Statistics for Data Science

MSc Data Science WiSe 2019/20

Prof. Dr. Dirk Ostwald

Frequentist inference

(8) Asymptotic estimator properties

Bibliographic remarks

The material presented in this section is based on Wasserman (2004), Held and Sabanés Bové (2014), and Casella and Berger (2002, Section 10.1). The formulation and proof of the consistency of the maximum likelihood estimator is based on Held and Sabanés Bové (2014, Section 4.2.2) and the formulation and proof of the asymptotic normality of the maximum likelihood estimator is based on Held and Sabanés Bové (2014, Section 4.2.3).

Overview

- This lecture is a brief introduction to “Asymptotic Statistics”.
- Asymptotic Statistics concerns estimator behavior for large sample sizes.
- Asymptotic Statistics is used to
  - Study the qualitative properties of estimators and
  - Derive estimator property approximations for large sample sizes.
- In “Big Data” sample sizes are very large (or are they?).
- Asymptotic Statistics is thus practically useful and justified.
- Van der Vaart (2000) provides a comprehensive introduction.

Overview

Let $\hat{\theta}_n$ be an estimator for the parameter θ of a statistical model based on a random sample $X_1, ..., X_n \sim p_\theta$ of size n. Then $\hat{\theta}_n$ is called

- *asymptotically unbiased*, if for large sample sizes $n \to \infty$ the expected value of $\hat{\theta}_n$ corresponds to the true, but unknown, parameter value,
- *consistent*, if for large sample sizes $n \to \infty$ the probability that $\hat{\theta}_n$ deviates from the true, but unknown, value becomes small,
- *asymptotically normally distributed*, if for large sample sizes $n \to \infty$, the distribution of $\hat{\theta}_n$ is given by a normal distribution,
- *asymptotically efficient*, if for large sample sizes $n \to \infty$, the distribution of $\hat{\theta}_n$ is given by a normal distribution with expectation corresponding to the true, but unknown, parameter value and variance corresponding to the reciprocal of the expected Fisher information, i.e., the Cramér-Rao lower bound.

Maximum likelihood estimators are asymptotically unbiased, consistent, asymptotically normally distributed, and asymptotically efficient.

Asymptotic estimator properties

- Asymptotic unbiasedness
- Consistency
- Asymptotic normality
- Asymptotic efficiency
- Maximum likelihood estimator properties

Asymptotic estimator properties

- **Asymptotic unbiasedness**
- Consistency
- Asymptotic normality
- Asymptotic efficiency
- Maximum likelihood estimator properties

Definition (Asymptotic unbiasedness)

Let $\mathcal{P}$ denote a parametric statistical model with PDF p_θ for $\theta \in \Theta$, let $X_1, ..., X_n \sim p_\theta$, and let $\hat{\theta}_n$ be an estimator for θ. Then $\hat{\theta}_n$ is called *asymptotically unbiased*, if

$$\lim_{n \to \infty} \mathbb{E}_\theta(\hat{\theta}_n) = \theta \text{ for all } \theta \in \Theta. \tag{1}$$

Remarks

- Asymptotically unbiased estimators are unbiased in the large sample limit.
- An unbiased estimator is necessarily asymptotically unbiased.

Example (An asymptotically unbiased estimator)

Let $X_1, ..., X_n \sim N(\mu, \sigma^2)$ be n i.i.d. Gaussian random variables and let

$$\hat{\sigma}_n^2 := \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 \tag{2}$$

denote the maximum likelihood estimator of the variance parameter σ^2. From the derivation of the unbiasedness of the sample variance (cf. Lecture (7)), we have

$$\mathbb{E}\left(\hat{\sigma}_n^2\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2\right) = \frac{n-1}{n}\sigma^2 \neq \sigma^2. \tag{3}$$

$\hat{\sigma}_n^2$ is thus a biased estimator for σ^2. However, because $(n-1)/n \to 1$ for $n \to \infty$, we have

$$\lim_{n\to\infty}\mathbb{E}\left(\hat{\sigma}_n^2\right) = \lim_{n\to\infty}\frac{n-1}{n}\sigma^2 = \sigma^2. \tag{4}$$

In the limit of large samples, the maximum likelihood estimator for the variance parameter of a Gaussian distribution is thus unbiased.

