tags: Theory2

NOTE:

field: Definition of Convergence

field: A sequence $\{a_n\}_{n>1}$ of real numbers is said to **converge** to a point $a \in \mathbb{R}$ if for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $m > N$ we have $|a_m - a| < \epsilon$

NOTE:

field: Example of convergence: $a_n = \frac{1}{n}$

field: For any $\epsilon > 0$, choose N such that $\frac{1}{N} < \epsilon$. Then for any $m > N$ we have that

$$a_n = \frac{1}{n} < \frac{1}{N} < \epsilon$$

and therefore $|a_m - 0| = \frac{1}{n} < \epsilon$

NOTE:

field: Given two convergent sequences $\{a_n\}$ and $\{b_m\}$ such that $a_m \to a$ and $b_m \to b$
$\lim_{n\to\infty} a_n b_n =$

field: Given two convergent sequences $\{a_n\}$ and $\{b_m\}$ such that $a_m \to a$ and $b_m \to b$
$\lim_{n\to\infty} a_n b_n = (\lim_{n\to\infty} a_n)(\lim_{n\to\infty} b_n) = ab$

NOTE:

field: Definition: Convergence in probability

field: A sequence of random variables $\{X_n\}_{n\geq 1}$ **converges in probability** to a random variable X, if for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$$

We write $X_n \xrightarrow{p} X$

Equivalently, $x_m \xrightarrow{p} x$ if $\lim_{n\to\infty} P(|x_n - x| < \epsilon) = 1$

NOTE:

field: Convergence in probability example: Let $\{x_n\}$ be a sequence of random variables such that $x_n \sim N(0, 1/m^2)$
Show that $x_n \xrightarrow{p} 0$:

field: Let $\epsilon > 0$. We obtain $P(|x_n - 0|) = P(x_n > \epsilon) + P(X_n < -\epsilon)$. ie we are looking at the tail probabilities.

Now,

$$\begin{aligned} P(X_n < -\epsilon) + P(x_n > \epsilon) &= P(nx_n < n\epsilon) + P(nx_n > n\epsilon) \\ &= \Phi(n\epsilon) + 1 - \Phi(n\epsilon) \\ &= 2\Phi(-n\epsilon) \underset{n\to\infty}{\rightarrow} 0 \end{aligned}$$

Therefore $x_n \xrightarrow{p} 0$

NOTE:

field: Example convergence in probability Let $W \sim N(0,1)$ and $U \sim Unif(0,1)$, and define the sequence $\{x_n\}_{n\geq 1}$ as $x_n = W$with prob $1 - 1/n, U$ with prob $1/n$

Show that $x_n \xrightarrow{p} W$

field: Let $\epsilon > 0$ Then.

$$\begin{aligned} P(|X_n - W| > \epsilon) &= P(|X_n - W| > \epsilon | X_n = W)P(X_n = W) \\ &+ P(|X_n - W| > \epsilon | X_n = U)P(X_n = U) \\ &= 0 \cdot (1 - 1/n) + p_n(1/n) \end{aligned}$$

Where p_n is a probability, and therefore $0 \leq p_n \leq 1$

It follows that $p_n \frac{1}{n} \underset{n\to\infty}{\to} 0$, and therefore $P(|X_n - W| > \epsilon) \underset{n\to\infty}{\to} 0$, for all $\epsilon > 0$, so that $X_n \xrightarrow{p} W$.

NOTE:

field: Does $X_n \xrightarrow{p} c$ imply $E(X_n) \to c$?

field: Let $X_n = 0$ with probability $1 - 1/n$, n^2 with probability $1/n$ Then $P(|X_n - 0| > \epsilon) \leq P(X_n = n^2) = 1/n \underset{n\to\infty}{\to} 0$ On the other hand, $E(X_n) = 0 \cdot P(X_n = 0) + n^2 P(X_n = n^2) = 0 + n^2 \frac{1}{n} = n \underset{n\to\infty}{\to} \infty$. Therefore $X_n \xrightarrow{p} c$ does not imply $E(X_n) \to c$

NOTE:

field: Does $E(X_n) \to c$ imply $X_n \xrightarrow{p} c$?

field: Let $X_n = 0$, with prob $1 - 1/n$, n with prob $1/n$. Then $E(X_n) = 0 \cdot P(X_n = 0) + nP(X_n = n) = 0 + n1/n = 1$ for all n. But $P(|X_n - 0| > \epsilon) \leq P(X_n = n) = \frac{1}{n} \underset{n\to\infty}{\to} 0$ It follows, $X_n \xrightarrow{p} 0$, and therefore we have $E(X_n) \to c$does not imply $X_n \xrightarrow{p} c$

NOTE:

field: Suppose $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ be two sequences of random variables such that $X_n \xrightarrow{p} x_0$ and $Y_n \xrightarrow{p} y_0$ as $n \to \infty$, where $x_o, y_0 \in \mathbb{R}$

What properties do we have?

field:

- $X_n \pm Y_m \xrightarrow{p} x_0 \pm y_0$ as n increases to ∞
- $X_n Y_n \xrightarrow{p} x_0 y_0$ as n increases to ∞
- $X_n/Y_n \xrightarrow{p} x_0/y_0$ as n increases to infinity, provided that $P(Y_n = 0) = 0$ fro all n and $y_0 \neq 0$

NOTE:

field: Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that $x_n \xrightarrow{p} x_0 \in \mathbb{R}$, as $n \to \infty$, and let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function . Then

$$g(X_n) \xrightarrow{p} \text{ as } n \to \infty$$

field: Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that $x_n \xrightarrow{p} x_0 \in \mathbb{R}$, as $n \to \infty$, and let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function . Then

$$g(X_n) \xrightarrow{p} g(x_0) \text{ as } n \to \infty$$

NOTE:

field: Proof of: Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that $x_n \xrightarrow{p} x_0 \in \mathbb{R}$, as $n \to \infty$, and let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function . Then

$$g(X_n) \xrightarrow{p} g(x_0) \text{ as } n \to \infty$$

field: Since g is continuous at $X = x_0$, we have that for any $\epsilon > 0$, there exits $\delta > 0$ such that $|g(x) - g(x_0)| > \epsilon$ implies $|x - x_0| > \delta$

We obtain

$$0 \leq P(|g(X_n) - g(x_0)| > \epsilon) \leq P(|X_n - x_0| > \delta) \underset{n\to\infty}{\to} 0$$

NOTE:

field: Weak Law of Large numbers

field: Let $X_1, X_2, X_3 \ldots$ Be a sequence of iid random variables with $E(X_1) = \mu$ (finite) and $V(X_1) = \sigma^2 < \infty$, and define $\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (the sample mean).

