STATISTICAL ASYMPTOTICS APTS/09

Example Sheet **GAY/April 2009**

[*Comments and corrections to alastair.young@imperial.ac.uk*]

1. Prove that random samples from the following distributions form (m, m) exponential families with either $m = 1$ or $m = 2$: Poisson, binomial, geometric, gamma (index known), gamma (index unknown). Identify the natural statistics and the natural parameters in each case.

The negative binomial distribution with both parameters unknown provides an example of a model that is not of exponential family form. Why?

2. Let $Y_1, \ldots, Y_n$ be IID $N(\mu, \mu^2)$. Show that this model is an example of a curved exponential family and find a minimal sufficient statistic.

3. Verify that the family of gamma distributions of known index constitutes a transformation model under the action of the group of scale transformations.

4. Verify that maximum likelihood estimators are equivariant with respect to the group of one-to-one transformations.

5. Verify directly that in the location-scale model the configuration has a distribution which does not depend on the parameters.

6. Suppose that $(y_1, \ldots, y_n)$ are generated by a stationary first-order Gaussian autoregression with correlation parameter ρ, mean μ and innovation variance τ. That is, $Y_1 \sim N(\mu, \tau/(1 - \rho^2))$ and for $j = 2, \ldots, n$,

$$Y_j = \mu + \rho(Y_{j-1} - \mu) + \epsilon_j,$$

where $(\epsilon_1, \ldots, \epsilon_n)$ are IID $N(0, \tau)$.

Find the log-likelihood function. Show that if μ is known to be zero, the log-likelihood has $(3, 2)$ exponential family form, and find the natural statistics.

7. Let $Y_1, \ldots, Y_n$ be IID Poisson (θ). Find the score function and the expected and observed information.

Consider the new parametrisation $\psi = \psi(\theta) = e^{-\theta}$. Compute the score function and the expected and observed information in the ψ-parametrisation.

8. Consider a multinomial distribution with four cells, the probabilities for which are

$$\pi_1(\theta) = \tfrac{1}{6}(1 - \theta), \pi_2(\theta) = \tfrac{1}{6}(1 + \theta),$$
$$\pi_3(\theta) = \tfrac{1}{6}(2 - \theta), \pi_4(\theta) = \tfrac{1}{6}(2 + \theta),$$

where θ is unknown, $|\theta| < 1$.

What is the minimal sufficient statistic? Show that $A' = (N_1 + N_2, N_3 + N_4)$ and $A'' = (N_1 + N_4, N_2 + N_3)$ are both ancillary.

If A is ancillary in the simple sense, we may write

$$P_Y(y;\theta) = P_{Y|A}(y \mid a;\theta) P_A(a).$$

The conditional expected information for θ given $A = a$ is

$$\begin{aligned} i_A(\theta \mid a) &= E\left\{ \frac{-\partial^2 \log P_{Y|A}(Y \mid a, \theta)}{\partial \theta^2} \middle| A = a; \theta \right\} \\ &= E\left\{ \frac{-\partial^2 \log P_Y(Y;\theta)}{\partial \theta^2} \middle| A = a; \theta \right\}. \end{aligned}$$

Now take expectations over the distribution of A:

$$E\{i_A(\theta \mid A)\} = i(\theta).$$

With two ancillaries competing,

$$E\{i_{A'}(\theta \mid A')\} = E\{i_{A''}(\theta \mid A'')\},$$

so that expected conditional information is no basis for choice between them. To discriminate between them it may be argued that A' is preferable to A'' if

$$\mathrm{var}\,\{i_{A'}(\theta \mid A')\} > \mathrm{var}\,\{i_{A''}(\theta \mid A'')\}.$$

Show that in the above example A' is preferable to A'' in these terms.

9. Consider again the model of Question 2. Let $T_1 = \bar{Y}$ and $T_2 = \sqrt{n^{-1}\sum_{i=1}^n Y_i^2}$. Show that $Z = T_1/T_2$ is ancillary. Why might inference on μ be based on the conditional distribution of $V = \sqrt{n}T_2$, given Z? Find the form of this conditional distribution.

10. Show that, if the parameters ψ and χ are orthogonal, any one-to-one smooth function of ψ is orthogonal to any one-to-one smooth function of χ.

