4 Mathematical Statistics

4.1 Estimation of parameters

4.1.1 Probability distributions of several random variables

Assume that the vector $(x_1, ..., x_n)$ originates from a sample of identically distributed data with the mean μ and variance σ^2.

Let $p(x_j)$ be the probability density of an individual variable x_j. Let $p_X(x_1, ..., x_n)$ be the joint probability density of all variables $(x_1, ..., x_n)$. If all variables in the data sample are independent,

$$p(x_1, ..., x_n) = p(x_1)...p(x_n)$$

and all pairs of variables are not correlated, such that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_i - \mu)(x_j - \mu)p(x_i)p(x_j)dx_i dx_j = 0.$$

Consider the scalar random variable

$$X = \frac{x_1 + ... + x_n}{n}$$

1. The mean value of X is μ
2. The variance of X is σ^2/n
3. The characteristic function of X is

$$\phi_X(s) = \phi^n\left(\frac{s}{n}\right)$$

Example: Normal distribution

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad p(X) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma}e^{-\frac{n}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

4.1.2 Central Limit Theorem

Let $(x_1, ..., x_n)$ be independent, identically distributed random variables with mean μ and variance σ^2. The scalar random variable

$$X = \frac{x_1 + ... + x_n}{n}$$

is normally distributed in the limit $n \to \infty$ with mean μ and variance σ^2/n.

4.1.3 Recipe # 11: How to construct probability density for a sum of independent random variables

1. Compute characteristic functions from probability densities of each random variable.
2. The characteristic function of the sum is the product of characteristic functions of each variable in the sum.
3. Compute the probability density of the sum of random variables from the characteristic function.

Examples:

- Probability distribution of a sum of two independent, normally distributed random variables

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}, \quad p(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}$$

- Probability distribution of a sum of two independent, Poisson distributed random variables

$$p_\infty(k) = \frac{\mu_x^k}{k!} e^{-\mu_x}, \qquad p_\infty(m) = \frac{\mu_y^m}{m!} e^{-\mu_y}$$

Main problem of statistics: Determine parameters of an object from parameters of a sample of finitely many data, which are extracted in finitely many, independent trials (experiments).

4.1.4 Recipe # 12: Point estimates for data sample

1. Given a data sample $(x_1, ..., x_n)$, compute the sample mean

$$\bar{x} = \frac{x_1 + ... + x_n}{n} = \frac{1}{n}\sum_{j=1}^{n} x_j$$

2. Compute the sample variance

$$s^2 = \frac{(x_1 - \bar{x})^2 + ... + (x_n - \bar{x})^2}{n-1} = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2$$

3. Approximate the unknown mean μ and variance σ^2 by the known sample mean $\bar{x}$ and sample variance s^2.

Example: A sample of 5 cans was measured against the weight. The result is $(203, 194, 210, 201, 196)$ in grams. Each can is supposed to be of 200 grams. Use point estimates to estimate if the standard weight of 200 grams fits to the main sample confidence interval $(\bar{x} - s, \bar{x} + s)$.

4.1.5 Recipe # 13: Maximum likelihood method for point estimates

1. Given a probability density $p(x_j; \lambda)$ for independent random variables $(x_1, ..., x_n)$ with a parameter λ, construct the joint probability density

$$p(x_1, ..., x_n; \lambda) = p(x_1; \lambda)...p(x_n; \lambda)$$

2. Simplify the natural logarithm of $p(x_1, ..., x_n; \lambda)$

3. Compute λ as a critical point of logarithm of $p(x_1, ..., x_n; \lambda)$

4. Check that the critical value of λ gives maximum.