Then

$$\bar{X_n} \xrightarrow{p} \mu \text{ as } n \to \infty$$

NOTE:

field: Proof of Weak Law of Large Numbers

field:

$$\begin{aligned} P(|\bar{X_n} - \mu| > \epsilon) &= P((\bar{X_n} - \mu)^2 > \epsilon^2) \\ &\leq \frac{E((\bar{X_n} - \mu)^2)}{\epsilon^2} \text{ by Chebyshev's Inequality} \\ &= \frac{V(\bar{X_n})}{\epsilon^2} \text{ by def of variance} \\ &= \frac{\sigma^2}{n\epsilon^2} \underset{n\to\infty}{\to} 0 \end{aligned}$$

Therefore $\bar{X_n} \xrightarrow{p} \mu$

NOTE:

field: Consistency

field: If our estimate converges in probability to the value of the parameter of interest as the sample size n increases

NOTE:

field: Consistency of S^2

field: Suppose $X_1, X_2, \ldots$ is a sequence of iid random varialbes with $E(X_1) = \mu$ finite and $V(X_1) = \sigma^2 < \infty$

and define

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X_n})^2 \quad \text{The sample variance}$$

Can we show that S_n^2 is a consistent estimate of σ^2? In other words, can we show taht $S_n^2 \xrightarrow{p} \sigma^2$ as $n \to \infty$

Using Chebychev's inequality, we obtain

$$\begin{aligned} P(|S_n^2 - \sigma^2| > \epsilon) &\leq \frac{E[(S_n^2 - \sigma^2)^2]}{\epsilon^2} \\ &= \frac{V(S_n^2)}{\epsilon^2} \end{aligned}$$

There fore, a sufficient condition that S_n^2 converges in probablility to σ^2 is that the variance of S_n^2 $V(S_n^2) \to 0$, as $n \to \infty$

NOTE:

field: $V(S_n^2) \to 0$ as long as

field: $V(S_n^2) \to 0$ as long as the fourth central moment $\mu_4 = E[(X_1 - \mu)^4]$ is finite.

NOTE:

field: Khinchin's WLLN

field: Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $E(X_1) = \mu$ (finite). Then, $\bar{X_n} \xrightarrow{p} \mu$ as $n \to \infty$

NOTE:

field: Let $X_1, X_2 \ldots$ be a sequence of random variables, such that for some $r > 0$ and $c \in \mathbb{R}$, $E[|X_n - c|^r] \underset{n\to\infty}{\to} 0$. Then $X_n \xrightarrow{p}$, as $n \to \infty$

field: (A general result to establish convergence in probability)

Let $X_1, X_2 \ldots$ be a sequence of random variables, such that for some $r > 0$ and $c \in \mathbb{R}$, $E[|X_n - c|^r] \underset{n\to\infty}{\to} 0$. Then $X_n \xrightarrow{p} c$, as $n \to \infty$

NOTE:

field: Consistent estimator for $X_1, X_2, \ldots X_n \sim$ iid Univorm$(0, \theta)$, $\theta > 0$. (and sketch of proof)

field: $X_{(n)} = \max(X_1, \ldots X_n)$ (the largest order statistic)

Proof

First recall that the pdf of $X_{(n)}$ is given by

$$f(x) = nx^{n-1}\theta^{-n}, 0 < x < \theta, 0 \text{otherwise}$$

We obtain

$$\begin{aligned} E(X_{(n)}) &= \int_0^\theta x f(x) dx \\ &= n\theta^{-n} \int_0^\theta x^n dx \\ &= \frac{n}{n-1}\theta \\ E(X_{(n)}^2) &= \int_0^\theta x^2 f(x) dx \\ &= n\theta^{-n} \int_0^\theta x^{n+1} dx \\ &= \frac{n}{n+2}\theta^2 \end{aligned}$$

We have

$$
\begin{aligned}
E[(X_{(n)} - \theta)^2] &= E(X_{(n)}^2) - 2\theta E(X_{(n)}) + \theta^2 \\
&= \frac{n}{n+2}\theta^2 - 2\theta \frac{n}{n+1}\theta + \theta^2 \\
& \dots \\
&= \frac{2\theta^2}{(n+1)(n+2)} \underset{n\to\infty}{\to} 0
\end{aligned}
$$

Hence, taking $c = 0$ and $r = 2$, from the previous theorem, we obtain $X_{(n)} \overset{p}{\to} \theta$ as $n \to \infty$

NOTE:

field: Definition Almost Sure Convergence

field: A sequence $\{X_n\}_{n\geq 1}$ of random variables is said to converge **Almost Surely** to a random variable X if for every $\epsilon > 0$,

$$P(\lim_{n\to\infty} |X_n - X| > \epsilon) = 0$$

We write $X_n \overset{a.s}{\to} X$ as $n \to \infty$

NOTE:

field: Strong Law of Large Numbers

field: Let $X_1, X_2, \dots$ be an iid sequence of random variables, with $E(X_1) = \mu$ (finite) and $V(X_1) = \sigma^2 < \infty$. Then,

$$\bar{X_n} \overset{a.s}{\to} \mu \quad \text{as } \mu \to \infty$$

NOTE:

field: Does convergence in probability imply convergence almost surely?

field: No. Let $\Omega = [0.1]$, with uniform probability distribution. Define the sequence $\{X_n\}_{n\geq 1}$ as:

$$\begin{aligned}
X_1(\omega) &= \omega + \mathbb{I}_{[0,1]}(\omega) \\
X_2(\omega) &= \omega + \mathbb{I}_{0,1/2}(\omega) \\
X_3(\omega) &= \omega + \mathbb{I}_{1/2,1}(\omega) \\
X_4(\omega) &= \omega + \mathbb{I}_{0,1/3}(\omega) \\
X_5(\omega) &= \omega + \mathbb{I}_{1/3,2/3}(\omega) \\
&\vdots
\end{aligned}$$

$X_5(\omega) = \omega + 1$

Let $X(\omega) = \omega$, then it is easy to show that $X_n \xrightarrow{p} X$ because $P(|X_n - X| \geq \epsilon) = P([a_n, b_n])$, where $l_n = \text{length}([a_n, b_n]) \underset{n\to\infty}{\rightarrow} 0$.

