Taras Bodnar
Department of Mathematics
Stockholm University
E-Mail: taras.bodnar@math.su.se

Theory of Statistical Inference

Re-exam, 2020/12/02

The answers to the tasks should be clearly formulated and structured. All non-trivial steps need to be commented. The solutions should be given in English or Swedish.

The written exam (home exam) is divided into two parts. The first part considers the most central of the course concepts and it is related to standard problems. The second part consists of problems that requires a higher level of understanding, the ability to generalize and to combine methods. Each part consists of three problems and will worth a maximum of 50 points. In order to receive grades A-E, a minimum of 35 points is required in the first part. The second part is only graded for students passing the first part. Given a minimum of 35 points in the first part, the final grade is determined by the sum of regular points in both parts of the exam and bonus points according to the following table:

Grade	A	B	C	D	E	F
Points	100-90]	(90-80]	(79-70]	(69-60]	<60 and ≥ 35 in Part I	<35 in Part I

Up to 10 bonus points (i.e., in addition to the ordinary 100 points) are given for the active participation in the problem sessions. The bonus points can be used only for the improvement of the grade conditionally that the exam is passed on the regular basis, i.e., minimum 35 points is received in its first part.

Rules applied for the home exam

1. As usual, you first have to register for the written exam through student.ladok.se at the latest one week before the written exam. In case you do not register for the exam, your written solutions will not be corrected.

2. The home exam will be available on the course webpage at 09:00 on December, 2nd. It should be handed in here on the course webpage on the same day, at the latest at 16:00 (deadline).

3. The home exam should be handed in PDF format as a **single PDF file**. There are no restrictions regarding what your PDF should contain. For example, the PDF may be based on a Word document, a Latex document, or scanned nicely handwritten solutions. If you plan on "scanning" handwritten solutions using your mobile phone, I suggest downloading and using a "scanning app". If you scan and thereby obtain several PDF files, then there are many programs that can be used to merge several PDF files into one PDF file.

4. When writing the home exam you may use any literature and computer program.

5. If you are a student that has the right to prolonged writing time (förlängd skrivtid), then your deadline is one hour later, i.e. at 17:00. If you have such a right, please inform me per e-mail about it at the beginning of the exam.

6. You will be asked to state on the exam that you have written the exam without the assistance of any other person. Do not forget to write solution to Problem 0. Without its solution your exam will not be corrected.

7. The home exam will be of the same character as the planned exam. Hence, your solution should be of the same type as for usual exams.

8. Do not forget to read carefully the title page of the home exam for further information.

Problem 0 [0P]

In order to confirm that you did this exam alone, the following sentence should be written as a solution to problem 0:

"I, the author of this document, hereby guarantee that I have produced these solutions to this home exam without the assistance of any other person. This means that I have for example not discussed the solutions or the home exam with any other person."

Without this sentence, it would not be possible for me to correct the exam.

Part I:

Problem 1 [14P]

Let $\mathbf{X} = (X_1, X_2, X_3, X_4)^\top$ have a multinomial distribution with probability mass function given by

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4; \pi) = \frac{n!}{x_1!x_2!x_3!x_4}\pi^{x_1}(\pi/2)^{x_2}(\pi/3)^{x_3}(1 - \pi - \pi/2 - \pi/3)^{x_4}$$

for $x_1 + x_2 + x_3 + x_4 = n$ with $\pi \in (0, 6/11)$ and known n.

(a) Derive the expression of the log-likelihood function and the expression of the score function. **[4P]**

(b) Compute the maximum likelihood estimator $\hat{\pi}_{ML}$ of π. **[3P]**

(c) Derive the expression of the observed Fisher information. **[3P]**

(d) Derive the expression of the expected Fisher information. **[4P]**

Problem 2 [18P]

Suppose that we have an iid sample $X_{1:n} = (X_1, X_2, ..., X_n)$ from a Rayleigh distribution with density of X_i given by

$$f(x) = \theta x \exp\left(-\frac{\theta x^2}{2}\right) \qquad \text{for} \qquad x \geq 0 \quad \text{and} \quad \theta > 0\,.$$

(a) Derive the expression of the log-likelihood function and present the maximum likelihood estimator $\hat{\theta}_{ML}$ of θ. **[4P]**

