Statistics Lecture Notes (2024/2025)

Griffin Reimerink

Contents

1 Probability Theory recap

1.1 Law of Large Numbers

Definition *Convergence in probability*

A sequence of random variables $X_1, X_2, \dots$ **converges in probability** to c (notation: $X_n \xrightarrow[n\to\infty]{P} c$) if

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P(|X_n - c| > \varepsilon) = 0$$

Proposition *Law of Large Numbers*

Let $X \sim \mathcal{F}_\theta$ be a random sample. Then $\overline{X}_n \xrightarrow[n\to\infty]{P} E[X_1]$.
More generally, for any $k \in \mathbb{N}$:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow[n\to\infty]{P} E[X_1^k]$$

Proposition

Let $X_1, \dots, X_n$ be an i.i.d. random sample and g a measurable function.
Then $g(X_1), \dots, g(X_n)$ is also an i.i.d. random sample, which the Law of Large Numbers also applies to.

1.2 Expectation, variance, covariance

Definition *Expectation and (co)variance*

If X is continuous:	$E[X] = \int_S x \cdot f_\theta(x)\,\mathrm{d}x$	$V(X) = \int_S (x - E[X])^2 \cdot f_\theta(x)\,\mathrm{d}x$
If X is discrete:	$E[X] = \sum_S x \cdot f_\theta(x)$	$V(X) = \sum_S (x - E[X])^2 \cdot f_\theta(x)$
Covariance:	$\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$	

Rules for expectation and (co)variance

Consider two random variables X, Y and constant $c \in \mathbb{R}$:
- $E[X + Y] = E[X] + E[Y]$
- $E[cX] = c \cdot E[X]$
- $V(X + Y) = V(X) + Y(Y) + 2 \cdot \mathrm{Cov}(X,Y)$
- $V(cX) = c^2 \cdot V(X)$
- If X, Y independent, then $\mathrm{Cov}(X,Y) = 0$
- If X, Y independent, then $V(X + Y) = V(X) + V(Y)$
- $V(X) = E[X^2] - E[X]^2$

Let $X_1, \dots, X_n$ be a random sample:
- $E[\overline{X}_n] = E[X_1]$
- $V(\overline{X}_n) = \frac{1}{n} V(X_1)$

2 Statistics

2.1 Random samples

Definition *Random sample, statistic*

A **random sample** of size n is a collection of n i.i.d. random variables $X_1, \ldots, X_n$.
For a size n random sample from a distribution $\mathcal{F}_\theta$ we use the notation $X_1, \ldots, X_n \sim \mathcal{F}_\theta$

A **statistic** is an observable function T of a collection of random variables such that T does not depend on any unknown parameters.
Let $T : \mathbb{R}^n \to \mathbb{R}$ be a statistic. Then $T(X_1, \ldots, X_n)$ is a random variable, with density $f_\theta(t(x_1, \ldots, x_n))$

Definition *Sample mean*

Given random variables $X_1, \ldots, X_n$ with realizations $x_1, \ldots, x_n$ we define:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Definition *Conditional density*

$$f_\theta(x_1, \ldots, x_k | x_{k+1}, \ldots, x_n) = \frac{f_\theta(x_1, \ldots, x_n)}{f_\theta(x_{k+1}, \ldots, x_n)}$$

$$f_\theta(x_1, \ldots, x_n | t(x_1, \ldots, x_n)) = \frac{f_\theta(x_1, \ldots, x_n, t(x_1, \ldots, x_n))}{f_\theta(t(x_1, \ldots, x_n))}$$

From now on we will use X and x as shorthand notations for $X_1, \ldots, X_n$ and $x_1, \ldots, x_n$ respectively.

