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

Statistical models

Exam, 2021/05/25

The solution should be given in English. The answers to the tasks should be clearly formulated and structured. All non-trivial steps need to be explained.

The grades will be given due to the following table

Grade	A	B	C	D	E	F
Points	100-90	89-80	79-70	69-60	59-50	< 50
Percent	100-90%	89-80%	79-70%	69-60%	59-50%	< 50%

The final grade is determined by the sum of regular points and bonus points. In order to pass the exam, students have to receive at least 50% of all points in both parts of the exam, i.e. at least 50% of all points for theoretical questions (Problems 5 and 6) and at least 50% of all points for computational problems (Problems 1-4).

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 after the exam is passed on the regular basis.

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 eight days 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 here on the course webpage at 08:00 on May, 25th. It should be handed in here on the course webpage on the same day, at the latest at 15:00 (deadline).

3 The home exam should be handed in in PDF format (i.e. one 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. For example, one can use the following webpage: https://jpg2pdf.com/

4 When writing the home exam you may use any literature and computer program. However, all non-trivial steps need to be explained.

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 16:00.

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.

Problem 1 [15P]

Find the expression of a minimal sufficient statistic for each of the following statistical models:

(a) Independent sample $Y_1, ..., Y_n$ from a Bernoulli distribution $\text{Be}(p)$ with $p \in (0, 1)$. **[2P]**

(b) Independent sample $Y_1, ..., Y_n$ from a Poisson distribution $\text{Po}(\lambda)$ with $\lambda > 0$. **[2P]**

(c) Single observation vector $\mathbf{Y} = (Y_1, Y_2, Y_3)$ from a multinomial distribution with probability mass function given by

$$\mathbb{P}(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3; p_1, p_2, p_3) = \frac{n!}{y_1! y_2! y_3!} p_1^{y_1} p_2^{y_2} p_3^{y_3} \quad \text{for} \quad y_1 + y_2 + y_3 = n$$

with known n and $p_1 + p_2 + p_3 = 1$ for $p_1, p_2, p_3 \in (0, 1)$. **[3P]**

(d) Independent sample $Y_1, ..., Y_n$ from a log-normal distribution with density of Y_i given by **[3P]**

$$f(y; \mu, \sigma) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln y - \mu)^2}{2\sigma^2}\right) \quad \text{for} \quad y > 0 \quad \text{and} \quad \mu \in \mathbb{R}, \sigma > 0.$$

(e) Independent sample $Y_1, ..., Y_n$ from a Laplace distribution with density of Y_i given by**[2P]**

$$f(y) = \frac{1}{2\sigma} \exp\left(-\frac{|y - \mu|}{\sigma}\right) \quad \text{for} \quad y \in \mathbb{R} \quad \text{and} \quad \sigma > 0\,,$$

where $\mu \in \mathbb{R}$ is assumed to be known.

(f) Single observation vector $\mathbf{Y} = (Y_1, Y_2, ..., Y_k)$ from a Dirichlet distribution with density given by**[3P]**

$$f(y_1, ..., y_k; \alpha_1, ..., \alpha_k) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k y_i^{\alpha_i - 1} \qquad \text{for} \qquad y_i \in (0, 1) \quad \text{with} \quad \sum_{i=1}^k y_i = 1$$

where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_k)^T$ is the vector of unknown parameters with $\alpha_i > 0$, $i = 1, ..., k$.

Problem 2 [15P]

Let Y be a normally distributed random variable with the density function given by

$$f(y; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right) \qquad \text{for} \qquad y \in \mathbb{R} \quad \text{and} \quad \mu \in \mathbb{R}, \sigma > 0\,.$$

Using that the normal distribution belongs to the exponential family, derive the expression of the skewness of the random variable Y.

