=1}^n X_i \sim Gamma(n, \theta)$ since $X_1, \dots, X_n$ iid $Exp(\theta)$. To prove S is sufficient,

$$f_{\underline{x},s}(\underline{x}, s;\theta) = \theta^{-n} e^{-s/\theta} \cdot I[x_1, \dots, x_n \ge 0, \sum_{i=1}^n x_i = s],$$

$$f_s(s;\theta) = \frac{s^{n-1} e^{-s/\theta}}{\Gamma(n)\theta^n} \cdot I[s \ge 0].$$

So dividing,

$$f_{\underline{x}|s}(\underline{x}, s; \Theta) = \frac{\Gamma(n)}{s^{n-1}} \cdot I[x_1, \dots, x_n \geq 0, \sum_{i=1}^{n} x_i = s]$$

which is free of Θ, proving S is sufficient for Θ.
Finally, we wish to construct $T^* = E[\frac{X_1^2}{2} \mid \sum_{i=1}^{n} X_i]$. We have that

$$T^* = \int_{supp(\underline{x})} t \cdot f_{\underline{x}|s} \, d\underline{x}$$

$$= \int_{supp(\underline{x})} \frac{x_1^2}{2} \cdot \frac{\Gamma(n)}{s^{n-1}} \cdot I[x_1, \dots, x_n \geq 0, \sum_{i=1}^{n} x_i = s] \, d\underline{x}$$

$$= \int_0^s \frac{x_1^2}{2} \int_{supp(x_2, \dots, x_n)} \frac{\Gamma(n)}{s^{n-1}} \cdot I[x_2, \dots, x_n \geq 0, \sum_{i=2}^{n} x_i = s - x_1] \, d\underline{x}$$

Well, by $f_{\underline{x}|s}$ being a density function,

$$\int_{supp(\underline{x})} \frac{\Gamma(n)}{s^{n-1}} \cdot I[x_1, \dots, x_n \geq 0, \sum_{i=1}^{n} x_i = s] \, d\underline{x} = 1$$

$$\Rightarrow \int_{supp(\underline{x}_{2:n})} \frac{\Gamma(n-1)}{(s - x_1)^{n-2}} \cdot I[x_2, \dots, x_n \geq 0, \sum_{i=2}^{n} x_i = s - x_1] \, d\underline{x}_{2:n} = 1$$

$$\Rightarrow \int_{supp(\underline{x}_{2:n})} \frac{\Gamma(n)}{s^{n-1}} \cdot I[x_2, \dots, x_n \geq 0, \sum_{i=2}^{n} x_i = s - x_1] \, d\underline{x}_{2:n} = \frac{(s - x_1)^{n-2} \cdot \Gamma(n)}{s^{n-1} \cdot \Gamma(n-1)}$$

$$= \frac{(n-1)(s - x_1)^{n-2}}{s^{n-1}}$$

$$\Rightarrow T^* = \int_0^s \frac{(n-1)(s - x_1)^{n-2} x_1^2}{2 s^{n-1}} \, dx_1$$

$u = x_1^2/2 \quad dv = (n-1)(s - x_1)^{n-2} dx_1$
$du = x_1 dx_1 \quad v = -(s - x_1)^{n-1}$

$$= \frac{1}{s^{n-1}} \int_0^s (s - x_1)^{n-1} x_1 \, dx_1$$

$u = x_1 \quad dv = (s - x_1)^{n-1} dx_1$
$du = dx_1 \quad v = \frac{-(s - x_1)^n}{n}$

$$= \frac{1}{s^{n-1}} \int_0^s \frac{(s - x_1)^n}{n} \, dx_1$$

$$= \frac{-1}{s^{n-1} n(n+1)} \left[(s - x_1)^{n+1} \right]_0^s$$

$$= \frac{s^2}{n(n+1)}.$$

Let $X_1, \dots, X_n$ be iid $f_{\underline{x}}(\underline{x}; \underline{\theta})$. The family $f_{\underline{x}}(\underline{x}; \underline{\theta})$ is called <u>complete</u> if

$$\int_{\mathrm{supp}(\underline{X})} u(\underline{x}) \cdot f_{\underline{x}}(\underline{x}; \underline{\theta})\, d\underline{x} = 0 \quad \forall\ \underline{\theta}$$

implies that $u(\underline{X}) \equiv 0$. I.e., there does not exist a nontrivial statistic $U(\underline{X})$ s.t. $E[U(\underline{X})] = 0 \ \forall\ \underline{\theta}$.

