Probability and Statistics

Exercise sheet 13

MC 13.1. Which of the following statements are true? (The number of correct answers is between 0 and 4.)

(a) If we reject the null hypothesis, the realized p-value must be less than or equal to the significance level α.

(b) If the realized p-value is less than or equal to the level α, we reject the null hypothesis.

(c) The realized p-value tells us the probability that H_0 is true.

(d) If the realized p-value is very low, it indicates that our data do not fit the null hypothesis well.

Exercise 13.2. Compute the realized p-values for the tests from Exercises 12.3, 12.4, 12.6, and 12.7.

Remark: You should write down an explicit formula (e.g., involving the standard normal CDF), but you do not need to compute its numerical value.

Remark: To compute the numerical value, you can use Wolfram Alpha to find values of the standard normal distribution function or Python:

```
from scipy.stats import norm
value = norm.cdf(x=1.96)  # Replace x with the value you're interested in
print(value)
```

You can find values of other distribution functions analogously.

Exercise 13.3. The average travel time from Zurich to Bellinzona by Intercity train is 146 minutes. The following times are recorded for the Cisalpino:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
152	145	141	137	145	146	139	147	138

We assume that these values are realizations of an i.i.d. sample $X_1, \dots, X_n$ with $X_i \sim \mathcal{N}(\mu, \sigma^2)$, where μ is an unknown parameter and $\sigma^2 = 9$ is known.

(a) Perform an appropriate test at the 5% level to determine whether the mean travel time of the Cisalpino differs from that of the Intercity.

(b) Compute the realized p-value.

(c) What is the lowest level α at which you would still reject the null hypothesis?

Exercise 13.4. Let $\{\mathbb{P}_\theta : \theta \in \Theta\}$, where $\Theta \subseteq \mathbb{R}$, be a family of models, and let $X_1, \dots, X_n$ be i.i.d. random variables from the distribution $\mathbb{P}_\theta$. Assume that we have a confidence interval for θ of the form $[A = a(X_1, \dots, X_n), B = b(X_1, \dots, X_n)]$ with coverage probability $1 - \alpha$.

Further, consider the hypotheses

$$H_0 : \theta = \theta_0, \qquad H_1 : \theta \neq \theta_0, \tag{1}$$

where $\theta_0 \in \Theta$ is fixed.

(a) Show that the test procedure

"We reject H_0 if and only if $\theta_0 \notin [A, B]$"

defines a test at level α.

(b) Conversely, for every $\theta_0 \in \Theta$, let $(T_{\theta_0}, K_{\theta_0})$ be a test for (1) at level α. Show that the random set

$$S(\omega) := \{\theta_0 \in \Theta : T_{\theta_0}(\omega) \notin K_{\theta_0}\}$$

is a confidence set for θ with coverage probability at least $1 - \alpha$. That is, show that

$$\mathbb{P}_{\theta_0}[\theta_0 \in S] \geq 1 - \alpha, \quad \theta_0 \in \Theta.$$

(c) How would you modify the confidence interval in (a) for the one-sided alternative

$$H_0 : \theta = \theta_0, \qquad H_1' : \theta > \theta_0?$$

Exercise 13.5. Find the p-values for the tests from Exercise 12.5.

Rewind

These are some additional exercises for you to review some of the material we covered during the semester.

MC 13.6. Let $X_1, X_2, \dots$ be a sequence of independent, identically distributed random variables with $\mathbb{E}[X_1^2] < \infty$. Let Z be a standard normally distributed random variable. We define $\mu := \mathbb{E}[X_1]$ and $\sigma^2 := \operatorname{Var}[X_1]$. Which of the following statements is correct? (Exactly one answer is correct.)

(a) $\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{\sigma^2 n}\sum_{i=1}^n X_i \le a\right] = \mathbb{P}[Z \le a],\ a \in \mathbb{R}.$

(b) $\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{\sigma^2 n}\sum_{i=1}^n (X_i - \mu) \le a\right] = \mathbb{P}[Z \le a],\ a \in \mathbb{R}.$

(c) $\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{\sqrt{\sigma^2 n}}\sum_{i=1}^n (X_i - \mu) \le a\right] = \mathbb{P}[Z \le a],\ a \in \mathbb{R}.$

(d) $\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{\sqrt{\sigma^2 n}}\sum_{i=1}^n X_i \le a\right] = \mathbb{P}[Z \le a],\ a \in \mathbb{R}.$

MC 13.7. Does the correct answer to MC 13.6 also hold exactly for finite n (i.e., if we omit $\lim_{n\to\infty}$)? (The number of correct answers is between 0 and 4.)

(a) The correct answer to MC 13.6 also holds without the limit for all distributions satisfying $\mathbb{E}[X_1^2] < \infty$.

(b) The correct answer to MC 13.6 also holds without the limit if the X_i's are normally distributed.

(c) The correct answer to MC 13.6 also holds without the limit if $\mu = 0$ and $\sigma = 1$.

(d) The correct answer to MC 13.6 never holds exactly for finite n if the limit is omitted.

Exercise 13.8. It costs \$1 to play a particular slot machine in Las Vegas. The machine is programmed so that it pays out \$2 with probability 0.45 (the player wins), and nothing with probability 0.55 (the casino wins). Let X_i be the net gain of the casino on the i-th round of the game. Let $S_n := \sum_{i=1}^n X_i$ be the casino's total gain after n rounds. Assuming that the outcomes of the games are independent, determine:

(a) $\mathbb{E}[S_n]$ and $\operatorname{Var}[S_n]$;

(b) the approximate probability that after $10'000$ rounds, the casino's gain lies between \$800 and \$1100.

Assume that we know the casino's gain after $10'000$ rounds is \$1200.

(c) Based on this observation, can we conclude that the probability the casino wins is greater than the stated value 0.55? Use significance level $\alpha = 0.05$.

Exercise 13.9. A team of three people is randomly selected from a group of six people. Among the six are three women (Anna, Elsa, and Helga) and three men (Franz, Mario, and Tobias). Let X be the number of women and Y the number of men in the selected team.

(a) What is the conditional probability that the team consists only of women, given that the team includes at least one woman?

(b) What is the conditional probability that the team consists only of women, given that Helga is in the team?

(c) Find the joint distribution of (X, Y). Compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

(d) Compute $\mathrm{Var}[X]$, $\mathrm{Var}[X+Y]$, $\mathrm{Cov}(X, Y)$, and $\mathrm{Corr}(X, Y)$.

Exercise 13.10. Let X and Y be independent random variables where X is uniformly distributed on $[0,1]$ and Y is exponentially distributed with parameter 1. Define $U = X + Y$ and $V = XY$.

(a) Compute $\mathbb{E}\left[\frac{V}{X^2+1}\right]$.

(b) Determine the distribution function and the density function of U.

Quantile table for the standard normal distribution

0.5	0.75	0.9	0.95	0.975	0.99	0.995	0.999
0	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902

For instance, $\Phi^{-1}(0.9) = 1.2816$, where Φ is the distribution function of $\mathcal{N}(0,1)$.

Table of standard normal distribution

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

For instance, $\mathbb{P}[Z \leq 1.96] = 0.975$.