Hint: The skewness of a random variable Y with finite third moment is given by

$$\text{Skewness}(Y) = \frac{\mathbb{E}\left((Y - \mathbb{E}(Y))^3\right)}{(\mathbb{V}ar(Y))^{3/2}}$$

Problem 3 [22P]

Let $y_1, y_2, \ldots, y_n$ be an iid. sample from a Borel distribution with probability mass function given by

$$\mathbb{P}(Y = y; \alpha) = \frac{1}{y!}(\alpha y)^{y-1} e^{-\alpha y} \qquad \text{for} \qquad y = 1, 2, ...$$

where $\alpha \in (0, 1)$ is an unknown parameter.

(a) Show that the above distribution is a one-parameter exponential distribution and find its canonical parameter θ.[**3P**]

(b) Prove that the maximum likelihood estimate for α is given by [**4P**]

$$\hat{\alpha}_{MLE} = 1 - \frac{1}{\bar{y}} \qquad \text{with} \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\,.$$

(c) Derive the asymptotic distribution of $\hat{\alpha}_{MLE}$ by using that [**4P**]

$$\sqrt{n}\left(\bar{y} - \frac{1}{1-\alpha}\right) \overset{as}{\to} \mathcal{N}\left(0, \frac{\alpha}{(1-\alpha)^3}\right) \qquad \text{as} \qquad n \to \infty.$$

(d) Construct a 95% two-sided confidence interval for α. [**2P**]

(e) Find the expected Fisher information $I(\hat{\alpha}_{MLE})$. [**3P**]

(f) Determine the likelihood ratio $L(\alpha_0)/L(\hat{\alpha}_{MLE})$ when α_0 is a given fixed value. [**4P**]

(g) Derive the saddlepoint approximation for the distribution of $\hat{\alpha}_{MLE}$ in a point α_0 without determining the normalization constant. [**2P**]

Problem 4 [23P]

Let Y_1 and Y_2 be two independent random variables with $Y_1 \sim MB(\alpha_1)$ (Maxwell–Boltzmann distribution with parameter α_1) and $Y_2 \sim MB(\alpha_2)$ whose densities are given by

$$f(y_1; \alpha_1) = \sqrt{\frac{2}{\pi}}\frac{y_1^2}{\alpha_1^3} e^{-y_1^2/(2\alpha_1^2)} \qquad \text{for} \qquad y_1 > 0 \quad \text{and} \quad \alpha_1 > 0$$

and

$$f(y_2; \alpha_2) = \sqrt{\frac{2}{\pi}}\frac{y_2^2}{\alpha_2^3} e^{-y_2^2/(2\alpha_2^2)} \qquad \text{for} \qquad y_2 > 0 \quad \text{and} \quad \alpha_2 > 0\,,$$

respectively.

(a) Assume that we have two random samples $Y_{1,i}$, $i = 1, ..., n$, and $Y_{2,j}$, $j = 1, ..., n$ from the distribution of Y_1 and Y_2, respectively. Derive the joint density of $Y_{1,i}$, $i = 1, ..., n$, and $Y_{2,j}$,

$j = 1, ..., n$ under the assumption that $Y_{1,1}$, ..., $Y_{1,n}$, $Y_{2,1}$, ..., $Y_{2,n}$ are mutually independent. **[2P]**

(b) Prove that a canonical statistic is $t(Y_{1,1}..., Y_{1,n}, Y_{2,1}, ..., Y_{2,n}) = (v, u)^T$ with $v = \sum_{i=1}^n Y_{1,i}^2/2$ and $u = \sum_{i=1}^n Y_{1,i}^2/2 + \sum_{j=1}^n Y_{2,j}^2/2$. Determine the canonical parameter $\boldsymbol{\theta}$. **[2P]**

(c) The aim is to test the model reduction hypothesis:

$$H_0 : \psi = 0 \quad \text{against} \quad H_1 : \psi \neq 0,$$

with $\psi = 1/\alpha_2^2 - 1/\alpha_1^2$. Calculate the marginal density $f_0(u)$ and specify the conditional distribution $f_0(v|u)$ under H_0. **[6P]**
Hint: *Use that if $Z \sim MB(\alpha)$, then Z^2 has the gamma distribution with shape $3/2$ and scale $2\alpha^2$, and the properties of the gamma distribution.*

