4 Point estimation and interval estimation of parameters and parametric functions

Let us consider a random sample $X_1, \dots, X_n \sim L(\vartheta)$, whose parameter ϑ is unknown and unobservable. However information about this constant is "hidden" and due to random influences "covered up" in the random sample. Point and interval estimation is aimed to "discover" this parameter.
A point estimator $T = T(X_1, \dots, X_n)$ is to serve as a "best guess" for an unknown parameter ϑ (which is called the estimand). It is a function of the observable sample data, thus it is a statistic. An estimate is the result from the actual application of the function to a particular sample of data. This estimate substitute the unknown parameter and we hope it is sufficiently close to this parameter.

Point estimation should be contrasted with interval estimation, which is the use of sample data to calculate an interval of possible (or probable) values of an unknown parameter ϑ. The most prevalent form of interval estimation is confidence interval (D, H), which is constructed with regard to in advance given confidence level. This level states how likely the interval is to cover the true parameter. The confidence limits $D = D(X_1, \dots, X_n)$, $H = H(X_1, \dots, X_n)$ are statistics, their numerical realizations depend on a particular sample of data.

Definition 4.1
Let $X_1, \dots, X_n$ be a random sample from a distribution $L(\vartheta)$. The set of all values, which the parameter ϑ gains, is called parametric space and is denoted as Θ. Arbitrary function $h(\vartheta)$ is called parametric function.

Definition 4.2
Let $X_1, \dots, X_n$ be a random sample from a distribution$L(\vartheta)$, let $h(\vartheta)$ be a parametric function and let $T, T_1, T_2, \dots$ be statistics.

(i.) Statistics T is said to be *unbiased estimator* of parametric function $h(\vartheta)$, if and only if
$\forall \vartheta \in \Theta: \; E(T) = h(\vartheta)$
[An unbiased estimator does not cause systematically overvalued or undervalued estimates of the parameter or the parametric function.]

(ii.) Let T_1, T_2 be two unbiased estimators of the identical parametric function $h(\vartheta)$. The estimator T_1 is said to be *better*, than estimator T_2 if and only if
$\forall \vartheta \in \Theta: \; D(T_1) < D(T_2)$

(iii.) A sequence of estimators $T_1, T_2, \dots, T_n, \dots$ is a sequence of *asymptotically unbiased* estimators of parametric function $h(\vartheta)$ if and only if
$\forall \vartheta \in \Theta: \; \lim_{n\to\infty} E(T_n) = h(\vartheta)$
[If $E(T) \neq h(\vartheta)$ then the estimates are distorted and the estimator is said to be biased. If distortion decrease as n increase statistic T is an asymptotically unbiased estimator of parameter or parametric function.]

(iv.) A sequence of estimators $T_1, T_2, \dots, T_n, \dots$ is a sequence of *consistent* estimators of parametric function $h(\vartheta)$ if and only if converges in probability toward $h(\vartheta)$, i.e.
$\forall \vartheta \in \Theta, \; \forall \varepsilon > 0: \; \lim_{n\to\infty} P(|T_n - h(\vartheta)| < \varepsilon) = 1$
[Increasing the sample size increases the probability of the estimator being close to the parameter.]

Previous definition listed desirable properties of point estimators of which consistency is most useful.

Corollary 4.3
If T_n is an unbiased estimator then it is asymptotically unbiased as well. If in addition $\lim_{n\to\infty} D(T_n) = 0$ then asymptotical unbiasness imply consistency. Thus:
If $\lim_{n\to\infty} E(T_n) = h(\vartheta) \wedge \lim_{n\to\infty} D(T_n) = 0$, then T_n is a consistent estimator of parametric function $h(\vartheta)$.

The statistics $M, S^2, S_{12}, R_{12}, \ldots$ introduced in 3rd section will be now explored whether they have some of mentioned desirable properties. Ve 3. kapitole jsem si zavedli statistiky $M, S^2, S_{12}, R_{12}, \ldots$. Nyní posoudíme, jestli tyto statistiky mají některé ze zmíněných žádoucích vlastností.

