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

Theory of Statistical Inference

Exam, 2020/10/29

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 October, 29th. 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 [13P]

Find the minimum sufficient statistic in the following situation:

(a) An independent sample $X_{1:n} = (X_1, X_2, ..., X_n)$ from a geometric distribution with probability mass function of X_i given by **[5P]**

$$f(x;\pi) = P(X_i = x) = (1-\pi)^{x-1}\pi \qquad \text{for} \qquad x = 1, 2, 3, ... \quad \text{and} \quad \pi \in (0,1)\,.$$

(b) An independent sample $X_{1:n} = (X_1, X_2, ..., X_n)$ from a Poisson distribution with probability mass function of X_i given by **[4P]**

$$f(x;\lambda) = P(X_i = x) = \frac{\lambda^x}{x!}\exp(-\lambda) \qquad \text{for} \qquad x = 0, 1, 2, ... \quad \text{and} \quad \lambda > 0.$$

(c) An independent sample $X_{1:n} = (X_1, X_2, ..., X_n)$ from an exponential distribution with density of X_i given by **[4P]**

$$f(x;\lambda) = \lambda \exp(-\lambda x) \qquad \text{for} \qquad x \geq 0 \quad \text{and} \quad \lambda > 0.$$

Problem 2 [16P]

Let $X \sim Bin(n, \pi)$ (binomial distribution with parameters n and π) with $\pi \in (0,1)$ and known n. Derive the expression of the Jeffreys prior for π in the following parametrization:

(a) Original parametrization with parameter π. **[4P]**

(b) Parametrization with $\phi = logit(\pi)$ as a parameter. **[4P]**

(c) Parametrization with $\eta = arcsin(\sqrt{\pi})$ as a parameter. **[4P]**

(d) Compute the posteriors for π, ϕ, and η (in total three posteriors should be computed) under each parametrization considered in parts (a), (b), and (c). **[4P]**

Problem 3 [21P]

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

$$f(x) = \beta\alpha x^{\alpha-1}\exp(-\beta x^\alpha) \qquad \text{for} \qquad x \geq 0 \quad \text{and} \quad \alpha, \beta > 0\,,$$

where the shape parameter α is assumed to be known.

(a) Derive the expression of the log-likelihood function and present the maximum likelihood estimator $\hat{\beta}_{ML}$ of β. **[3P]**

(b) Determine the asymptotic distribution of $\hat{\beta}_{ML}$ as $n \to \infty$ and provide the analytical expressions of its parameters expressed as functions of α and β. **[3P]**

(c) Use the results of part (b) to construct the Wald statistic T_1 for testing the null hypothesis $H_0: \beta = \beta_0$ against $H_1: \beta \neq \beta_0$. **[3P]**

(d) Compute the score statistic T_2 for testing the hypotheses in part (c). **[5P]**

(e) Compute the likelihood ratio statistic T_3 for testing the hypotheses in part (c). **[5P]**

(f) What are the asymptotic distributions of the T_1, T_2, and T_3 under the null hypothesis in part (c)? **[2P]**

Part II:

Problem 4 [13P]

Let X be uniformly distributed on $[0, \theta]$.

(a) Prove that $Z = X/\theta$ is a pivot quantity for θ. Derive the distribution of Z. **[3P]**

(b) Construct a two-sided $(1-\alpha)$ confidence interval for parameter θ.**[6P]**

(c) Let Y be uniformly distributed on $[\theta, 1]$. Find a pivot quantity for parameter θ and derive its distribution.**[4P]**

Problem 5 [17P]

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

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4; \pi_1, \pi_2, \pi_3) = \frac{n!}{x_1!x_2!x_3!x_4}\pi_1^{x_1}\pi_2^{x_2}\pi_3^{x_3}(1-\pi_1-\pi_2-\pi_3)^{x_4}$$

for $x_1 + x_2 + x_3 + x_4 = n$ with $\pi_1, \pi_2, \pi_3 \in (0,1)$ and known n. We want to test the null hypothesis:

$$H_0: \pi_1 = \pi_2 = \pi_3. \tag{1}$$

(a) Formulate the alternative hypothesis H_1. **[1P]**

(b) Derive the generalized likelihood ratio statistic for testing H_0 in (1). **[12P]**

