Statistics II: Exercise Session 4

1 Theory of estimation

1.1 Basic concepts

True parameter of the population	ϑ
Estimator	$\widehat{\theta} = g(X_1, \ldots, X_n)$
Estimated value	$\widehat{\vartheta} = g(x_1, \ldots, x_n)$

Unbiased estimator

$$\mathrm{E}[\widehat{\theta}] = \mathrm{E}[g(X_1, \ldots, X_n)] = \vartheta$$

Unbiasedness means, that given a large number of samples the average over all estimations lies near the true parameter.

Mean squared error (MSE)

$$\mathrm{MSE}(\widehat{\theta}) = \mathrm{E}[(\widehat{\theta} - \vartheta)^2] = \underbrace{\mathrm{E}[(\widehat{\theta} - \mathrm{E}[\widehat{\theta}])^2]}_{=\mathrm{Var}(\widehat{\theta})} + \underbrace{(\mathrm{E}[\widehat{\theta}] - \vartheta)^2}_{\mathrm{Bias}^2}$$

Efficient estimator

Assume unbiased estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$.

- Estimator $\widehat{\theta}_1$ is **relatively efficient** compared to $\widehat{\theta}_2$, if $\mathrm{Var}(\widehat{\theta}_1) \leq \mathrm{Var}(\widehat{\theta}_2)$.
- Estimator $\widehat{\theta}_1$ is **absolutely efficient** for parameter ϑ, if it has the smallest variance among all unbiased estimators of ϑ.

1.2 Estimation methods

Maximum likelihood method

Likelihood function	$L(\vartheta) = L(\vartheta \mid x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i \mid \vartheta)$	$\to$ maximize
Log-likelihood function	$\log\{L(\vartheta)\} = \sum_{i=1}^{n} \log\{f(x_i \mid \vartheta)\}$	$\to$ maximize

Least squares method

Quadratic form $\quad Q(\vartheta) = \sum_{i=1}^{n} (x_i - \mathrm{E}[X_i])^2 = \sum_{i=1}^{n} (x_i - g_i(\vartheta))^2 \quad \to$ minimize

Exercises

7-1: *Unbiasedness*

A population has the mean μ and the variance σ^2. Let (X_1, X_2, X_3) be a simple (theoretical) random sample from this population. The following three estimators are given:

$$\begin{aligned}
\widehat{\theta}_1 &= \frac{1}{3}(X_1 + X_2 + X_3); \\
\widehat{\theta}_2 &= \frac{1}{4}(2X_1 + 2X_3); \\
\widehat{\theta}_3 &= \frac{1}{3}(2X_1 + X_3).
\end{aligned}$$

a) Which of these estimators are unbiased?

b) Which of them would you prefer according to the efficiency criterion? (Reason!)

7-3: *Lamps*

A supply of $N = 1000$ lamps is examined by means of a simple random sample of size $n = 20$. With help of the random variable X: "Number of defective lamps in the random sample of size $n = 20$" the number d of defective lamps in the supply is estimated.

a) Determine an unbiased estimator $\widehat{\theta} = f(X)$ for d and show that $\mathrm{E}(\widehat{\theta}) = d$ holds.

b) In a given sample the number of defective lamps is 3. How many defective lamps would you estimate to be in the supply altogether?

7-5: *Gambling machine*

A gambling machine has the following probability distribution for a win X per game (in EUR):

x	-1	0	+1
$\mathrm{P}(X = x)$	p	p	$1 - 2p$

The producer of these gambling machines asked a statistician to perform an estimate of p in order to know whether p is changing during the usage of the machines.

a) The statistician draws a random sample of size $n = 6$, i.e. he plays the machines exactly 6 times and writes down the wins. The sample $(X_1,X_2,X_3,X_4,X_5,X_6)$ had the following realization: $(-1, 1, -1, 0, 1, 1)$. Verbalize this sample result.

b) State the following probabilities: $\mathrm{P}(X = 0)$; $\mathrm{P}(X = -1)$; $\mathrm{P}(X = 1)$.

c) How would you determine the probabilities for the win X per game, given the sample above, if you did not have any information about the probability distribution of X?

d) What is the probability $\mathrm{P}\{(X_1,X_2,X_3,X_4,X_5,X_6) = (-1, 1, -1, 0, 1, 1)\}$ given the probability distribution above?

e) Determine the maximum likelihood estimator for p in this problem.

f) Estimate p by maximum likelihood method using the given sample result.

g) Estimate p by least squares method using the given sample result.