(b) Determine the asymptotic distribution of $\hat{\theta}_{ML}$ as $n \to \infty$ and provide the analytical expressions of its parameters expressed as functions of θ. **[4P]**

(c) Compute the probability $\tau = P(X_1 \geq 2)$ as a function of θ. **[3P]**

(d) Derive the maximum likelihood estimator $\hat{\tau}_{ML}$ of τ. **[2P]**

(e) Determine the asymptotic distribution of $\hat{\tau}_{ML}$ as $n \to \infty$ and provide the analytical expressions of its parameters expressed as functions of τ. **[5P]**

Problem 3 [18P]

Suppose that we have an iid sample $X_{1:n} = (X_1, X_2, ..., X_n)$ from a Laplace distribution with density of X_i given by

$$f(x) = \frac{\theta}{2} \exp\left(-\theta|x|\right) \qquad \text{for} \qquad x \in \mathbb{R} \quad \text{and} \quad \theta > 0\,.$$

(a) Show that the conjugate prior for θ is given by the gamma distribution with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$, that is

$$f(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\beta\theta)$$

Determine the parameters in the corresponding posterior distribution. **[5P]**

(b) Compute the posterior mean of θ when the conjugate prior is used. **[2P]**

(c) Find the expression of the Jeffreys prior for θ and compute the corresponding posterior distribution. **[7P]**

(d) Calculate the posterior mean when the Jeffreys prior is used and compare it to the expression of the MLE estimator for θ. **[4P]**

Part II:

Problem 4 [13P]

Let X be a Gumbel-distributed random variable with the density function given by

$$f(x) = \frac{1}{\beta} \exp\left(-\frac{x}{\beta} - \exp\left(-\frac{x}{\beta}\right)\right) \qquad \text{for} \qquad x \in \mathbb{R} \quad \text{and} \quad \beta > 0$$

(a) Prove that the distribution function of X is given by **[4P]**

$$F(x) = \exp\left(-\exp\left(-\frac{x}{\beta}\right)\right) \qquad \text{for} \qquad x \in \mathbb{R} \quad \text{and} \quad \beta > 0$$

(b) Prove that $Z = X/\beta$ is a pivot quantity for β. **[3P]**

(c) Construct a two-sided 95% confidence interval for parameter β, when we observe $x = 0.5$ as a realization of X.**[6P]**

Problem 5 [20P]

Suppose that X has a Zero Inflated Geometric (ZIG) distribution with probability mass function given by

$$f(0) = p + (1-p)\eta \quad \text{and} \quad f(x) = (1-p)\eta(1-\eta)^x \qquad \text{for} \qquad x = 1, 2, \ldots \quad ,$$

where $p, \eta \in (0, 1)$.

(a) Reparameterize the ZIG distribution in terms of the new parameters **[2P]**

$$\pi = p + (1-p)\eta \quad \text{and} \quad \beta = \eta.$$

(b) In the rest of the problem, assume that an iid sample $X_{1:n} = (X_1, X_2, \ldots, X_n)$ is available from the ZIG distribution whose realizations are denoted by $x_{1:n} = (x_1, x_2, \ldots, x_n)$ where $x_1 = x_2 = \ldots = x_k = 0$ and $x_{k+1}, x_{k+2}, \ldots, x_n > 0$. Construct the log-likelihood function for $\boldsymbol{\theta} = (\pi, \beta)^\top$. **[4P]**

(c) Find the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{ML} = (\hat{\pi}_{ML}, \hat{\beta}_{ML})^\top$ of $\boldsymbol{\theta}$. **[4P]**

(d) Derive the observed Fisher information matrix. **[3P]**

(e) Construct an asymptotic 90% two-sided confidence interval for π and an asymptotic 90% two-sided confidence interval for β. **[4P]**

(f) Are the estimators $\hat{\pi}_{ML}$ and $\hat{\beta}_{ML}$ asymptotic independent? Explain your answer. **[3P]**

Problem 6 [17P]

