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

Statistical models

Re-exam, 2019/08/23

The only allowed aid is a pocket calculator provided by the department. The solution should be given in English. The answers to the task should be clearly formulated and structured. All non-trivial steps need to be commented. The answers of the text questions should cover the corresponding material presented during lectures.

The post exam review will take place on Monday, September 2, 2019 from 12:00 to 13:00 in room 329 (house 6).

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%	$< 50\%$

Problem 1 [14P]

Show that the following distributions belong to the exponential family. Find the canonical statistic and the canonical parameter in the minimal representation for each distribution:

(a) Bernoulli distribution Be(p) with $p \in (0,1)$. **[3P]**

(b) Multinomial distribution with probability mass function given by **[3P]**

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

with $p_1, p_2, p_3 \in (0,1)$ and known n.

(c) Borel distribution with probability mass function **[3P]**

$$\mathbb{P}(Y = y; \alpha) = \frac{1}{y!}(\alpha y)^{y-1} e^{-\alpha y} \qquad \text{for} \qquad y = 1, 2, \ldots \qquad \text{with} \qquad \alpha \in (0,1).$$

(d) Laplace distribution with density **[3P]**

$$f(y) = \frac{1}{2\sigma} \exp\left(-\frac{|y-\mu|}{\sigma}\right) \qquad \sigma > 0\,,$$

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

(e) What are the minimal sufficient statistics in parts (a)-(d)?**[2P]**

Problem 2 [16P]

Assume $\mathbf{Y} = (Y_1, \ldots, Y_k)^T$ to be Dirichlet distributed with probability density function given by

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

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

(a) Show that $\mathbf{Y}$ belongs to the exponential family and compute its canonical statistics $\mathbf{t}(\mathbf{y})$ as well as canonical parameter $\boldsymbol{\theta}$. **[4P]**

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

(c) Compute $\mathbb{E}\left(\ln\left(Y_1\right)\right)$. **[2P]**

(d) Compute $\mathbb{V}\left(\ln\left(Y_1\right)\right)$. **[2P]**

(e) Compute $\mathbb{C}ov\left(\ln\left(Y_1\right), \ln\left(Y_2\right)\right)$. **[2P]**

(f) Compute $\mathbb{E}\left(\ln\left(Y_1/Y_2\right)^2\right)$. **[3P]**

Hint: The solutions to part (d)-(f) should be presented in the terms of the polygamma function of order m (probably using several m's) given by

$$\psi^{(m)}(x) = \frac{\partial^m}{\partial x^m}\left(\ln \Gamma(x)\right).$$

Problem 3 [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 $\boldsymbol{\theta}$. **[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 $y_2 = 8$ 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 $y_2 = 8$ as realizations of Y_1 and Y_2, 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 4 [22P]

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

$$f(y) = \frac{(y+k-1)!}{y!(k-1)!}\pi^k(1-\pi)^y \qquad \text{for} \qquad y = 0, 1, 2, ..., \quad \pi \in (0, 1)$$

and known integer k.

(a) Show that Y belongs to the exponential family and compute its canonical statistics $t(Y)$ as well as canonical parameter θ. **[3P]**

(b) Determine the norming constant $C(\theta)$. **[2P]**

(c) Compute $\mu = E(Y)$. **[3P]**

(d) Show that this distribution satisfies the demands for use as ingredient in a generalized linear model. Find the canonical link function. **[4P]**

(e) Let $Y_1, ..., Y_n$ be independent observations with density of Y_i given by

$$f(y_i) = \frac{(y_i + k - 1)!}{y_i!(k-1)!} \pi_i^k (1 - \pi_i)^{y_i} \qquad \text{for} \qquad y_i = 0, 1, 2, ..., \quad \pi \in (0, 1)$$

and known integer k. Consider the canonical link function and the linear predictor $\eta_i = \alpha + \beta x_i$ where x_i is a deterministic variable. Derive the likelihood equation system for α and β. **[5P]**

(f) Find an expression of the deviance, and provide an expression of the square deviance residuals of the generalized linear model from part (e). **[5P]**

Problem 5 [10P]

Provide the definition of the completeness of a test statistic. Is the canonical statistic complete in a full exponential family? Explain your answer. Formulate Basu's theorem (without the proof).

Problem 6 [15P]

Prove that $\log C(\theta)$ is strictly convex and derive the following two equalities

$$\frac{\partial \log C(\theta)}{\partial \theta} = E_\theta(t),$$
$$\frac{\partial^2 \log C(\theta)}{\partial \theta^2} = Var_\theta(t)$$

for the canonical statistic $t(y) \in \mathbb{R}$ of a regular exponential family with canonical parameter $\theta \in \mathbb{R}$ and norming constant $C(\theta)$.

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
$$\begin{aligned} I(\beta) &= \left(\frac{\partial\theta}{\partial\beta}\right)^T V_y(\mu(\beta))\left(\frac{\partial\theta}{\partial\beta}\right) \\ &= X^T G'(\mu(\beta))^{-1} V_y(\mu(\beta))^{-1} G'(\mu(\beta))X. \end{aligned}$$

- 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

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