6. Statistical mean and moments of probability distributions

Statistical mean

Statistical mean $E[X]$ of a random variable X:

$$\text{Discrete } X\text{: } E[X] = \sum_{i=1}^{n} x_i \, P(X = x_i)\,; \quad \text{continuous } X\text{: } E[X] = \int_{-\infty}^{\infty} x \, f_X(x) dx.$$

Properties: $E[a] = a, \;$ for $a = \text{const.}$,

$E[aX + bY] = aE[X] + bE[Y]$,

$E[XY] = E[X]E[Y], \;$ for independent X and Y,

$E[Y] = E[g(X)], \;$ for $Y = g(X)$.

Moments of probability distributions

Raw moments of random variable X:

$$\text{Discrete } X\text{: } m_k = E[X^k] = \sum_{i=1}^{n} x_i^k \, P(X = x_i)\,; \quad \text{continuous } X\text{: } m_k = E[X^k] = \int_{-\infty}^{\infty} x^k \, f_X(x) dx.$$

Central moments of random variable X:

$$\text{Discrete } X\text{: } \mu_k = E[(X - E[X])^k] = \sum_{i=1}^{n} (x_i - E[X])^k \, P(X = x_i),$$

$$\text{Continuous } X\text{: } \mu_k = E[(X - E[X])^k] = \int_{-\infty}^{\infty} (x - E[X])^k \, f_X(x) dx.$$

The first raw moment, m_1, is called **mean**, m or *expected value*, while the second central moment, μ_2, is called **variance**, $\text{Var}(X) = \sigma_X^2$. To calculate the variance, usually the following formula is used:

$$\text{Var}(X) = E[X^2] - (E[X])^2.$$

Joint raw and central moments of a vector random variable with two components:

$$E\left[X^j Y^k\right] = \iint x^j y^k f_{XY}(x, y) \; dx \, dy,$$

$$E\left[(X - E[X])^j \, (Y - E[Y])^k\right] = \iint (x - E[X])^j \, (y - E[Y])^k f_{XY}(x, y) \; dx \, dy.$$

Mostly, the first joint raw moment named **correlation** $R_{XY} = E[XY]$ and the first joint central moment named **covariance** $\text{Cov}[X, Y] = E[(X - m_X)(Y - m_Y)]$ are used. The two moments are related by:

$$\text{Cov}[X, Y] \; = \; R_{XY} - E[X]E[Y].$$

For independent X and Y: $R_{XY} = E[X]E[Y]$ and $\text{Cov}[X, Y] = 0$.

6. Statistical mean and moments of probability distributions - problems

1. A player throws the dice. If the result is an even number, the player gets the same amount of money units. If the result is an odd number, the player pays the same amount of money units. Calculate the number of money units the player can expect to have after 100 throws. R: 50
2. A player throws two dices and gets a number of money units equal to the product of both dice results. How many money units can he expect to have after 100 throws? R: 1225
3. A worker tends to N^2 machines. Due to small failures he must intervene now at this and then on another machine. Since the machines are identical it is assumed that the probability of failure is the same for all machines. Machines are placed in m rows by n machines so that the distance between the two machines in the adjacent column or adjacent row is equal to a. The worker reaches the machine with failure by the shortest route, walking around the (square-footprint) machines. In how many rows should the machines be placed to ensure the shortest average route for the worker? R: N rows by N machines, $E[s] = 2a(N^2 - 1)/(3N)$
4. Determine the mean and variance of a binomial random variable. R: $E[X] = np$, $\mathrm{Var}[X] = np(1-p)$
5. Determine the mean and variance of a standard normal random variable. R: $E[X] = 0$, $\mathrm{Var}[X] = 1$

6. Statistical mean and moments of probability distributions – additional problems

1. The telephone office monitors length of calls. 40 % of calls last a minute, 30 % of calls last two minutes, 20 % of calls last three minutes and 10 % of calls last four minutes. What are the mean and variance of the call length? R: $E[X] = 2\ \mathrm{min}$, $\mathrm{Var}[X] = 1\ \mathrm{min}^2$
2. Determine the mean and variance of number of dots in throwing dice.
R: $E[X] = 7/2$, $\mathrm{Var}[X] = 35/12$
3. Determine the mean and variance of a Poisson random variable. R: $E[X] = \lambda$, $\mathrm{Var}[X] = \lambda$
4. Determine the mean and variance of a uniformly distributed random variable in the interval $[a, b]$. R: $E[X] = (a+b)/2$, $\mathrm{Var}[X] = (a-b)^2/12$
5. Determine the mean and variance of an exponentially distributed random variable.
R: $E[X] = 1/\lambda$, $\mathrm{Var}[X] = 1/\lambda^2$

7. Basic concepts of engineering statistics, point estimation

Basic concepts of engineering statistics

Population, X, is a set of all possible outcomes, with a probability distribution $f_X(x)$.

Sample, $\boldsymbol{V} = (X_1, X_2, \dots, X_n)$, is a subset of n measurements from the population X. It is a vector random variable. If $X_1, X_2, \dots, X_n$ are independent, then $\boldsymbol{V}$ is a **random sample**.

Statistics or **sample characteristics**, $Z_n = Z(\boldsymbol{V})$, is any scalar function of vector variable $\boldsymbol{V}$.