(d) Calculate the p-value of the test from (c) if $n = 2$ and the realizations of $\sum_{i=1}^n Y_{1,i}^2$ and $\sum_{j=1}^n Y_{2,j}^2$ are $\sum_{i=1}^n y_{1,i}^2 = 1.5$ and $\sum_{j=1}^n y_{2,i}^2 = 0.5$, respectively. Is the null hypothesis rejected at significance level 0.1? **[6P]**

(e) Derive the statistic of the deviance test for the null hypothesis from (c). What is the asymptotic null distribution of this test statistic? **[6P]**

(f) Perform the deviance test from (e) at significance level 0.1 by using that $n = 200$, and the realizations of $\sum_{i=1}^n Y_{1,i}^2$ and $\sum_{j=1}^n Y_{2,j}^2$ are $\sum_{i=1}^n y_{1,i}^2 = 110$ and $\sum_{j=1}^n y_{2,j}^2 = 90$, respectively. **[1P]**
Hint: *Important quantiles of the χ^2-distribution at various degrees of freedom are:*

x	1	2	3	4	5
$\chi^2_{0.9}(\text{df} = x)$	2.71	4.61	6.25	7.78	9.24
$\chi^2_{0.95}(\text{df} = x)$	3.84	5.99	7.81	9.49	11.07
$\chi^2_{0.975}(\text{df} = x)$	5.02	7.38	9.35	11.14	12.83

Problem 5 [15P]

Let $Y_1, Y_2, ..., Y_n$, $n > 2$, be an iid. sample of a normally distributed random variable Y with the density function given by

$$f(y; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \qquad \text{for} \qquad y \in \mathbb{R} \quad \text{and} \quad \mu \in \mathbb{R}, \sigma > 0\,.$$

Let

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2.$$

Prove that

$$\left(\frac{Y_1 - \bar{Y}}{S}, \frac{Y_2 - \bar{Y}}{S}, ..., \frac{Y_n - \bar{Y}}{S}\right)^\top$$

is independent of $\bar{Y}$ and S.

Problem 6 [10P]

Let $Y_1, Y_2, ..., Y_n$, $n > 2$, be an iid. sample of an inverse-gamma distributed random variable Y with density given by

$$f(y;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} \exp\left(-\beta/y\right) \qquad \text{for} \qquad y > 0 \quad \text{and} \quad \alpha, \beta > 0.$$

Find the profile likelihood function for α.

Some formulas

- *Hölder's Inequality*: If S is a measurable subset of $\mathbb{R}^n$ with the Lebesgue measure, and f and g are measurable real- or complex-valued functions on S, then Hölder's inequality is

$$\int_S |f(x)g(x)|dx \leq \left(\int_S |f(x)|^p dx\right)^{\frac{1}{p}} \left(\int_S |g(x)|^q dx\right)^{\frac{1}{q}}.$$

- *Moment-generating function* of the canonical statistics t:

$$M(\psi) = \mathrm{E}_\theta(\exp(\psi^T t)) = \frac{C(\theta+\psi)}{C(\theta)}.$$

- The *saddlepoint approximation* of a density $f(t) = f(t;\theta_0)$ in an exponential family is

$$f(t;\theta_0) = (2\pi)^{-\frac{k}{2}} \det(V_t(\hat{\theta}(t)))^{-\frac{1}{2}} \frac{C(\hat{\theta}(t))}{C(\theta_0)} \exp\left((\theta_0 - \hat{\theta}(t))^T t\right).$$

  The corresponding approximation of the structure function is

$$g(t) \approx (2\pi)^{-\frac{k}{2}} \det(V_t(\hat{\theta}(t)))^{-\frac{1}{2}} C(\hat{\theta}(t)) \exp\left(-\hat{\theta}(t)^T t\right).$$

- The saddlepoint approximation for the density of the ML estimator $\hat{\psi} = \hat{\psi}(t)$ in any smooth parametrization of a regular exponential family is

$$f(\hat{\psi};\psi_0) \approx (2\pi)^{-\frac{k}{2}} \sqrt{\det I(\hat{\psi})} \cdot \frac{L(\psi_0)}{L(\hat{\psi})}.$$