However X_n does not converge to X almost surely, because for every $\omega \in [0,1]$, alternates between ω and $\omega + 1$, infinetly often as $n \to \infty$

NOTE:

field: Convergence in Distribution

field: A sequence $\{X_n\}_{n\geq 1}$ of random variables converges in distribution to a random variable X if,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous

We write $X_n \xrightarrow{d} X$

NOTE:

field: Example of convergence in distribution

Let $X_n \sim N(0, \frac{n+1}{n})$, and $X \sim N(0,1)$. We want to show that $X_n \xrightarrow{d} X$.

field:

$$P(X_n \leq X) = P(\sqrt{\frac{n}{n+1}}X_n \leq \sqrt{\frac{n}{n+1}}x)$$
$$= \Phi(\sqrt{\frac{n}{n+1}}x) \underset{n\to\infty}{\to} \Phi(x)$$

And we obtain that $F_{X_n} \to \Phi(x) = F_X(x), \forall x$, and therefore $X_n \xrightarrow{d} X$

NOTE:

field: Does Convergence in probability imply convergence in distribtuion?

field: Yes

NOTE:

field: Does Convergence in distribution imply convergence in probability?

field: No - unless converges in distribution to a constant

NOTE:

field: A sequence $\{X_n\}_{n\geq 1}$ of random variables converges in probability to a constant $c \in \mathbb{R}$ if and only if

field: A sequence $\{X_n\}_{n\geq 1}$ of random variables converges in probability to a constant $c \in \mathbb{R}$ if and only if the sequence converges in distribution to c

NOTE:

field: If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ we have that

1. $X_n \pm Y_n$
2. X_nY_n

field: In general it is not true that if $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ we have that

1. $X_n \pm Y_n \xrightarrow{d} X + Y$
2. $X_n Y_n \xrightarrow{d} XY$

NOTE:

field: Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that $X_n \xrightarrow{d} X$, for some random variable X (possibly a constant). Then for any continuous funciton $g : \mathbb{R} \to \mathbb{R}$, we have $g(X_n) \xrightarrow{d}$

field: Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that $X_n \xrightarrow{d} X$, for some random variable X (possibly a constant). Then for any continuous funciton $g : \mathbb{R} \to \mathbb{R}$, we have $g(X_n) \xrightarrow{d} g(X)$

NOTE:

field: Let $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ be two sequences of random variables such that $X_n \xrightarrow{d} X$ for some random variable X (possibly a constant) and $Y_n \xrightarrow{p} c \in \mathbb{R}$

Then, as $n \to \infty$,

1. $X_n \pm Y_n \xrightarrow{d}$
2. $X_n Y_n \xrightarrow{d}$
3. $X_n / Y_n \xrightarrow{d}$ provided $P(Y_n = 0) = 0 \forall n$ and $c \neq 0$

field: Slutsky's Theorem Let $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ be two sequences of random variables such that $X_n \xrightarrow{d} X$ for some random variable X (possibly a constant) and $Y_n \xrightarrow{p} c \in \mathbb{R}$

Then, as $n \to \infty$,

1. $X_n \pm Y_n \xrightarrow{d} X \pm c$
2. $X_n Y_n \xrightarrow{d} cX$
3. $X_n / Y_n \xrightarrow{d} X/c$ provided $P(Y_n = 0) = 0 \forall n$ and $c \neq 0$

NOTE:

field: Central Limit Theorem

field: Let $X_1, X_2, \ldots$ be an iid sequence of random variables, with $E(X_1) = \mu$(finite) and $V(X_1) = \mu^2 < \infty$

Then, for $\bar{X_n} = \frac{1}{n}\sum_{i=1}^{\infty} X_i$ (the sample mean), we have that

$$\frac{\sqrt{n}(\bar{X_n} - \mu)}{\sigma} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty$$

NOTE:

field: Equivalent results of CLT

field:

- $\frac{(\bar{X_n} - \mu)}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0,1)$
- $\sqrt{n}(\bar{X_n} - \mu) \xrightarrow{d} N(0, \sigma^2)$
- $\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1)$
- $\bar{X_n} \xrightarrow{d} N(\mu, \sigma^2/n)$

NOTE:

field: Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that the mgf $M_{X_n}(t)$ of X_n exists in a neighborhood of 0, for all , and suppose that

$$\lim_{n\to\infty} M_{X_n}(t) = M_X(t) \quad \text{for all } t \text{ in a neighborhood of } 0$$

where $M_X(t)$ is the mgf for some random variable X. Then,

field: Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that the mgf $M_{X_n}(t)$ of X_n exists in a neighborhood of 0, for all , and suppose that

$$\lim_{n\to\infty} M_{X_n}(t) = M_X(t) \quad \text{for all } t \text{ in a neighborhood of } 0$$

where $M_X(t)$ is the mgf for some random variable X. Then, there exists a unique cdf $F_x(x)$ whose moments are determined by $M_y(t)$ and for all x, where $F_x(x)$ is continuous we have $\lim_{n\to\infty} F_{X_n}(x) = F_x(x)$

NOTE:

field: $\frac{\sqrt{n}(\bar{X}-\mu)}{S_n} \xrightarrow{d}$

field: Using the CLT, and slutsky's theorem, we have

$$\frac{\sqrt{n}(\bar{X}-\mu)}{S_n} = \frac{\sigma}{S_n} \cdot \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \xrightarrow{d} N(0,1)$$

NOTE:

field: $g(X) \approx$
$E(g(X)) \approx$, $V(g(X)) \approx$

field:

$$g(X) \approx g(\mu) + g'(X)(X-\mu)$$

Using a first order taylor approximation $E(g(X)) \approx g(\mu)$, $V(g(X)) \approx [g'(\mu)]^2 V(X)$

NOTE:

field: Delta Method

field: Let $\{Y_n\}_{n\geq 1}$ be a sequence of random variables such that $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$ as $n \to \infty$. Suppose that for a given function g and a specific value of θ, $g'(\theta)$ exists and is not equal to zero. Then

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{d} N(0, \sigma^2[g'(\theta)]^2)$$

as $n \to \infty$

NOTE:

field: Second Order delta method

field: Let $\{Y_n\}_{n\geq 1}$ be a sequence of random variables such that $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$ as $n \to \infty$. And that for a given function g as specific value of θ, we hvae $g'(\theta) = 0$, but $g''(\theta)$ Exists and is not equal to 0. Then

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{d} \sigma^2 \frac{g''(\theta)}{2} \chi_1^2 \quad \text{as } n \to \infty$$

NOTE:

field: $\chi_n^2 \dot\sim$ for sufficiently large n

field: $\chi_n^2 \dot\sim N(n, 2n)$

NOTE:

field: Definition Statistic

field: Let $X_1, \ldots, X_n$ be a random sample from a given population. Then, any <u>observable</u> real-valued (or vector-valued) function $T(\mathbf{X}) = T(X_1, \ldots, X_n)$ of the random variables $X_1, \ldots, X_n$ is called a **Statistic**

NOTE:

field: Sampling Distribution

field: The probability distribution of the statitic $T(\mathbf{X})$ is called the **Sampling Distribution** of $T(\mathbf{X})$

NOTE:

field: Sufficient Statistic

field: A statistic $T(\mathbf{X})$ is a **Sufficient Statistic** for θ, if the conditional distribution of the sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ

NOTE:

field: Determine if $T(\mathbf{X}) = \sum X_i$ where $X_i \sim Bern(p)$ is sufficient for p using definition of sufficiency

field:

$$\begin{aligned}
P(\mathbf{X} = \mathbf{x} \big| T = t) &= \frac{P(\cap_{i=1}^n X_i = x_i)}{P(T = t)} \\
&= \prod_{i=1}^n \frac{P(X_i = x_i)}{P(T = t)} \quad \text{by independence} \\
&= \frac{p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}}{\binom{n}{t} p^t (1-p)^{n-t}} \quad \text{Because } T \sim \text{Binom}(n, p) \\
&= \frac{p^t (1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}} \quad \text{because } t = \sum_{i=1}^n x_i \\
&= \frac{1}{\binom{n}{t}} \quad \text{which is free of } p
\end{aligned}$$

NOTE:

field: How to show sufficiency (not using factorization)

field: Let $p(\mathbf{X}|\theta)$ be the joint PDF or PMF of $\mathbf{X}$ and $q(t|\theta)$ the PDF or PMF of the statistic $T(\mathbf{X})$. Then $T(\mathbf{X})$ is a sufficient statistic for θ if for every $\mathbf{X}$ in the sample space, the ratio

$$\frac{p(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)}$$

is constant as a function of θ

NOTE:

field: Suppose that $X_1, \dots X_n$ are iid $N(\mu, \sigma^2)$ where σ^2 is known. If the statistic $T(\mathbf{X}) = \bar{X}_n$ sufficient for μ?

field:

$$\begin{aligned}
\frac{f(\mathbf{x}|\mu)}{q(T(\mathbf{X})|\mu)} &= \frac{(2\pi\sigma^2)^{n/2} e^{-\frac{1}{2\sigma^2}[\sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2]}}{(2\pi\sigma/n)^{-1/2} e^{-\frac{1}{2\sigma^2}(\bar{x}-\mu)^2}} \\
&= n^{-1/2}(2\pi\sigma^2)^{-(n-1)/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\bar{x})^2}
\end{aligned}$$

Which does not depend on μ, and therefore $\bar{X}_n$ is suffient for μ as long as σ^2 is known

NOTE:

field: The joint pdf of the sample $\mathbf{X} = (X_1, X_2, \dots X_n)$ isSuppose that $X_1, \dots X_n$ are iid $N(\mu, \sigma^2)$ where σ^2 is known.

field:

$$\begin{aligned}
f(\mathbf{x}|\mu) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(x_i-\mu)^2} \\
&= (2\pi\sigma^2)^{n/2} e^{-1/2\sigma^2 \sum_{i=1}^n (x_i-\mu)^2} \\
&= (2\pi\sigma^2)^{n/2} e^{-1/2\sigma^2 \sum_{i=1}^n (x_i-\bar{x}+\bar{x}-\mu)^2} \\
&= (2\pi\sigma^2)^{n/2} e^{-1/2\sigma^2 \sum_{i=1}^n (x_i-\bar{x})^2 + 2(\bar{x}-\mu)\sum_{i=1}^n (x_i-\bar{x}) + n(\bar{x}-\mu)^2} \\
&= (2\pi\sigma^2)^{n/2} e^{-1/2\sigma^2 (\sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2)}
\end{aligned}$$

NOTE:

field: Show a statistic $T(\mathbf{X})$ is sufficient

field: Neyman factorization theorem Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of the sample $\mathbf{X}$, A statistic $T(\mathbf{X})$ is a sufficient statistic for θ if and only if there exists functions $g(t|\theta)$ and $h(\mathbf{x})$ such that for all sample points $\mathbf{x}$ and all values of θ we can write

$$f(\mathbf{x}|\theta) = g(T(x)|\theta)h(\mathbf{x})$$

Note, in the theorem

- The function $g(T(\mathbf{X})|\theta)$ depends on $\mathbf{x} = (x_1, \dots x_n)$ only through the statistic $T(\mathbf{X})$.
- The function $h(\mathbf{X})$ does not depend on θ

NOTE:

field: Exponential Family

field:

$$f(\mathbf{X}|\theta) = \mathbf{h}(\mathbf{x})\mathbf{c}(\theta)\mathbf{e}^{\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \mathbf{w}_{\mathbf{i}}((\theta))\mathbf{t}_{\mathbf{i}}(\mathbf{x})}$$

NOTE:

field: Sufficiency in the exponential family

field: Let $X_1, \dots, X_n$ be iid observations from a PDF or PMF, $f(x|\boldsymbol{\theta})$ that belongs to an exponential family of the form

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})e^{\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)}$$

Where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$, $d \leq k$. Then

$$T(\mathbf{X}) = \big(\sum_{j=1}^{k} t_i(x_j), \cdots, \sum_{j=1}^{k} t_k(x_j)\big)$$

NOTE:

field: Minimal Sufficient Statistic

field: A sufficient statistic $T(\mathbf{X})$ is called a **Minimal Sufficient Statistic** if for any other sufficient statistic $T'(\mathbf{X})$, $T(\mathbf{X})$ is a function of $T'(\mathbf{X})$

NOTE:

field: Determining if a statistic is minimal sufficient

field: Let $f(x|\theta)$ be the PDF or PMF of a sample $\mathbf{X}$. Suppose there exists a function $T(x)$ such that, for every two sample points, $\mathbf{x}$ and $\mathbf{y}$, the ratio $\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)}$ is constant as a fucntion of θ iff and only if $T(\mathbf{x}) = T(\mathbf{y})$. Then $T(\mathbf{x})$ is a minimal sufficient statistic for θ.

NOTE:

field: Example of finding a minimal sufficient statistic: Suppose that $X_1, \ldots, X_n$ are idd Bernoulli(p). What is a minimal sufficient statistic for p?

field:

$$\begin{aligned} f(\mathbf{x}|p) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \end{aligned}$$

And therefore for any two sample points $\mathbf{x}$ and $\mathbf{y}$, we obtain

$$\begin{aligned} \frac{f(\mathbf{x}|p)}{f(\mathbf{y}|p)} &= \frac{p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}}{p^{\sum_{i=1}^{n} y_i}(1-p)^{n-\sum_{i=1}^{n} y_i}} \\ &= p^{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i}(1-p)^{\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i} \end{aligned}$$

Which is constant as a function of p iff $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$

Hence it follows from Lehman-Sheffe that $T(\mathbf{x}) = \sum_{i=1}^{n} x_i$ is minimal sufficient for p

NOTE:

field: Minimal sufficient statistic for μ, σ^2, where the Xs are $N(\mu, \sigma^2)$

field: $T(\mathbf{x}) = (\bar{x}, S_x^2)$ by Lehmann-Schaffe is minimal sufficient.