Point estimation

Point estimator, $\hat{q}_n = Z_n = Z(\boldsymbol{V})$, is statistics Z, used for estimating a parameter q of distribution $f_X(x)$ of population X.

Estimator $\hat{q}_n$ is **consistent**, if: $\lim_{n\to\infty} P[|\hat{q}_n - q| < \varepsilon] = 1$ for arbitrary small ε.
Estimator $\hat{q}_n$ is **unbiased**, if: $E[\hat{q}_n] = q$.
Estimator $\hat{q}_n$ is **asymptotically unbiased**, if: $\lim_{n\to\infty} E[\hat{q}_n] = \lim_{n\to\infty} \left(q + O(1/n)\right) = q$.

Important point estimators

Sample mean	$\langle X\rangle_n = \frac{1}{n}\sum_{i=1}^{n} X_i$	consistent, unbiased
Sample variance	$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \langle X\rangle_n)^2$	consistent, asymptotically unbiased
Corrected sample variance	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \langle X\rangle_n)^2$	consistent, unbiased
Sample raw moments	$m_{k,n} = \langle X^k\rangle_n = \frac{1}{n}\sum_{i=1}^{n} X_i^k$	consistent, unbiased
Sample central moments	$\mu_{k,n} = \langle (X_i - \langle X\rangle_n)^k\rangle_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \langle X\rangle_n)^k$	consistent, biased
Sample relative frequency	$p_n(A) = \frac{n_A}{n}$	consistent, unbiased

Method of moments: parameters of distribution are expressed by distribution moments. So many of the lowest moments are used as many parameters have to be estimated. The estimators are then obtained by replacing the moments of distribution with the corresponding sample moments.

Method of maximal likelihood: a *likelihood function*, corresponding to probability of getting the sample in the volume $d\boldsymbol{v}$ around $\boldsymbol{v}$, is constructed: $L(\boldsymbol{v}; q) = f_X(x_1; q) \cdot f_X(x_2; q) \cdot \ldots \cdot f_X(x_n; q)$. Here, $\boldsymbol{v}$ is a sample and q a parameter of a probability distribution $f_X(x; q)$ which we estimate. The estimator $\hat{q}$ is determined by maximizing the value of the likelihood function:

$$\frac{\partial L(\boldsymbol{v}; q)}{\partial q} = 0 \quad \text{or, equivalently,} \quad \frac{\partial\left(\ln\left(L(\boldsymbol{v}; q)\right)\right)}{\partial q} = 0 \to \text{the solution for } q \text{ is } \hat{q}.$$

The assumption of the method is: the sample $\boldsymbol{v}$ was selected because it is the most probable.

7. Basic concepts of engineering statistics, point estimation - problems

1. Synthetic fibers used to make carpets have normally distributed tensile strength with an average of 520 kPa and standard deviation of 25 kPa.
    a. What is the probability that the average tensile strength of a random sample of six fibers is greater than 522 kPa? R: P=0.422
    b. What is the probability of the first problem case if the sample size is increased from six to 50 fibers? R: P=0.286
2. Using the method of moments find the estimator of the parameter λ of the exponential probability distribution displaced for x_0: $f(x) = \lambda \exp(-\lambda(x - x_0))$. R: $\hat{\lambda} = 1/(\langle X \rangle - x_0)$
3. Using the method of maximum likelihood find the estimator of the parameter λ of the Poisson probability distribution. R: $\hat{\lambda} = \langle X \rangle$
4. Using the method of maximum likelihood find the estimator of the parameter q of the probability distribution with a probability density function $f(x) = (q+1)x^q$ for $0 \le x \le 1$. R: $\hat{q} = -1 - n/\sum_{i=1}^{n} \ln x_i$
5. Flow time of certain product has been measured in a workshop for ten selected pieces. The resulting values are (in minutes): 17, 21, 14, 23, 20, 24, 19, 19, 25 and 18. It is assumed that the flow time of the studied product is normally distributed. Determine point estimators of mean and standard deviation of the flow time. R: $\hat{\mu} = 20$ min, $\hat{\sigma} = 3.37$ min

7. Basic concepts of engineering statistics, point estimation – additional problems

1. Noise of compressors is normally distributed with an average of 34 dB and a standard deviation of 2 dB.
    a. What is the probability that the average noise of a sample of five compressors is less than 33 dB? R: P = 0.132
    b. What is the probability of the first part of the problem, if the sample size is increased from five to 20 compressors? R: P = 0.013
2. Using the method of moments find the estimator of parameter δ of the Rayleigh probability distribution $f(x) = (2x/\delta^2)\exp(-x^2/\delta^2), x > 0$. R: $\hat{\delta} = 2\langle X \rangle/\sqrt{\pi}$
3. Using the method of maximum likelihood find the estimator of proportion p of the binomial probability distribution. R: $\hat{p} = \langle X \rangle/n$
4. The roughness (R_a) of a turned surface has been measured at nine locations of a shaft. The following values were obtained (in µm): 7, 12, 11, 5, 12, 9, 7, 8 and 10. Assume that the surface roughness is normally distributed. Point estimate the mean and the standard deviation of roughness. R: $\hat{\mu} = 9$ µm, $\hat{\sigma} = 2.45$ µm