Asymptotic estimator properties

- Asymptotic unbiasedness
- **Consistency**
- Asymptotic normality
- Asymptotic efficiency
- Maximum likelihood estimator properties

Definition (Consistency)

Let $\mathcal{P}$ denote a parametric statistical model with PDF p_θ for $\theta \in \Theta$, let $X_1, ..., X_n \sim p_\theta$, and let $\hat{\theta}_n$ be an estimator for θ. Then a sequence of estimators $\hat{\theta}_1, \hat{\theta}_2, ...$ is called a *consistent sequence of estimators*, if for every $\epsilon > 0$ and every $\theta \in \Theta$,

$$\lim_{n\to\infty} \mathbb{P}_\theta \left(|\hat{\theta}_n - \theta| < \epsilon \right) = 1 \Leftrightarrow \lim_{n\to\infty} \mathbb{P}_\theta \left(|\hat{\theta}_n - \theta| \geq \epsilon \right) = 0. \tag{5}$$

If $\hat{\theta}_1, \hat{\theta}_2, ...$ is a consistent sequence of estimators, then $\hat{\theta}_n$ is referred to as a *consistent estimator*.

Remarks

- As $n \to \infty$, the probability that $\hat{\theta}_n$ is arbitrarily close to θ becomes high.
- As $n \to \infty$, the probability that $\hat{\theta}_n$ deviates from θ becomes small.
- These properties hold for all possible true, but unknown, parameter values.
- The convergence type used here is *convergence in probability* (cf. Lecture (5)).
- Consistency can be shown directly or based on mean square error criteria.

Theorem (Mean squared error criterion for consistency)

Let $\mathcal{P}$ denote a parametric statistical model with PDF p_θ for $\theta \in \Theta$, let $X_1, ..., X_n \sim p_\theta$, and let $\hat{\theta}_n$ be an estimator for θ. If

$$\lim_{n\to\infty} \text{MSE}(\hat{\theta}_n) = \lim_{n\to\infty} \mathbb{E}_\theta\left((\hat{\theta}_n - \theta)^2\right) = 0, \tag{6}$$

then $\hat{\theta}_n$ is a consistent estimator.

Proof
From Chebychev's inequality (cf. Lecture (5)), it follows directly that

$$\mathbb{P}_\theta\left(|\hat{\theta}_n - \theta| \geq \epsilon\right) \leq \frac{\mathbb{E}_\theta\left((\hat{\theta}_n - \theta)^2\right)}{\epsilon^2}$$

Taking limits then completes the proof. □

Theorem (Bias and variance criterion for consistency)

Let $\mathcal{P}$ denote a parametric statistical model with PDF p_θ for $\theta \in \Theta$, let $X_1, ..., X_n \sim p_\theta$, and let $\hat{\theta}_n$ be an estimator for θ. If

$$\lim_{n\to\infty} \mathrm{B}(\hat{\theta}_n) = 0 \text{ and } \lim_{n\to\infty} \mathbb{V}_\theta(\hat{\theta}_n) = 0 \tag{7}$$

then $\hat{\theta}_n$ is a consistent estimator.

<u>Proof</u>

- If for $n \to \infty$ it holds that $\mathrm{B}(\hat{\theta}_n) \to 0$, then also $\mathrm{B}(\hat{\theta}_n)^2 \to 0$.
- If for $n \to \infty$ both $\mathrm{B}(\hat{\theta}_n)^2 \to 0$ and $\mathbb{V}_\theta(\hat{\theta}_n) \to 0$, then also $\mathrm{MSE}(\hat{\theta}_n) \to 0$.
- Hence $\mathbb{P}_\theta\left(|\hat{\theta}_n - \theta| \geq \epsilon\right) \to 0$ for $n \to \infty$.

