Theorem

Alternative Formula for Expected Value

Intermediate Statistics

Theorem

Transformation of Random Variable

Intermediate Statistics

Theorem

Sample Mean and Variance

Intermediate Statistics

Theorem

Delta Method

Intermediate Statistics

Theorem

Multivariate Delta Method

Intermediate Statistics

Theorem

Geometric Series

Intermediate Statistics

Theorem

Gaussian Tail Inequality

Intermediate Statistics

Theorem

Markov's Inequality

Intermediate Statistics

Theorem

Chebyshev's Inequality

Intermediate Statistics

Theorem

Hoeffding's Inequality

Intermediate Statistics

Let X be a random variable and define $Y = g(X)$, where g must be a monotonic function. Then Y has pdf $$p_Y(y) = p_X\left(g^{-1}(y)\right)\left\lvert\frac{dg^{-1}(y)}{dy}\right\rvert$$	Let X be a non-negative continuous random variable. Then $$\mathbb{E}(X) = \int_0^\infty P\left(X > t\right)dt.$$ Analogously, if X is a discrete non-negative random variable, we have $$\mathbb{E}(X) = \sum_{k=1}^\infty P(X \geq k).$$
if $X \sim N\left(\mu, \sigma^2\right)$ and $Y = g(X)$ with σ^2 small, we have $$Y \approx N\left(g(\mu), \sigma^2(g'(\mu))^2\right)$$ Proof: Start with Taylor Expansion of $g(X)$ around μ.	Given a random sample $X_1, \ldots, X_n \sim N\left(\mu, \sigma^2\right)$, the following statements are true: • $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ • $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ • $\bar{X}$ and S^2 are independent
For $r \in (0,1)$, $$a + ar + ar^2 + \ldots = \frac{a}{1-r}.$$ A partial geometric series $a + ar + ar^2 + \ldots + ar^{n-1}$ sums up to $\frac{a(1-r^n)}{1-r}$	Suppose that $Y_n = (Y_{n1}, \ldots, Y_{nk})$ is a sequence of random variables such that $$\sqrt{n}\left(Y_n - \mu\right) \rightsquigarrow N(0, \Sigma).$$ For a function $g : \mathbb{R}^k \to \mathbb{R}$ and the gradient of g with respect to y be $\nabla g(y) = \left(\frac{\partial g}{\partial y_1}, \ldots, \frac{\partial g}{\partial y_k}\right)^\intercal$. Denoting with ∇_μ the gradient evaluated at $y = \mu$ where all elements are assumed to be non-zero, we have $$\sqrt{n}\left(g(Y_n) - g(\mu)\right) \rightsquigarrow N\left(0, \nabla_\mu^\intercal \Sigma \nabla_\mu\right).$$
If X is a non-negative random variable with existing expectation $\mathbb{E}(X)$, then $$P\left(X > \varepsilon\right) \leq \frac{\mathbb{E}(X)}{\varepsilon}$$ Proof: Trivially, the inequality $\varepsilon\mathbb{1}(X > \varepsilon) \leq X$ holds. Taking the expectation on both sides and rearranging yields Markov's inequality.	If $X \sim N(0,1)$, then $$P(\lvert X\rvert > \varepsilon) \leq \frac{2}{\varepsilon}e^{-\varepsilon^2/2}.$$ In general, for a sample $X_1, \ldots, X_n$ where $X_i \sim N\left(\mu, \sigma^2\right)$, we have $$P\left(\left\lvert\bar{X}_n - \mu\right\rvert > \varepsilon\right) \leq \frac{2\sigma}{\varepsilon\sqrt{n}}\exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right)$$
Let $X_1, \ldots, X_n$ be independent observations such that $\mathbb{E}[X_i] = \mu$ and $a \leq X_i \leq b \quad \forall i$. Then, for any $\varepsilon > 0$, we have $$P\left(\left\lvert\bar{X}_n - \mu\right\rvert > \varepsilon\right) \leq 2\exp\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right)$$	Let X be a random variable with mean μ and variance σ^2. Then $$P\left(\lvert X - \mu\rvert > \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon}$$ Proof: Define $Z = (X-\mu)^2$. By Markvo's Inequality, we have $P(Z > \varepsilon^2) \leq \frac{\mathbb{E}(Z)}{\varepsilon^2}$ which is equivalent to $$P\left(\lvert X - \mu\rvert > \varepsilon\right) \leq \frac{\mathbb{E}\left[(X-\mu)^2\right]}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}$$

