STATISTICAL ASYMPTOTICS APTS/10

Example Sheet (with sketch solutions) **GAY/April 2010**

[*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?

[**A1.** *Just a matter of rewriting the density functions in exponential family form. With the negative binomial distribution, the sample space $\mathcal{X}$ depends on the unknown parameter, not allowed in exponential family, if both parameters are unknown.*]

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.

[**A2.** *Immediate from writing joint density in exponential family form:*

$$P_Y(y; \mu) \propto \exp\left\{-\frac{1}{2\mu^2}\sum y_i^2 + \frac{1}{\mu}\sum y_i - n\log\mu\right\}.$$

Minimal sufficient statistic is $(\sum Y_i, \sum Y_i^2)$.]

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

[**A3.**

Let $Y \sim \text{Gamma}\,(k, \lambda)$ *denote* Y *has a gamma distribution of known index* k *and pdf*

$$f_Y(y; \lambda) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{\Gamma(k)}.$$

Then $\sigma Y \sim \text{Gamma}\,(k, \lambda/\sigma)$.]

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

[**A4.**

This follows from the transformation property, that if $\phi = \phi(\theta)$ is a one-to-one transformation of the parameter θ, then $\hat{\phi} = \phi(\hat{\theta})$, where $\hat{\theta} = s(Y)$ is the maximum likelihood estimator of θ. If the transformation $\phi(\cdot)$ corresponds to $\bar{g}_\phi \in \bar{G}$, then $g_\phi(Y)$ is the transformation of Y whose MLE is $\hat{\phi}$. Then $\hat{\phi} = s(g_\phi(Y))$, while $\phi(\hat{\theta}) = \bar{g}_\phi(s(Y))$. Hence $s(g_\phi(Y)) = \bar{g}_\phi(s(Y))$, for all such g_ϕ, which is the requirement of equivariance.]

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

[**A5.** *This is just a matter of writing the configuration in terms of* $\epsilon_1, \ldots, \epsilon_n$*, where* $Y_j = \eta + \tau\epsilon_j$.]

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.

[**A6.** *The distribution of* Y_j *given* $Y_{(j-1)} = \{Y_1, \ldots, Y_{j-1}\}$ *depends only on* Y_{j-1} *and contributes the term*

$$-\frac{1}{2}\log(2\pi\tau) - \frac{\{y_j - \mu - \rho(y_{j-1} - \mu)\}^2}{2\tau}$$

to the log-likelihood. The full log-likelihood is

$$-\frac{n}{2}\log(2\pi\tau) + \frac{1}{2}\log(1-\rho^2) - \frac{(y_1-\mu)^2(1-\rho^2)}{2\tau} - \sum_{j=2}^{n}\frac{\{y_j - \mu - \rho(y_{j-1}-\mu)\}^2}{2\tau}$$

$$= -\frac{n}{2}\log(2\pi\tau) + \frac{1}{2}\log(1-\rho^2) - \frac{(y_1-\mu)^2 + (y_n-\mu)^2}{2\tau}$$

$$-\sum_{j=2}^{n-1}\frac{(y_j-\mu)^2(1+\rho^2)}{2\tau} + \frac{\rho}{\tau}\sum_{j=2}^{n}(y_j-\mu)(y_{j-1}-\mu).$$

If $\mu = 0$ *this has (3,2) exponential family form with natural statistics*

$$y_1^2 + y_n^2, \quad \sum_{j=2}^{n-1} y_j^2, \quad \sum_{j=2}^{n} y_j y_{j-1}. \qquad]$$

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.