- *Principle of exact tests of $H_0: \psi = 0$ vs. $H_1: \psi \neq 0$*

  1. Use v as test statistic, with null distribution density $f_0(v|u)$
  2. Reject H_0, if the probability to observe $v_{obs}|u_{obs}$ or a more extreme value (towards the alternative) is too unlikely. One general approach to formulate this p-value is

$$p = Pr(f_0(v|u_{obs}) \leq f_0(v_{obs}|u_{obs})),$$

     and reject if, say, $p < \alpha$. Note: p can be calculated as

$$\int_{\{v: f_0(v|u_{obs}) \leq f_0(v_{obs}|u_{obs})\}} f_0(v|u_{obs}) dv.$$

     If v is discrete the integration is replaced by a summation.

- *Large sample approximation of the exact test*: In an exponential family, with parametrization using (θ_u, ψ), canonical statistic $t = (u, v)$ and null-hypothesis $H_0: \psi = 0$ the score test is

$$W_u = (v - \mu_v(\hat{\theta}_u, 0))^T \left(I(\hat{\theta}_u, 0)^{-1}\right)_{vv} (v - \mu_v(\hat{\theta}_u, 0))$$

- *Asymptotically equivalent tests*:
  - Deviance
$$W = 2\log\frac{L(\hat\theta)}{L(\hat\theta_0)},$$
    where $\hat\theta = (\hat\psi, \hat\lambda)$ and $\hat\theta_0 = (\psi_0, \hat\lambda_0 = \hat\lambda(\psi_0))$.
  - Quadratic form
$$W_e^* = (\hat\theta_0 - \hat\theta)^T I(\hat\theta_0)(\hat\theta_0 - \hat\theta)$$
  - Score test
$$W_u = U(\hat\theta_0)^T I(\hat\theta_0)^{-1} U(\hat\theta_0)$$
  - Wald test
$$W_e = (\hat\psi - \psi_0)^T I^{\psi\psi}(\hat\theta)^{-1}(\hat\psi - \psi_0)$$
- *Likelihood equations in the GLM*: The likelihood equation system for a GLM with canonical link function $\theta \equiv \eta = X\beta$ is
$$X^T[y - \mu(\beta)] = 0.$$
  For a model with non-canonical link, the equation system is
$$X^T G'(\mu(\beta))^{-1} V_y(\mu(\beta))^{-1}[y - \mu(\beta)] = 0,$$
  where $G'(\mu)$ and $V_y(\mu)$ are $n \times n$ diagonal matrices with diagonal elements $g'(\mu_i)$ and $v_y(\mu_i) = \mathrm{Var}(y_i; \mu_i)$, respectively.
- Deviance (or residual deviance) for a GLM
$$D = D(\mathbf{y}, \boldsymbol{\mu}(\hat{\boldsymbol{\beta}})) = 2[\log(L(\mathbf{y};\mathbf{y})) - \log(L(\boldsymbol{\mu}(\hat{\boldsymbol{\beta}});\mathbf{y}))]$$
- The *observed and expected information matrices for a GLM* with canonical link function are identical and are given by
$$J(\beta) = I(\beta) = X^T V_y(\mu(\beta))X,$$
  which is a weighted sums of squares of the regressors. With non-canonical link the Fisher information is given by
$$\begin{aligned} I(\beta) &= \left(\frac{\partial\theta}{\partial\beta}\right)^T V_y(\mu(\beta)) \left(\frac{\partial\theta}{\partial\beta}\right) \\ &= X^T G'(\mu(\beta))^{-1} V_y(\mu(\beta))^{-1} G'(\mu(\beta))^{-1} X. \end{aligned}$$
- Exponential family with an additional *dispersion parameter*:
$$f(y_i; \theta_i, \phi) = \exp\left(\frac{\theta_i y_i - \log C(\theta_i)}{\phi}\right) h(y_i; \phi),$$
  where $C(\theta_i)$ is the normalization factor in the special linear exponential family where $\phi = 1$.