Theorem 4.4
Let $X_1, \ldots, X_n$ be a random sample from a distribution with expected value μ, variance σ^2 and distribution function $F(x)$. Let M_n is a sample mean, S_n^2 is a sample variance and $F_n(x)$ is a value of sample distribution function in a point x. Then:

1. – M_n is an unbiased estimator of a parameter μ. [Thus $\forall \mu \in \mathbf{R} : E(M_n) = \mu$]
   – S_n^2 is an unbiased estimator of a parameter σ^2. [Thus $\forall \sigma \geq 0 : E(S_n^2) = \sigma^2$]
   – $F_n(x)$ is an unbiased estimator of $F(x)$ for any given $x \in \mathbf{R}$.
   [Thus $\forall x \in \mathbf{R} : E(F_n(x)) = F(x)$]
2. – $M_1, \ldots, M_n, \ldots$ is a sequence of consistent estimators of the parameter μ.
   – $S_1^2, \ldots, S_n^2, \ldots$ is a sequence of consistent estimators of the parameter σ^2.
   – $F_1(x), \ldots, F_n(x), \ldots$ is a sequence of consistent estimators of the parameter $F(x)$ for any given $x \in \mathbf{R}$.

Remark 4.5
Sample standard deviation S is not !! unbiased estimator of the parameter σ with only exception - when S follows degenerate distribution, thus it is certain to take the constant value.
[If S would be unbiased estimator of parameter σ, then $E(S) = \sigma$. Hence
$D(S) = E(S^2) - (E(S))^2 = \sigma^2 - \sigma^2 = 0$. Zero variance implies degenerate distribution.]

Theorem 4.6
Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from two-dimensional distribution with covariance σ_{12}. Then sample covariance S_{12} is an unbiased estimator of parameter σ_{12}.
[Thus $\forall \sigma_{12} \in \mathbf{R} : E(S_{12}) = \sigma_{12}$]

So far we have attended to point estimators and their properties. Now let us turn to interval estimators. Interval estimates may be contrasted with point estimates and have the advantage over these as they convey more information not just a "best estimate" of a parameter but an indication of the precision with which the parameter is between confidence limits.

Definition 4.7
Let $X_1, \ldots, X_n$ be a random vector from distribution $L(\vartheta)$, $h(\vartheta)$ be parametric function, number $\alpha \in (0,1)$ and $D = D(X_1, \ldots, X_n)$, $H = H(X_1, \ldots, X_n)$ be statistics.

(i.) Interval (D, H) is called $100(1-\alpha)\%$ (two-sided) confidence interval for $h(\vartheta)$, if
$\forall \vartheta \in \Theta : \; P(D < h(\vartheta) < H) \geq 1 - \alpha$
[It is almost certain that the confidence interval contains parameter$h(\vartheta)$.]

(ii.) Interval (D, ∞) is called $100(1-\alpha)\%$ (left-sided) confidence interval for $h(\vartheta)$, if
$\forall \vartheta \in \Theta : \; P(D < h(\vartheta)) \geq 1 - \alpha$

(iii.) Interval $(-\infty, H)$ is called $100(1-\alpha)\%$ (right-sided) confidence interval for $h(\vartheta)$, if
$\forall \vartheta \in \Theta: \ P(h(\vartheta) < H) \geq 1-\alpha$

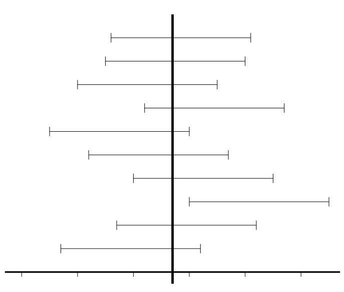
(iv.) The number α is called significance level. [The convention is that α is a small number close to zero; the most common α is 0,05; 0,01; 0,1.] The number $1-\alpha$ (sometimes reported as a percentage $100\%(1-\alpha)$) is called the confidence level. Statistic D is called lower bound (lower limit), statistic H is called upper bound (upper limit).