11. Suppose that Y is distributed according to a density of the form

$$p(y;\theta) = \exp\{s(y)^T c(\theta) - k(\theta) + D(y)\}.$$

Suppose that θ may be written $\theta = (\psi, \lambda)$, where ψ denotes the parameter of interest, possibly vector valued, and that $c(\theta) = (c_1(\psi), c_2(\theta))^T$, for functions c_1, c_2, where $c_1(\cdot)$ is a one-to-one function of ψ. Then, writing $s(y) = (s_1(y), s_2(y))^T$, the log-likelihood function is of the form

$$l(\psi, \lambda) = s_1(y)^T c_1(\psi) + s_2(y)^T c_2(\theta) - k(\theta).$$

Let ϕ be the *complementary mean parameter* given by

$$\phi \equiv \phi(\theta) = E\{s_2(Y);\theta\}.$$

Show that ψ and ϕ are orthogonal parameters.

Let Y have a gamma distribution with shape parameter ψ and scale parameter ϕ, and density

$$f(y;\psi,\phi) = \phi^{-\psi} y^{\psi-1} \exp(-y/\phi)/\Gamma(\psi).$$

Show that $\psi\phi$ is orthogonal to ψ.

12.* *Dispersion models.* The defining property of dispersion models is that their model function is of the form

$$a(\lambda, y) \exp\{\lambda t(y;\gamma)\},$$

where $\lambda \in \mathbb{R}$ and $\gamma \in \mathbb{R}^k$ are parameters. Show that λ and γ are orthogonal.

Exponential dispersion models are a subclass of dispersion models where

$$t(y;\gamma) = \gamma . y - K(\gamma).$$

Let Y be a *1-dimensional* random variable with density belonging to an exponential dispersion family. Show that the cumulant generating function of Y is

$$K_Y(t;\gamma,\lambda) = \lambda \left\{ K\left(\gamma + \frac{t}{\lambda}\right) - K(\gamma) \right\}$$

and that Y has mean

$$E(Y) = \mu(\gamma) = \frac{\partial K(\gamma)}{\partial \gamma}.$$

Show also that $\mathrm{var}\,(Y) = \frac{1}{\lambda} V(\mu)$ where

$$V(\mu) = \left. \frac{\partial^2 K(\gamma)}{\partial \gamma^2} \right|_{\gamma=\gamma(\mu)},$$

and $\gamma(\mu)$ indicates the inverse function of $\mu(\gamma)$.

The notation

$$Y \sim ED\big(\mu, \sigma^2 V(\mu)\big)$$

is used to indicate that Y has density $P(y;\gamma,\lambda)$ which belongs to an exponential dispersion family with $\gamma = \gamma(\mu)$, $\lambda = 1/\sigma^2$ and variance function $V(\mu)$.

Let Y have the inverse Gaussian distribution $Y \sim IG(\phi, \lambda)$ with density

$$P(y;\phi,\lambda) = \frac{\sqrt{\lambda}}{\sqrt{2\pi}} y^{-3/2} e^{\sqrt{\lambda\phi}} \exp\left\{ -\frac{1}{2} \left(\frac{\lambda}{y} + \phi y \right) \right\}, \qquad y > 0, \lambda > 0, \phi \geq 0.$$

Show that $Y \sim ED(\mu, \sigma^2 V(\mu))$ with $V(\mu) = \mu^3$.

Let $Y_1, \ldots, Y_n$ be independent random variables with

$$Y_i \sim ED\left(\mu(\gamma), \frac{\sigma^2}{w_i} V\big(\mu(\gamma)\big) \right), \quad i = 1, \ldots, n,$$

where $w_1, \dots, w_n$ are known constants. Let $w_+ = \sum w_i$.

Show that

$$\frac{1}{w_+}\sum_{i=1}^{n} w_i Y_i \sim ED\left(\mu(\gamma), \frac{\sigma^2}{w_+} V\big(\mu(\gamma)\big)\right).$$

Deduce that, if $Y_1, \dots, Y_n$ are IID $IG(\phi, \lambda)$, then $\overline{Y} = n^{-1}\sum_{i=1}^{n} Y_i \sim IG(n\phi, n\lambda)$.

