Probability and Statistics

Exercise sheet 13

Exercise 13.1 In a new study on the reliability of ball-bearings (in German: Kugellager), two samples of two different types of ball-bearings were tested, one piece for each of two different types in one of 10 different scenarios. The resulting numbers of rotations until breakdown were

type I	3.03	5.53	5.60	9.30	9.92	12.51	12.95	15.21	16.04	16.84
type II	3.19	4.26	4.47	4.53	4.67	4.69	12.78	6.79	9.37	12.75

Each column represents one of the testing scenarios. Before the realization of this study, it was not clear which type was more reliable.

(a) Are we dealing with a paired sample? Please explain your answer.

(b) Perform a t-test for the null hypothesis "the expected number of rotations until break-down is the same for the two types of ball-bearings for each testing scenario" with level 5%. (What are the model assumptions of a t-test?)

(c) Which other test would be an alternative if you do not want to assume a normal distribution?

Hint: (Clicking the following link will reveal the solution of (c).) You can find the R-code at `https://www.kaggle.com/jakobheiss/sol13-1/edit`.

(d) Compare your results with Exercise 12.4 (the numbers in the table are the same), and discuss your conclusion.

Solution 13.1

(a) This is a paired sample—the connection between the data is the testing scenario, which does give natural pairs.

(b) The model is given by $X_1 - Y_1, \dots, X_{10} - Y_{10}$ i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$, where μ and σ are unknown and the $X_i - Y_i$ are all independent. The null and alternative hypotheses are given by

$$H_0 : \mu = 0 \quad \text{und} \quad H_A : \mu \neq 0.$$

The test statistic is

$$T := \frac{\overline{X - Y}_n}{S_n \sqrt{1/n}},$$

where the estimator S_n for σ is given by

$$S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - Y_i - \overline{X - Y}_n \right)^2},$$

and under H_0, T is t-distributed with $n - 1 = 9$ degrees of freedom. With a level of 5%, the null hypothesis will be rejected when $|T| > t_{9,0.975} = 2.262$. From the data, we obtain $\overline{x - y}_{10} = 3.943$ and $S_n(\omega) = 3.18$, so $T(\omega) = 3.92$, i.e., H_0 is rejected. (The realized value of the P-value is $\pi((\mathbf{X}, \mathbf{Y})(\omega)) = 2\,(1 - t_9\,(3.92)) = 0.0035$, so the P-value is even 10 times smaller than necessary to reject H_0.)

(c) The sign test does not assume normal distributions. The null hypothesis of this test is that all samples $X_1 - Y_1, \dots, X_n - Y_n$ are i.i.d. with an arbitrary (unknown) continuous distribution F with $F(0) = \frac{1}{2}$. Under the null hypothesis, the test statistic

$$T := \sum_{i=1}^{n} I_{\{\mathcal{X}_i - X_i > 0\}}$$

follows a $\mathrm{Bin}(n, p = \frac{1}{2})$-distribution. For the given data, we obtain $T(\omega) = 9$ by counting how many times type I performed better than type II. The realized P-value

$$\pi((\mathbf{X}, \mathbf{Y})(\omega)) = \mathbb{P}_{H_0}\left[\left|T - \frac{10}{2}\right| \geq 9 - \frac{10}{2}\right] = 2\sum_{k=9}^{10} \binom{n}{k} 2^{-n} = 0.0215$$

can be easily computed with the following R-code:

```
binom.test(9, 10,alternative="two.sided")
```

or equivalently:

```
2*(1-pbinom(8,10,prob=0.5))
```

So we reject H_0.

Alternatively, a ranked sign test (which has not been covered in the lecture) would be a good option, as it also does not assume normality. It needs basically the same assumptions as the sign test and additionally symmetry of the distribution of $X_i - Y_i$ around zero. Running this test in R

```
wilcox.test(c(3.03, 5.53, 5.60, 9.30, 9.92, 12.51, 12.95, 15.21, 16.04, 16.84),
            c(3.19, 4.26, 4.47, 4.53, 4.67,4.69, 12.78, 6.79, 9.37, 12.75),paired=TRUE)
```

results in a P-value of 0.0039. So this test rejects H_0, too.

All three tests agree that the observed data would be significantly more likely if type I ball-bearings are more reliable than their type II counterparts.