$\square$

Example (Consistency of the sample mean)

With the bias and variance criterion for consistency, the consistency of the sample mean for $X_1, ..., X_n \sim N(\mu, \sigma^2)$ follows directly from

$$\mathrm{B}(\bar{X}_n) = 0 \text{ and } \lim_{n\to\infty} \mathbb{V}_\theta(\bar{X}_n) = \lim_{n\to\infty} \frac{1}{n}\sigma^2 = 0. \tag{8}$$

Asymptotic estimator properties

- Asymptotic unbiasedness
- Consistency
- **Asymptotic normality**
- Asymptotic efficiency
- Maximum likelihood estimator properties

Definition (Asymptotic normality)

Let $\mathcal{P}$ denote a parametric statistical model with PDF p_θ, let $X_1, ..., X_n \sim p_\theta$, and let $\hat{\theta}_n$ be an estimator for θ. Let further $\tilde{\theta} \sim N(\mu, \sigma^2)$ be a normally distributed random variable with expectation parameter μ and variance parameter σ^2. Then, if $\hat{\theta}_n$ converges to $\tilde{\theta}$ in distribution, $\hat{\theta}_n$ is said to be *asymptotically normally distributed* and we write

$$\hat{\theta}_n \overset{a}{\sim} N(\mu, \sigma^2). \tag{9}$$

Remark

- Convergence in distribution means that $\lim_{n\to\infty} P_n(\hat{\theta}_n) = P(\tilde{\theta})$ (cf. Lecture (5)).

Asymptotic estimator properties

- Asymptotic unbiasedness
- Consistency
- Asymptotic normality
- **Asymptotic efficiency**
- Maximum likelihood estimator properties

Definition (Asymptotic efficiency)

Let $\mathcal{P}$ denote a parametric statistical model with PDF p_θ, $X_1, ..., X_n \sim p_\theta$ be random sample, $\hat{\theta}_n$ be an estimator for θ, and let $J_n(\theta)$ denote the expected Fisher information of the random sample $X_1, ..., X_n$. If

$$\hat{\theta}_n \overset{a}{\sim} N\left(\theta, J(\theta)_n^{-1}\right), \tag{10}$$

then $\hat{\theta}_n$ is said to be *asymptotically efficient*.

Remarks

- Because the asymptotic distribution is a normal distribution, asymptotic efficiency implies asymptotic normality.
- Because the expectation of the asymptotic distribution is the true, but unknown, parameter value θ, asymptotic efficiency implies asymptotic unbiasedness.
- The variance of the asymptotic distribution is called the *asymptotic variance*.
- The asymptotic variance of an asymptotically efficient estimator attains the Cramér-Rao bound.
- The term *efficiency* is used with variations in the literature.

Asymptotic estimator properties

- Asymptotic unbiasedness
- Consistency
- Asymptotic normality
- Asymptotic efficiency
- **Maximum likelihood estimator properties**

Maximum likelihood estimators are

(1) not necessarily unbiased,

(2) consistent,

(3) asymptotically normally distributed,

(4) asymptotically unbiased, and

(5) asymptotically efficient.

Example (Non-necessary unbiasedness of MLEs)

Let $X_1, X_2, ..., X_n \sim N(\mu, \sigma^2)$ and for $\mu_{ML} := n^{-1}\sum_{i=1}^n X_i$ consider the maximum likelihood estimator

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^n (x_i - \mu_{ML})^2 \tag{11}$$

of the variance parameter σ^2. Recall that

$$\mathbb{E}\left(\sum_{i=1}^n (X_i - \bar{X}_n)^2\right) = (n-1)\sigma^2. \tag{12}$$

Hence

$$\mathbb{E}(\hat{\sigma}^2_{ML}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n (x_i - \mu_{ML})^2\right) = \frac{(n-1)}{n}\sigma^2 \neq \sigma^2, \tag{13}$$

and $\hat{\sigma}^2_{ML}$ is not unbiased.

