Formulae used for exercises in Random phenomena course - 5th set

Basics of mathematical statistics
From the *population* S a *sample* of n objects is selected. On the sample objects the random variable X is measured and the measurements are gathered in a random vector variable $\mathbf{V} = (X_1, X_2, ..., X_n)$. *Statistic* is any function of the observation $\mathbf{V}$: $Z(\mathbf{V})$. In general, any statistic is also a random variable.
Sample mean of a sample with n elements is:

$$\langle X \rangle_n = \frac{1}{n} \sum_{i=1}^{n} X_i \,. \tag{1}$$

Sample mean is unbiased and consistent estimator of the population mean m:

$$\mathrm{E}\left[\langle X \rangle_n\right] = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left[X_i\right] = \frac{1}{n} nm = m \,, \qquad P\left(|\langle X \rangle_n - m| \geq \varepsilon\right) \leq \frac{\sigma_x^2}{n\varepsilon^2} \underset{n\to\infty}{\longrightarrow} 0 \,, \text{ for } \varepsilon > 0. \tag{2}$$

If the random variable X is normally distributed with probability density $\mathcal{N}\left(X; m, \sigma\right)$, then the sample mean is also normally distributed with probability density $\mathcal{N}\left(\langle X \rangle_n; m, \sigma/\sqrt{n}\right)$. If X is not normally distributed, then the probability distribution of the sample mean approaches the normal distribution with increasing size n of the sample:

$$\lim_{n\to\infty} f_{\langle X \rangle_n}(\langle x \rangle_n) = \mathcal{N}\left(\langle X \rangle_n; m, \sigma/\sqrt{n}\right) \,. \tag{3}$$

Usually, it is assumed that for $n \geq 30$ the probability distribution of $\langle X \rangle_n$ is approximately normal regardless of the probability distribution of X.
Sample variance of a sample with n elements is:

$$s^2 = \left\langle (X - \langle X \rangle_n)^2 \right\rangle_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \langle X \rangle_n)^2 \,. \tag{4}$$

Sample variance is biased estimator of the population variance σ^2:

$$\mathrm{E}\left[s^2\right] = \left(1 - \frac{1}{n}\right) \sigma^2 = \sigma^2 + \mathcal{O}(1/n) \,. \tag{5}$$

Corrected sample variance of a sample with n elements is:

$$S^2 = \frac{ns^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \langle X \rangle_n)^2 \,. \tag{6}$$

Corrected sample variance is unbiased and consistent estimator of population variance σ^2:

$$\mathrm{E}\left[S^2\right] = \sigma^2 \qquad \mathrm{Var}\left[S^2\right] \underset{n\to\infty}{\longrightarrow} 0 \,. \tag{7}$$

Concepts of parameter estimation - point estimation
The parameters of probability distributions (eg. m, σ, p, λ etc.) can be point estimated or interval estimated. Among the methods for point estimation there are *method of moments* and *method of maximal likelihood.*
In the **method of moments** the population moments are equated to the corresponding sample moments. As the population moments are functions of probability distribution parameters, the equations are solved to yield estimators of the unknown parameters as functions of the sample moments. The

number of population-sample moments needed is equal to the number of unknown parameters. The estimator of parameter θ is denoted by $\hat{\theta}$.

In the **method of maximal likelihood** a *likelihood function* is constructed:

$$L(\mathbf{v};\theta) = f(x_1;\theta) \cdot f(x_2;\theta) \cdot \dots \cdot f(x_n;\theta)\,, \tag{8}$$

where $f(x;\theta)$ is probability density function of the random variable X with parameter θ. The likelihood function equals the probability of getting the sample in the volume $d\mathbf{v}$ around $\mathbf{v}$. The parameter θ of the probability density function is estimated by maximizing the value of $L(\mathbf{v};\theta)$. That is, we have chosen the sample $\mathbf{v}$ because it is most probable. To achieve this, the function $L(\mathbf{v};\theta)$ is differentiated by θ and the derivative is set equal to zero. The resulting equation is solved to determine the estimator $\hat{\theta}$. If there is more than one parameter to estimate, the likelihood function is differentiated by each single parameter and set equal to zero to get a system of equations. The solutions of the system are the estimators of the unknown parameters.