2.2 Sufficient statistics

Definition *Sufficient statistic*

A statistic T is called **sufficient** for θ if the conditional density of X given $T(X)$ does not depend on θ:

$$f_\theta(x|t(x)) = f(x|t(x))$$

Proposition *Factorization theorem*

Given a random sample $X \sim \mathcal{F}_\theta$, then
T is a sufficient statistic for θ if and only if the joint density $f_\theta(x)$ of X can be factorized into:

$$f_\theta(x) = g(t(x); \theta) \cdot h(x) \qquad \text{for all } x = (x_1, \ldots, x_n) \in S_x$$

Definition *Exponential family*

A distribution $\mathcal{F}_\theta$ with θ containing d parameters ($|\theta| = d$) belongs to the **exponential family**
if the density f_θ of $\mathcal{F}_\theta$ can be decomposed into:

$$f_\theta(x) = h(x) \cdot \exp\left\{\sum_{j=1}^{d} \eta_j(\theta) T_j(x) - A(\theta)\right\}$$

Proposition *Sufficient statistics of exponential families*

Let $X_1, \ldots, X_n \sim \mathcal{F}_\theta$ be a random sample from a distribution of an exponential family with d parameters.
Then the sufficient statistics for θ are:

$$\left(\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_d(X_i)\right)$$

3 Estimators

3.1 Estimators

MM (Method of Moments) estimator

Consider a distribution $\mathcal{F}_\theta$, where θ covers d unknown parameters ($|\theta| = d$) and a random sample from this distribution $X_1, \ldots, X_n \sim \mathcal{F}_\theta$.
We solve the following system of equations for θ to find the **MM estimator** for θ.

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} X_i &= E[X_1] \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 &= E[X_1^2] \\ \vdots \qquad & \qquad \vdots \\ \frac{1}{n}\sum_{i=1}^{n} X_i^n &= E[X_1^n] \end{aligned}$$

The estimator converges to the true parameters for large n. (Law of Large Numbers)

***Definition** Likelihood*

The **likelihood** function is defined as:

$$L : \Theta \to \mathbb{R}_0^+ \qquad L(\theta) := f_\theta(x_1, \ldots, x_n)$$

ML (Maximum Likelihood) estimator

Given a sample $X_1, \ldots, X_n \sim \mathcal{F}_\theta$, the **ML estimator** of $\theta \in \Theta$ is defined as:

$$\hat{\theta}_{ML} := \underset{\theta \in \Theta}{\operatorname{argmax}}\{L(\theta)\}$$

To compute this maximum we use derivatives. To make it easier to compute, we can use logarithms:

$$\hat{\theta}_{ML} := \underset{\theta \in \Theta}{\operatorname{argmax}}\{L(\theta)\} = \underset{\theta \in \Theta}{\operatorname{argmax}}\{\log(L(\theta))\}$$

***Definition** Bias*

The estimator $\hat{\theta}_n$ is an **unbiased estimator** if for all $n \in \mathbb{N} : E[\hat{\theta}_n] = \theta$.
The estimator $\hat{\theta}_n$ is an **asymptotically unbiased estimator** of θ if for $n \to \infty : E[\hat{\theta}_n] \to \theta$.
The **bias** of the estimator $\hat{\theta}_n$ is defined as:

$$B(\hat{\theta}_n) := E[\hat{\theta}_n] - \theta$$

The **Mean Squared Error** (MSE) of $\hat{\theta}_n$ is defined as:

$$\mathrm{MSE}(\hat{\theta}_n) := E[(\hat{\theta}_n - \theta)^2]$$

Proposition

$$\mathrm{MSE}(\hat{\theta}_n) = V(\hat{\theta}_n) + B(\hat{\theta}_n)^2$$

***Proposition** Cramér-Rao theorem*

Consider a random sample $X \sim \mathcal{F}_\theta$ of size n and an unbiased estimator $\hat{\theta} = \hat{\theta}(X)$ of θ.
Then (under certain regulatory conditions):

$$V(\hat{\theta}(X)) \geq \frac{1}{E\left[\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log(f_\theta(X_1, \ldots, X_n))\right)^2\right]}$$