Remark

- The *restricted maximum likelihood* framework results in unbiased variance parameter estimates for Gaussian models.

Theorem (Consistency of maximum likelihood estimators)

Let $\mathcal{P}$ denote a parametric statistical model with PDF p_θ for $\theta \in \Theta$, let $X_1, ..., X_n \sim p_\theta$, and assume that the Fisher regularity conditions hold. Let $\hat{\theta}_n^{ML}$ denote the maximum likelihood estimator for the true, but unknown, value θ and let $\hat{\theta}_n^{ML}$ be defined here as a local maximizer of the likelihood function L_n based on a sample of size n. Then there exists a consistent sequence of maximum likelihood estimators $\hat{\theta}_1, \hat{\theta}_2, ...$ and $\hat{\theta}_n^{ML}$ is said to be a consistent estimator of θ.

Remarks

- $\hat{\theta}_n^{ML}$ converges in probability to the true, but unknown value, θ.
- Formally, $\lim_{n\to\infty} \mathbb{P}_\theta \left(|\hat{\theta}_n^{ML} - \theta| < \epsilon\right) = 1$ for every $\epsilon > 0$ and $\theta \in \Theta$.
- There exists proofs of varying mathematical generality and depth.
- Put bluntly, the theorem states that for large sample sizes maximum likelihood estimates equal the true, but unknown, parameter value with certainty.

Proof (Consistency of maximum likelihood estimators)

The idea of the proof provided by Held and Sabanés Bové (2014, Section 4.2.2) is to show that there exists a (possibly local) maximum of the likelihood function L_n in an interval $]\theta-\epsilon, \theta+\epsilon[$ around the true, but unknown, parameter value θ for arbitrary small ϵ with probability 1 as $n \to \infty$. Because the likelihood function is assumed to be continuous and $\hat{\theta}_n^{ML}$ is understood as a local maximizer of the likelihood function, this entails that $|\hat{\theta}_n^{ML} - \theta| < \epsilon$ with probability 1 as $n \to \infty$, which in turn corresponds to the convergence in probability of $\hat{\theta}_n^{ML}$ to the true, but unknown, parameter value θ.

To show that there exists a maximum of the likelihood function L_n in $]\theta - \epsilon, \theta + \epsilon[$ with probability 1 as $n \to \infty$ it suffices (with the continuity of the likelihood function) to show that

$$\lim_{n\to\infty} \mathbb{P}\left(L_n(\theta) > L_n(\theta - \epsilon)\right) = 1 \text{ and } \lim_{n\to\infty} \mathbb{P}\left(L_n(\theta) > L_n(\theta + \epsilon)\right) = 1 \tag{14}$$

for arbitrary small $\epsilon > 0$. Because analogous arguments can be made to show either of these statements, we consider only the first. The aim is thus to show that

$$\lim_{n\to\infty} \mathbb{P}\left(L_n(\theta) > L_n(\theta - \epsilon)\right) = 1, \tag{15}$$

which can be argued with the weak law of large numbers and the positivity of the Kullback-Leibler divergence between two non-identical PDFs with equal support.

