CompStat Exam I summer term 2024

Students from the class

July 26, 2024

The exam is much more mathematically (statistics & probability theory) challenging than the lecture & exercises.

1 Multiple Choice

The multiple choice section contained 10 questions with four options for answers where only one option was correct. We don't include all options, mostly as we don't remember them all. Further, don't take the questions/answer options for granted, we might misremember some, especially the specific numbers. This part should provide a general overview over what kind of questions to expect. This old exam should be highly valuable as we did not have such a document and were taken by surprise by some of the questions that differed from the lecture/exercises.

1. We were given $\hat{\theta}_{ML} = \frac{1}{\bar{X}n}$ and were asked to compute $I(\theta_0))^{-1}$ for a given random sample of five observations.

2. θ_0 is an estimator for a Bernoulli-distributed random variable with probability $P(Y = 1) = \theta_0$. What is θ_0^2 an estimator for?

   (a) $P(X^2 = 1)$

   (b) $P(\sqrt{X} = 1)$

   (c) For nothing

   (d) For the joint distribution of X with an independent random variable Y with the a Bernoulli distribution and $P(Y = 1) = \theta_0$

3. Determine $P(Y = 1|X = x)$ for the different intervals of x. We have the prior probability $P(Y = 0) = 0.25$ and the conditional density functions.

$$f_{X|Y}(Y = 0) = \frac{1}{0.6} * 1_{(x_i \in [-0.5, 0.1])} \text{ and } f_{X|Y}(Y = 1) = \frac{1}{0.6} * 1_{(x_i \in [-0.1, 0.5])}$$

   (a) $m(X_i) = \begin{cases} 0.00 & x \in [-0.5, 0] \\ 1.00 & x \in [0, 0.5] \end{cases}$

(b) $m(X_i) = \begin{cases} 1.00 & x \in [-0.5, 0] \\ 0.00 & x \in [0, 0.5] \end{cases}$

(c) $m(X_i) = \begin{cases} 1.00 & x \in [-0.5, -0.1] \\ 0.25 & x \in [-0.1, 0.1] \\ 0.00 & x \in [0.1, 0.5] \end{cases}$

(d) $m(X_i) = \begin{cases} 0.00 & x \in [-0.5, -0.1] \\ 0.75 & x \in [-0.1, 0.1] \\ 1.00 & x \in [0.1, 0.5] \end{cases}$

4. We have a Poisson distribution (the distribution function was given) with observed $x = \{1, 3, 4, 5, 7\}$. What is the correct expression of the likelihood function?

   (a) $L(\lambda) = \frac{\lambda^{20} exp(-\lambda)}{87091200}$

   (b) $L(\lambda) = \frac{\lambda^{20} exp(-5\lambda)}{87091200}$

   (c) $L(\lambda) = \frac{\lambda^{4} exp(-5\lambda)}{87091200}$

   (d) $L(\lambda) = \frac{\lambda^{20} exp(-5\lambda)}{420}$

5. Fill in the missing term for the NR algorithm. (R code was provided where the last part of $\theta_{(m)} = \theta_{(m-1)} -$... was missing. The answer options were written in mathematical expresssions, not in R code.)

6. We have $n^{\frac{1}{3}} Z_n \to_d N(0, 4)$. Which of the expressions is true (Landau symbols).

7. When is the basic bootstrap method advantageous to bootstrap-t?

8. Which of the statements is true (four options on regression splines).

9. What holds for the estimator for large p in non-parametric regression?

   (a) Large bias, large variance

   (b) Small bias, small variance

   (c) Small bias, large variance

   (d) Large bias, small variance

10. What is bootstrap consistency? (four definitions given)

2 Open Questions

1. We have the probability functions for a random variable X in table 1 below. Further, we draw a random sample $\{x_1, ..., x_5\}$ with $x_1 = 1, x_2 = 2, x_3 = 2, x_4 = 1, x_5 = 2$.

   (a) Write down the log-likelihood function for our random sample.

   (b) Determine $\hat{\theta}_{ML}$.

	$x=0$	$x=1$	$x=2$	$x=3$
$P(X=x)$	$\frac{2}{3}\theta$	$\frac{1}{3}\theta$	$\frac{2}{3}(1-\theta)$	$\frac{1}{3}(1-\theta)$

Table 1: probability density function of our random variable X

(c) Compute the second derivative of the log-likelihood function for the random sample $l_n''(\theta_0)$.

(d) Compute $E[l_n''(\theta_0)]$.

2. Fill the empty spaces. Your answers for 2. and 3. should include $\hat{\theta}_{ML}$.

(a) $E[-\frac{1}{n}l_n''(\theta_0)]$ is ...

(b) $E[-\frac{1}{n}l_n''(\theta_0)] = \lim_{n\to\infty}$ ...

(c) $E[-l_n''(\theta_0)] =$ ...

3. We use a spline function to estimate the unknown regression function. All the assumptions from the lecture hold. We know that $Y_i = m(X_i) + \epsilon_i$, where (Y_i, X_i) are independent of the ϵ_i

(a) We now draw a new i.i.d. observation (Y_{new}, X_{new}). Show that $E_\epsilon[\sum_{i=1}^n (Y_{new} - \hat{m}_p(X_{new}))^2]$ can be disposed into an adjustable and a fixed part.

(b) Compare GCV to C_p and give one advantage and one disadvantage of C_p over GCV. (The equations for GCV and C_p were given).

(c) Assume the true function $m(X_i)$ is linear, i.e. $m(X_i) = \beta_0 + \beta_1 X_i$. What are the best knots to choose? What degree should k be?