Let $\boldsymbol{X}_{1:n} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)$ be an independent sample from a multivariate 3-dimensional normal distribution with density function of $\boldsymbol{X}_i = (X_{1,i}, X_{2,i}, X_{3,i})^\top$ given by

$$f(\boldsymbol{x}; \boldsymbol{\mu}) = (2\pi)^{-3/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right),$$

where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu \\ \mu \\ \mu \end{pmatrix} \in \mathbb{R}^3 \quad \text{and} \quad \boldsymbol{\Sigma} = \sigma^2 \mathbf{Q} \quad \text{with} \quad \mathbf{Q} = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

and ρ is assumed to be known. The notation $|\boldsymbol{\Sigma}|$ stands for the determinant of matrix $\boldsymbol{\Sigma}$ and $(\boldsymbol{x} - \boldsymbol{\mu})^\top$ denotes the transpose of $(\boldsymbol{x} - \boldsymbol{\mu})$.

(a) Compute the Jeffreys prior for μ and σ^2. **[5P]**

(b) Derive the expression of the posterior distribution for μ and σ^2 observing data $\boldsymbol{X}_{1:n}$ when the Jeffreys prior from part (a) is used. **[5P]**
Hint: *If you will not be able to compute the Jeffreys prior for* $\boldsymbol{\mu}$ *in part (a), then assume that it is a constant, that is use that* $f(\mu, \sigma^2) \propto \sigma^{-2}$.

(c) Compute a Bayesian point estimate for μ. **[2P]**

(d) Calculate the marginal posterior distribution for σ^2. **[5P]**

Taras Bodnar
Department of Mathematics
Stockholm University
E-Mail: taras.bodnar@math.su.se

Statistisk inferensteori

Omtentamen, 2020/12/02

Lösningarna på uppgifterna skall vara välstrukturerade och tydligt redovisade. Alla icke-triviala steg skall motiveras. Lösningarna får endast vara skrivna på engelska eller svenska.

Den här tentamen är skriftlig och är indelad i två delar. Den första delen fokuserar på de mest centrala delarna av kursen och består av standardproblem. Den andra delen består av problem som kräver en djupare förståelse, samt förmågan att generalisera och kombinera metoder. De två delarna består var och en av tre frågor och kan som mest ge 50 poäng. Totalt kan de två delarna alltså som mest ge 100 poäng. För att uppnå betygen A-E krävs minst 35 poäng på den första delen. Den andra delen rättas enbart om 35 poäng uppnåtts på den första delen, och betyget ges då av summan av antalet poäng på de två delarna enligt:

Betyg	A	B	C	D	E	F
Poäng	[100-90]	(90-80]	(79-70]	(69-60]	< 60 och ≥ 35 i Del I	< 35 i Del I

Bonusuppgifterna kan bidra med upp till 10 ytterligare poäng om betygen A-E redan uppnåtts. Bonuspoängen kan alltså endast användas för att höja ett redan godkänt betyg.

Regler för hemtentamen

1 Tentamensregistreringen sker som vanligt via student.ladok.se senast en vecka innan tentamensdagen. Om du inte registrerat dig så kommer dina lösningar inte att rättas.

2 Tentamen kommer läggas upp på kurshemsidan klockan 09:00 den 2:e december. Lösningarna skall lämnas in via kurshemsidan senast klockan 16:00 samma dag.

3 Lösningarna skall lämnas in i **pdf-format som en enda pdf-fil**. Inom ramarna för detta krav får du själv välja hur dokumentet med lösningarna ska genereras. Pdf-filen kan till exempel baseras på ett Worddokument, ett Latexdokument eller skannade lättlästa handskrivna lösningar. Om du föredrar att fotografera lösningarna med din mobiltelefon så rekommenderar vi att du använder en skanningsapp. Om du skannar lösningarna och får flera pdf-filer så finns det program som kan användas för att slå ihop flera pdf-filer till en enda pdf-fil.

4 Det går bra att använda valfri litteratur och datorprogram för att lösa uppgifterna.

5 Om du har fått förlängd skrivtid så infaller inlämningsdeadline en timme senare, det vill säga klockan 17:00.

6 I uppgift 0 kommer du ombedas att intyga att du har löst uppgifterna utan hjälp från någon annan. Om du inte gör detta kommer din lösning inte att rättas.

7 Hemtentamen kommer att utformas på samma sätt som den ordinarie tentamen. Din lösning bör därför utformas på samma sätt som för en vanlig salstentamen.