Proof (Consistency of maximum likelihood estimators)

To set up this line of reasoning, we first note that the event of interest

$$L_n(\theta) > L_n(\theta - \epsilon) \tag{16}$$

can be reformulated as

$$\begin{aligned}
L_n(\theta) > L_n(\theta - \epsilon) &\Leftrightarrow \frac{1}{n} \ln L_n(\theta) > \frac{1}{n} \ln L_n(\theta - \epsilon) \\
&\Leftrightarrow \frac{1}{n} \ln L_n(\theta) - \frac{1}{n} \ln L_n(\theta - \epsilon) > 0 \\
&\Leftrightarrow \frac{1}{n} \sum_{i=1}^{n} \ln p_\theta(x_i) - \frac{1}{n} \sum_{i=1}^{n} \ln p_{\theta-\epsilon}(x_i) > 0 \\
&\Leftrightarrow \frac{1}{n} \sum_{i=1}^{n} \ln \left(\frac{p_\theta(x_i)}{p_{\theta-\epsilon}(x_i)} \right) > 0
\end{aligned} \tag{17}$$

Recall that the weak law of large numbers states that the sample mean of n independent and identically distributed random variables $X_i, i = 1, ..., n$ converges in probability to the expected value of X_i. Applying this theorem to the random variable $\ln(p_\theta(x_i)/p_{\theta-\epsilon}(x_i))$ thus entails that

$$\frac{1}{n} \sum_{i=1}^{n} \ln \left(\frac{p_\theta(x_i)}{p_{\theta-\epsilon}(x_i)} \right) \xrightarrow[n\to\infty]{P} \mathbb{E}_\theta \left(\ln \left(\frac{p_\theta(x)}{p_{\theta-\epsilon}(x)} \right) \right) \tag{18}$$

Proof (Consistency of maximum likelihood estimators)

Finally, the right-hand side of the above corresponds to the Kullback-Leilber divergence between $p_\theta(x)$ and $p_{\theta-\epsilon}(x)$ (which, with the Fisher regularity conditions, are distinct PDFs with equal support), such that with the non-negativity of the Kullback-Leibler divergence, it follows that

$$\mathbb{E}_\theta\left(\ln\left(\frac{p_\theta(x)}{p_{\theta-\epsilon}(x)}\right)\right) > 0 \tag{19}$$

We have thus shown that $\frac{1}{n}\sum_{i=1}^n \ln\left(\frac{p_\theta(x_i)}{p_{\theta-\epsilon}(x_i)}\right)$ converges in probability to a value larger than 0, which corresponds to a limiting probability of 1 for $\frac{1}{n}\sum_{i=1}^n \ln\left(\frac{p_\theta(x_i)}{p_{\theta-\epsilon}(x_i)}\right)$ to be larger than zero.

In summary, we thus have

$$\lim_{n\to\infty}\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \ln\left(\frac{p_\theta(x_i)}{p_{\theta-\epsilon}(x_i)}\right) > 0\right) = 1 \Leftrightarrow \lim_{n\to\infty}\mathbb{P}\left(L_n(\theta) > L_n(\theta-\epsilon)\right) = 1. \tag{20}$$

as required for the proof's main argument. □

Theorem (Asymptotic efficiency of maximum likelihood estimators)

Let $\mathcal{P}$ be a parametric statistical model with PDF p_θ, let $X_1, X_2, ..., X_n \sim p_\theta$ be a random sample, and let $\hat{\theta}_n^{ML}$ denote the maximum likelihood estimator for the true, but unknown, parameter θ. Let $J_n(\theta)$ denote the expected Fisher information of the random sample $X_1, X_2, ..., X_n$ and assume that the Fisher regularity conditions hold. Then $\hat{\theta}_n^{ML}$ is asymptotically efficient, i.e.,

$$\hat{\theta}_n^{ML} \overset{a}{\sim} N\left(\theta, J_n(\theta)^{-1}\right). \tag{21}$$

Remarks

- The asymptotic efficiency of $\hat{\theta}_n^{ML}$ implies its asymptotic normality and unbiasedness.

Proof (Asymptotic efficiency of maximum likelihood estimators)

The idea of the proof provided by Held and Sabanés Bové (2014, Section 4.2.3) is to show that

$$\sqrt{J_n(\theta)}\left(\hat{\theta}_n^{ML}\right) \overset{a}{\sim} N(0,1) \tag{22}$$

and then, informally, concluding that $\hat{\theta}_n^{ML} \overset{a}{\sim} N\left(\theta, J_n(\theta)^{-1}\right)$ by means of the transformation of normal random variables under linear-affine functions (cf. Lecture (4)).