(c) Determine the distribution of the test statistics derived in part (b). **[1P]**

(d) Perform the generalized likelihood ratio test when the following data $(x_1, x_2, x_3, x_4) = (12, 19, 15, 20)$ are observed. **[3P]**

Problem 6 [20P]

Let $\boldsymbol{X}_{1:n} = (\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n)$ 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

$$f(\boldsymbol{x}; \boldsymbol{\mu}) = (2\pi)^{-3/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right), \quad \boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^\top \in \mathbb{R}^3,$$

with known covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1,\ldots,3}$. 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 $\boldsymbol{\mu}$. **[5P]**

(b) Derive the expression of the posterior distribution for $\boldsymbol{\mu}$ 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 $f(\boldsymbol{\mu}) \propto 1$.*

(c) Calculate two Bayesian point estimates for $\boldsymbol{\mu}$. **[4P]**

(d) Using the posterior of part (b), derive the marginal posterior $f(\mu_1|\boldsymbol{X}_{1:n})$ for the first component μ_1 of $\boldsymbol{\mu}$ and use it to construct a $(1-\alpha)$ credible interval for μ_1.**[6P]**

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

Statistisk inferensteori

Tentamen, 2020/10/29

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 ske 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 29:e oktober. 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 [13P]

Hitta en minimalt tillräcklig statistika för följande modeller:

(a) Vi har ett oberoende stickprov $X_{1:n} = (X_1, X_2, ..., X_n)$ från en geometrisk fördelning vars sannolikhetsfunktion ges av **[5P]**

$$f_{X_i}(x;\pi) = P(X_i = x) = (1-\pi)^{x-1}\pi \qquad \text{för} \qquad x = 1, 2, 3, ... \quad \text{och} \quad \pi \in (0,1)\,.$$

(b) Vi har ett oberoende stickprov $X_{1:n} = (X_1, X_2, ..., X_n)$ från en Poissonfördelning vars sannolikhetsfunktion ges av **[4P]**

$$f_{X_i}(x;\lambda) = P(X_i = x) = \frac{\lambda^x}{x!}\exp(-\lambda) \qquad \text{för} \qquad x = 0, 1, 2, ... \quad \text{och} \quad \lambda > 0.$$

(c) Vi har ett oberoende stickprov $X_{1:n} = (X_1, X_2, ..., X_n)$ från en exponentialfördelning vars täthetsfunktion ges av **[4P]**

$$f_{X_i}(x;\lambda) = \lambda \exp(-\lambda x) \qquad \text{for} \qquad x \geq 0 \quad \text{and} \quad \lambda > 0.$$

Problem 2 [16P]

Låt $X \sim Bin(n, \pi)$ (här betecknar $Bin(n, \pi)$ binomialfördelningen med parametrar n och π) med $\pi \in (0, 1)$ och känt n. Hitta Jeffrey's apriorifördelning för π under följande parametriseringar:

(a) Den ursprungliga parametriseringen (med parametern π). **[4P]**

(b) En omparametrisering till $\phi = logit(\pi)$. **[4P]**

(c) En omparametrisering till $\eta = arcsin(\sqrt{\pi})$. **[4P]**

(d) Bestäm aposteriorifördelingarna för π, ϕ, och η under motsvarande parametrisering i (a), (b), och (c) (d.v.s. totalt efterfrågas tre aposteriorifördelningar). **[4P]**

Problem 3 [21P]

Antag att vi har ett oberoende likafördelat stickprov $X_{1:n} = (X_1, X_2, ..., X_n)$ från en Weibullfördelning vars täthetsfunktion ges av

$$f(x) = \beta\alpha x^{\alpha-1}\exp(-\beta x^\alpha)\,, \quad \alpha, \beta > 0\,.$$

där shape-parametern α är en känd konstant.