Remark 4.8
The choice of two-sided, left-sided or right-sided confidence interval depends on particular situation. For instance two-sided interval may be of use for constructor who is interested in lower and upper bound for the true size μ of particular product. Doing purchase of precious metals we would be interested in lower bound for true gold content μ in bought ingot. Right-sided confidence interval is of use for chemist, who needs to know the upper bound for content of foreign matters μ in analyzed sample.

Poznámka 4.9 Process of confidence interval derivation

1. Find out a statistic V which is an unbiased point estimator of parametric function $h(\vartheta)$.
2. Find out a pivotal quantity (statistic) W which is a monotone function of both point estimator V and parametric function $h(\vartheta)$, further its distribution is known and does not depend on unknown parametric function $h(\vartheta)$.
   Then find its quantiles $w_{\alpha/2}$ and $w_{1-\alpha/2}$ such that:
   $\forall \vartheta \in \Theta: \ P(w_{\alpha/2} < W < w_{1-\alpha/2}) \geq 1-\alpha$.
3. Rearrange equivalently an inequality $w_{\alpha/2} < W < w_{1-\alpha/2}$ toward an inequality $D < h(\vartheta) < H$.
4. Statistics D and H replace with their numerical realization d and h and thus obtain $100(1-\alpha)\%$ empirical confidence interval for $h(\vartheta)$. This interval covers unknown parametric function $h(\vartheta)$ with probability at least $1-\alpha$.

Remark 4.10
Imagine 100 mutually independent random samples from distribution with expected value μ and for each of them derive corresponding 95% empirical confidence interval for parameter μ. If we randomly choose one interval, the probability is 95% we end up having chosen an interval that contains the parameter; however we may be unlucky and have picked the wrong one. We will never know. (A number of right intervals is approximately 95, a number of wrong intervals is approximately 5.)



Process of confidence interval derivation is shown in following example.

Example 4.11
Let $X_1, \ldots, X_n$ be a random sample from normal distribution $N(\mu, \sigma^2)$, where $n \geq 2$ and numeric value of parameter σ is known. Set a $100(1-\ \alpha)\%$ confidence interval on μ.

Solution

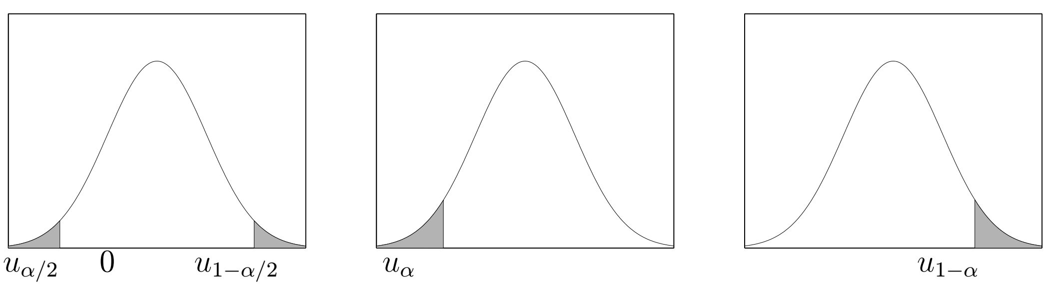
1. Statistic V is an unbiased poin estimator for parameter μ.
   $V = M = \frac{1}{n}\sum\limits_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$.
   $[E(M) = \mu,\ D(M) = \frac{\sigma^2}{n}$ and linear combination of normal random sample keeps normality.]

2. Pivotal quantity $W = U$ is convenient to our task;
   $W = U = \frac{M-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
   In this case quantiles $w_{\alpha/2}$, $w_{1-\alpha/2}$ are standard normal, thus $u_{\alpha/2} = -u_{1-\alpha/2}$ and $u_{1-\alpha/2}$. Hence
   $\forall \vartheta \in \Theta : 1-\alpha \leq P(u_{\alpha/2} < U < u_{1-\alpha/2})$.