Proposition *Cauchy-Schwarz inequality*

$$|\operatorname{Cov}(Y,Z)| \leq \sqrt{V(Y)\cdot V(Z)}$$

3.2 Rao-Blackwell theorem

Definition *Joint, marginal and conditional densities*

For two random variables X and Y with sample spaces S_X and S_Y we have:

- The **joint density** $f(x,y)$.
- The **marginal densities** $f(x) = \int_{S_Y} f(x,y)\,\mathrm{d}y$ and $f(y) = \int_{S_X} f(x,y)\,\mathrm{d}x$
- The **conditional densities** $f(x \mid y) = \frac{f(x,y)}{f(y)}$ and $f(y \mid x) = \frac{f(x,y)}{f(x)}$

If X and Y are (statistically) **independent**, we have

$$f(x,y) = f(x)\cdot f(y) \qquad f(x\mid y) = f(x) \quad f(y\mid x) = f(y)$$

Definition *Conditional expectation and variance*

Consider two random variables X and Y with sample spaces S_X and S_Y.
The **conditional expectation** of X given $Y=y$ (with $y \in S_Y$) is:

$$E[X \mid Y=y] = \int_{S_x} x\cdot f(x\mid y)\,\mathrm{d}x$$

The **conditional variance** of X given $Y=y$ (with $y \in S_Y$) is:

$$V(X\mid Y=y) = E[X^2 \mid Y=y] - E[X\mid Y=y]^2$$

Note: $E[X\mid Y]$ and $V(X\mid Y)$ are random variables.

Proposition

$$E[X] = E[E[X\mid Y]] \qquad V(X) = E[V(X\mid Y)] + V(E[X\mid Y])$$

Definition *Unbiased estimator of $g(\theta)$*

Let $X \sim \mathcal{F}_\theta$ and $g:\Theta \to \mathbb{R}$. The statistic $T(X)$ is called an **unbiased estimator** of $g(\theta)$ if

$$E[T(X)] = g(\theta) \quad \forall\theta\in\Theta$$

Proposition *Rao-Blackwell Theorem*

Consider a random sample $X \sim \mathcal{F}_\theta$ and a function $g:\Theta\to\mathbb{R}$. If:

1. The statistic $W = W(X)$ is unbiased estimator of $g(\theta)$.
2. The statistic $T = T(X)$ is sufficient for θ.

we can define a new estimator

$$\phi(T) := E[W \mid T]$$

with the properties:

1. $E[\phi(T)] = g(\theta)$, i.e. $\phi(T)$ is also an unbiased estimator of $g(\theta)$.
2. $V(\phi(T)) \leq V(W)$, i.e. the variance of $\phi(T)$ is potentially smaller than the variance of W

4 Statistical tests

4.1 Introduction

Definition *Hypothesis test*

Consider a random sample $X \sim \mathcal{F}_\theta$ with a **parameter space** Θ, which is the space of all possible parameters θ.
We consider a partition of Θ:

$$\Theta = \Theta_0 \cup \Theta_1 \qquad \Theta_0 \cap \Theta_1 = \varnothing$$

A **statistical hypothesis** H is a statement about θ:

$$\textbf{Null hypothesis } H_0 : \theta \in \Theta_0 \qquad \textbf{Alternative hypothesis } H_1 : \theta \in \Theta_1$$

The test decision is typically based on a **test statistic** $W = W(X)$ with $W : S_X \to \mathbb{R}$.
We define a partition of $\mathbb{R}$ into $\mathbb{R} = R \cup R^C$, where R is called the **rejection region**.
And then we define the **decision rule** $D : S_X \to \{H_0, H_1\}$:

$$D(x) = \begin{cases} H_0 \text{ if } W(x) \in R^C \\ H_1 \text{ if } W(x) \in R \end{cases}$$

A good statistical test should fulfill:

$$P_\theta(W(X) \in R \mid H_1) \text{ is close to } 1 \qquad P_\theta(W(X) \in R \mid H_0) \text{ is close to } 0$$

Definition *Power function, test level, p-value*

The **power function** of a statistical test is defined as:

$$\beta : \Theta \to [0, 1] \qquad \beta(\theta) = P_\theta(W(X) \in R)$$

$\beta(\theta)$ is the probability to decide for the alternative hypothesis H_1, given that $\theta \in \Theta$ is the true parameter.
$\beta(\theta)$ should be low for $\theta \in \Theta_0$ and high for $\theta \in \Theta_1$.
A statistical test is called a test to the **level** $\alpha \in [0, 1]$ if $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$
The **p-value** is the lowest test level α to which H_0 could have been rejected.

