EXERCISE SET 3,
MS-A0503, FIRST COURSE IN PROBABILITY AND STATISTICS

Explorative exercises

I will expect that you study the explorative problems BEFORE the first lecture of the week. It is VERY STRONGLY RECOMMENDED that you work on them in groups. I LIKE CAPITAL LETTERS, apparently.

Problem 1. The mean score on a certain test is known to be 100. Ten students take the test and get the scores 99, 102, 111, 105, 107, 100, 96, 141, 99, 92.

(1) Compute the sample variance of the test scores.
(2) Compute an interval $[a, b]$ such that you can say with 95% confidence that the standard deviation σ satisfies $a \leq \sigma \leq b$.
(3) Discuss what is meant by "with 95% confidence" in the sentence above.

Problem 2. We are now informed that the test score in Problem 1 is indeed normally distributed, with mean 100.

(1) What value of σ would maximize the "likelihood"

$$f(99) \cdot f(102) \cdot f(111) \cdot f(105) \cdot f(107) \cdot f(100) \cdot f(96) \cdot f(141) \cdot f(99) \cdot f(92)$$

of the observed data? (Hint: It is probably easier to work with abstract values $X_1, \ldots, X_n$, rather than concrete numbers.)
(2) How does the number you arrived at in part 1 compare to the sample variance? Can you generalize this observation?

Problem 3. Authorities think the test results seen are suspiciously good, and are getting suspicious, as to whether the mean score (i.e. expected value) of the test is really 100. What is the probability (assuming $\mu = 100$) of seeing results that are as least as good as the ones observed? (First: discuss what "at least as good" means!) Do you agree that there is cause for suspicion?

Homework problems

The homework problems are reported during the second exercise session of the week. You are allowed and encouraged to work in groups, but every student should be prepared to present the solutions individually. During the last exercise session of the week, the teacher will ask you to mark what problems you have solved, and you get points according to how many problems you marked as solved. If you mark a problem as solved, however, you should also be prepared to present your solution in front of the class.

Homework 1. Let $X_1, \ldots, X_n$ be a sample from the distribution whose density function is

$$f(x) = \begin{cases} e^{-(x-\theta)} & \text{if } x \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

(1) Sketch the distribution function $f(x)$.
(2) Determine the maximum likelihood estimator of θ.

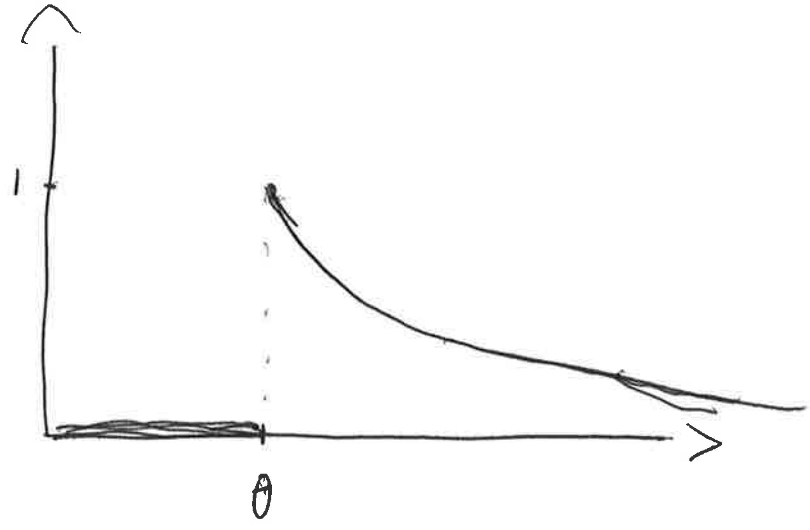
Homework 2. The average salary of newly graduated students from a certain study program is 31 000€, with a standard deviation of 3 200€. 15 recent graduates from the program are sampled randomly. Approximate the probability that their average exceeds 33 000€.

Homework 3. Alex (good, gender neutral name) just graduated from the study program in Problem 2.

(1) Is the information given enough to approximate or bound the probability that Alex gets a salary of at least 33 000€?
(2) Is the information given enough to approximate or bound the probability that Alex gets a salary of at least 39 000€?

Week 5, Homework 1

1)



2)

$$L(\theta) = \prod_{i=1}^{n} f(x_i) = \begin{cases} 0 & \text{if } x_i < \theta \text{ for some } i \\ e^{n\theta - \sum x_i} & \text{otherwise} \end{cases}$$

On the region $\theta \leq \min x_i$, $L(\theta)$ is strictly ~~decre~~ increasing in θ, so it is maximized in $\min x_i$.

$$\hat{\theta} = \min x_i \quad .$$

Week 5, Homework 2

The average salary $\bar{X}$ has approximate distribution $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$, where

$\mu = 31000$, $\sigma = 3200$, $n = 15$.

The probability that the average is ≥ 33000 is then

$$P[\bar{X} \geq 33000] = P\left[\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \geq \frac{33000-\mu}{\sigma/\sqrt{n}}\right]$$

$$= 1 - \Phi\left(\frac{33000-\mu}{\sigma/\sqrt{n}}\right)$$

$$\approx 1 - \Phi(2{,}42)$$

$$= \Phi(-2.42) = 0.0078.$$

i.e. 0.78%

Week 5, Homework 3

1) For Alex's salary, we cannot use the central limit theorem (too small sample size). So the best bound of $\mathbb{P}[|X-\mu| > a]$ is the Chebyshev's bound, which is non-vacuous only if $a > \sigma$.

So we cannot use it to bound

$\mathbb{P}[X \geq 33000] = \mathbb{P}[X - 31000 > 2000]$, as $\sigma = 3200$.

2) Chebyshev's inequality:

$$\mathbb{P}[X \geq 39000] \leq \mathbb{P}[|X-31000| \geq 8000]$$

$$= \mathbb{P}\left[|X-\mu| \geq \frac{8000}{3200}\sigma\right]$$

$$\leq \left(\frac{1}{8000/3200}\right)^2 = 0.16.$$