Tentamen Statistiek AN

Shota Gugushvili

January 14, 2014

1 Guidelines

This is an open book exam: use of the textbook, handouts and notes you have taken during the classes is permitted. Use of electronic devices, including calculators, is not allowed. Write legibly and give concise and clear answers to the questions. Good luck!

2 Problems

Problem 1. Let A, B and C be three events, such that $\mathbb{P}(C) > 0$. Prove that

$$\mathbb{P}(A \cup B|C) = \mathbb{P}(A|C) + \mathbb{P}(B|C) - \mathbb{P}(A \cap B|C).$$

Problem 2. Let the PDF f be given by

$$f(x) = \begin{cases} \frac{4}{5}x^3, & \text{if } 0 \leq x < 1; \\ \frac{4}{5}e^{1-x} & \text{if } 1 \leq x < \infty; \\ 0 & \text{otherwise.} \end{cases}$$

(a) Determine the corresponding CDF F.

(b) Suppose $X \sim F$. Compute the probability

$$\mathbb{P}\left(\frac{1}{2} < X < 2\right).$$

Problem 3. Let $X \sim N(0,1)$ and $Y = |X|$. Give an explicit expression for the CDF F_Y of Y (in terms of the CDF corresponding to the standard normal distribution $N(0,1)$) and next determine the corresponding PDF f_Y.

Problem 4. Let $X_1, X_2, \ldots X_\ell$ be IID observations from the $N(\mu_X, 1)$ distribution and let $Y_1, Y_2, \ldots, Y_m$ be IID observations from the $N(\mu_Y, 1)$ distribution. Assume additionally that the two samples of X's and Y's are independent.

(a) Find an $1 - \alpha$ confidence interval $C_n = (a, b)$ for the parameter

$$\theta = 3\mu_X - 5\mu_Y.$$

The length of a confidence interval is $b - a$. For obvious reasons one is interested in $1 - \alpha$ confidence intervals that are shortest possible in length. Suppose a statistician can only afford to select a *total* sample size $n = \ell + m$ in order to make statistical inferences on θ. So n is *fixed* and then ℓ and m are chosen by a statistician subject to a restriction $\ell + m = n$ (note that $\ell = n - m$ and so the choice of ℓ also determines m).

(b) Find the length of the $1 - \alpha$ confidence interval for θ from part (a). Rewrite it in terms of ℓ and n by eliminating m from the expression.

(c) Suppose there indeed exists ℓ_0, such that $n = \ell_0 + m_0$ and the length of the confidence interval from part (a) with this $\ell = \ell_0$ is minimal. Show how ℓ_0 can be found. *Hint:* Use the expression for the length of the confidence interval from part (b). This is a function of ℓ. Minimise this function over ℓ. You will obtain that ℓ_0 is a solution of a certain equation (rounded off to the nearest integer, if needed).

Problem 5. Let $X_1, \ldots, X_n$ be IID observations from a distribution with mean μ and variance σ^2 (both unknown parameters). Suppose one wants to estimate μ. An obvious estimator of μ is the sample mean

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Consider another estimator of μ, defined by $\widetilde{\mu}_n = \kappa \hat{\mu}_n$, where $\kappa > 0$ is a constant one can choose.

(a) Compute the mean squared errors $\mathrm{MSE}_{\hat{\mu}_n}$ and $\mathrm{MSE}_{\widetilde{\mu}_n}$ of the two estimators $\hat{\mu}_n$ and $\widetilde{\mu}_n$. Next use this to determine a condition on κ guaranteeing that

$$\mathrm{MSE}_{\widetilde{\mu}_n} < \mathrm{MSE}_{\hat{\mu}_n}.$$

(b) An estimator with a smaller MSE is thought to be a better estimator. Part (a) would suggest that for some value of κ, chosen to satisfy a condition you derived in part (a), $\widetilde{\mu}_n$ is a better estimator than $\hat{\mu}_n$. What is an obvious problem with this 'estimator'?

Problem 6. Let the PDF f be given by

$$f(x;\theta) = \begin{cases} \theta x^{\theta - 1}, & \text{if } 0 < x < 1; \\ 0 & \text{otherwise.} \end{cases}$$

Here $\theta > 0$ is a parameter.

(a) Verify that for every fixed $\theta > 0$, $f(x;\theta)$ is indeed a PDF.

(b) Suppose one has IID observations $X_1, \ldots, X_n \sim f(x;\theta)$. Find the method of moments estimator of θ.

Problem 7. Let $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$.

(a) Show that the maximum likelihood estimator is

$$\hat{\lambda}_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

(b) Find the Fisher information $I_n(\lambda)$ in n observations.

(c) Find the maximum likelihood estimator $\hat{\tau}$ of

$$\tau = \mathbb{P}(X \leq 1).$$

Here $X \sim \text{Poisson}(\lambda)$.

(d) Use the delta method to determine the estimated standard error $\widehat{\text{se}}[\hat{\tau}]$ of $\hat{\tau}$ and obtain an (asymptotic) $1 - \alpha$ confidence interval for τ.

(e) Suppose we want to test the null hypothesis $\tau = 0.5$ versus the alternative hypothesis $\tau \neq 0.5$. Suggest a reasonable test statistic to that end. Next, as a toy example, assume that $n = 4$ and that the observations are $X_i = 1, i = 1, \ldots, 4$. Compute explicitly the (approximate) p-value based on this sample.

Problem 8. Let $X_1, \ldots, X_n$ be an IID sample from the $N(\mu, \sigma^2)$ distribution, where μ and σ^2 are unknown parameters. Suppose one wants to test the null hypothesis $\sigma^2 = 4$ at level α.

(a) Derive the likelihood ratio test for this case.

(b) Let $\lambda(x^n)$ denote the observed value of the likelihood ratio test statistic. What is the (approximate) p-value?

Problem 9. Consider the simple linear regression model with the intercept omitted (β_0 known and equal to zero),

$$Y_i = \beta_1 X_i + \epsilon_i$$

This is the regression through the origin model. The observations are the pairs

$$(X_1, Y_1), \ldots, (X_n, Y_n).$$

(a) Give an expression for the least squares estimator $\hat{\beta}_1$ in this setting. Reminder: this estimator is a minimiser over β_1 of the sum of squares

$$\sum_{i=1}^{n} (Y_i - \beta_1 X_i)^2.$$

(b) Assume that the measurement errors ϵ_i are IID with the $N(0, \sigma^2)$ distribution. Treating covariate values X_i's as constants, show that the least squares estimator $\hat{\beta}_1$ is normally distributed. Find its mean and the variance.

3 Norming

The following norming is used:

1: 2
2(a): 2, **2(b)**: 2
3: 2
4(a): 2, **4(b)**: 2, **4(c)**: 2
5(a): 2, **5(b)**: 2
6(a): 1, **6(b)**: 2
7(a): 2, **7(b)**: 2, **7(c)**: 1, **7(d)**: 3, **7(e)**: 2
8(a): 2, **8(b)**: 1
9(a): 2, **9(b)**: 2

The final grade **G** is determined according to the formula **Q=0.3*H+0.7*E**, where **H** is the average grade for two homework assignments and **E=10*P**, with **P** the number of points scored in the written exam according to the above norming. **G** is then obtained by rounding off **Q** (to the nearest integer or a number ending on .5).