All the calculations done in this solution can be found in the R-code https://www.kaggle.com/jakobheiss/sol13-1/edit.

(d) In Exercise 12.4, the null hypothesis was not rejected, because there we had less information available. The additional information that each of the columns corresponds to a specific testing scenario is a very valuable information. An intuitive explanation for this is that part of the "distracting noise" is due to the choice of testing scenarios. So conditioned on the choice of testing scenario, the "distracting noise" is much smaller.

Exercise 13.2 Consider the null hypothesis H_0: X has the density $f_0(x) = f(x)$ and the alternative H_A: X has the density X $f_1(x) = f(x-1)$ for the following cases:

(a) f is a standard normal density,

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

(b) f is a Cauchy density,

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Compute in both cases the form of the rejection region of the most powerful test (also known as the likelihood ratio test; see the Neyman-Pearson lemma). Which differences do you find.

Solution 13.2 Using the Neyman-Pearson test with the hypothesis

$$H_0 : f_0(x) = f(x),$$
$$H_A : f_1(x) = f(x-1),$$

the likelihood ratio is given by

$$R(x) := \frac{L(\theta_1; x)}{L(\theta_0; x)} = \frac{f_1(x)}{f_0(x)} = \frac{f(x-1)}{f(x)}.$$

(a) In the case of a normal distribution, $R(x) = e^{x-\frac{1}{2}}$, and we reject when $R(x) > c$, i.e., $x > \ln c + \frac{1}{2}$. So the rejection region is of the form (a, ∞).

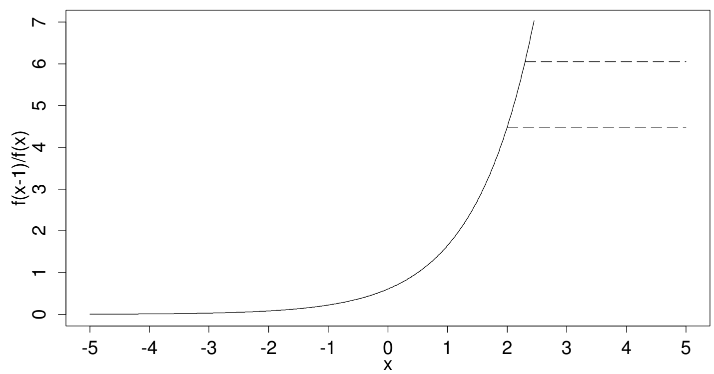

Figure 1: Rejections regions for the normal case.

(b) In the case of a Cauchy distribution, the likelihood ratio is given by $R(x) = \frac{x^2+1}{x^2-2x+2}$, and then we have an interesting behavior as one can see in Figure 2. If we put $c = 1$, we have an unbounded interval; but if we put $c > 1$, we get a bounded interval. (Of course, the choice of c depends on the desired level α of the the test. E.g. $c = 1$ corresponds to a rejection region of $x \geq 1/2$, wich corresponds to $\alpha \approx 0.3524$.) This happens because the Cauchy distribution

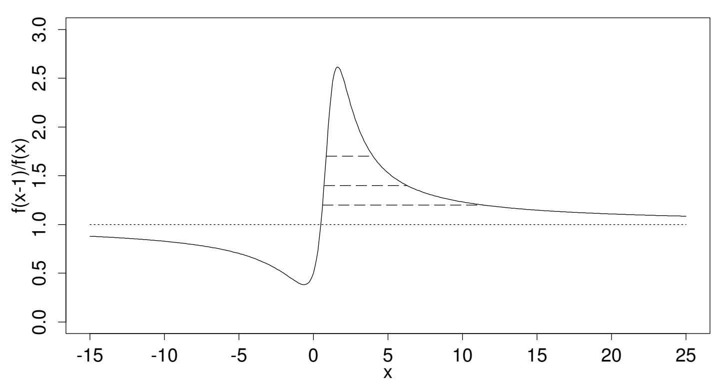

Figure 2: Rejections regions for the Cauchy case.

is heavy-tailed.

Exercise 13.3 Consider $X_1, ..., X_n$ i.i.d. $\sim \text{Exp}(\lambda)$, $\lambda \in \Theta = (0, +\infty)$. Recall that the density of $X_i \sim \text{Exp}(\lambda)$ is given by $f_\lambda(x) = \lambda e^{-\lambda x} I_{(0,+\infty)}(x)$. We want to test $H_0 : \lambda = 1$ versus $H_A : \lambda = 2$.