13. Let $Y_1, \dots, Y_n$ be independent random variables such that Y_j has a Poisson distribution with mean $\exp\{\lambda + \psi x_j\}$, where $x_1, \dots, x_n$ are known constants.

Show that the conditional distribution of $Y_1, \dots, Y_n$ given $S = \sum Y_j$ does not depend on λ. Find the conditional log- likelihood function for ψ, and verify that it is equivalent to the profile log-likelihood.

14. Verify that in general the likelihood ratio and score tests are invariant under a reparameterization $\psi = \psi(\theta)$, but that the Wald test is not.

Write $\theta = (\theta_1, \theta_2)$, where θ_1 is parameter of interest. Suppose $\psi = \psi(\theta) = (\psi_1, \psi_2)$ is an interest respecting transformation, with $\psi_1 \equiv \psi_1(\theta) = \theta_1$. Show that the profile log-likelihood is invariant under this reparameterization.

15. Let $Y_1, \dots, Y_n$ be IID $N(\mu, \sigma^2)$, and let the parameter of interest be μ. Obtain the form of the profile log-likelihood.

Show how to construct a confidence interval with asymptotic coverage $1 - \alpha$ based on the profile log-likelihood.

16. Verify that the rth degree Hermite polynomial H_r satisfies the identity

$$\int_{-\infty}^{\infty} e^{ty} H_r(y)\phi(y)dy = t^r e^{\frac{1}{2}t^2}.$$

Verify that the moment generating function of S_n^* has the expansion

$$\begin{aligned} M_{S_n^*}(t) &= \exp\{K_{S_n^*}(t)\} \\ &= e^{\frac{1}{2}t^2} \exp\left\{\frac{1}{6\sqrt{n}}\,\rho_3 t^3 + \frac{1}{24n}\,\rho_4 t^4 + O(n^{-3/2})\right\} \\ &= e^{\frac{1}{2}t^2}\left\{1 + \frac{\rho_3}{6\sqrt{n}}\,t^3 + \frac{\rho_4}{24n}\,t^4 + \frac{\rho_3^2}{72n}\,t^6 + O(n^{-3/2})\right\}. \end{aligned}$$

On using the above identity, this latter expansion may be written

$$\begin{aligned} M_{S_n^*}(t) = \int_{-\infty}^{\infty} e^{ty}\Big\{1 &+ \frac{1}{6\sqrt{n}}\,\rho_3 H_3(y) \\ &+ \frac{1}{24n}\,\rho_4 H_4(y) + \frac{1}{72n}\,\rho_3^2 H_6(y) + O(n^{-3/2})\Big\}\phi(y)dy. \end{aligned}$$

Comparison with the definition

$$M_{S_n^*}(t) = \int_{-\infty}^{\infty} e^{ty} f_{S_n^*}(y)dy,$$

provides a heuristic justification for the Edgeworth expansion.

17. Verify that integration of the Edgeworth expansion for the density of S_n^* yields the distribution function expansion given in lecture notes.

18. Let $Y_1, \ldots, Y_n$ be IID $N(\mu, \sigma^2)$. Obtain the saddlepoint approximation to the density of $S_n = \sum_{i=1}^n Y_i$, and comment on its exactness.

19. Let $Y_1, \ldots, Y_n$ be IID exponential random variables with pdf $f(y) = e^{-y}$. Obtain the saddlepoint approximation to the density of $S_n = \sum_{i=1}^n Y_i$, and show that it matches the exact density except for the normalizing constant.

20. Fill in the details of the statistical derivation of the saddlepoint approximation to the density of S_n.

21. Verify the calculations leading to the Laplace approximation (3.11) of lecture notes.

22. Let $Y_1, \ldots, Y_n$ be IID exponential random variables of mean μ. Verify that the p^*-formula for the density of $\hat{\mu}$ is exact.