The proof rests on

- the *continuous mapping theorem*,
- *Slutsky's theorem*
- $\frac{1}{\sqrt{n}} S_n(\theta) \xrightarrow[n\to\infty]{\text{D}} N\left(0, J(\theta)\right)$,
- $\frac{1}{n} I_n(\theta) \xrightarrow[n\to\infty]{\text{P}} J(\theta)$.

We first first introduce these building blocks individually and then consider their joint application to show the central result.

Proof (Asymptotic efficiency of maximum likelihood estimators)

Continuous mapping theorem

The continuous mapping theorem (Mann and Wald, 1943) states that convergence in probability and convergence in distribution are preserved under the application of a continuous univariate real-valued function f, i.e.,

- if $X_n \xrightarrow[n\to\infty]{\mathrm{P}} X$, then also $f(X_n) \xrightarrow[n\to\infty]{\mathrm{P}} f(X)$, and
- if $X_n \xrightarrow[n\to\infty]{\mathrm{D}} X$, then also $f(X_n) \xrightarrow[n\to\infty]{\mathrm{D}} f(X)$.

Slutsky's theorem

Slutsky's theorem (Slutsky, 1926) states that for

$$X_n \xrightarrow[n\to\infty]{\mathrm{D}} X \text{ and } Y_n \xrightarrow[n\to\infty]{\mathrm{P}} a \in \mathbb{R} \tag{23}$$

it holds that

$$X_n + Y_n \xrightarrow[n\to\infty]{\mathrm{D}} X + a \tag{24}$$

and

$$X_n Y_n \xrightarrow[n\to\infty]{\mathrm{D}} aX \tag{25}$$

Proof (Asymptotic efficiency of maximum likelihood estimators)

Proof of $\frac{1}{\sqrt{n}} S_n(\theta) \xrightarrow[n\to\infty]{D} N\left(0, J(\theta)\right)$

With the properties of the score function (cf. Lecture (7)), we first note that

$$\frac{1}{\sqrt{n}} S_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S(\theta), \quad \mathbb{E}(S) = 0, \quad \mathbb{V}(S) = J(\theta). \tag{26}$$

We next note that for n i.i.d. random variables $Y_1, ..., Y_n$ with expectation $\mu := \mathbb{E}(Y_i)$ and variance $\sigma^2 := \mathbb{V}(Y_i)$, the central limit theorem in the Lindenberg and Lévy form states that (cf. Lecture (5))

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{\mathrm{D}} N(n\mu, \sigma^2). \tag{27}$$

Substitution thus yields

$$\frac{1}{\sqrt{n}} S_n(\theta) \xrightarrow[n\to\infty]{\mathrm{D}} N(0, J(\theta)). \tag{28}$$

Proof (Asymptotic efficiency of maximum likelihood estimators)

Proof of $\frac{1}{n} I_n(\theta) \xrightarrow[n\to\infty]{P} J(\theta)$

With the properties of the Fisher information (cf. Lecture (7)), we first note that

$$\frac{1}{n} I_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} I(\theta), \quad \mathbb{E}(I(\theta)) = J(\theta). \tag{29}$$

We next note that for n i.i.d. random variables $Y_1, ..., Y_n$ with expectation $\mu := \mathbb{E}(Y_i)$ the weak law of large numbers states that (cf. Lecture (5))

$$\frac{1}{n} \sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{\mathrm{P}} \mu. \tag{30}$$

Substitution thus yields

$$\frac{1}{n} I_n(\theta) \xrightarrow[n\to\infty]{\mathrm{P}} J(\theta). \tag{31}$$

Proof (Asymptotic efficiency of maximum likelihood estimators)