NOTE:

field: Facts about sufficiency

field:

- The entire sample $\mathbf{X}$ is always sufficeint.
- Any one-to-one funciton of a minimal sufficient statisitc is also a minimal sufficient statistic

NOTE:

field: Ancillary Statistic

field: A statistic $S(\mathbf{X})$ whose distribution does not depend on the parameter θ is called an ancillary statistic for θ

NOTE:

field: Complete statistic

field: Let $f(t|\theta)$ be the family of pdf's or pmfs for a statistic $T = T(\mathbf{x})$.

The family of probability distributions is called **complete** (with respect to θ) if $E_\theta(g(t)) = 0$ for all θ, implies $P_\theta(g(T) = 0) = 1$ for all θ

Equivalently, we say that $T = T(\mathbf{X})$ is a complete statistic.

In short, a statistic $T = T(\mathbf{x})$ is complete, if $E_\theta(g(T)) = 0$ for all θ implies $g(t) = 0$ with probability 1

NOTE:

field: (Binomial complete sufficient statistic)

field: Suppose the statistic $T \sim Binom(n, p)$, $0 < p < 1$, and let g be a function such that $E_p(g(T)) = 0$ for all p.

Then, with $r = (\frac{p}{1-p})^t$

$$\begin{aligned}
0 &= E_p(g(T)) \\
&= \sum_{t=0}^{n} g(t) \binom{n}{t} p^t (1-p)^{n-1} \\
&= (1-p)^n \sum_{t=0}^{n} g(t) \binom{n}{t} (\frac{p}{1-p})^t \\
&= (1-p)^n \sum_{t=0}^{n} g(t) \binom{n}{t} r^t \\
&= \neq 0 \cdot \text{This is a polynomial of degree } n \text{ in } r \text{ with coefficients } g(t) \binom{n}{t}
\end{aligned}$$

For the polynomial to be 0 for all r (and consequently for all p) each coefficient must be zero and therefore it must be the case that $g(t) = 0$ for $t = 0, 1, 2, \cdots, n$ Since $T \sim Binom(n, p)$, we have that T takes on the values $t = 0, 1, 2, \ldots n$ with probability 1 and therefore, we obtain $P_p(g(T) = 0) = 1$. Hence T is a complete statistic.

NOTE:

field: Uniform complete sufficient statistic

field: Suppose that $X_1, \ldots, X_n$ are iid Uniform$(0, \theta)$, $\theta > 0$. We know that $T(\mathbf{X}) = X_{(n)}$ (the max order statistic) is sufficient for θ. Furtheremore ,

$$f(t|\theta) = nt^{n-1}\theta^{-n} \quad 0 < t < \theta$$

Now suppose that $g(t)$ is a function satisfying $E_\theta(g(T)) = 0, \forall\theta$ Differentiating on both sides with respect to θ,

$$
\begin{aligned}
0 &= \frac{d}{d\theta}E_\theta(g(t)) \\
&= \frac{d}{d\theta}\int_0^\theta g(t)nt^{n-1}\theta^{-n}dt \\
&= \theta^{-n}\frac{d}{d\theta}\int_0^\theta g(t)nt^{n-1}dt + (\frac{d}{d\theta}\theta^{-n})\int_0^\theta g(t)nt^{n-1}dt \\
&= \theta^{-n}g(\theta)n\theta^{n-1} + 0
\end{aligned}
$$

Since $n\theta^{-1} \neq 0$, we must have that $g(\theta) = 0 \quad \forall\theta > 0$. And therefore T is complete.

NOTE:

field: Does minimal sufficent imply complete?

field: No

Suppose that $X_1, \ldots X_n$ are iid $N(\theta, \theta^2)$ where $\theta \in \mathbb{R}$ is the unknown parameter of interest.

We have

$$
\begin{aligned}
\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} &= \frac{(2\phi\sigma^2)^{-n/2}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\theta)^2}}{(2\phi\sigma^2)^{-n/2}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-\theta)^2}} \\
&= \frac{e^{-\frac{1}{2\sigma^2}[\sum_{i=1}^n x_i^2 - 2\theta\sum_{i=1}^n x_i]}}{e^{-\frac{1}{2\sigma^2}[\sum_{i=1}^n y_i^2 - 2\theta\sum_{i=1}^n y_i]}}
\end{aligned}
$$

Which is free of θ if $\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$ and $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$
It follows that $T(\mathbf{X}) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ is minimal sufficient for θ
Now observe that $T_1(\mathbf{X}) = \sum_{i=1}^n x_i \sim N(n\theta, n\theta^2)$ and therefore

$$
\begin{aligned}
E(T_1^2) &= V(T_1) + [E(T_1)]^2 \\
&= n\theta^2 + n^2\theta^2 \\
&= n\theta^2(1+n)
\end{aligned}
$$

On the other hand, for $T_2 = \sum_{i=1}^n x_i^2$,

$$\begin{aligned} E(T_2) &= nE(X_1)^2 \\ &= n[V(X_1) + [E(X_1)]^2] \\ &= n\theta^2 + n\theta^2 \\ &= 2n\theta^2 \end{aligned}$$

Then, taking $h(t_1, t_2) = 2t_1^2 - (n+1)t_2$, we have

$$\begin{aligned} E_\theta[h(T_1, T_2)] &= E_\theta[2T_1^2 - (n+1)T_2] \\ &= 2E_\theta(T_1^2) - (n+1)E(T_2) \\ &= 2n(n+1)\theta^2 - 2n(n+1)\theta^2 \\ &= 0 \quad \forall\theta \end{aligned}$$

But because $h(\mathbf{t}) \neq 0 \quad \forall\theta$, we have that $T(\mathbf{X})$ is not complete.

NOTE:

field: Complete statistics in the exponential family

field: Let $X_1, \ldots, X_n$ be iid observations from an exponential family. with PDF or PMF of the form

$$f(x|\theta) = h(x)c(\theta)e^{\sum_{j=1}^k \omega_j(\theta_j)t_j(x)}$$

Where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$

Then, the statistic $T(\mathbf{X}) = (\sum_{i=1}^n t_1(x_i), \sum_{i=1}^n t_2(x_i), \ldots, \sum_{i=1}^n t_k(x_i))$ is complete, as long as the parameter space Θ contains an open set in $\mathbb{R}^k$

NOTE:

field: Suppose that a statistic T is complete and let g be a one-to-one function. Is the statistic $U = g(T)$ also complete?

field: Yes

NOTE:

field: Does complete statistic imply minimial sufficient statistic?

field: If a minimal sufficient statistic exists, then any complete statistic is also a minimal sufficient statistic

NOTE:

field: Basu's Theorem

field: If $T(\mathbf{x})$ is a complete and minimal sufficient statistic, then $T(\mathbf{x})$ is an independent of every ancillary statistic.