[**A7.** *I reckon*

$$u^{(\theta)}(\theta; y) = \frac{1}{\theta}\sum_{j=1}^{n} y_j - n,$$

$$i^{(\theta)}(\theta) = \frac{n}{\theta}, \quad j^{(\theta)}(\theta) = \frac{1}{\theta^2}\sum_{j=1}^{n} y_j,$$

$$u^{(\psi)}(\psi; y) = \frac{1}{\psi}\left(\frac{\sum_{j=1}^n y_j}{\log\psi} + n\right), \quad i^{(\psi)}(\psi) = \frac{n}{(-\log\psi)\psi^2},$$

$$j^{(\psi)}(\psi) = \frac{1}{\psi^2}\left\{\frac{\sum_{j=1}^n y_j}{(\log\psi)^2} + \frac{\sum_{j=1}^n y_j}{\log\psi} + n\right\}.$$

Note that $\hat{\psi} = e^{-\hat{\theta}}$. *We can proceed either directly, or use the formulae for the way the score and information change under a reparameterisation.*]

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

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

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

[**A8.** *The minimal sufficient statistic is* (N_1, N_2, N_3, N_4). *Let* $N_1 + N_2 + N_3 + N_4 = n$. *Write* $A' = (a_1, a_2)$, $A'' = (b_1, b_2)$. *Then* $a_1 \sim \text{Binomial}\,(n, 1/3)$ *and* $b_1 \sim \text{Binomial}\,(n, 1/2)$, *showing that* A' *and* A'' *are ancillary.*

Simple calculations give

$$\begin{aligned} i_{A'}(\theta \mid a') &= \frac{3a_1 + n(1-\theta^2)}{(1-\theta^2)(4-\theta^2)}, \\ i_{A''}(\theta \mid a'') &= \frac{2\theta b_1 + n(1-\theta)(2+\theta)}{(1-\theta^2)(4-\theta^2)}. \end{aligned}$$

Then

$$\text{var}\{i_{A'}(\theta \mid A')\} = \frac{2n}{\{(1-\theta^2)(4-\theta^2)\}^2},$$
$$\text{var}\{i_{A''}(\theta \mid A'')\} = \frac{n\theta^2}{\{(1-\theta^2)(4-\theta^2)\}^2},$$

leading immediately to the required conclusion.]

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.

[**A9.** *Write* $U = \bar{Y}$ *and* $W = \sum_{i=1}^n (Y_i - \bar{Y})^2/\mu^2$. *Then* U *and* W *are independent* $N(\mu, \mu^2/n)$ *and* χ^2_{n-1} *respectively, so the joint density of* (U, W) *is*

$$f_{U,W}(u, w) = c_1\mu^{-1}\exp\{-\frac{n}{2\mu^2}(u-\mu)^2\}w^{(n-3)/2}\exp(-\frac{w}{2}).$$

Here and below, $c_1, c_2, \ldots$ *denote generic constants, not depending on* μ.

Transform to obtain the joint density of (V, Z). *We have* $W = V^2(1 - Z^2)/\mu^2, U = ZV/\sqrt{n}$, *and the Jacobian is* $2V^2/(\mu^2\sqrt{n})$, *so the joint density of* (V, Z) *is*

$$c_2\mu^{-1}\mu^{-(n-3)}v^{n-3}\mu^{-2}v^2h_1(z)\exp\{-\frac{n}{2\mu^2}(zv/\sqrt{n}-\mu)^2 - v^2(1-z^2)/(2\mu^2)\},$$

where $h_1(z)$ *is some function of* z, *not depending on* μ.

Simplification shows this is

$$c_3h_2(z)\mu^{-n}v^{n-1}\exp\{-\frac{1}{2}(v/\mu - z\sqrt{n})^2\},$$

for some $h_2(z)$.

Observe that

$$\int \mu^{-n}v^{n-1}\exp\{-\frac{1}{2}(v/\mu - z\sqrt{n})^2\}dv = h_3(z),$$

say, not depending on μ *(write* $t = v/\mu$ *and substitute), so we see that the marginal density of* Z, *obtained by integrating out* V *from the joint density, does not depend on* μ, Z *is ancillary.*

The minimal sufficient statistic $(T_1, T_2) \equiv (V, Z)$. *Since* Z *is ancillary, the Conditionality Principle implies that we should base inference on the conditional distribution of* V *given* Z. *The conditional density is obtained by dividing the joint density of* V *and* Z *by the marginal density of* Z, *and is*

$$f(v \mid z; \mu) = c_4\mu^{-n}v^{n-1}\exp\{-\frac{1}{2}(v/\mu - z\sqrt{n})^2\},$$

directly from the above.]