8 Glöm inte att läsa instruktionerna på hemtentamens titelsida noggrant för ytterligare information.

Problem 0 [0P]

För att intyga att du löst uppgifterna på egen hand, skriv följande mening som svar till uppgift 0:

"Jag, författaren till detta dokument, intygar härmed att jag har producerat lösningarna utan hjälp från någon annan. Detta innebär bland annat att jag inte på något sätt har diskuterat lösningarna till hemtentamen med någon annan person." Utan detta intygande kommer dina lösningar inte att rättas.

Del I:

Problem 1 [14P]

Låt $\mathbf{X} = (X_1, X_2, X_3, X_4)^\top$ vara multinomialfördelad med sannolikhetsfunktion

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4; \pi) = \frac{n!}{x_1!x_2!x_3!x_4}\pi^{x_1}(\pi/2)^{x_2}(\pi/3)^{x_3}(1 - \pi - \pi/2 - \pi/3)^{x_4}$$

för $x_1 + x_2 + x_3 + x_4 = n$ med $\pi \in (0, 6/11)$ and känt n.

(a) Ange loglikelihoodfunktionen och scorefunktionen. **[4P]**

(b) Hitta maximum-likelihood-skattaren $\hat{\pi}_{ML}$ av π. **[3P]**

(c) Hitta Fishers observerade informationsmatris. **[3P]**

(d) Hitta Fishers förväntade informationsmatris. **[4P]**

Problem 2 [18P]

Antag att $X_{1:n} = (X_1, X_2, ..., X_n)$ utgör ett stickprov med oberoende och likafördelade observationer från en Rayleighfördelning med täthetsfunktion

$$f(x) = \theta x \exp\left(-\frac{\theta x^2}{2}\right) \qquad \text{för} \qquad x \geq 0 \quad \text{och} \quad \theta > 0\,.$$

(a) Hitta loglikelihood-funktionen och bestäm maximum-likelihood-skattaren $\hat{\theta}_{ML}$ av θ. **[4P]**

(b) Hitta den asymptotiska fördelningen för $\hat{\theta}_{ML}$ då $n \to \infty$ och ange de analytiska uttrycken för gränsfördelningens parametrar som funktioner av θ. **[4P]**

(c) Bestäm sannolikheten $\tau = P(X_1 \geq 2)$ som ett uttryck av θ. **[3P]**

(d) Hitta maximum likelihood-skattaren $\hat{\tau}_{ML}$ av τ. **[2P]**

(e) Bestäm den asymptotiska fördelningen för $\hat{\tau}_{ML}$ då $n \to \infty$ och ange de analytiska uttrycken för gränsfördelningens parametrar som funktioner av τ. **[5P]**

Problem 3 [18P]

Låt $X_{1:n} = (X_1, X_2, ..., X_n)$ utgöra ett stickprov med oberoende likafördelade observationer från en Laplacefördelning med täthet

$$f(x) = \frac{\theta}{2} \exp\left(-\theta|x|\right) \qquad \text{for} \qquad x \in \mathbb{R} \quad \text{and} \quad \theta > 0\,.$$

(a) Visa att den konjugerande apriorifördelningen för θ ges av en Gammafördelning med shape-parameter $\alpha > 0$ och rate-parameter $\beta > 0$ vars täthet ges av

$$f(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\beta\theta),$$

och bestäm den motsvarande aposteriorifördelningens parametrar. **[5P]**

(b) Hitta aposterioriväntevärdet av θ om apriorifördelningen ges av Gammafördelningen i (a). **[2P]**

(c) Bestäm Jeffrey's apriorifördelning för θ och hitta motsvarande aposteriorifördelning. **[7P]**

(d) Ange aposterioriväntevärdet om vi använder Jeffrey's apriorifördelning. Jämför ditt svar med maximum-likelihood-skattaren av θ. **[4P]**

Del II:

Problem 4 [13P]

Låt X vara en Gumbelfördelad slumpvariabel med täthet

$$f(x) = \frac{1}{\beta} \exp\left(-\frac{x}{\beta} - \exp\left(-\frac{x}{\beta}\right)\right) \qquad \text{för} \qquad x \in \mathbb{R} \quad \text{och} \quad \beta > 0$$