NOTE:

field: Likelihood function

field: Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of the sample $\mathbf{X} = (X_1, \ldots, X_n)$, then given that $\mathbf{X} = \mathbf{x}$ is observed, the function of θ defined as

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$$

is called the Likelihood Function

NOTE:

field: Idea of likelihood function

field: Suppose that $\mathbf{X}$ is a discrete random vector (so we can interpret probabilities easier)

Then $L(\theta|\mathbf{x}) = P_\theta(\mathbf{X} = \mathbf{x})$. Now if we compare the likelihood function at two parameter values θ_1, θ_2 and we observe that

$$P_{\theta_1}(\mathbf{X} = \mathbf{x}) = L(\theta_1|\mathbf{x}) > L(\theta_2|\mathbf{x}) = P_{\theta_2}(\mathbf{X} = \mathbf{x}$$

Then, the sample point $\mathbf{x}$ that we actually observed is more likely to have occured if $\theta = \theta_1$, than if $\theta = \theta_2$, which can be interpreted as that θ_1, is a more plausible value for the true value of θ than θ_2 is.

NOTE:

field: Fisher information - one parameter case

field: Let X be a random variable with pdf or pmf $f(x|\theta)$ where $\theta \in \Theta \subseteq \mathbb{R}$ (Fisher) information about θ contained in X is

$$I_X(\theta) = E_\theta[\big(\frac{\partial}{\partial\theta}\log f(x|\theta)\big)^2]$$

NOTE:

field: Example of one parameter case Fisher information Suppose that $X \sim Bern(p)$ What is the information that X contains about the parameter p?

field: We have that $f(x|p) = p^x(1-p)^{1-x}$. Then

$$\log f(x|p) = x\log p + (1-x)\log(1-p)$$
$$\frac{\partial}{\partial p}\log f(x|p) = \frac{x}{p} - \frac{1-x}{1-p}$$

We obtain

$$\begin{aligned}\big(\frac{\partial}{\partial p}\log f(x|p)\big)^2 &= \big(\frac{x}{p} - \frac{1-x}{1-p}\big)^2 \\ &= \frac{x^2}{p^2} - \frac{2x(1-x)}{p(1-p)} + \frac{(1-x)^2}{(1-p)^2} \\ &= \frac{x^2}{p^2} - \frac{2(x-x^2)}{p(1-p)} + \frac{(1-2x+x^2)}{(1-p)^2}\end{aligned}$$

Therefore,

$$
\begin{aligned}
I_x(p) &= E_p[(\frac{\partial}{\partial p}\log f(x|p))^2] \\
&= \frac{p}{p^2} - \frac{2(p-p)}{p(1-p)} + \frac{1-2p+p}{(1-p)^2} \\
&= \frac{1}{p} + \frac{1}{1-p} \\
&= \frac{1}{p(1-p)}
\end{aligned}
$$

NOTE:

field:

$$I_x(\theta) = E_\theta\big[(\frac{\partial}{\partial\theta}\log f(x|\theta))^2\big] =$$

field: If $f(x|\theta)$ satisfies

$$\frac{\partial}{\partial\theta}E_\theta\big(\frac{\partial}{\partial\theta}\log f(x|\theta)\big) = \int \frac{\partial}{\partial\theta}\big[\frac{\partial}{\partial\theta}\log f(x|\theta)\big]f(x|\theta)dx$$

$$I_x(\theta) = E_\theta\big[(\frac{\partial}{\partial\theta}\log f(x|\theta))^2\big] = -E_\theta(\frac{\partial^2}{\partial\theta^2}\log f(x|\theta))$$

NOTE:

field: Suppose that $X_1, \ldots, X_n$ are iid observations with common pdf or pmf $f(x|\theta)$. Then, the information about θ contained in the sample $\mathbf{X} = (X_1, \ldots, X_n)$ is

field:

$$I_{\mathbf{X}}(\theta) = nI_{X_1}(\theta)$$

NOTE:

field: Fisher Information - multiparameter case

field: Let X be a random variable with pdf or pmf $f(x|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \Theta \subseteq \mathbb{R}^2$. Denote by

$$I_{ij}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\big[(\frac{\partial}{\partial \theta_i} \log f(x|\boldsymbol{\theta}))(\frac{\partial}{\partial \theta_j} \log f(x|\boldsymbol{\theta}))\big] = -E_{\boldsymbol{\theta}}[\frac{\partial}{\partial \theta_i \theta_j} \log f(x|\boldsymbol{\theta})]$$

For $i, j = 1, 2$. Then the (fisher) information matrix about $\boldsymbol{\theta}$ is

$$I_x(\boldsymbol{\theta}) = \begin{pmatrix} I_{11}(\boldsymbol{\theta}) & I_{12}(\boldsymbol{\theta}) \\ I_{21}(\boldsymbol{\theta}) & I_{12}(\boldsymbol{\theta}) \end{pmatrix}$$

NOTE:

field: Find Fisher information for Normal RVs

field: We have that $\boldsymbol{\theta} = (\mu, \sigma^2)$ and $f(x|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

Then,

$$\frac{\partial}{\partial \mu} \log f(x|\boldsymbol{\theta}) = \frac{\partial}{\partial}[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2] = \frac{(x-\mu)}{\sigma^2}$$

$$\frac{\partial}{\partial \sigma^2} = \frac{1}{2\sigma^2}[\frac{(x-\mu)^2}{\sigma^2} - 1]$$

Therefore $I_{11} = E_\theta[(\frac{\partial}{\partial \mu} \log f(x|\boldsymbol{\theta}))^2] = E_\theta[\frac{(x-\mu)^2}{\sigma^4}] = \frac{1}{\sigma^4}\sigma^2 = \frac{1}{\sigma^2}$

$$\begin{aligned}
I_{22}(\boldsymbol{\theta}) &= E_\theta[\frac{\partial}{\partial \sigma^2} \log f(x|\boldsymbol{\theta})^2] \\
&= E_\theta\big\{[\frac{1}{2\sigma^2}(\frac{(x-\mu)^2}{\sigma^2} - 1)]^2\big\} \\
&= \frac{1}{4\sigma^4} E_\theta[(\frac{(x-\mu)^2}{\sigma^2} - 1)^2] \\
&= \frac{1}{4\sigma^4 \cdot 2} \\
&= \frac{1}{2\sigma^4} \quad \text{Since } = V(\chi_1^2)
\end{aligned}$$

Now for the off diagonal elements,

$$\begin{aligned}
I_{12}(\boldsymbol{\theta}) = I_{22}(\boldsymbol{\theta}) &= E_\theta\big[(\frac{\partial}{\partial\mu}\log f(x|\theta)(\frac{\partial}{\partial\sigma^2}\log f(x|\theta))\big] \\
&= E_\theta\big[\frac{(x-\mu)}{\sigma^2}\frac{1}{2\sigma^2}[\frac{x-\mu}{\sigma^2}\cdot 1]\big] \\
&= \frac{1}{2\sigma^4}E_\theta[\frac{(x-\mu)^3}{\sigma^3}-(x-\mu)]
\end{aligned}$$

