Review - Random variables III: Important distributions

Normal distribution

Normal (or Gaussian) distribution $\mathsf{N}(\mu, \sigma^2)$ has the density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R},$$

where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ are parameters.

- When $\mu = 0$ and $\sigma^2 = 1$, the density is $f(x) = \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\}$ and the distribution is called the **standard** normal distribution and is denoted as $\mathsf{N}(0,1)$.
- If $X \sim \mathsf{N}(\mu, \sigma^2)$, then $\mathrm{E}\, X = \mu$, $\operatorname{var} X = \sigma^2$ and

$$Z = \frac{X-\mu}{\sigma} \sim \mathsf{N}(0,1).$$

- The **distribution function** of $\mathsf{N}(0,1)$ is denoted as Φ with

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\{-t^2/2\} dt$$

  and can be computed only numerically. Values of Φ can be found in statistical tables.
- From the symmetry we get

$$\Phi(x) + \Phi(-x) = 1 \quad \text{for all } x \in \mathbb{R}.$$

- Rule of one, two and three σ:

$$\mathrm{P}\left(\mu - \sigma \le X \le \mu + \sigma\right) \approx 0.68$$
$$\mathrm{P}\left(\mu - 2\sigma \le X \le \mu + 2\sigma\right) \approx 0.95$$
$$\mathrm{P}\left(\mu - 3\sigma \le X \le \mu + 3\sigma\right) \approx 0.997.$$

- The quantiles $u_\alpha = \Phi^{-1}(\alpha)$ of $\mathsf{N}(0,1)$ are also in tables Further $u_\alpha = -u_{1-\alpha}$.

Quantile function:

The quantile function F^{-1} of a random variable X is defined for $\alpha \in (0,1)$ as

$$F^{-1}(\alpha) = \inf\{x : F(x) \ge \alpha\}.$$

If F is continuous and increasing, then F^{-1} is the inverse function of F. The values of $F^{-1}(\alpha)$ are called α-quantiles and denoted as q_α.

Median:

The **median** of the random variable X is such a value $\hat{x}$, for which

$$\mathrm{P}(X \le \hat{x}) = \mathrm{P}(X \ge \hat{x}) = 1/2$$

(it is not necessarily unique). When F is continuous and increasing, then $\hat{x} = q_{0.5} = F^{-1}(0.5)$.

Exercises 5 - Random variables III: Important distributions

1. Consider a random variable X with an infinite but countable set of possible results. Prove that the X cannot have a discrete uniform distribution.

2. We are rolling a six-sided dice until a 6 occurs. Denote the roll when 6 occurs for the first time in the series as the T-th attempt.

a) What is the distribution of the random variable T?

b) Find $\mathrm{P}(T > 3)$.

c) Find $\mathrm{P}(T > 7 | T > 3)$.

d) Find $\mathrm{E}\, T$ and $\mathrm{var}\, T$.

3. Consider a random variable T with the geometric distribution $T \sim \mathrm{Geom}(p)$. Show that $\mathrm{P}(T = n + k | T > n) = \mathrm{P}(T = k)$ (meaning that the distribution is memoryless).

4. On average there are two faulty pixels among every 10 million produced on LCD screens. Suppose we buy a 1280×1024 pixel screen.

a) What is the probability that there will be no faulty pixels?

b) What is the probability that there will be exactly one faulty pixel?

c) What is the probability that there will be at least two faulty pixels?

d) Find the expected number of faulty pixels on the screen.

Hint: use the Poisson approximation.

5. The operation time until a failure of an engine (in days) is a random variable X with the exponential distribution $\mathrm{Exp}(\lambda)$ with $\lambda = 1/50$.

a) What is the mean time until a failure?

b) What is the probability that the engine will be operational for more than 50 days?

c) The engine lasted already for 50 days. What is the probability that it will be operational altogether for longer than 70 days?

Remark 1. Only geometric and exponential distributions are memoryless.

6. The IQ score among the population follows the normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 15$.

a) Find the probability that the IQ of a randomly selected individual will be larger than 130.

b) Which IQ value separates the top 1% of the population?

c) Find the probability that the IQ of a randomly selected individual will be between 85 and 115.

d) Find such a symmetric interval around the mean value of the IQ, which would cover 95% of the population.

Hint: The table is for standard normal distribution $N(0, 1)$. To solve the tasks you need to standardized the random variable. Further the table does not give us the values of distribution function of $N(0, 1)$ but something slightly different.

Additional exercises - Random variables III: Important distributions

Discrete random variables

7. A test is composed of 5 questions with four possible answers, a, b, c and d. For each question there is exactly one correct answer. Suppose that a student did not prepare for the test and picks the answers at random. Denote X the number of correctly answered questions.

a) What is the distribution of X? How is it called?

b) At least 4 correct answers are needed to pass. Find the probability that the student passes.

c) What is the expected number of correctly answered questions ($\mathrm{E}\,X$)? What is its variance?

8. Let X be the expected number of calls incoming to a police station in one hour. Suppose that k calls come with the probability of $\frac{\lambda^k}{k!}e^{-\lambda}$ for $k = 0, 1, 2, \dots$, with $\lambda > 0$.

a) Verify that the proposed distribution is indeed a probability distribution. How is it called?

b) Compute the expected number of incoming calls in one hour and its variance.

Continuous random variables

9. The length of a phone call (in minutes) of a helpdesk operator is a random variable with the density

$$f(x) = \begin{cases} ce^{-x/5} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

a) Compute the constant $c > 0$, so that f truly forms a density of a random variable.

b) What is the expected mean length of a call?

c) Find the distribution function F and draw its graph.

d) What is the probability, that the phone call will be longer than 3 minutes?

e) The call has already lasted for 3 minutes. What is the probability that it will be altogether longer than 10 minutes?

f) Find the variance of X.

Transformations of random variables

10. Let X be a random variable with $\mathrm{E}\,X = \mu$ and $\operatorname{var} X = \sigma^2$.

a) Compute the expectation and variance of

$$Z = \frac{X - \mu}{\sigma}.$$

b) Suppose that Z is continuous with a density f_Z and distribution function F_Z. Find the density of X as a function of f_Z.

Normal distribution

11. A factory line is producing steel nails with the mean length of 200 mm and a standard error (standard deviation) of 0.5 mm governed by the normal distribution.

a) Find the probability that a randomly selected nail will be shorter than 199 mm.

b) Find the probability that a randomly selected nail will be longer than 201.5 mm.

c) Which length will a nail exceed with a 99% probability?

d) Find such a symmetric bounds around the mean length, which cover 95% of the produced nails.