Theorem

Bernstein's Inequality

Intermediate Statistics

Theorem

McDiarmid's Inequality

Intermediate Statistics

Theorem

Jensen's Inequality

Intermediate Statistics

Theorem

Cauchy-Schwartz Inequality

Intermediate Statistics

Definition

Little o

Intermediate Statistics

Definition

Big O

Intermediate Statistics

Definition

Little o_p

Intermediate Statistics

Definition

Big o_p

Intermediate Statistics

Definition

Consistent Estimator

Intermediate Statistics

Theorem

Consistency Conditions

Intermediate Statistics

Let $X_1, \ldots, X_n$ be independent random variables. Suppose that

$$\sup_{x_1,\ldots,x_n,x_i'} |g(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_n) - g(x_1,\ldots,x_{i-1},x_i',x_{i+1},\ldots,x_n)| \leq c_i \quad (1)$$

for $i = 1, \ldots, n$. Then

$$P\left(g\left(X_1,\ldots,X_n\right) - \mathbb{E}\left[g\left(X_1,\ldots,X_n\right)\right] \geq \varepsilon\right) \leq \exp - \frac{2\varepsilon^2}{\sum_{i=1}^n c_i^2}.$$

Let $X_1, \ldots, X_n$ be independent observations such that $\mathbb{E}[X_i] = 0$, $|X_i| \leq M$, and $\mathbb{V}(X_i) \leq \sigma^2$. Then, for every $\varepsilon > 0$, we have

$$P\left(\left|\bar{X}_n\right| \geq \varepsilon\right) \leq 2\exp\left(-\frac{n\varepsilon^2}{2\sigma^2 + \frac{2}{3}M\varepsilon}\right).$$

Let X and Y be two random variables with finite variance. Then

$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$

Let X be a random variable and g a convex function. Then $\mathbb{E}g(X) \geq g\left(\mathbb{E}(X)\right)$. On the other hand, if g is concave, we have $\mathbb{E}g(X) \leq g\left(\mathbb{E}(X)\right)$. Example: From Jensen's inequality, it follows that $\mathbb{E}(X^2) \geq \mathbb{E}(X)^2$, since $g(x) = x^2$ is convex.

$a_n = O(b_n)$ if for large $n > n_0$, there exists some constant $C > 0$ such that $a_n < Cb_n$

$a_n = o(b_n)$ means that $\forall C$ and $n > n_0$,

$$a_n < Cb_n$$

(a_n is bounded from above by b_n)

Y_n is $O_p(1)$ if, for any $\varepsilon > 0$, there exists some finite constant $C > 0$ such that

$$P\left(|Y_n| > C\right) \leq \varepsilon$$

for all n (stochastically bounded from above). $Y_n = O_p(a_n)$ means that $\frac{Y_n}{a_n} = O_p(1)$.

Y_n is $o_p(1)$ if, for every $\varepsilon > 0$, we have

$$P\left(|Y_n| > \varepsilon\right) \to 0$$

or equivalently

$$P\left(|Y_n| \leq \varepsilon\right) \to 1.$$

$Y_n = o_p(a_n)$ means that $\frac{Y_n}{a_n} = o_p(1)$.

Let θ_n be a sequence of estimators of parameter θ satisfying

$$\lim_{n\to\infty} \mathbb{V}\left(\hat{\theta}_n\right) = 0$$

and Then $\hat{\theta}_n$ is a consistent sequence of estimators of θ.

A sequence of estimators θ_n is consistent of the parameter θ if, for every $\epsilon > 0$ and every $\theta \in \Theta$ we have

$$\lim_{n\to\infty} P_\theta\left(|\theta_n - \theta| < \epsilon\right) = 1$$

or equivalently

$$\lim_{n\to\infty} P_\theta\left(|\theta_n - \theta| \geq \epsilon\right) = 0,$$

i.e. θ_n converges in probability to θ.

Theorem

Consistency of MLE

Intermediate Statistics

Definition

Shattering

Intermediate Statistics

Definition

Shatter Coefficient

Intermediate Statistics

Definition

VC Dimension

Intermediate Statistics

Definition

Uniform Distribution

Intermediate Statistics

Definition

Multinomial Distribution

Intermediate Statistics

Definition

Statistic

Intermediate Statistics

Definition

Almost Sure Convergence

Intermediate Statistics

Definition

Convergence in Probability

Intermediate Statistics

Definition

Convergence in Quadratic Mean

Intermediate Statistics

Let $\mathcal{A}$ be a class of sets and F be a finite set $\{x_1, \ldots, x_k\}$. Let G be some subset of F. $\mathcal{A}$ picks out G if $A \cap F = G$ for some $A \in \mathcal{A}$. The set F is shattered if $s(\mathcal{A}, F) = 2^k$, i.e. if all subsets can be picked out by $\mathcal{A}$.

Let $\hat{\theta}$ be the MLE of θ and let $\tau(\theta)$ be a continuous function of θ. Under regularity conditions, for every $\epsilon > 0$ and $\theta \in \Theta$,

$$lim_{n\to\infty} P_\theta \left(\left| \tau(\hat{\theta}) - \tau(\theta) \right| \geq \epsilon \right) = 0,$$

i.e. $\tau(\hat{\theta})$ is a consistent estimator of θ. The conditions are a) an iid random sample, b) identifiability of the parameter, c) common support and differentiability of the density and d) a parameter space which contains an open set of which the true parameter is an interior point.

The Vapnik-Chervonenkis (VC) Dimension is defined as

$$d = d(\mathcal{A}) = \text{largest k such that } s_k(\mathcal{A}) = 2^k.$$

This means that d is the size of the largest set that can be shattered.

The shatter coefficient is defined as

$$s_k(\mathcal{A}) = \sup_{F \in \mathcal{F}_k} s(\mathcal{A}, F),$$

where $\mathcal{F}_k$ denotes all finite sets with k elements. Fact: $s_k(\mathcal{A}) \leq 2^k$.

Multivariate version of Binomial. Draw all from urn with balls colored in$k$different colors. $p = (p_1, \ldots, p_k)$ where $\sum_j p_j = 1$ and p_j is probability of drawing color j. Draw n balls from the urn with replacement and let $X = (X_1, \ldots, X_n)$be the count of the number of balls of each color. Then X has a Multinomial distribution with pdf

$$p(x) = \begin{pmatrix} n \\ x_1 \ldots x_k \end{pmatrix} p_1^{x_1} \ldots p_k^{x_k}.$$

A continuous random variable X has a Uniform(a, b) distribution if its pdf is

$$\begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

and CDF

$$\begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } x \in [a, b) \\ 1 & \text{for } x \geq b \end{cases}.$$

The mean of X is $\frac{1}{2}(a + b)$ and the variance $\frac{1}{12}(b - a)^2$.

X_n converges almost surely to X, written $X_n \xrightarrow{a.s.} X$, if, for every $\varepsilon > 0$,

$$P\left(\lim_{n\to\infty} |X_n - X| < \varepsilon \right) = 1.$$

Almost sure convergence of X_n to X is equivalent to

$$\forall \varepsilon > 0 : \lim_{n\to\infty} P\left(\sup_{m \geq n} |X_m - X| \leq \varepsilon \right) = 1.$$

A statistic T is any function of the data $X_1, \ldots, X_n$, i.e. $T = g(X_1, \ldots, X_n)$.

A sequence of random variables X_n converges to X in quadratic mean (L_2 convergence) if

$$\mathbb{E}(X_n - X)^2 \to 0$$

as $n \to \infty$. We write $X_n \xrightarrow{q.m.} X$.

Xn converges to X in probability ($X_n \xrightarrow{P} X$), if

$$\forall \varepsilon > 0 : P(|X_n - X| > \varepsilon) \to 0$$

as $n \to \infty$ (notice that we thus have $X_n - X = o_P(1)$).

Definition

Convergence in Distribution

Intermediate Statistics

Theorem

Convergence Relationships

Intermediate Statistics

Theorem

Continuous Mapping Theorem

Intermediate Statistics

Theorem

Slutsky's Theorem

Intermediate Statistics

Theorem

The Weak Law of Large Numbers (WLLN)

Intermediate Statistics

Theorem

The Strong Law of Large Numbers (SLLN)

Intermediate Statistics

Theorem

The Central Limit Theorem (CLT)

Intermediate Statistics

Theorem

Estimate σ in CLT

Intermediate Statistics

Theorem

Multivariate Central Limit Theorem

Intermediate Statistics

Definition

Loss Function

Intermediate Statistics

Between the different convergence definitions, the following relationships hold:

- $X_n \xrightarrow{a.s.} X$ implies that $X_n \xrightarrow{P} X$.
- $X_n \xrightarrow{q.m.} X$ implies that $X_n \xrightarrow{P} X$.
- $X_n \xrightarrow{P} X$ implies that $X_n \rightsquigarrow x$.
- If $X_n \rightsquigarrow X$ and if X has a point mass distribution, i.e. $P(X = c) = 1$ for some c, then $X_n \xrightarrow{P} X$.

X_n converges to X in distribution if

$$\lim_{n\to\infty} F_n(t) = F(t)$$

at all t for which F is continuous. We write $X_n \rightsquigarrow X$.

Let X_n and Y_n be sequences of random variables and let X be a simple random variable and c a constant. We have

- If $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$, then $X_n + Y_n \rightsquigarrow X + c$.
- If $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$, then $X_n Y_n \rightsquigarrow cX$.

In general, $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow Y$ does not imply that $X_n + Y_n \rightsquigarrow X + Y$.

Let X_n and Y_n be sequences of random variables. Also, let X and Y be simple random variables. For a continuous function g, we have

1. If $X_n \xrightarrow{P} X$, then $g(X_n) \xrightarrow{P} g(X)$.
2. If $X_n \rightsquigarrow X$, then $g(X_n) \rightsquigarrow g(X)$.

Let $X_1, \ldots, X_n$ be iid with mean μ. Then we have $\bar{X}_n \xrightarrow{a.s.} \mu$.

Given a random sample $X_1, \ldots, X_n$ iid, the sample mean $\bar{X}_n$ converges in probability to μ. Therefore, $\bar{X}_n - \mu = o_p(1)$.

Let $X_1, \ldots, X_n$ be an iid sample where $\mathbb{E}(X_i) = \mu$ and $\mathbb{V}(X_i) = \sigma^2$. Let $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. Denote the sample variance with $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2$. Then

$$T_n = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{S_n} \rightsquigarrow N(0,1).$$

Proof: We have that $T_n = Z_n W_n$, where $Z_n = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \rightsquigarrow N(0,1)$ and $W_n = \frac{\sigma}{S_n} \xrightarrow{P} 1$. The result then follows from Slutsky's Theorem.

Let $X_1, \ldots, X_n$ be an iid sample where $\mathbb{E}(X_i) = \mu$ and $\mathbb{V}(X_i) = \sigma^2$. Let $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. Then

$$Z_n \equiv \frac{\bar{X}_n - \mu}{\sqrt{\mathbb{V}\left(\bar{X}_n\right)}} = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \rightsquigarrow Z,$$

where $Z \sim N(0,1)$.

A loss function $L\left(\theta, \hat{\theta}\right) : \Theta^2 \to [0, \infty)$ measures the cost associated with the value of an estimator $\hat{\theta}$ not being equal to the true parameter θ. Common loss functions are

1. Squared Loss
2. Absolute Loss
3. Zero-One Loss

Let $X_1, \ldots, X_n$ be a sample of iid random vectors where $X_i = (X_{1i}, \ldots, X_{ki})^\intercal$ with mean $\mu = (\mu_1, \ldots, \mu_k)^\intercal$ and covariance matrix Σ. Let $\bar{X} = \left(\bar{X}_1, \ldots, \bar{X}_k\right)^\intercal$ where $\bar{X}_j = n^{-1} \sum_{i=1}^n X_{ji}$. Then,

$$\sqrt{n}\left(\bar{X} - \mu\right) \rightsquigarrow N(0, \Sigma)$$

Definition

Risk of an Estimator

Intermediate Statistics

Definition

Minimax Risk

Intermediate Statistics

Definition

Bayes Risk

Intermediate Statistics

Definition

Posterior Risk

Intermediate Statistics

Theorem

Bayes Risk (in terms of posterior risk)

Intermediate Statistics

Theorem

Common Bayes Estimators

Intermediate Statistics

Theorem

Minimax of Bayes Estimator

Intermediate Statistics

Theorem

Bayes Estimator with Constant Risk

Intermediate Statistics

Theorem

p-value

Intermediate Statistics

Definition

Likeliood Function

Intermediate Statistics

The minimax risk is defined as

$$R_n = \inf_{\hat{\theta}} \sup_{\theta} R\left(\theta, \hat{\theta}\right).$$

It is the risk of the estimator whose maximal risk is lowest among all competing estimators $\hat{\theta}$. It follows that an estimator $\hat{\theta}$ is minimax if

$$\sup_{\theta} R\left(\theta, \hat{\theta}\right) = \inf_{\hat{\theta}} \sup_{\theta} R\left(\theta, \hat{\theta}\right).$$

The risk of an estimator $\hat{\theta}$ is the expected value of the associated loss function, where the expectation is taken over all sample variables:

$$R\left(\theta, \hat{\theta}\right) = \mathbb{E}\left(L\left(\theta, \hat{\theta}\right)\right) = \int L\left(\theta, \hat{\theta}\left(x^n\right)\right) p(x^n; \theta) dx^n$$

Under squared error loss, the risk is equal to the mean squared error.

The posterior risk of an estimator $\hat{\theta}(x^n)$ is

$$r(\hat{\theta}|x^n) = \int L(\theta, \hat{\theta}(x^n))\pi(\hat{\theta}|x^n) d\theta$$

The Bayes risk of an estimator $\hat{\theta}$ with prior distribution π is

$$B_\pi(\hat{\theta}) = \int R\left(\theta, \hat{\theta}\right) \pi\left(\theta\right) d\theta.$$

Notice that the remaining uncertainty of the risk lies in different values for θ: The risk already has dealt with uncertainty in the data, as

$$R\left(\theta, \hat{\theta}\right) = \mathbb{E}_\theta\left(L(\theta, \hat{\theta})\right)$$

Estimators which minimize the Bayes risk are called *Bayes estimators*.

- Under squared error loss, the Bayes estimator is the poseterior mean $\mathbb{E}(\theta|X = x^n)$.
- Under absolute loss, the Bayes estimator is the posterior median $F^{-1}_{\theta|X}(\frac{1}{2})$.
- Under 0-1-loss, the Bayes estimator is the posterior mode of $\pi(\theta|x^n)$.

The Bayes risk $B_\pi(\hat{\theta})$ can also be expressed as

$$B_\pi(\hat{\theta}) = \int r(\hat{\theta}|x^n) m(x^n) dx^n,$$

where $m(x^n)$ is the marginal distribution of the data (sometimes called the *evidence*). An estimator $\hat{\theta}$ which minimizes the posterior risk is therefore a Bayes estimator since the integrand in $B_\pi(\hat{\theta})$ will be minimal at all x.

Let $\hat{\theta}$ be the Bayes estimator under some prior distribution π. If the risk is constant (with respect to θ) then this estimator is minimax. Proof: We have that $R\left(\theta, \hat{\theta}\right) = c$, where c is some constant. It follows that $B_\pi(\hat{\theta}) = \int r(\hat{\theta}|x^n) m(x^n) dx^n = c$ as well and hence $R\left(\theta, \hat{\theta}\right) \leq B_\pi(\hat{\theta})$ holds for all θ. By the "Minimax of Bayes Estimator" Theorem, this implies that the estimator is minimax.

Let $\hat{\theta}$ be the Bayes estimator under some prior π. If its risk is always smaller than the Bayes risk, i.e. if

$$R(\theta, \hat{\theta}) \leq B_\pi(\hat{\theta}) \quad \forall \theta,$$

then $\hat{\theta}$ is the minimax estimator and π is called a least favorable prior. Proof: By contradiction. Assume that $\hat{\theta}$ was not minimax. Then show that this would imply that the estimator did not minimize the Bayes risk in the first place (Hint: The average of a function is always less than or equal to its maximum).

Let $X^n = (X_1, \ldots, X_n)$ have joint density $p(x^n; \theta)$ where $\theta \in \Theta$. The likelihood function $L : \Theta \to [0, \infty)$ is the joint density regarded as a function of parameter θ, i.e.

$$L(\theta) = p(x^n; \theta).$$

The likelihood is not a pdf and defined only up to a constant of proportionality.

Suppose we have a test of the form: reject when $W(X^n) > c$. Then the p-value when $X^n = x^n$ is

$$p(x^n) = \sup_{\theta \in \Theta_0} P_\theta\left(W(X^n) \geq W(x^n)\right)$$

Theorem

Equivariance Property of MLE

Intermediate Statistics

Theorem

Mean Squared Error (MSE)

Intermediate Statistics

Theorem

Rao-Blackwell Theorem

Intermediate Statistics

Definition

Sufficiency

Intermediate Statistics

Theorem

Factorization Theorem

Intermediate Statistics

Definition

Minimal Sufficiency

Intermediate Statistics

Theorem

Find Minimal Sufficient Statistic

Intermediate Statistics

Definition

Empirical CDF

Intermediate Statistics

Definition

Kernel Density Estimator

Intermediate Statistics

The mean squared error (MSE) is

$$MSE = \mathbb{E}_\theta\left[(\hat{\theta}-\theta)^2\right] = \int (\hat{\theta}(x^n)-\theta)^2 p(x^n;\theta)dx^n.$$

The MSE can be decomposed into variance and bias squared, i.e.

$$MSE = \mathbb{V}_\theta(\hat{\theta}) + Bias^2,$$

where $Bias = \mathbb{E}_\theta(\hat{\theta}) - \theta$.

Let $\hat{\theta}$ be the MLE. If $\eta = g(\theta)$, then the MLE of η is $\hat{\eta} = g(\hat{\theta})$. Proof: Suppose g is invertible so $\eta = g(\theta)$ and $\theta = g^{-1}(\eta)$. Define $L^\star(\eta) = L(\theta)$ where $\theta = g^{-1}(\eta)$. Hence,

$$L^\star(\hat{\eta}) = L(\hat{\theta}) \geq L(\theta) = L^\star(\eta)$$

and thus $\hat{\eta}$ maximizes $L^\star(\eta)$. For non-invertible functions, this still holds if we define

$$L^\star(\eta) = \sup_{\theta:\tau(\theta)=\eta} L(\theta).$$

Suppose that we have a random sample $X_1,\ldots,X_n \sim p(x;\theta)$. An estimator T is sufficient for θ if the conditional distribution of $X_1,\ldots,X_n|T$ does not depend on θ. Thus $p(x_1,\ldots,x_n|t,\theta) = p(x_1,\ldots,x_n|t)$.

Let W be an unbiased estimator of $\tau(\theta)$ and let T be a sufficient statistic. Define $W' = \mathbb{E}(W|T)$. Then W' is unbiased with variance $\mathbb{V}_\theta(W') \leq \mathbb{V}_\theta(W) \quad \forall\theta$.

T is a minimal sufficient statistic for θ if it is sufficient and if it is a function of any other sufficient statistic U, i.e. $T = g(U)$ for some function g.

An estimator $T(X^n)$ is sufficient for θ if the joint pdf of X^n can be factored as

$$p(x^n|\theta) = h(x^n)g\left(T(x^n);\theta\right).$$

The empirical cumulative distribution function (ECDF) puts mass $\frac{1}{n}$ at each data point. It is defined as

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i \leq x).$$

Notice that $\hat{F}_n(x) \sim \text{Bernoulli}(F_X(x))$. We also have that

$$P\left(\sup_x |\hat{F}(x) - F(x)| > \varepsilon\right) \leq 2e^{-2n\varepsilon^2},$$

that is $\sup_x |\hat{F}(x) - F(x)| \xrightarrow{P} 0$.

An estimator T is minimal sufficient if and only if it has the following property:

$$T(y^n) = T(x^n) \leftrightarrow \frac{p(y^n;\theta)}{p(x^n;\theta)} \text{ does not depend on } \theta$$

The kernel density estimator is a non-parametric estimator of the density function. It is defined a

$$\hat{p}_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h}K\left(\frac{x-X_i}{h}\right),$$

where $h > 0$ is the bandwidth and K, the kernel, is a symmetric density with mean zero.