- *Jacobian matrix*: Let $g : \mathbb{R}^n \to \mathbb{R}^n$ and $y = g(x) = (g_1(x), \dots, g_n(x))^T$ with $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ then

$$\left(\frac{\partial y}{\partial x}\right) = \begin{bmatrix} \frac{\partial g_1(x)}{\partial x_1} & \dots & \frac{\partial g_1(x)}{\partial x_n} \\ & \ddots & \\ \frac{\partial g_n(x)}{\partial x_1} & \dots & \frac{\partial g_n(x)}{\partial x_n} \end{bmatrix}$$

- *Score function:*

$$U(\theta) = \frac{d}{d\theta} \log L(\theta),$$

where $L(\theta)$ is the likelihood function.

- *Observed information:*

$$J(\theta) = -\frac{d^2}{d\theta d\theta^T} \log L(\theta)$$

- *Expected information:*

$$I(\theta) = -E_\theta \left(\frac{d^2}{d\theta d\theta^T} \log L(\theta)\right)$$

- *Reparametrization lemma*: If ψ and $\theta = \theta(\psi)$ are two equivalent parametrizations of the same model then the score functions are related by

$$U_\psi(\psi; y) = \left(\frac{\partial \theta}{\partial \psi}\right)^T U_\theta(\theta(\psi); y).$$

Furthermore, the expected information matrices are related by

$$I_\psi(\psi) = \left(\frac{\partial \theta}{\partial \psi}\right)^T I_\theta(\theta(\psi)) \left(\frac{\partial \theta}{\partial \psi}\right)$$

and the observed information at the MLE by

$$J_\psi(\hat{\psi}) = \left(\frac{\partial \theta}{\partial \psi}\right)^T J_\theta(\theta(\hat{\psi})) \left(\frac{\partial \theta}{\partial \psi}\right).$$

- *Change of variables in multivariate density:* Let $\mathbf{X}$ has a density $f_{\mathbf{X}}(\mathbf{x})$ and let $\mathbf{Y} = g(\mathbf{X})$ with $g : \mathbb{R}^k \to \mathbb{R}^k$. Then

$$f_{\mathbf{Y}}(\mathbf{y}) = det \left(\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}\right)^{-1} f_{\mathbf{X}}(\mathbf{x}(\mathbf{y}))$$

- *Taylor's theorem in several variables:* Suppose $f : \mathbb{R}^n \to \mathbb{R}$ be a k times differentiable function at the point $\boldsymbol{a} \in \mathbb{R}^n$. Then

$$f(\boldsymbol{x}) = \sum_{|\alpha| \le k} \frac{D_\alpha f(\boldsymbol{a})}{\alpha!} (\boldsymbol{x} - \boldsymbol{a})^\alpha + R_{\boldsymbol{a},k}(\mathbf{h}),$$

where $R_{\mathbf{a},k}$ denotes the remainder term and $|\alpha|$ denotes the sum of the derivatives in the n components (i.e. $|\alpha| = \alpha_1 + \dots + \alpha_n$).

In the above notation

$$D_\alpha f(\boldsymbol{x}) = \frac{\partial^{|\alpha|} f(\boldsymbol{x})}{\partial x_1^{\alpha_1} \cdot \partial x_n^{\alpha_n}}, \quad |\alpha| \leq k.$$

- *Multivariate Newton-Raphson*:
  Input: Gradient function $g'(\theta)$, Hesse matrix $g''(\theta)$ and start value $\theta^{(0)}$.

  While not converged, do

$$\theta^{(k+1)} = \theta^{(k)} - \left[g''(\theta^{(k)})\right]^{-1} g'(\theta^{(k)})$$

- *Inverse of partitioned matrix:*
  Let $\mathbf{A}$ be symmetric and positive definite and let

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \quad \text{and} \quad \mathbf{A}^{-1} = \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}$$

  Then

$$\begin{aligned} \mathbf{B}_{11} &= (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}, \\ \mathbf{B}_{12} &= -\mathbf{B}_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{B}_{21} &= \mathbf{B}_{12}^T, \\ \mathbf{B}_{22} &= (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}. \end{aligned}$$