23.* Let $y_1, \ldots, y_n$ be independent realisations of a continuous random variable Y with density belonging to a location-scale family,

$$p(y; \mu, \sigma) = \frac{1}{\sigma} p_0 \left(\left(\frac{y - \mu}{\sigma} \right) \right),$$

$(y - \mu)/\sigma \in \mathcal{X}$, $\mu \in \mathbb{R}, \sigma > 0$. Assume that the maximum likelihood estimate $(\hat{\mu}, \hat{\sigma})$ of (μ, σ) based on $y = (y_1, \ldots, y_n)$ exists and is finite and that p_0 is suitably differentiable. Define the sample configuration a by

$$a = \left(\frac{y_1 - \hat{\mu}}{\hat{\sigma}}, \ldots, \frac{y_n - \hat{\mu}}{\hat{\sigma}} \right).$$

Show that the p^*-formula for the conditional density of $(\hat{\mu}, \hat{\sigma})$ given a is

$$p^*(\hat{\mu}, \hat{\sigma}; \mu, \sigma \mid a) = c(\mu, \sigma, a) \frac{\hat{\sigma}^{n-2}}{\sigma^n} \prod_{i=1}^n p_0 \left(\frac{\hat{\sigma}}{\sigma} a_i + \frac{\hat{\mu} - \mu}{\sigma} \right),$$

and is exact.

24. Let $X_1, \ldots, X_n$ be independent exponential random variables with mean $1/\lambda$ and let $Y_1, \ldots, Y_n$ be an independent sample of independent exponential random variables of mean $1/(\psi\lambda)$.

Find the p^* approximation to the density of $(\hat{\psi}, \hat{\lambda})$, and hence find an approximation to the marginal density of $\hat{\psi}$. The exact distribution of $\hat{\psi}/\psi$ is an F-distribution with degrees of freedom $(2n, 2n)$, so that the exact density of $\hat{\psi}$ is given by

$$\frac{\Gamma(2n)}{\Gamma(n)} \frac{1}{\psi} \left(\frac{\hat{\psi}}{\psi}\right)^{n-1} \left(\frac{\hat{\psi}}{\psi} + 1\right)^{-2n}.$$

Comment on the exactness of the marginal density approximation.

25. As in question 15, let $Y_1, \ldots, Y_n$ be IID $N(\mu, \sigma^2)$, but suppose the parameter of interest is the variance σ^2.

Obtain the form of the profile log-likelihood. Show that the profile score has an expectation which is non-zero.

Find the modified profile log-likelihood for σ^2 and examine the expectation of the modified profile score.

26. Let $Y_1, \ldots, Y_n$ be independent exponential random variables, such that Y_j has mean $\lambda \exp(\psi x_j)$, where $x_1, \ldots, x_n$ are known scalar constants and ψ and λ are unknown parameters.

In this model the maximum likelihood estimators are not sufficient and an ancillary statistic is needed. Let

$$a_j = \log Y_j - \log \hat{\lambda} - \hat{\psi} x_j,$$

$j = 1, \ldots, n$, and take $a = (a_1, \ldots, a_n)$ as the ancillary.

Find the form of the profile log-likelihood function and of the modified profile log-likelihood function for ψ.

27. Let $Y_1, \ldots, Y_n$ be IID $N(\mu, \sigma^2)$ and consider testing $H_0 : \mu = \mu_0$. Show that the likelihood ratio statistic for testing H_0 may be expressed as

$$w = n \log\{1 + t^2/(n-1)\},$$

where t is the usual Student's t statistic.

Show directly that

$$Ew = 1 + \frac{3}{2n} + O(n^{-2})$$

in this case, so that the Bartlett correction factor $b \equiv 3/2$.

Examine numerically the adequacy of the χ^2, approximation to w and to $w' = w/(1 + 3/2n)$.

28. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent pairs of independently normally distributed random variables such that, for each j, X_j and Y_j each have mean μ_j and variance σ^2.

Find the maximum likelihood estimator of σ^2 and show that it is not consistent.

Find the form of the modified profile log-likelihood function for σ^2 and examine the estimator of σ^2 obtained by its maximization.

Let $S = \sum_{i=1}^{n}(X_i - Y_i)^2$. What is the distribution of S? Find the form of the marginal log-likelihood for σ^2 obtained from S and compare it with the modified profile likelihood.

[*This is the 'Neyman-Scott problem' which typifies situations with large numbers of nuisance parameters. Note, however, that the model falls outside the general framework that we have been considering, in that the dimension of the parameter* $(\mu_1, \ldots, \mu_n, \sigma^2)$ *depends on the sample size, and tends to* ∞ *as* $n \to \infty$.]