Repetition and important formulas from modul 7

This module was all about point estimators. The overall assumption is that we have observations $x_1, \ldots, x_n$ that come from an independent sample, i.e. the underlying random variables $X_1, \ldots, X_n$ are iid (independent and identically distributed). We also say $x_1, \ldots, x_n$ is a sample from X in this case.

→ We often know or have an idea about the distribution of the random variables (e.g. is it discrete or continuous, symmetric or asymmetric, some prior information, ...?).

→ Assuming a distribution, we want to estimate the parameters of it (e.g. μ and σ for the normal or n and p for the binomial distribution).

→ AIM: Find a function that gives an estimation of the parameters from data.

1 Examples of point estimators

Normal distribution Consider a sample $x_1, \ldots, x_n$ from $X \sim \mathcal{N}(\mu, \sigma^2)$. Point estimation for μ:

$$\hat{\mu} = \bar{x} = \frac{1}{n}(x_1 + \cdots + x_n).$$

Point estimation for σ^2:

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Binomial distribution Consider an observation x from $X \sim \text{Bin}(n, p)$ (e.g. number of successes in n experiments) where n is know but p unknown. Point estimation for p:

$$\hat{p} = \frac{x}{n}.$$

In general, consider a distribution family $F(x; \theta)$ (think of F as the normal or binomial distribution for example) with unknown parameter θ (e.g. μ or p). Let $x_1, \ldots, x_n$ be observations from independent random variables $X_i \sim F$. A point estimation (punktskattning) $t = \hat{\theta}$ of θ is a function g of the sample, i.e. $\hat{\theta} = t = g(x_1, \ldots, x_n)$.

We can consider this estimation again as a random variable $T = g(X_1, \ldots, X_n)$, i.e. by inserting the random variables into the statistic instead of the observations. This new random variable is called point estimator and has statistical properties like $\mathrm{E}[T]$ or $\mathrm{V}[T]$.
Example: In the case of estimating μ in the normal distribution, $\theta = \mu$ and the function of the estimation is

$$t = g(x_1, \ldots, x_n) = \frac{1}{n}(x_1 + \cdots + x_n).$$

(A note on the terminology: with concrete values inserted this is then called an *estimate* of the parameter μ.) If we instead insert the random variables into this function,

$$T = g(X_1, \ldots, X_n) = \frac{1}{n}(X_1 + \cdots + X_n),$$

then T is a random variable, the *estimator*, and we can for example calculate its expected value

$$\mathrm{E}[T] = \mathrm{E}\left[\frac{1}{n}(X_1 + \cdots + X_n)\right] = \frac{1}{n}(\mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]) = \frac{1}{n}n\mu = \mu.$$

Observe that the expected value of the estimator is exactly what we wanted to estimate with it. This means on average the estimator gives us the exact parameter value. This is a nice and often desirable property of an estimator, called unbiasedness.

2 Properties of estimators

Unbiasedness (väntevärdesriktighet) An estimator is *unbiased* if

$$\mathrm{E}[T] = \theta \quad (\mathrm{E}[\hat{\theta}] = \theta).$$

Consistency (konsistens) An unbiased point estimator is *consistent* if

$$\mathrm{V}[T] = \mathrm{V}[g(X_1, \ldots, X_n)] \to 0, \quad \text{as } n \to \infty.$$

This means, the estimator gets more reliable the more observations are used for the estimation.

Efficiency (effektivitet) For two unbiased estimators T_1 and T_2, we say T_1 is more efficient than T_2 if

$$\mathrm{V}[T_1] < \mathrm{V}[T_2].$$

This is a relative measure to compare the two estimators.

Note that often there are many unbiased estimators available, see e.g. example 6.2 in JR. However, of course not all combinations give an unbiased estimator. Choose for example $\mu_3 = \frac{1}{5}(X_1 + 2X_2 + 3X_3)$ as a third estimator in example 6.2 in JR. Do you see what has to hold for estimators of the mean as in 6.2 to make it an unbiased estimator? (Look at the coefficients and the scaling factor ...)

3 Standard error (medelfel)

Another measure of a random variable is the variance that tells us how big the spread around the mean can be. So we can also ask here: How far away from the true value can our estimation lie? So let's determine the variance of the estimators:

$$\mathrm{V}[\hat{\mu}] = \mathrm{V}\left[\frac{1}{n}(X_1 + \cdots + X_n)\right] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$
$$\mathrm{V}[\hat{p}] = \mathrm{V}\left[\frac{X}{n}\right] = \frac{1}{n^2}\mathrm{V}[X] = \frac{1}{n^2} np(1-p) = \frac{1}{n} p(1-p).$$

Hence, the standard deviation is

$$D[\hat{\mu}] = \frac{\sigma}{\sqrt{n}} \quad \text{and} \quad D[\hat{p}] = \sqrt{\frac{1}{n} p(1-p)}.$$

As we see, the red parameters are unknown, it's the one that we want and need to estimate. So we also need to put the estimates in here and by doing so, we obtain the so-called *standard error (medelfel)*:

$$d[\hat{\mu}] = d[\bar{X}] = \frac{s}{\sqrt{n}} \quad \text{and} \quad d[\hat{p}] = \sqrt{\frac{1}{n} \hat{p}(1-\hat{p})}.$$