Exercise Sheet 3

Exercise 0

Karl's Conjecture: Let $x_1, \dots, x_n \in [a, b]$ be i.i.d. random variables with mean $\mu = \mathbb{E}[x_i]$. Let

$$z = \sum_{i=1}^{n} (x_i - \mu).$$

Then

1. $\mathbb{E}[z] = 0$
2. For all $s > 0$, we have

$$\mathbb{E}[\exp(sz)] \leq \exp\left(\frac{n^2}{8} s^2 (b-a)^2\right)$$

Task: Prove the second statement directly, without decomposing z into individual random variables $z_i = x_i - \mu$ as in Step 2 of the proof of Hoeffding's inequality. There is no constraint on the proof of item (1).

Do we need independence of the random variables $x_1, \dots, x_n$? What is the price we pay for directly proving the second statement?

Exercise 1

Let $x_1, \dots, x_n$ be i.i.d. random variables that follow a Bernoulli distribution with parameter $p = \mathbb{P}(x_i = 1)$. Define $\mu_n = (x_1 + \cdots + x_n)/n$. For any $\varepsilon > 0$, Hoeffding's inequality states

$$\mathbb{P}(|\mu_n - p| \geq \varepsilon) \leq 2 \exp(-2n\varepsilon^2),$$

Use Hoeffding's inequality to derive the inequality

$$\mathbb{P}(S_n \geq t) \leq 2 \exp\left(-2n \left(\frac{t}{n} - p\right)^2\right)$$

for all $t > 0$, where $S_n = (x_1 + \cdots + x_n)$ is a random variable that follows the Binomial distribution $\mathrm{Binom}(n, p)$.

Exercise 2

Suppose the annual rainfall in a region follows a normal distribution with a mean of $\mu = 800$ mm and a standard deviation of $\sigma = 100$ mm.

a) Estimate the probability that the annual rainfall in a randomly selected year will be at least $1,000$ mm.

b) Compute the bounds provided by Markov's and Chebyshev's inequality.

c) Compute Chernoff's bound. For $s > 0$, use

$$\mathbb{E}[\exp(sx)] = \exp\left(s\mu + \frac{1}{2}s^2\sigma^2\right).$$

Compare all bounds against the actual probability.

Exercise 3

Suppose a pharmaceutical company is conducting clinical trials for a new drug. Each patient has an independent probability $p = 0.02$ of exhibiting adverse side effects. The company administers the drug to n patients, where n is a specific number.

Plot the bounds provided by Chebyshev's and Hoeffding's inequalities, as well as the actual probability that at least 50 out of n patients will exhibit side effects. The plot should be a function of the number n of patients tested. If any of the bounds exceed 1, set them equal to 1.