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.

[**A10.** *This is a consequence of the definition of orthogonality of ψ and χ, together with the formula for the way the information matrix transforms under reparametrisation, noting the special form of the transformation being made. The Jacobian term in the formula for the new information matrix is block diagonal, so the new information matrix itself is.*]

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.

[**A11.** *The log-likelihood function for (ψ, ϕ) is given by*

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

We know that $\frac{\partial \tilde{l}(\theta)}{\partial \phi}$ has mean 0,

$$E\{s_2(Y)^T \frac{\partial c_2(\theta)}{\partial \phi}; \theta\} = \frac{\partial k(\theta)}{\partial \phi}.$$

So

$$\phi^T \frac{\partial c_2(\theta)}{\partial \phi} = \frac{\partial k(\theta)}{\partial \phi},$$

for all θ. Differentiating this with respect to ψ gives

$$\frac{\partial^2 k(\theta)}{\partial \psi \partial \phi} - \phi^T \frac{\partial^2 c_2(\theta)}{\partial \psi \partial \phi} = 0.$$

But

$$\frac{\partial^2 \tilde{l}}{\partial\psi\partial\phi}(\psi,\phi) = s_2(y)^T\frac{\partial^2 c_2(\theta)}{\partial\psi\partial\phi} - \frac{\partial^2 k(\theta)}{\partial\psi\partial\phi}$$

and

$$\tilde{i}_{\psi\phi}(\theta) = E\{-\frac{\partial^2\tilde{l}}{\partial\psi\partial\phi}\} = 0,$$

as required to verify orthogonality.

For the Gamma example, the natural parameters are $(\psi, -1/\phi)$*, corresponding to* $(\log y, y)$*, so that* $E(Y) = \psi\phi$ *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 $w_1, \ldots, w_n$ 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 $Y_1, \ldots, Y_n$ are IID $IG(\phi, \lambda)$, then $\overline{Y} = n^{-1} \sum_{i=1}^{n} Y_i \sim IG(n\phi, n\lambda)$.

[**A12.** *We have*

$$\frac{\partial l}{\partial \gamma} = \frac{\lambda \partial t(y;\gamma)}{\partial \gamma}$$
$$\frac{\partial^2 l}{\partial \lambda \partial \gamma} = \frac{\partial t(y;\gamma)}{\partial \gamma},$$

but *the general result*

$$E\left(\frac{\partial l}{\partial \gamma}\right) = 0 \Rightarrow E\left(\frac{\partial t}{\partial \gamma}\right) = 0$$

and so, in the obvious partitioning of the expected information,

$$i_{\lambda\gamma} = E\left(\frac{-\partial^2 l}{\partial \lambda \partial \gamma}\right) = 0.$$

To obtain the cumulant generating function, integrate directly to obtain the moment generating function, take logs.

The mean follows directly from differentiating the m.g.f. or c.g.f. Also, by the same technique,

$$\operatorname{var} Y = \frac{1}{\lambda} \frac{\partial^2 K(\gamma)}{\partial \gamma^2}.$$

The function $\mu(\cdot)$ is one-to-one and smooth because of general results about exponential families. It is therefore possible to reparametrise using the mean $\mu = \mu(\gamma)$ instead of the natural parameter γ. The question makes this notationally explicit.

If $Y \sim IG(\phi, \lambda)$ the cumulant generating function of Y is easily calculated as

$$K_Y(t; \phi, \lambda) = -\sqrt{\lambda(\phi - 2t)} + \sqrt{\lambda\phi}.$$

Then

$$\mu \equiv EY = K'_Y(0; \phi, \lambda) = \sqrt{\lambda/\phi},$$
$$K''_Y(0; \phi, \lambda) = \frac{1}{\lambda}\left(\frac{\lambda}{\phi}\right)^{3/2} = \frac{1}{\lambda}\mu^3.$$

With the reparametrisation $(\mu, \sigma^2) = \left(\sqrt{\lambda/\phi}, 1/\lambda\right)$, *the density of* Y *can be written as*

$$P(y; \mu, \sigma^2) = \exp\left\{\frac{1}{\sigma^2}\left(\frac{1}{\mu^2}\, y + \frac{1}{\mu}\right)\right\} \frac{1}{\sqrt{2\pi\sigma}}\, y^{-3/2} e^{-1/(2\sigma^2 y)},$$

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

The rest here is just a straightforward exercise in manipulation of cumulant generating functions.

