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

Statistical models

Exam, 2023/05/17

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. Mathematical expressions in the answers should be simplified as far as possible.

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 course. The bonus points can be used in the exams which will take place in academic year 2022/2023 only.

Problem 1 [19P]

Let Y be a negatively binomial distributed random variable with probability mass function given by

$$f(y;\pi) = \binom{y+k-1}{k-1} \pi^k (1-\pi)^y, \pi \in (0,1) \quad \text{and} \quad y \in \{0,1,2,...\}$$

where k is a known integer.

(a) Show that Y belongs to the exponential family. What is the canonical statistic $t(Y)$ and the canonical parameter θ in the minimal representation? **[3P]**

(b) Determine the norming constant $C(\theta)$. **[3P]**

(c) Derive $\mu = \mathbb{E}(Y)$. **[4P]**

(d) Show that $\mathbb{V}ar(Y) = \mu + \frac{\mu^2}{k}$. **[5P]**

(e) Let $y_1, y_2, ..., y_n$ be realizations of $Y_1, Y_2..., Y_n$ (a sample with independent and identically distributed observations from the distribution of Y). Derive the maximum likelihood estimator of μ. **[2P]**

(f) Using the results of part (e), provide the maximum likelihood estimator of π. **[2P]**

Problem 2 [17P]

Let $Y_1, Y_2, ..., Y_n$ be a sample with independent and identically distributed observations from the gamma distribution with known mean $\mu > 0$, unknown shape parameter $\nu > 0$ and density of Y_i given by

$$f(y_i;\nu) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^{\nu} y_i^{\nu-1} \exp\left(-\frac{\nu}{\mu} y_i\right), \quad y_i > 0, \ i = 1, ..., n$$

(a) Show that the distribution of Y_i belongs to a one-parameter exponential family with canonical parameter ν and canonical statistics $t(Y_i) = \log(Y_i) - Y_i/\mu$.**[3P]**

(b) Knowing that $\mathbb{E}(Y_i) = \mu$ and $\mathbb{E}(\log(Y_i)) = \psi(\nu) - \log(\nu) + \log(\mu)$, prove that the maximum likelihood estimator of ν satisfies the following equation

$$n\psi(\nu) - n\log\nu + n\log\mu - n = \sum_{i=1}^{n} \log y_i - \frac{1}{\mu}\sum_{i=1}^{n} y_i,$$

where $y_1, y_2, ..., y_n$ are realizations of $Y_1, Y_2..., Y_n$ and $\psi(\nu) = \partial \log \Gamma(\nu)/\partial\nu$ is the digamma function. **[4P]**

(c) Let $\hat{\nu}_{ML}$ be the maximum likelihood estimator of ν, the solution of the likelihood equation in part (b). Show that the observed Fisher information is given by **[3P]**

$$J(\hat{\nu}_{ML}) = n\{\psi'(\hat{\nu}_{ML}) - 1/\hat{\nu}_{ML}\}.$$

(d) Determine an expression for the likelihood ratio $L(\nu_0)/L(\hat{\nu}_{ML})$ in terms of ν_0 and $\hat{\nu}_{ML}$. **[3P]**

(e) Derive the saddlepoint approximation for the distribution of $\hat{\nu}_{ML}$ in a point ν_0. **[4P]**

Problem 3 [16P]

Let $\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n$, $n > 2$, be a sample with independent and identically distributed observations from a p-dimensional multivariate normal distribution with density of $\mathbf{Y}_i$ given by

$$f(\mathbf{y}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}_0) = (2\pi)^{-p/2} \det(\boldsymbol{\Sigma}_0)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}_0^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right), \; \boldsymbol{\mu} \in \mathbb{R}^p, \; \mathbf{y}_i \in \mathbb{R}^p, \; i = 1, ..., n$$

where $\boldsymbol{\Sigma}_0$ is a positive definite known covariance matrix. Let

$$\bar{\mathbf{Y}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Y}_i \quad \text{and} \quad \mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})^\top$$

be the sample mean vector and the sample covariance matrix, respectively. Prove that $\bar{\mathbf{Y}}$ and $\mathbf{S}$ are independent.