(a) Visa att fördelningsfunktionen för X ges av **[4P]**

$$F(x) = \exp\left(-\exp\left(-\frac{x}{\beta}\right)\right) \qquad \text{för} \qquad x \in \mathbb{R} \quad \text{och} \quad \beta > 0$$

(b) Visa att $Z = X/\beta$ är en pivotvariabel för β. **[3P]**

(c) Konstruera ett tvåsidigt konfidensintervall för parametern β baserat på realisationen $x = 0.5$ av X.**[6P]**

Problem 5 [20P]

Låt X följa en Zero-Inflated Geometrisk fördelning (ZIG-fördelning) med sannolikhetsfunktion

$$f(0) = p + (1-p)\eta \quad \text{and} \quad f(x) = (1-p)\eta(1-\eta)^x \qquad \text{för} \qquad x = 1, 2, ... \quad ,$$

där $p, \eta \in (0, 1)$.

(a) Omparametrisera ZIG-fördelningen till de nya parametrarna **[2P]**

$$\pi = p + (1-p)\eta \quad \text{and} \quad \beta = \eta.$$

(b) I deluppgifterna (b)-(f) antar vi att vi har tillgång till realisationen $x_{1:n} = (x_1, x_2, ..., x_n)$ av ett iid stickprov $X_{1:n} = (X_1, X_2, ..., X_n)$ från ZIG-fördelningen. Elementen i $x_{1:n}$ antas uppfylla $x_1 = x_2 = ... = x_k = 0$ och $x_{k+1}, x_{k+2}, ..., x_n > 0$. Hitta loglikelihood-funktionen för $\boldsymbol{\theta} = (\pi, \beta)^\top$. **[4P]**

(c) Hitta maximum-likelihood-skattaren $\hat{\boldsymbol{\theta}}_{ML} = (\hat{\pi}_{ML}, \hat{\beta}_{ML})^\top$ av $\boldsymbol{\theta}$. **[4P]**

(d) Hitta Fisher's observerade informationsmatris. **[3P]**

(e) Konstruera ett asymptotiskt 90% tvåsidigt konfidensintervall för π och ett asymptotiskt 90% tvåsidigt konfidensintervall för β. **[4P]**

(f) är skattarna $\hat{\pi}_{ML}$ och $\hat{\beta}_{ML}$ asymptotiskt oberoende? Motivera din slutsats. **[3P]**

Problem 6 [17P]

Låt $\boldsymbol{X}_{1:n} = (\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n)$ vara ett oberoende stickprov från en multivariat 3-dimensionell normalfördelning, det vill säga täthetsfunktionen för $\boldsymbol{X}_i = (X_{1,i}, X_{2,i}, X_{3,i})^\top$ ges av

$$f(\boldsymbol{x}; \boldsymbol{\mu}) = (2\pi)^{-3/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) ,$$

där

$$\boldsymbol{\mu} = \begin{pmatrix} \mu \\ \mu \\ \mu \end{pmatrix} \in \mathbb{R}^3 \quad \text{och} \quad \boldsymbol{\Sigma} = \sigma^2 \mathbf{Q} \quad \text{med} \quad \mathbf{Q} = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

och ρ är en känd konstant. Här betecknar $|\boldsymbol{\Sigma}|$ determinanten av matrisen $\boldsymbol{\Sigma}$ och $(\boldsymbol{x}-\boldsymbol{\mu})^\top$ betecknar transponatet av $(\boldsymbol{x}-\boldsymbol{\mu})$.

(a) Hitta Jeffrey's apriorifördelning för μ och σ^2. **[5P]**

(b) Härled aposteriorifördelningen för μ och σ^2 baserat på observerad data $\boldsymbol{X}_{1:n}$ om vi använder Jeffreys apriorifördelning från del (a). **[5P]**
Hälp: *Om du inte lyckas hitta Jeffrey's apriorifördelning för* $\boldsymbol{\mu}$ *i (a) så kan du anta att dess täthet är konstant, det vill säga* $f(\mu, \sigma^2) \propto \sigma^{-2}$.

(c) Bestäm en Bayesiansk punktskattare för μ. **[2P]**

(d) Ange den marginella aposteriorifördelningen för σ^2.**[5P]**