(a) Härled ett uttryck för log-likelihood-funktionen av β och hitta maximum-likelihood-skattaren $\hat{\beta}_{ML}$ av β. **[3P]**

(b) Ange den asymptotiska fördelningen för $\hat{\beta}_{ML}$ då $n \to \infty$. Den asymptotiska fördelningens parametrar skall anges som uttryck på sluten form. **[4P]**

(c) Använd del (b) för att konstruera Wald-statistikan T_1 för testet av $H_0: \beta = \beta_0$ mot $H_1: \beta \neq \beta_0$. **[3P]**

(d) Beräkna score-statistikan T_2 för testet av hypoteserna i (c). **[4P]**

(e) Beräkna likelihood-ratio-statistikan T_3 för testet av hypoteserna i (c). **[5P]**

(f) Hita de asymptotiska fördelningarna för T_1, T_2, och T_3 under nollhypotesen i (c). **[2P]**

Del II:

Problem 4 [13P]

Låt X vara uniformt fördelad på $[0, \theta]$.

(a) Visa att $Z = X/\theta$ är en pivot-variabel för θ. Hitta fördelningen för Z. **[4P]**

(b) Konstruera ett tvåsidigt $(1-\alpha)$ konfidensintervall för parametern θ.**[6P]**

(c) Låt Y vara uniformt fördelad på $[\theta, 1]$. Hitta en pivot-variabel för θ och ange dess fördelning.**[3P]**

Problem 5 [17P]

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

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4; \pi_1, \pi_2, \pi_3) = \frac{n!}{x_1!x_2!x_3!x_4}\pi_1^{x_1}\pi_2^{x_2}\pi_3^{x_3}(1 - \pi_1 - \pi_2 - \pi_3)^{x_4}$$

för $x_1 + x_2 + x_3 + x_4 = n$ med $\pi_1, \pi_2, \pi_3 \in (0, 1)$ och känt n.Vi vill testa nollhypotesen:

$$H_0: \pi_1 = \pi_2 = \pi_3. \tag{2}$$

(a) Formulera alternativhypotesen H_1. **[1P]**

(b) Hitta den generalisade likelihood-ratio-statistikan för H_0 i (2). **[12P]**

(c) Hitta fördelningen för test-statistikan från (b). **[1P]**

(d) Utför ett generaliserat likelihood-ratio-test av H_0 baserat på observationen $(x_1, x_2, x_3, x_4) = (12, 19, 15, 20)$. **[3P]**

Problem 6 [20P]

Låt $\boldsymbol{X}_{1:n} = (\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n)$ utgöra 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

$$f(\boldsymbol{x}; \boldsymbol{\mu}) = (2\pi)^{-3/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right), \quad \boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^\top \in \mathbb{R}^3,$$

med känd variansmatris $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1,...,3}$. Här betecknar $|\boldsymbol{\Sigma}|$ determinanten av matrisen $\boldsymbol{\Sigma}$ och $(\boldsymbol{x} - \boldsymbol{\mu})^\top$ betecknar transponatet av $(\boldsymbol{x} - \boldsymbol{\mu})$.

(a) Bestäm Jeffrey's apriorifördelning för $\boldsymbol{\mu}$. **[5P]**

(b) Hitta aposteriorifördelningen för $\boldsymbol{\mu}$ baserad på observerad data $\boldsymbol{X}_{1:n}$ om apriorifördelningen ges av Jeffrey's apriorifördelning från del (a). **[5P]**
Hjälp: *Om du inte lyckas hitta Jeffrey's apriorifördelning i del (a) så kan du anta att aprioritätheten är konstant, det vill säga $f(\boldsymbol{\mu}) \propto 1$.*

(c) Ange två Bayesianska punktskattare av $\boldsymbol{\mu}$. **[4P]**

(d) Hitta den marginella aposteriori-tätheten $f(\mu_1|\boldsymbol{X}_{1:n})$ för den första komponenten μ_1 av $\boldsymbol{\mu}$ baserat på aposteriorifördelningen från (b). Använd sedan ditt svar för att konstruera ett $(1 - \alpha)$ kredibilitetsintervall för μ_1.**[6P]**