Problem 4 [23P]

Let Y_1 and Y_2 be two independent random variables with $Y_1 \sim Po(\lambda)$ (Poisson distribution with parameter λ) and $Y_2 \sim Po(c\lambda)$, respectively.

(a) Derive the joint probability mass function of Y_1 and Y_2. **[2P]**

(b) Prove that the canonical statistic is $t(Y_1, Y_2) = (v, u)^T$ with $v = Y_2$ and $u = Y_1 + Y_2$. Determine the canonical parameter vector $\boldsymbol{\theta}$. **[2P]**

(c) Calculate the marginal probability mass function $f(u)$. **[2P]**

(d) Specify the conditional distribution $f(v|u)$. **[2P]**

(e) Using the conditional principle derive the exact test of the hypothesis $c = 1$. Present the conditional distribution $f_0(v|u)$ under H_0. **[2P]**

(f) Calculate the p-value of the test from (e) if $y_1 = 2$ and $y_2 = 6$ are realizations of Y_1 and Y_2, respectively. Is the null hypothesis rejected at significance level 0.1? **[6P]**

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

(h) Perform the deviance test from (g) at significance level 0.1 by using $y_1 = 2$ and $y_2 = 6$ as realizations of Y_1 and Y_2, respectively. **[1P]**
Hint:

- If $Y \sim Po(\beta)$, then its probability mass function is given by

$$f(y; \beta) = \frac{\beta^y}{y!}e^{-\beta} \quad \text{for} \quad \beta > 0 \quad \text{and} \quad y = 0, 1, ...$$

- 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 [10P]

Let $\mathbf{Y}$ be a p-dimensional multivariate normally distributed random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, that is $\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\mathbf{X} = \mathbf{A}\mathbf{Y} + \mathbf{b}$ where $\mathbf{A} : q \times p$ and $\mathbf{b} : q \times 1$ are deterministic. Using the expression of moment generating function of $\mathbf{Y}$ given by

$$M_{\mathbf{Y}}(\mathbf{t}) = \exp\left(\mathbf{t}^\top \boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma}\mathbf{t}\right), \mathbf{t} \in \mathbb{R}^p,$$

where $\mathbf{t}^\top$ denotes the transpose of $\mathbf{t}$, prove that $\mathbf{X} \sim \mathcal{N}_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$.

Problem 6 [15P]

Provide the definition of the completeness of a test statistic. Using the fact that the canonical statistic in a full exponential family is complete, state and prove Basu's theorem.

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 $v_y(\mu_i) = \text{Var}(y_i; \mu_i)$, 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 $y = g(x) = (g_1(x), \ldots, g_n(x))^T$ with $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ then

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

- *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

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

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

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

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

- *Vector/matrix derivatives:*
  - $\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$ for a symmetric matrix $\mathbf{A}$ and a vector $\mathbf{x}$;
  - $\frac{\partial \mathrm{tr}[\mathbf{X}^\top \mathbf{A}]}{\partial \mathbf{X}} = 2\mathbf{A} - \mathrm{diag}(\mathbf{A})$ for symmetric matrices $\mathbf{A}$ and $\mathbf{X}$ where $\mathrm{diag}(\mathbf{A})$ denotes the diagonal matrix consisting of the diagonal elements of $\mathbf{A}$;
  - $\frac{\partial \log(\det(\mathbf{X}))}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - \mathrm{diag}(\mathbf{X}^{-1})$ for a symmetric matrix $\mathbf{X}$ where $\mathrm{diag}(\mathbf{X}^{-1})$ denotes the diagonal matrix consisting of the diagonal elements of $\mathbf{X}^{-1}$.