Definition *Test outcomes, type 1 and 2 error*

- If H_0 is true and $D(x) = H_0$, then we stay with the null hypothesis, which is a good decision.
- If H_0 is true and $D(x) = H_1$, then we incorrectly reject H_0, which is a **type 1 error**.
- If H_1 is true and $D(x) = H_0$, then we fail to reject H_0, which is a **type 2 error**.
- If H_1 is true and $D(x) = H_1$, then we reject the null hypothesis, which is the purpose of the test.

Type 1 errors are a lot more critical than type 2 errors. The probability of a type 1 error is bounded by α.
If H_0 is rejected, we call it a **significant test result**.

Definition *Quantiles*

Consider a random variable $X \sim \mathcal{F}_\theta$. The α-**quantile** q_α of the distribution F_θ is defined as:

$$P_\theta(X \leq q_\alpha) = \alpha \quad \text{or} \quad F_\theta(q_\alpha) = \alpha$$

where F_θ is the CDF of X. To determine q_α, one could compute the inverse CDF:

$$F_\theta(q_\alpha) = \alpha \iff q_\alpha = F_\theta^{-1}(\alpha)$$

4.2 Gaussian and Likelihood Ratio tests

Proposition *Properties of the Gaussian distribution*

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$.

$$a \cdot (X + b) \sim \mathcal{N}(a \cdot (\mu + b), a^2 \cdot \sigma^2) \qquad \frac{1}{\sigma} \cdot (X - \mu) \sim \mathcal{N}(0, 1)$$

Consider a random sample $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$.

$$\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu) \sim \mathcal{N}(0, 1)$$

Proposition

The sum of exponential distributed random variables is Gamma distributed.

Definition *Likelihood ratio test*

Consider a random sample $X \sim \mathcal{F}_\theta$ and test problem $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$.
The **likelihood ratio (LR) test statistic** is defined as:

$$\lambda(X) := \frac{\sup\limits_{\theta \in \Theta_0} L_x(\theta)}{\sup\limits_{\theta \in \Theta_0 \cup \Theta_1} L_x(\theta)} \quad \text{where } L_x(.) \text{ is the likelihood}$$

A **likelihood ratio test (LRT)** makes use of the LR test statistic, with the decision rule:

$$D_\lambda = \begin{cases} H_0 & \text{if } \lambda(X) > c \\ H_1 & \text{if } \lambda(X) \leq c \end{cases}$$

The test level α depends on the value of c.

4.3 UMP tests

Definition *Uniform most powerful (UMP) test*

Consider a random sample $X \sim \mathcal{F}_\theta$ and test problem $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$.
A test $D(X)$ is the **UMP test** if all other tests $\widetilde{D}(X)$ to the same level α have less power on Θ_1:

$$P_\theta(D(X) = H_1) \geq P_\theta(\widetilde{D}(X) = H_1) \quad \text{for all } \theta \in \Theta_1$$

Proposition *Neyman-Pearson lemma*

Consider a random sample $X \sim \mathcal{F}_\theta$ and the <u>simple</u> test problem $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$.
A test with the following test statistic, rejection region and decision rule:

$$W(X) = \frac{f_{\theta_0}(X)}{f_{\theta_1}(X)} \qquad R = \{x \in S_x : W(x) < k\} \qquad D(X) = \begin{cases} H_1 & \text{if } W(X) < k \\ H_0 & \text{if } W(X) \geq k \end{cases}$$

is the UMP test of level $\alpha := P_{\theta_0}(W(X) < k)$.
This lemma also holds with test statistic $W(T(X))$ instead of $W(X)$, where $T(X)$ is a sufficient statistic.
$W(X)$ is called the **density ratio** and $W(T(X))$ is called the **sufficient statistic density ratio**.