(a) Apply the Neyman-Pearson lemma to find a most powerful test of level α based on $X = (X_1, ..., X_n)$.

Hint: We recall that if $Y_1, ..., Y_n$ are i.i.d. $\sim \text{Exp}(\lambda_0)$, then $\sum_{i=1}^n Y_i \sim G(n, \lambda_0)$.

(b) What is the power of the Neyman-Pearson test you have found?

Hint: You can express your answer in terms of F_n and F_n^{-1}, the cdf and inverse cdf of a Gamma distribution with parameters n and 1, that we denote by $\text{G}(n, 1)$.

(c) For $n = 10$, we observe the following sample:

1.009	0.132	0.384	0.360	0.206	0.588	0.872	0.398	0.339	1.079

What decision do you take if you want the level of the test to be equal to $\alpha = 0.05$? What about $\alpha = 0.01$?

Hint: The quantiles of the $G(10, 1)$ distribution of order 5% and 1% are 5.425 and 4.130, respectively.

Solution 13.3

(a) The NP test is given in the form

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \frac{f_{\lambda_1}(\mathbf{x})}{f_{\lambda_0}(\mathbf{x})} > c(\alpha) \\ \gamma_\alpha, & \frac{f_{\lambda_1}(\mathbf{x})}{f_{\lambda_0}(\mathbf{x})} = c(\alpha) \\ 0, & \frac{f_{\lambda_1}(\mathbf{x})}{f_{\lambda_0}(\mathbf{x})} < c(\alpha), \end{cases}$$

for some suitable $c(\alpha) > 0$ and $\gamma_\alpha \in [0, 1]$ such that $\mathbb{E}_{\lambda_0}[\varphi(\mathbf{X})] = \alpha$. A value of 1 corresponds to rejecting the null hypothesis, and a value of 0 corresponds to not rejecting the null hypothesis. Here we only consider $\mathbf{x} = (x_1, ..., x_n)^T$ such that $x_i > 0$ for each $i \in \{1, ..., n\}$, since the X_i are positive almost surely.

The likelihood ratio is given by

$$\begin{aligned} R(\mathbf{x}) := \frac{f_{\lambda_1}(\mathbf{x})}{f_{\lambda_0}(\mathbf{x})} &= \frac{\prod_{i=1}^n \lambda_1 e^{-\lambda_1 x_i}}{\prod_{i=1}^n \lambda_0 e^{-\lambda_0 x_i}} \\ &= \left(\frac{\lambda_1}{\lambda_0}\right)^n e^{-\lambda_1 \sum_{i=1}^n x_i + \lambda_0 \sum_{i=1}^n x_i} \\ &= \text{const.}(\lambda_0, \lambda, n) \exp\left((\lambda_0 - \lambda_1) \sum_{i=1}^n x_i\right), \end{aligned}$$

and $\lambda_0 - \lambda_1 = -1$ is negative. Moreover, $\text{const.}(\lambda_0, \lambda, n) = 2^n$. So

$$\begin{aligned} & R(\mathbf{x}) > c \\ \Leftrightarrow\ & 2^n e^{-\sum_{i=1}^n x_i} > c \\ \Leftrightarrow\ & g\big(T(x_1, ..., x_n)\big) > c \\ \Leftrightarrow\ & T(x_1, ..., x_n) < t =: g^{-1}(c), \end{aligned}$$

where $T(x_1, ..., x_n) = \sum_{i=1}^n x_i$, $g(s) = 2^n \exp(-s)$ and so $t = -\log c + n \log 2$. The equivalence holds since g is strictly decreasing.

Under $H_0 : \lambda = \lambda_0 = 1$, $\sum_{i=1}^n X_i \sim \mathrm{G}(n, 1)$ (by independence) has a continuous distribution. Therefore the case $\frac{f_{\lambda_1}(\mathbf{x})}{f_{\lambda_0}(\mathbf{x})} = c(\alpha)$ (which is equivalent to $\sum_{i=1}^n x_i = t_\alpha$) has probability 0, and in particular, the middle branch of the NP test does not affect the condition $\mathbb{E}_{\lambda_0}[\varphi(\mathbf{X})] = \alpha$. Therefore, we can arbitrarily choose $\gamma_\alpha = 0$.