But $E_\theta[(x-\mu)^3] = E_\theta[(x-\mu)] = 0$, because X is symmetric around μ, and we obtain $I_{12}(\boldsymbol{\theta}) = I_{21}(\boldsymbol{\theta}) = 0$

We obtain that

$$\begin{aligned}
I_{x_1}(\boldsymbol{\theta}) &= \begin{pmatrix} I_{11}(\boldsymbol{\theta}) & I_{12}(\boldsymbol{\theta}) \\ I_{21}(\boldsymbol{\theta}) & I_{22}(\boldsymbol{\theta}) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}
\end{aligned}$$

And hence

$$I_{\mathbf{x}}(\boldsymbol{\theta}) = nI_{X_1}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

NOTE:

field: $I_T(\theta) \leq$

field: $I_T(\theta) \leq I_{\mathbf{X}}(\theta)$ (The information of the statistic is less than or equal to the information of the sample)

NOTE:

field: Let $\mathbf{X} = X_1, \ldots, X_n$ denote the entire data, and let $T = T(\mathbf{X})$ be some statistic. Then, for all $\theta \in \Theta \subseteq \mathbb{R}$, $I_{\mathbf{X}}(\theta) \geq I_t(\theta)$ Where the equality is attained...

field: Let $\mathbf{X} = X_1, \ldots, X_n$ denote the entire data, and let $T = T(\mathbf{X})$ be some statistic. Then, for all $\theta \in \Theta \subseteq \mathbb{R}$, $I_{\mathbf{X}}(\theta) \geq I_t(\theta)$ Where the equality is attained if and only iff $T(\mathbf{X})$ is sufficient for θ

NOTE:

field: Let $\mathbf{X} = (X_1, \ldots, X_n)$, denote a sample of iid observations and suppose the statistic $T(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X}))$ is such that T_1 and T_2 are independent. Then

$$I_T(\boldsymbol{\theta}) =$$

field: Let $\mathbf{X} = (X_1, \ldots, X_n)$, denote a sample of iid observations and suppose the statistic $T(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X}))$ is such that T_1 and T_2 are independent. Then

$$I_T(\boldsymbol{\theta}) = I_{T_1}(\boldsymbol{\theta}) + I_{T_2}(\boldsymbol{\theta})$$

NOTE:

field: Point estimator

field: Any statistic $T(\mathbf{X})$ that is used to estimate the value of a parameter is called a point estimator of θ. We write $\hat{\theta} = T(\mathbf{X})$

NOTE:

field: Method of moments

field:

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i^1, \quad \mu_1 = E(X^1)$$
$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \quad \mu_2 = E(X^2)$$
$$\vdots$$
$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \quad \mu_k = E(X^k)$$

Equating and solving for θ gives the MoM estimators

NOTE:

field: Example Method of Moments Suppose that $X_1, \ldots, X_n$ are iid Binomial(k, p), where both k and p are unknown.

field: We have that

$$P(X_i = x|k, p) = \binom{k}{x} p^x (1-p)^{k-x}, x = 0, 1, \ldots, k$$

and we obtain $E(X_1) = kp$, $E(X_1^2) = kp(1-p) + k^2p^2$

Solving the sytem of equations we obtain

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = kp$$
$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = kp(1-p) + k^2p^2$$

Sovling the system of equations:

$$\tilde{p} = \frac{\bar{x}}{\tilde{k}}$$
$$\tilde{k} = \frac{\bar{x}^2}{\bar{x} - \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Possible problems: k has to be an integer, and not negative. (Estimates of parameters that are outside of the parameter space.)

NOTE:

field: Maximum Likelihood Estimator

field: In this context, we define the **Maximum Likelihood Estimator (MLE)** of θ as the parameter value $\hat{\theta}_{ML} = \hat{\theta}(\mathbf{x})$ that satisfies

$$L(\hat{\theta}_{ML}|\mathbf{x}) = \sup_{\theta \in \Theta} L(\theta|\mathbf{x})$$

Note this often proceedes as taking the derivative of the log likelihood function and setting to zero to solve for parameters - not always

NOTE:

field: Example of MLE Suppose that $X_1, \ldots, X_n$ are iid Exponential(λ). Find the MLE $\hat{\lambda}_{ML}$ of λ

field: Suppose that $X_1, \ldots, X_n$ are iid Exponential(λ). Find the MLE $\hat{\lambda}_{ML}$ of λ

We have that $f(x|\lambda) = \frac{1}{\lambda} e^{x/\lambda}$, $x > 0$, and therefore

$$L(\lambda|x) = \prod_{i=1}^{n} \frac{1}{\lambda} e^{x_i/\lambda} = \lambda^{-n} e^{-\frac{1}{\lambda}\sum_{i=1}^{n} x_i}$$

Since $\log(\cdot)$ is a strictly monotone (one-to-one) and increasing, we consider instead the maximization of the log-likelihood

$$l(\lambda|\mathbf{x}) = \log L(\lambda|\mathbf{x}) = -n \log \lambda - \frac{1}{\lambda} \sum_{i=1}^{n} x_i$$

$$\frac{\partial}{\partial \lambda} l(\lambda|\mathbf{x}) = \frac{-n}{\lambda} + \frac{1}{\lambda^2} \sum_{i=1}^{n} x_i$$

Solving $\frac{\partial}{\partial\lambda} l(\lambda|\mathbf{x}) = 0$, we obtain

$$\frac{-n}{\lambda} + \frac{1}{\lambda^2}\sum_{i=1}^{n} x_i = 0$$

$$-n\lambda + n\bar{x} = 0$$

$$\lambda = \bar{x}$$

NOTE:

field: Example of MLE when can't differemtiate

Suppose that $X_1, \ldots, X_n$ are iid Uniform$(0, \theta)$, $\theta > 0$. Find the MLE of θ

field: We have that $f(x|\theta) = \frac{1}{\theta} I(0 < x < \theta)$

And therefore

$$\begin{aligned} L(\theta|\mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\theta} I(0 < x_i < \theta) \\ &= \frac{1}{\theta^n} I(X_{(1)} > 0) I(X_{(n)} < \theta) \end{aligned}$$

In this case, the support of X depends on θ and the maximization problem only makes sense whenever $L(\theta|\mathbf{x}) > 0$. We cannot simply approach the problem by taking partial derivatives, but assuming the likelihood is positive, we notice that $L(\theta|\mathbf{x})$ is decerasing as a function of θ, for $\theta > X_{(n)}$