The c.g.f. of Y_i *is*

$$K_{Y_i}\left(t; \gamma, \frac{\sigma^2}{w_i}\right) = \frac{w_i}{\sigma^2}\left\{K\left(\gamma + \frac{\sigma^2}{w_i}\, t\right) - K(\gamma)\right\}.$$

Then $w_i Y_i$ *has c.g.f.*

$$K_{w_i Y_i}\left(t; \gamma, \frac{\sigma^2}{w_i}\right) = \frac{w_i}{\sigma^2}\left\{K(\gamma + \sigma^2 t) - K(\gamma)\right\},$$

and $\sum_i w_i Y_i$ *has c.g.f.*

$$\frac{w_+}{\sigma^2}\left\{K(\gamma + \sigma^2 t) - K(\gamma)\right\}.$$

Finally $\sum w_i Y_i / w_+$ *has c.g.f.*

$$\frac{w_+}{\sigma^2}\left\{K\left(\gamma + \frac{\sigma^2}{w_+}\, t\right) - K(\gamma)\right\},$$

giving the conclusion. The inverse Gaussian case provides an example.]

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

Show that the conditional distribution of $Y_1, \ldots, Y_n$ 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.

[**A13.** *That the conditional distribution does not depend on* λ *is easy: the distribution of* S *is Poisson with mean* $\sum \exp\{\lambda + \psi x_j\}$. *The definition of conditional log- likelihood gives*

$$\psi \sum x_j y_j - s \log[\sum \exp\{\psi x_j\}].$$

The maximum likelihood estimator of λ *for fixed* ψ *is given by*

$$\hat{\lambda}_\psi = \log\left(\frac{\sum y_j}{\sum \exp\{\psi x_j\}}\right)$$

and plugging into the expression for the full log-likelihood shows that the profile log-likelihood is equivalent to the conditional 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.

[**A14.** *Parameterization invariance of the likelihood function ensures invariance of the likelihood ratio and score tests.*

We have $l^{(\psi)}(\psi) = l^{(\theta)}(\theta(\psi))$*. To test* $H_0 : \theta = \theta_0$*, the LRT rejects for large values of* $l^{(\theta)}(\hat{\theta}) - l^{(\theta)}(\theta_0)$*. The LRT in the* ψ*-parametrisation is* $H_0 : \psi \equiv \psi(\theta) = \psi_0 \equiv \psi(\theta_0)$*, based on*

$$l^{(\psi)}(\hat{\psi}) - l^{(\psi)}(\psi_0) \equiv l^{(\theta)}\big(\theta(\hat{\psi})\big) - l^{(\theta)}\big(\theta(\psi_0)\big) = l^{(\theta)}(\hat{\theta}) - l^{(\theta)}(\theta_0),$$

since $\hat{\psi} = \psi(\hat{\theta}), \hat{\theta} = \theta(\hat{\psi})$*.*

That the score test is parametrisation invariant follows on showing that

$$U^{(\theta)}(\theta_0)^T i^{(\theta)}(\theta_0)^{-1} U^{(\theta)}(\theta_0) \equiv U^{(\psi)}(\psi_0)^T i^{(\psi)}(\psi_0)^{-1} U^{(\psi)}(\psi_0),$$

using the formulae for the way U, i *transform under reparametrisation.*

The Wald statistic is not (generally) invariant, as is checked by working out the formula for it in a new parametrisation. For example, in the one-dimensional case, to test $H_0 : \theta = \theta_0$ *the Wald statistic might be taken as*

$$(\hat{\theta} - \theta_0)^2 i^{(\theta)}(\hat{\theta}), \qquad *$$

say. If we consider a reparametrisation $\psi = \psi(\theta)$ *we have the statistic*

$$\big(\psi(\hat{\theta}) - \psi(\theta_0)\big)^2 i^{(\theta)}\big(\theta(\hat{\psi})\big) \left(\frac{\partial \theta(\psi)}{\partial \psi} \bigg|_{\psi = \hat{\psi}} \right)^2, \qquad **$$

on noting that

$$i^{(\psi)}(\psi) = i^{(\theta)}\big(\theta(\psi)\big) \left(\frac{\partial}{\partial \psi} \theta(\psi) \right)^2.$$

The two quantities $*$ *and* $**$ *don't necessarily coincide.*

An example: let $Y_1, \ldots, Y_n$ *be IID Poisson* (θ)*. Consider testing* $H_0 : \theta = 1$*. We have* $\hat{\theta} = \bar{Y}$*,* $l^{(\theta)}(\hat{\theta}) = n\bar{Y}(\log \bar{Y} - 1)$ *and* $l^{(\theta)}(1) = -n$*. In the parametrisation* $\psi = e^{-\theta}, H_0 : \psi = e^{-1}$*. We have* $l^{(\psi)}(\psi) = n(\bar{Y}\log(-\log\psi) + \log\psi)$*, and* $\hat{\psi} = e^{-\bar{Y}}$*. Then*

$$l^{(\psi)}(\hat{\psi}) = n\bar{Y}(\log \bar{Y} - 1), l^{(\psi)}(e^{-1}) = -n.$$

The Wald statistic for $H_0 : \theta = 1$ *is* $w_p^{(\theta)}(\theta_0) = n(\bar{Y} - 1)^2/\bar{Y}$*. In the parametrisation* $\psi = e^{-\theta}$ *we have*

$$w_p^{(\psi)}(\psi_0) = n(e^{-\bar{Y}} - e^{-1})^2 e^{2\bar{Y}}/\bar{Y}.$$

For example, if $n = 50$, $\bar{Y} = 1.3$, $w_p^{(\theta)} \doteq 3.45$, $w_p^{(\psi)} \doteq 4.70$. *Referring to* χ^2, *the significance level is* 0.06 *for* $w_p^{(\theta)}$ *and* 0.03 *for* $w_p^{(\psi)}$.

Quite similar considerations show the profile log-likelihood is invariant under interest respecting reparameterizations.

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* $\psi_1 \equiv \psi_1(\theta) = \theta_1$.

We have

$$\begin{aligned} l^{(\psi)}(\psi) &\equiv l^{(\psi)}(\psi_1, \psi_2) \\ &= l^{(\theta)}\big(\theta_1(\psi), \theta_2(\psi)\big) \\ &= l^{(\theta)}\big(\psi_1, \theta_2(\psi_1, \psi_2)\big). \end{aligned}$$

Then

$$\sup_{\psi_2} l^{(\psi)}(\psi_1, \psi_2) = \sup_{\theta_2} l^{(\theta)}(\psi_1, \theta_2).$$

An instructive example to look at is the case of $Y_1, \ldots, Y_n$ *IID* $N(\mu, \sigma^2)$. *Check directly that the profile log-likelihood for* μ *is unchanged under either of the reparameterizations* $(\mu, \sigma) \to (\mu, \log \sigma)$ *or* $(\mu, \sigma) \to (\mu, \log(\sigma/\mu))$.]

15. Let $Y_1, \ldots, Y_n$ 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.

[**A15.** *The maximum likelihood estimate of* σ^2 *for fixed* μ *is* $\hat{\sigma}^2_\mu = \sum_{i=1}^n (Y_i - \mu)^2/n$. *Hence*

$$\begin{aligned} l_p(\mu) &= l(\mu, \hat{\sigma}^2_\mu) \\ &= -\tfrac{1}{2} n \log \hat{\sigma}^2_\mu - \tfrac{1}{2} n \\ &= -\tfrac{1}{2} n \log \left\{ \frac{\Sigma(Y_j - \bar{Y})^2 + n(\bar{Y} - \mu)^2}{n} \right\} - \tfrac{1}{2} n, \end{aligned}$$

since

$$l(\mu, \sigma^2) = -\tfrac{1}{2} n \log \sigma^2 - \frac{\Sigma(Y_j - \mu)^2}{2\sigma^2},$$

apart from a constant. Now, $\hat{\mu} = \bar{Y}$, *so* $l_p(\hat{\mu}) = -\frac{1}{2} n \log \hat{\sigma}^2 - \frac{1}{2} n$ *where* $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$.

Then

$$\begin{aligned} 2\{l_p(\hat{\mu}) - l_p(\mu)\} &= n \log \frac{\hat{\sigma}^2_\mu}{\hat{\sigma}^2} \\ &= n \log \left[1 + \frac{(\mu - \hat{\mu})^2}{\hat{\sigma}^2} \right], \end{aligned}$$

since $\hat{\sigma}^2_\mu = \hat{\sigma}^2 + (\mu - \hat{\mu})^2$.

An approximate CI of coverage $1 - \alpha$ *for* μ *is*

$$\big\{\mu : 2\{l_p(\hat{\mu}) - l_p(\mu)\} \le c\big\}$$

where c is s.t. $P(\chi_1^2 \le c) = 1 - \alpha$. *This is*

$$
\begin{aligned}
&\left\{\mu : n\log\left(1+\frac{(\mu-\hat\mu)^2}{\hat\sigma^2}\right) \le c\right\} \\
\equiv &\left\{\mu : 1+\frac{(\mu-\hat\mu)^2}{\hat\sigma^2} \le \exp\left\{\frac{c}{n}\right\}\right\} \\
\equiv &\left\{\mu : \frac{(\mu-\hat\mu)^2}{\hat\sigma^2} \le \exp\left\{\frac{c}{n}\right\} - 1\right\} \\
\equiv &\left\{\mu : -\hat\sigma\sqrt{\exp\left(\frac{c}{n}\right)-1} \le \mu - \hat\mu \le \hat\sigma\sqrt{\exp\left(\frac{c}{n}\right)-1}\right\} \\
\equiv &\left(\hat\mu - \hat\sigma\sqrt{\exp\left(\frac{c}{n}\right)-1}, \hat\mu + \hat\sigma\sqrt{\exp\left(\frac{c}{n}\right)-1}\right). \qquad]
\end{aligned}
$$

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

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

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.

[**A16.** *Note first that*

$$\kappa_r(S_n) = n\kappa_r(Y) = n\kappa_r,$$

since $K_{S_n}(t) = nK_Y(t)$. *Also*

$$\kappa_r(Y/b) = \kappa_r(Y)/b^r = \kappa_r/b^r,$$

so, since cumulants of order higher than 2 are unaffected by location transformations,

$$\kappa_r(S_n^*) = \kappa_r\left(\frac{S_n}{\sqrt{n\sigma^2}}\right) = \frac{\kappa_r(S_n)}{(\sqrt{n\sigma^2})^r}$$
$$= \frac{n\kappa_r}{n^{r/2}\sigma^r} = n^{1-r/2}\rho_r.$$

Also, because of standardisation, $\kappa_1(S_n^*) = 0$, $\kappa_2(S_n^*) = 1$. *Then*

$$K_{S_n^*}(t) = \kappa_1(S_n^*)t + \kappa_2(S_n^*)\frac{t^2}{2!} + \kappa_3(S_n^*)\frac{t^3}{3!} + \kappa_4(S_n^*)\frac{t^4}{4!} + O(n^{-3/2})$$
$$= \tfrac{1}{2}t^2 + \frac{\rho_3}{6\sqrt{n}}t^3 + \frac{\rho_4}{24n}t^4 + O(n^{-3/2}).$$

The rest of the derivation sketched in the question follows on noting that $e^x = 1 + x + x^2/2 + O(x^3)$. *The identity is easily proved by induction on* r. *It certainly holds for* $r = 0$. *Assume it holds for* $r-1$. *By the definition of* H_r *and by integrating by parts,*

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

by the inductive hypothesis.]

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

[**A17.** *This follows immediately on noting that*

$$\int_{-\infty}^{x} H_r(y)\phi(y)dy = (-1)^r\int_{-\infty}^{x}\phi^{(r)}(y)dy$$
$$= (-1)^r\phi^{(r-1)}(x)$$
$$= -H_{r-1}(x)\phi(x). \qquad]$$

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.

[**A18.** *We have* $K_Y(t) = \mu t + \frac{1}{2}\sigma^2t^2$. *Now the saddlepoint equation is*

$$nK_Y'(\hat\phi) = x,$$
$$n\mu + \sigma^2\hat\phi n = x,$$
$$\hat\phi = (x - n\mu)/(\sigma^2 n) = \frac{x}{\sigma^2 n} - \frac{\mu}{\sigma^2}.$$

Also, $K''_Y(\hat{\phi}) = \sigma^2$*, so the saddlepoint approximation to the density of* S_n *is*

$$\begin{aligned} f_{S_n}(x) &= \frac{1}{\sqrt{2\pi}} \frac{1}{(n\sigma^2)^{\frac{1}{2}}} \exp\{n\mu\hat{\phi} + \tfrac{1}{2}n\sigma^2\hat{\phi} - \hat{\phi}x\} \\ &= \frac{1}{\sqrt{2\pi n\sigma^2}} \exp\left\{\frac{x\mu}{\sigma^2} - \frac{n\mu^2}{\sigma^2} + \tfrac{1}{2}n\sigma^2\left(\frac{x-n\mu}{n\sigma^2}\right)^2 - \frac{x^2}{n\sigma^2} + \frac{x\mu}{\sigma^2}\right\} \\ &= \frac{1}{\sqrt{2\pi n\sigma^2}} \exp\left\{\tfrac{1}{2}\frac{(x-n\mu)^2}{n\sigma^2} - \frac{(x-n\mu)^2}{n\sigma^2}\right\} \\ &= \frac{1}{\sqrt{2\pi n\sigma^2}} \exp\left\{-\frac{1}{2n\sigma^2}(x-n\mu)^2\right\}, \end{aligned}$$

which is exact, since $S_n \sim N(n\mu, n\sigma^2)$.]

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.

[**A19.** *If* Y *has pdf* $f(y) = e^{-y}$*, we have* $M_Y(t) = (1-t)^{-1}$*, so* $K_Y(t) = -\log(1-t)$*. The saddlepoint equation is*

$$\begin{aligned} nK'_Y(\hat{\phi}) &= x, \\ \frac{n}{1-\hat{\phi}} &= x, \\ n &= x - x\hat{\phi} \\ -x\hat{\phi} &= n - x, \end{aligned}$$

so that

$$\begin{aligned} K_Y(\hat{\phi}) &= -\log(1-\hat{\phi}) = -\log\left(1 - \left(1 - \frac{n}{x}\right)\right) = \log\left(\frac{x}{n}\right), \\ K''_Y(\hat{\phi}) &= x^2/n^2. \end{aligned}$$

So the saddlepoint approximation to the density of S_n *is*

$$f_{S_n}(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{(x^2/n)^{\frac{1}{2}}} \exp\left\{n \log\frac{x}{n} + n - x\right\} = c_n x^{n-1} e^{-x},$$

where

$$c_n = (2\pi)^{-\frac{1}{2}} e^n n^{-n+\frac{1}{2}}.$$

This agrees with the exact gamma pdf

$$x^{n-1}e^{-x}/\Gamma(n),$$

except for replacement of $\Gamma(n)$ *by a term which is asymptotically equivalent to the leading term of Stirling's approximation.*]

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

[**A20.** *We have*

$$f_{S_n}(s;\lambda) = \exp\{s\lambda - nK_Y(\lambda)\} f_{S_n}(s).$$

To see this note that

$$f_{S_n}(s;\lambda) = \int f(y_1;\lambda)\dots f(y_n;\lambda)dy_1\dots dy_n,$$

where the integral is over all $(y_1,\dots,y_n)$ *such that* $\sum y_i = s$, *and substitute* $f(y;\lambda) = \exp\{y\lambda - K_Y(\lambda)\}f(y)$.