Suppose that $T_1(\underline{X})$ and $T_2(\underline{X})$ are both unbiased estimators of $\tau(\theta)$. Then $U(\underline{X}) := T_1(\underline{X}) - T(\underline{X}_2)$ will have an expectation of 0 for all $\underline{\theta}$. If the family is complete, then $U \equiv 0 \Rightarrow T_1 \equiv T_2$. I.e., a family being complete implies that there exists only one unbiased estimator.

<u>Lehmann-Scheffe Theorem:</u> Suppose we have an unbiased estimator $\underline{T}(\underline{X})$ of $\underline{\tau}(\underline{\theta})$, and we have a sufficient statistic $\underline{S}$ for $\underline{\theta}$, and that $f_{\underline{x}}(\underline{x}; \underline{\theta})$ is complete. Then, $\underline{T}^* = E[\underline{T} | \underline{S}]$ is the unique unbiased estimator of $\underline{\tau}(\underline{\theta})$ that is also a function of the sufficient statistic $\underline{S}$. Hence $\underline{T}^*$ is automatically an UMVUE for $\underline{\tau}(\underline{\theta})$.

<u>12/5/23:</u>

A single parameter family comes from the <u>exponential class</u> if its pdf (or pmf) can be expressed as

$$f(x; \theta) = h(x) g(\theta) e^{\eta(\theta) T(x)}$$

where its support does not depend on θ.

<u>Ex:</u> Bernoulli: $f(x; p) = p^x (1-p)^{1-x} \cdot I[x \in \{0, 1\}]$

$$= \left(\frac{p}{1-p}\right)^x (1-p) \cdot I[x \in \{0, 1\}]$$

$$= (1-p)\exp\left(x\log\left(\frac{p}{1-p}\right)\right)\cdot I[x\in\{0,1\}].$$

Similarly, for a vector of parameters $\underline{\theta}$, we say it belongs to the exponential family if it can be expressed as

$$f(x;\underline{\theta}) = h(x)g(\underline{\theta})e^{\underline{\eta}(\theta)\cdot\underline{T}(x)}.$$

<u>Ex:</u> $N(\mu,\sigma^2)$ $\quad f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

$$= (2\pi\sigma^2)^{-1/2}e^{\frac{-\mu^2}{2\sigma^2}}e^{x^2\cdot\frac{-1}{2\sigma^2}+x\cdot\frac{\mu}{\sigma^2}}.$$

By the factorization criterion, given $f(x;\underline{\theta}) = h(x)g(\underline{\theta})\exp(\underline{\eta}(\underline{\theta})\cdot\underline{T}(\underline{x}))$, we have that $\underline{T}(X)$ is a vector of sufficient statistics. We can then find a statistic that is unbiased for $\underline{\theta}$ and a function of $\underline{T}(X)$. Finally, by completeness of exponential class & Lehmann-Scheffé, that statistic is automatically UMVUE for $\underline{\theta}$ (or $\tau(\underline{\theta})$).

<u>Ex:</u> Suppose $X_1,\dots,X_n$ are iid Bernoulli(p) with p unknown. Find a UMVUE of $\tau(p) = p(1-p)$.

From previous work, since $f(x;p) = (1-p)e^{x\log(\frac{p}{1-p})}\cdot I[x\in\{0,1\}]$, we have that

$$f_{\underline{X}}(\underline{x};p) = (1-p)^n e^{\sum_{i=1}^n x_i\cdot\log(\frac{p}{1-p})}\cdot I[\underline{x}\in\{0,1\}^n].$$

Hence, this is a part of the exponential class, and the sufficient statistic is $S = \sum_{i=1}^n X_i$. We recognize that $S\sim$ Binomial(n,p) so $E[S] = np$, $\mathrm{var}(S) = np(1-p)$. Hence,

$$np(1-p) = \mathrm{var}(S) = E[S^2] - E[S]^2 = E[S^2] - (np)^2$$

$\Rightarrow E[S^2] = np - np^2 + n^2p^2$. From this,

$$p - p^2 = \tau(p) = E[AS + BS^2]$$

$$= Anp + B(np - np^2 + n^2p^2)$$

$$\Rightarrow (1 - An - Bn)p + (-1 + Bn - Bn^2)p^2 = 0$$

$$\Rightarrow \begin{cases} -1 + Bn - Bn^2 = 0 \Rightarrow B = \frac{-1}{n(n-1)} \\ 1 - An - Bn = 0 \Rightarrow A = \frac{1}{n} - B = \frac{n-1--1}{n(n-1)} = \frac{1}{n-1} \end{cases}$$

$$\Rightarrow AS + BS^2 = \frac{1}{n-1} \sum_{i=1}^{n} X_i - \frac{1}{n(n-1)} \left(\sum_{i=1}^{n} X_i \right)^2$$

is UMVUE for $\tau(p) = p(1-p)$.