Picture with $L(\theta)$ as zero untill $X_{(n)}$ on x axis, goes up to $1/X_{(n)}$ there and decreases with $\frac{1}{\theta^n}$

It follows the MLE of θ is $\hat{\theta}_{ML} = X_{(n)}$

NOTE:

field: If $\hat{\theta}_{ML}$ is the MLE of θ, then for any function $\tau(\theta)$, the MLE of $\eta = \tau(\theta)$ is $\hat{\eta}_{ML} =$

field: If $\hat{\theta}_{ML}$ is the MLE of θ, then for any function $\tau(\theta)$, the MLE of $\eta = \tau(\theta)$ is $\hat{\eta}_{ML} = \tau(\hat{\theta}_{ML})$

NOTE:

field: Bias

field: Let $\hat{\theta} = T(\mathbf{X})$ be an estimatro of θ. Then the Bias of $\hat{\theta}$ as an estimator of θ is defined as

$$B_\theta(\hat{\theta}) = E_\theta(\hat{\theta} - \theta) = E_\theta(\hat{\theta}) - \theta$$

That is the difference between the expected value of $\hat{\theta}$ and θ.
An estimator $\hat{\theta}$ of θ is said to be unbiased if $B_\theta(\hat{\theta}) = 0 \quad \forall\theta$

NOTE:

field: Mean Squared Error

field: Let $\hat{\theta} = T(\mathbf{X})$ be an estimate of θ. Then, the **Mean Squared Error** (MSE) of $\hat{\theta}$ as an estimator of θ is defined as:

$$MSE(\hat{\theta}) = E_\theta[(\hat{\theta} - \theta)^2] = V_\theta(\hat{\theta}) + [B_\theta(\hat{\theta})]^2$$

NOTE:

field: Do unbiased estimators always exist?

field: No, Suppose that $X \sim$ Binomial(n, p) and let $\theta = 1/p$ be the parameter of interest. Can we find an unbiased estimator for θ?- No

NOTE:

field: UMVUE

field: An estimator W^* is called a best unbiased estimator of $\tau(\theta)$ if it satisfies $E_\theta(W^*) = \tau(\theta)$, for all θ, and for any other estimator W with $E_\theta(W) = \tau(\theta)$, we have $V_\theta(W^*) \leq V_\theta(W), \forall\theta$. Equivalently W^* is also called a **Uniform Minimal Variance Unbiased Estimator** (UMVUE) of $\tau(\theta)$

NOTE:

field: Finding a UMVUE

field: Start with a complete statistic, (find min suff statistic, prove completeness), Find bias (ie $E(T(\mathbf{X}))$). Then adjust $T(\mathbf{X})$ to be unbiased. (ie center or scale)

NOTE:

field: Cramer-Rao Inequality

field: Let $X_1, \ldots, X_n$ be a sample with joint pdf or pmf $f(\mathbf{x}|\theta)$ and let $W(\mathbf{X}) = W(X_1, \ldots, X_n)$ be any estimator satisfying

$$\frac{d}{d\theta} E_\theta(W(X)) = \int \frac{d}{d\theta}[W(\mathbf{X}) f(\mathbf{x}|\theta)] d\mathbf{x}$$

and $V_\theta(W(\mathbf{X})) < \infty$
Then,

$$V_\theta(W(\mathbf{X})) \geq \frac{(\frac{d}{d\theta} E_\theta(W(\mathbf{X})))^2}{E_\theta[(\frac{\partial}{\partial\theta} \log f(\mathbf{x}|\theta))^2]}$$

Observe that if the sample $X_1, \ldots, X_n$ is iid with common pdf or pmf $f(x|\theta)$, we obtain

$$V_\theta(W(\mathbf{X})) \geq \frac{[\frac{d}{d\theta} E_\theta(W(\mathbf{X}))]^2}{n E_\theta[(\log f(\mathbf{x}|\theta))^2]}$$

The denominator is the information in the sample about θ

We have that as the information number gets bigger we have a smaller bound for the variance. of the best unbiased estimator and therefore more information is available.

NOTE:

field: Cramer-Rao and UMVUE example UMVUE of λ for Poisson

field: Poisson example, we have $\tau(\lambda) = \lambda$, so $\frac{d}{d\lambda}\tau(\lambda) = 1$

On the other hand,

$$\begin{aligned}
nE_\lambda[(\frac{d}{d\lambda}\log f(x|\lambda))^2] &= -nE_\lambda(\frac{\partial^2}{\partial\lambda^2})\log f(x|\lambda)) \\
&= -nE_\lambda(\frac{\partial^2}{\partial\lambda^2}\log(\frac{e^{-\lambda}\lambda^x}{x!})) \\
&= -nE_\lambda[\frac{\partial^2}{\partial\lambda^2}(-\lambda + x\log\lambda - \log(x!))] \\
&= -nE_\lambda(\frac{-x}{\lambda^2}) \\
&= \frac{n}{\lambda}
\end{aligned}$$

Therefore, for any unbiased estimator W of λ, we must have $V_\lambda(W) \geq \lambda/n$. Since $V_\lambda(\bar{X}) = \frac{\lambda}{n}$, we have that $\bar{X}$ is an UMVUE of λ

NOTE:

field: Does S^t for Normal attain cramer rao?

field: No - Suppose that $X_1, \ldots, X_n$ are iid N(μ, σ^2) and consider the estimation of σ^2 when μ is unknown.

We have that

$$\frac{\partial^2}{\partial(\sigma^2)^2}\log[\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}] = \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}$$

and

$$\begin{aligned}
-E[\frac{\partial^2}{\partial(\sigma^2)^2}\log f(x|\mu,\sigma^2)] &= -E(\frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}) \\
&= -\frac{1}{2\sigma^4} + \frac{\sigma^2}{\sigma^6} \\
&= \frac{1}{2\sigma^4}
\end{aligned}$$

and therefore, any unbiased estimator W of σ^2 must satisfy $V(W) \geq \frac{2\sigma^4}{n}$. Recall that for S^2 we have

$$V(S^2) = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n}$$

and therefore S^2 does not attain the cramer-rao lower bound.

NOTE:

field: Rao-Blackwell

field: Let W be any unbiased estimator $\tau(\theta)$ and let T be a sufficient statistic for θ. Define $\phi(T) = E(W|T)$. Then $E_\theta(\phi(T)) = \tau(\theta)$ and $V_\theta(\phi(T)) \leq V_\theta(W)$, for all θ That is, $\phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$

NOTE:

field: Use of Rao-Blackwell

field: Estimators can be improved (their MSE) using sufficiency (already sufficient statistics, or functions of sufficient statistics cannot be improved)

NOTE:

field: Are unbiased estimators based on complete sufficient statistics unique.

field: Unbiased estimators based on complete sufficient statistics are unique.