With these building blocks in place, we are now in the position to show

$$\sqrt{J_n(\theta)}\left(\hat{\theta}_n^{ML}\right) \overset{a}{\sim} N(0,1). \tag{32}$$

To this end, we first consider a first-order Taylor approximation of the score function with expansion point corresponding to the true, but unknown, parameter value,

$$S_n(\tilde{\theta}) \approx S_n(\theta) - I_n(\theta)(\tilde{\theta} - \theta). \tag{33}$$

For $\tilde{\theta} = \hat{\theta}_n^{ML}$, we then have

$$\begin{aligned} S_n\left(\hat{\theta}_n^{ML}\right) &\approx S_n(\theta) - I_n(\theta)\left(\hat{\theta}_n^{ML} - \theta\right) \\ \Leftrightarrow 0 &\approx S_n(\theta) - I_n(\theta)\left(\hat{\theta}_n^{ML} - \theta\right) \\ \Leftrightarrow \sqrt{n}S_n(\theta) &\approx \sqrt{n}I_n(\theta)\left(\hat{\theta}_n^{ML} - \theta\right) \\ \Leftrightarrow \sqrt{n}\left(\hat{\theta}_n^{ML} - \theta\right) &\approx \sqrt{n}\frac{S_n(\theta)}{I_n(\theta)} \end{aligned} \tag{34}$$

and with $n^{1/2} = n^1 n^{-1/2}$ obtain

$$\sqrt{n}\left(\hat{\theta}_n^{ML} - \theta\right) \approx \left(\frac{I_n(\theta)}{n}\right)^{-1}\frac{S_n(\theta)}{\sqrt{n}} \tag{35}$$

Proof (Asymptotic efficiency of maximum likelihood estimators)

With the results on the convergences of the two product terms and by application of Slutsky's theorem, it then follows that

$$\sqrt{n}\left(\hat{\theta}_n^{ML} - \theta\right) \xrightarrow[n\to\infty]{\text{D}} J(\theta)^{-1} N(0, J(\theta)). \tag{36}$$

and with the transformation of normal random variables under linear-affine functions (cf. Lecture (4)), it follows that

$$\sqrt{n}\left(\hat{\theta}_n^{ML} - \theta\right) \xrightarrow[n\to\infty]{\text{D}} N\left(0, J(\theta)^{-1}\right). \tag{37}$$

□

Summary

Let $\hat{\theta}_n$ be an estimator for the parameter θ of a statistical model based on a random sample $X_1, ..., X_n \sim p_\theta$ of size n. Then $\hat{\theta}_n$ is called

- *asymptotically unbiased*, if for large sample sizes $n \to \infty$ the expected value of $\hat{\theta}_n$ corresponds to the true, but unknown, parameter value,
- *consistent*, if for large sample sizes $n \to \infty$ the probability that $\hat{\theta}_n$ deviates from the true, but unknown, value becomes small,
- *asymptotically normally distributed*, if for large sample sizes $n \to \infty$, the distribution of $\hat{\theta}_n$ is given by a normal distribution,
- *asymptotically efficient*, if for large sample sizes $n \to \infty$, the distribution of $\hat{\theta}_n$ is given by a normal distribution with expectation corresponding to the true, but unknown, parameter value and variance corresponding to the reciprocal of the expected Fisher information, i.e., the Cramér-Rao lower bound.

Maximum likelihood estimators are asymptotically unbiased, consistent, asymptotically normally distributed, and asymptotically efficient.

References

Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

Held, L. and Sabanés Bové, D. (2014). *Applied statistical inference*, volume 10. Springer.

Mann, H. B. and Wald, A. (1943). On stochastic limit and order relationships. *The Annals of Mathematical Statistics*, 14(3):217–226.

Slutsky, E. (1926). Über stochastische asymptoten und grenzwerte (abdruck). *Tohoku Mathematical Journal, First Series*, 27:67–70.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer.