

Statistical models

Exam, 2016/05/24

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 Wednesday, June 8, 2016 from 10:00 to 11:00 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 [20P]

Let Y_1, Y_2, \dots, Y_n be an iid. sample of a Weibull distributed random variable Y , i.e.,

$$Y_1, \dots, Y_n \stackrel{iid}{\sim} \text{Wei}(\alpha, \beta)$$

with the density of Y given by

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

- (a) Derive the expression of the log-likelihood function. [2P]
- (b) Calculate the score vector $U(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\alpha, \beta)^T$. [3P]
- (c) Compute the observed Fisher information matrix $J(\boldsymbol{\theta})$. [4P]
- (d) Let $\hat{\alpha}_{MLE}$ denote the maximum likelihood estimator for α . Construct the 95% Wald confidence interval for α using the results of part (c). [2P]
- (e) Derive the expression of the profile log-likelihood function for α . [5P]
- (f) In this part of Problem 1 we assume that $\alpha = 5$. Provide the minimal sufficient statistic for β and explain your answer. [4P]

Problem 2 [15P]

Let $Y \sim F_{d,d}$ (an extended F -distribution with both degrees of freedom equal $d > 0$), i.e., the density of Y is given by

$$f(y; d) = \frac{1}{B\left(\frac{d}{2}, \frac{d}{2}\right)} y^{d/2-1} (1+y)^{-d} \quad \text{for } y \geq 0,$$

where $d > 0$ is an unknown parameter; $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ denotes the *beta function*.

- (a) Show that Y belongs to the exponential family and compute its canonical statistics $t(y)$ as well as canonical parameter θ . [4P]
- (b) Is this a regular exponential family? Explain your answer. [1P]
- (c) Determine the norming constant $C(\theta)$. [2P]
- (d) Compute $E\left(\ln\left(\frac{\sqrt{Y}}{1+Y}\right)\right)$. [3P]
- (e) Calculate $E\left(\frac{1}{4}[\ln(Y)]^2 - \ln(1+Y)\ln(1+1/Y)\right)$. [5P]

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

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

Problem 3 [18P]

Let Y_1, Y_2, \dots, Y_n be an iid. sample of a normally distributed random variable $Y \sim N(\mu, \sigma^2)$.

- (a) Show that Y belongs to the exponential family? What is the canonical statistics $\mathbf{t}(y)$ and the canonical parameter vector $\boldsymbol{\theta}$? [4P]
- (b) Calculate the canonical statistic \mathbf{t}_n for the whole sample and the norming constant $C_n(\boldsymbol{\theta})$ in this case. [2P]
- (c) Compute the expected Fisher information $I(\boldsymbol{\theta})$ for the canonical parametrization from part (b), i.e., when the whole sample is used. [5P]
- (d) Present the density of $(Y_1, Y_2, \dots, Y_n)^T$ by using the mean value parametrization. What is the expected Fisher information matrix in this case? [4P]
- (e) Consider the parametrization of the model with $\boldsymbol{\eta} = (\mu, \sigma^2)^T$ as a parameter vector. Compute the expected Fisher information matrix $I(\boldsymbol{\eta})$ by using the results of part (d) (or by using the results of part (c)) and the reparametrization lemma. [3P]

Problem 4 [22P]

Let Y_1 and Y_2 be two independent random variables with $Y_1 \sim Po(\lambda_1)$ (Poisson distribution with parameter λ_1) and $Y_2 \sim Po(c\lambda_1)$, 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 $f(v)$. [2P]
- (d) Specify the conditional distribution $f(v|u)$. [2P]
- (e) Using the conditional principle derive the exact test of the hypothesis $c = 4$. Present the test statistic and its distribution $f_0(v|u)$ under H_0 . [3P]
- (f) Calculate the p -value of the test from (e) if $y_1 = 2$ and $y_2 = 7$ are realizations of Y_1 and Y_2 , respectively. Is the null hypothesis rejected at significance level 0.05? [4P]
- (g) Derive the statistic of the deviance test for the null hypothesis from (e). What is the null distribution of this test statistic? [5P]
- (h) Perform the deviance test from (f) at significance level 0.05 by using $y_1 = 24$ and $y_2 = 78$ as realizations of Y_1 and Y_2 , respectively. [2P]

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 [10P]

Provide the definition of sufficient and minimal sufficient statistics? Formulate a criteria of a sufficient statistics to be minimal sufficient. Present two examples of the minimal sufficient statistic with the corresponding explanation (excluding the example considered in Problem 1).

Problem 6 [15P]

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

$$\begin{aligned}\frac{\partial \log C(\theta)}{\partial \theta} &= E_{\theta}(t), \\ \frac{\partial^2 \log C(\theta)}{\partial \theta^2} &= \text{Var}_{\theta}(t)\end{aligned}$$

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 :

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

- 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.

- 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 $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ 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

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

- *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

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

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

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

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

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

- *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

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