Definition *Monotone likelihood ratio*

Consider a random sample $X \sim \mathcal{F}_\theta$ and a sufficient statistic $T(X)$.
$T(X)$ has a **monotone likelihood ratio** if $W(t) := \dfrac{f_{T,\theta_0}(t)}{f_{T,\theta_1}(t)}$ is a monotone function of $t \in S_T$.

Proposition *Karlin-Rubin theorem*

Consider a random sample $X \sim \mathcal{F}_\theta$, a sufficient statistic $T(X)$ with a monotone likelihood ratio, and the composite test problem $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$

- If $T(X)$ has a monotonically decreasing likelihood ratio, then the test that rejects H_0 if $T > t_0$ is UMP of the level $\alpha = P_{\theta_0}(T(X) > t_0)$.
- If $T(X)$ has a monotonically increasing likelihood ratio, then the test that rejects H_0 if $T < t_0$ is UMP of the level $\alpha = P_{\theta_0}(T(X) < t_0)$.

4.4 Student's t-test

Degrees of freedom

Let $X_1, \dots, X_n$ be a random sample. Introducing an estimator (for example the arithmetic mean or empirical variance) leads to the loss of 1 **degree of freedom**, since 1 variable from the sample will depend on the others. This makes the effective sample size $n-1$ instead of n.

Definition *Chi-squared distribution*

Consider a sample from a standard Gaussian distribution $X_1, \dots, X_n \sim \mathcal{N}(0,1)$. Then the random variable:

$$S := \sum_{i=1}^{n} X_i^2 \sim \chi_n^2$$

is **Chi-Square distributed** with n degrees of freedom, with expectation n and variance $2n$.

Definition *t-distribution*

Consider a standard Gaussian distributed random variable $X \sim \mathcal{N}(0,1)$ and a Chi-square distributed random variable $S \sim \chi_n^2$ with n degrees of freedom. If X and S are statistically independent, then the random variable:

$$T = \frac{X}{\sqrt{\frac{1}{n}S}} \sim t_n$$

is **t-distributed** with n degrees of freedom, with expectation 0 and for $n > 2$, variance $\frac{n}{n-2}$.

For $n \to \infty$ we have: $t_n \xrightarrow{D} \mathcal{N}(0,1)$. For $n > 30$, $\mathcal{N}(0,1)$ is a good approximation for the t-distribution.

A Gaussian distribution turns into a t-distribution with $n-1$ degrees of freedom if we replace σ^2 with an estimator.

Definition *F-distribution*

Consider two Chi-square distributed random variables $S_1 \sim \chi_{n_1}^2$ and $S_2 \sim \chi_{n_1}^2$ with n_1 and n_2 degrees of freedom. If S1 and S2 are statistically independent, then the random variable:

$$F = \frac{\frac{1}{n_1} \cdot S_1}{\frac{1}{n_2} \cdot S_2} \sim F_{n_1,n_2}$$

is **F-distributed** with parameters n_1 and n_2.

4.5 Confidence intervals

Definition *Confidence interval*

Consider a random sample $X \sim \mathcal{F}_\theta$ with $\theta \in \Theta$. An interval $[L(X), U(X)]$ that contains the unknown parameter θ with probability $1-\alpha$ is called a $1-\alpha$ **confidence interval** for θ. We have:

$$P_\theta(L(X) \leq \theta \leq U(X) \geq 1-\alpha \quad \forall \theta \in \Theta \iff \inf_{\theta \in \Theta}\{P_\theta(L(X) \leq \theta \leq U(X)\} \geq 1-\alpha$$

$L(X)$ and $U(X)$ are statistics, and $1-\alpha$ is called the **confidence coefficient**.