The NP test can then be equivalently given by

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \sum_{i=1}^n x_i < t_\alpha \\ 0, & \sum_{i=1}^n x_i \geq t_\alpha. \end{cases}$$

We still need to enforce the condition $\mathbb{E}_{\lambda_0}[\varphi(\mathbf{X})] = \alpha$ by choosing a suitable value of t_α. This is equivalent to

$$\mathbb{P}_{\lambda_0}\left[\sum_{i=1}^n X_i < t_\alpha\right] = \alpha,$$

or also to

$$\mathbb{P}_{\lambda_0}\left[\sum_{i=1}^n X_i \leq t_\alpha\right] = \alpha.$$

Since $\sum_{i=1}^n X_i \sim \mathrm{G}(n, 1)$ under H_0, this means that $t_\alpha = F_n^{-1}(\alpha)$, for F_n the cdf of the $\mathrm{G}(n, 1)$-distribution.

(b) By the definition of the power, we have

$$\beta = \mathbb{E}_{\lambda_1}[\varphi(\mathbf{X})] = \mathbb{P}_{\lambda_1}\left[\sum_{i=1}^n X_i \leq F_n^{-1}(\alpha)\right].$$

Recall that if $Y \sim \mathrm{Exp}(\lambda)$, then $\lambda Y \sim \mathrm{Exp}(1)$. Thus, under $H_1 : \lambda = \lambda_1 = 2$, $2X_1, ..., 2X_n$ are i.i.d $\sim \mathrm{Exp}(1)$, and therefore, by independence, $\sum_{i=1}^n 2X_i \sim \mathrm{G}(n, 1)$. It follows that

$$\beta = \mathbb{P}_{\lambda_1}\left[2\sum_{i=1}^n X_i \leq 2F_n^{-1}(\alpha)\right] = F_n\left(2F_n^{-1}(\alpha)\right).$$

(c) We compute $\sum_{i=1}^{10} x_i = 5.367$.

- For $\alpha = 0.05$, $F_{10}^{-1}(\alpha) = F_{10}^{-1}(0.05) \approx 5.425 > \sum_{i=1}^{10} x_i$. Therefore, we reject H_0 at a level of 5%.
- For $\alpha = 0.01$, $F_{10}^{-1}(0.01) \approx 4.130$. Therefore, we cannot reject H_0 at a level of 1%—these data do not present a compelling enough evidence against the null hypothesis.

Exercise 13.4 Again in the setup of Exercise 13.3, it turns out that the Neyman-Pearson test you found there in (a) is actually UMP at the level α for testing $H_0 : \lambda = 1$ versus $H'_A : \lambda > 1$. More precisely, the same NP test is the most powerful among all tests of level α for the alternative $H''_A : \lambda = \lambda_1$ for any $\lambda_1 \in \Theta'_A = (1, +\infty)$, not only for $\lambda \in \Theta_A = \{2\}$.

How can you see why this is true?

Solution 13.4 The explicit form of the NP test for this problem is

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \sum_{i=1}^n x_i < F_n^{-1}(\alpha), \\ 0, & \sum_{i=1}^n x_i \geq F_n^{-1}(\alpha). \end{cases}$$

We know from the Neyman-Pearson lemma applied in Exercise 13.3 that φ is a UMP test of level α for testing $H_0 : \lambda = 1$ versus $H_1 : \lambda = 2$. In other words, for any other test φ' such that $\mathbb{E}_{\lambda_0}[\varphi'(\mathbf{X})] \leq \alpha$, we have a lower power, i.e.,

$$\mathbb{E}_{\lambda_1}[\varphi'(\mathbf{X})] \leq \mathbb{E}_{\lambda_1}[\varphi(\mathbf{X})].$$

However, φ does not depend on the particular value of $\lambda_1 = 2$—all we use is that $\lambda_0 - \lambda_1$ is negative. More specifically, if we had to test $H_0 : \lambda = 1$ versus $H''_A : \lambda = \lambda'_1$ for some $\lambda'_1 > 1$, we should obtain exactly the same test as above. Since this same test is again UMP of level α, this implies that it is actually UMP of level α for the testing problem $H_0 : \lambda = 1$ versus $H'_A : \lambda > 1$, in the sense given in the exercise.

If you have feedback regarding the exercise sheets, please send a mail to Jakob Heiss.