3. Stated inequality will be rearranged equivalently so that estimand μ will be isolated between lower and upper limit statistic.
   $$1-\alpha \leq P(u_{\alpha/2} < U < u_{1-\alpha/2}) = P(\underbrace{M - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}}_{D} < \mu < \underbrace{M + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}}_{H}).$$

4. After observing the sample we can find values m for M and n, from which we calculate the empirical confidence interval with fixed numbers d and h as endpoints.

In case of left-sided or right-sided confidence interval the significance level α does not to be halved, it is concentrated on one tail of distribution instead. Thus in previous example
the left-sided confidence interval would be expressed as $(D, \infty) = (M - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha}, \infty)$
the right-sided confidence interval would be expressed as $(-\infty, H) = (-\infty, M + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha})$.



Example 4.12
Consider a random sample of size 10 from a distribution $N(\mu; 0,04)$. Based on the data set the realization of sample mean m was calculated, $m = 2,06$. Set a 95% empirical confidence interval on μ : a) two-sided, b) left-sided, c) right-sided.

Solution
$\sigma = 0,02;\ n = 10;\ m = 2,06;\ \alpha = 0,05;\ u_{1-\alpha/2} = u_{0,975} = 1,96;\ u_{1-\alpha} = u_{0,95} = 1,64$
<u>ad a)</u>
$d = m - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2} = 2,06 - \frac{0,2}{\sqrt{10}} \cdot 1,96 = 1,94$
$h = m + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2} = 2,06 + \frac{0,2}{\sqrt{10}} \cdot 1,96 = 2,18$
$P(1,94 < \mu < 2,18) \geq 0,95$
Thus $\mu \in (1,94; 2,18)$ with probability at least 0,95.

<u>ad b)</u>

$d = m - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha} = 2,06 - \frac{0,2}{\sqrt{10}} \cdot 1,64 = 1,96$
$P(1,96 < \mu) \geq 0,95$
Thus $\mu > 1,96$ with probability at least 0,95.

ad c)
$h = m + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha} = 2,06 + \frac{0,2}{\sqrt{10}} \cdot 1,64 = 2,16$
$P(\mu < 2,16) \geq 0,95$
Thus $\mu < 2,16$ with probability at least 0,95.

Remark 4.13
Let (d,h) be a $100(1-\alpha)\%$ empirical confidence interval for parametric function $h(\vartheta)$. Let us denote $\Delta = h - d$. A number Δ is called the width of confidence interval.

a) As the significance level α remains constant the width of confidence interval Δ decreases with increasing size of sample n.

b) As the size of sample n remains constant the width of confidence interval Δ decreases with increasing significance level α

Example 4.14
Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$, where σ^2 is known. Find the minimum sample size such that the width of $100(1-\alpha)\%$ empirical confidence interval on parameter μ does not exceed the number δ.

Solution
We require that the confidence interval width $\Delta \leq \delta$. Thus
$\delta \geq \Delta = h - d = m + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2} - (m - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}) = \frac{2\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}$

$\sqrt{n} \geq \frac{2\sigma}{\delta} \cdot u_{1-\alpha/2}$

$n \geq \frac{4\sigma^2 u^2_{1-\alpha/2}}{\delta^2}$
We choose the minimum natural number which satisfies the last inequality to be the sample size.

Example 4.15
In an example 4.12 a) an user is not satisfied with the with of 95% confidence interval on μ (1,94; 2,18). He would appreciate the width which would not exceed the umber 0.16, further he do not want increase the significance level α. What would you suggest him?

Solution
To constrict the width of interval we have to change the sample size n.
$\delta = 0,16,\ n = ?,\ \sigma = 0,2,\ u_{0,975} = 1,96$
$n \geq \frac{4\sigma^2 u^2_{1-\alpha/2}}{\delta^2} = \frac{4 \cdot 0,04 \cdot 1,96^2}{0,16^2} = 24,01$

The sample size $n = 25$ meets the requirements of the user.