Confidence intervals can be one-sided, i.e. $L(X) = -\infty$ or $U(X) = \infty$

Analytically, we can only compute exact confidence intervals for Gaussian distributions.

5 Asymptotic statistics

5.1 Consistent estimators

Definition *Convergence in probability*

A sequence of random variables $X_1, X_2, \dots$ **converges in probability** to c (notation: $X_n \xrightarrow[n\to\infty]{P} c$) if

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P(|X_n - c| > \varepsilon) = 0$$

Definition *Consistent estimator*

Consider a random sample $X \sim \mathcal{F}_\theta$ with $\theta \in \Theta$ and $n \in \mathbb{N}$. The estimator $\hat{\theta}_n$ is a **consistent estimator** if:

$$\forall \theta \in \Theta : \lim_{n\to\infty} \hat{\theta}_n \xrightarrow{P_\theta} \theta$$

Proposition

Consider a random sample $X \sim \mathcal{F}_\theta$ and an estimator $\hat{\theta}_n$ of θ.
The estimator $\hat{\theta}_n$ is consistent if $\lim_{n\to\infty} E[\hat{\theta}_n] = \theta$ (i.e. $\hat{\theta}_n$ is asymp. unbiased) and $\lim_{n\to\infty} V(\hat{\theta}_n) = 0$.

Proposition *Markov inequality*

For a single random variable $X \sim \mathcal{F}_\theta$ with sample space $S_X \subseteq \mathbb{R}_0^+$, we have for all $r > 0$:

$$P_\theta(X \geq r) \leq \frac{E[X]}{r}$$

and for any function $g : S_x \to \mathbb{R}_0^+$:

$$P_\theta(g(X) \geq r) \leq \frac{E[g(X)]}{r}$$

Proposition *Chebyshev inequality*

For any random variable Y (with $V(Y) \leq \infty$) and any $\varepsilon > 0$:

$$P(|Y - E[Y]| > \varepsilon) \leq \frac{V(Y)}{\varepsilon^2}$$

Proposition *Jensen's inequality*

Let $X \sim \mathbb{F}_\theta$ be a random variable on the possibly infinite interval (a, b) and let the function $g(.)$ be differentiable and convex on (a, b). If $E[X]$ and E $[g(X)]$ both exist, then:

$$E[g(X)] \geq g(E[X])$$

Proposition *Information inequality*

Let $X \sim \mathcal{F}_\theta$ be a random variable with $\theta \in \Theta$ and density $f_\theta(.)$. Moreover let θ_0 be the true parameter. Then:

$$E_{\theta_0}[\log(f_{\theta_0}(X))] \geq E_{\theta_0}[\log(f_\theta(X))]$$

Proposition *Consistency of the ML estimator*

Let $X \sim \mathcal{F}_\theta$ be a random sample with $\theta \in \Theta$, and let θ_0 be the true parameter. Under the following conditions:

- The sample space S_X does not depend on θ.
- θ_0 is an interior point of Θ.
- The log-likelihood $l_X(\theta)$ is differentiable in θ.
- θ_0 is the unique solution of $l'_X(\theta) = 0$

the Maximum Likelihood estimator is consistent for θ_0, i.e. it converges in probability to θ_0 for $n \to \infty$.

Definition *Convergence in distribution*

A sequence of random variables $X_1, X_2, \ldots$ **converges in distribution** to X (notation: $X_n \xrightarrow[n\to\infty]{D} X$), if for all $x \in \mathbb{R}$ at which the CDF F_X of X is continuous:

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

5.2 Convergence in distribution

Proposition *Central Limit Theorem*

Given a random sample $X \sim \mathcal{F}_\theta$ with expectation $E[X_1] = \mu$ and variance $V(X_1) = \sigma^2 < \infty$, then:

$$\sqrt{n} \cdot \frac{\overline{X}_n - \mu}{\sigma} \xrightarrow[n\to\infty]{D} \mathcal{N}(0,1)$$

Proposition

For a sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ we have:

$$X_n \xrightarrow[n\to\infty]{P} X \implies X_n \xrightarrow[n\to\infty]{D} X$$

Proposition *Continuous Mapping Theorem*

Given $\{X_n\}_{n\in\mathbb{N}}$ and and a continuous function g, we have:

$$X_n \xrightarrow[n\to\infty]{P} X \implies g(X_n) \xrightarrow[n\to\infty]{P} X \qquad X_n \xrightarrow[n\to\infty]{D} X \implies g(X_n) \xrightarrow[n\to\infty]{D} X$$

Proposition *Slutsky's theorem*

For two sequences of random variables $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ with $X_n \xrightarrow[n\to\infty]{D} X$ and $Y_n \xrightarrow[n\to\infty]{P} c$, we have:

1. $X_n + Y_n \xrightarrow[n\to\infty]{D} X + c$
2. $X_n \cdot Y_n \xrightarrow[n\to\infty]{D} c \cdot X$
3. $\frac{X_n}{Y_n} \xrightarrow[n\to\infty]{D} \frac{1}{c} \cdot X$ (if $c \neq 0$)

5.3 Asymptotic efficiency

Definition *Expected Fisher information for n=1*

Given a random sample $X \sim \mathcal{F}_\theta$, we define the **expected Fisher information** (of a sample of size $n = 1$) as

$$I(\theta) = E\left[\left(\frac{\mathrm{d}}{\mathrm{d}\theta} l_{X_1}(\theta)\right)^2\right] = -E\left[\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} l_{X_1}(\theta)\right] \qquad \text{where } l_{X_1} \text{ is the log-likelihood}$$

Definition *Asymptotically efficient estimator*

Given a random sample $X \sim \mathcal{F}_\theta$ with parameter space Θ, the estimator $\hat{\theta}$ of θ is an **efficient estimator** if:

$$\text{for all } \theta \in \Theta \qquad \sqrt{n} \cdot (\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{D} \mathcal{N}\left(0, I(\theta)^{-1}\right)$$

Proposition *Asymptotic efficiency of the ML estimator*

Given a random sample $X \sim \mathcal{F}_\theta$, the ML estimator is asymptotically efficient under the following conditions:

- The parameter space $\Theta \subseteq \mathbb{R}$ must be open.
- The density f_θ must be 3-times differentiable w.r.t. θ.
- The sample space S_X is not allowed to depend on θ.

Proposition *Asymptotic Likelihood Ratio test*

Consider a random sample $X \sim \mathcal{F}_\theta$ and the test problem $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$ where Θ_0, Θ_1 is a partition. Consider the Likelihood Ratio test statistic:

$$\lambda_n(X) := \lambda(X_1, \dots, X_n) = \frac{\sup_{\theta \in \Theta_0} \{L_{X_1,\dots,X_n}(\theta)\}}{\sup_{\theta \in \Theta_0 \cup \Theta_1} \{L_{X_1,\dots,X_n}(\theta)\}}$$

Then under the following conditions:

- $\Theta \subseteq \mathbb{R}$ must be an open set.
- The density f_θ must be 3-times differentiable w.r.t. θ.
- The sample space S_X is not allowed to depend on θ.

we have under H_0,

$$-2 \cdot \log(\lambda_n(X)) \xrightarrow[n \to \infty]{D} \chi_1^2$$

Asymptotic confidence interval

Using the asymptotic efficiency of the ML estimator, the $1 - \alpha$ **asymptotic confidence interval** for θ is:

$$\left[\hat{\theta}_{ML,n} - \frac{q_{1-\alpha/2}}{\sqrt{n \cdot I(\theta)}}, \hat{\theta}_{ML,n} + \frac{q_{1-\alpha/2}}{\sqrt{n \cdot I(\theta)}}\right]$$

$q_{1-\alpha/2}$ is a quantile of the Gaussian distribution and $I(\theta)$ is the Fisher information.
Since θ is unknown, we replace $I(\theta)$ by the observed Fisher information $I(\hat{\theta}_{ML,n})$.

Index