The associated moment generating function is

$$M_{S_n}(t;\lambda) = \exp\{n\big(K_Y(\lambda+t) - K_Y(\lambda)\big)\},$$

and the cumulant generating function is

$$K_{S_n}(t;\lambda) = n\big(K_Y(\lambda+t) - K_Y(\lambda)\big).$$

The cumulants of S_n, *under* $f(y;\lambda)$, *are*

$$\begin{aligned} E_\lambda(S_n) &= K'_{S_n}(0;\lambda) = nK'_Y(\lambda),\\ \mathrm{var}_\lambda(S_n) &= K''_{S_n}(0;\lambda) = nK''_Y(\lambda),\\ \kappa_r(S_n;\lambda) &= K^{(r)}_{S_n}(0;\lambda) = nK^{(r)}_Y(\lambda). \end{aligned}$$

The corresponding standardised cumulants are

$$\rho_r(S_n;\lambda) = \frac{1}{n^{r/2-1}}\,\rho_r(\lambda),$$

where

$$\rho_r(\lambda) = K^{(r)}_Y(\lambda)/\{K''_Y(\lambda)\}^{r/2}.$$

We have

$$f_{S_n}(s) = \exp\{nK_Y(\lambda) - \lambda s\} f_{S_n}(s;\lambda).$$

Now the Edgeworth expansion for $f_{S_n}(s;\lambda)$ *is*

$$\begin{aligned} f_{S_n}(s;\lambda) = &\frac{1}{\sqrt{\mathrm{var}_\lambda(S_n)}}\,\phi\left(\frac{s-E_\lambda(S_n)}{\sqrt{\mathrm{var}_\lambda(\mathrm{S_n})}}\right)\\ &\times\left\{1 + \frac{1}{6\sqrt{n}}\,\rho_3(\lambda)H_3\left(\frac{s-E_\lambda(S_n)}{\sqrt{\mathrm{var}_\lambda(\mathrm{S_n})}}\right)\right.\\ &+\frac{1}{24n}\,\rho_4(\lambda)H_4\left(\frac{s-E_\lambda(S_n)}{\sqrt{\mathrm{var}_\lambda(\mathrm{S_n})}}\right)\\ &\left.+\frac{1}{72n}\,\rho_3^2(\lambda)H_6\left(\frac{s-E_\lambda(S_n)}{\sqrt{\mathrm{var}_\lambda(\mathrm{S_n})}}\right) + O(n^{-3/2})\right\}. \end{aligned}$$

Now choose $\hat\lambda$, as a function of s, so that $E_{\hat\lambda}(S_n) = s$. Then $nK_Y'(\hat\lambda) = s$. Then

$$\begin{aligned} f_{S_n}(s;\hat\lambda) &= \frac{1}{\sqrt{nK_Y''(\hat\lambda)}}\,\phi(0) \\ &\times \left\{1 + \frac{1}{24n}\,\rho_4(\hat\lambda)H_4(0) + \frac{1}{72n}\,\rho_3^2(\hat\lambda)H_6(0) + O(n^{-2})\right\}. \end{aligned}$$

Note that we can assert that the error is of order $O(n^{-2})$ since $H_{2r+1}(0) = 0$.

Since $H_4(0) = 3$ and $H_6(0) = -15$, with the notation $\hat\rho_r = \rho_r(\hat\lambda)$, $r = 3, 4$, we have

$$f_{S_n}(s;\hat\lambda) = \frac{1}{\sqrt{2\pi nK_Y''(\hat\lambda)}}\left\{1 + \frac{1}{24n}\,(3\hat\rho_4 - 5\hat\rho_3^2) + O(n^{-2})\right\},$$

from which the result follows, since

$$f_{S_n}(s) = \exp\{nK_Y(\hat\lambda) - \hat\lambda s\} f_{S_n}(s;\hat\lambda). \qquad]$$

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

[**A21.** *Taylor expansion of $g(y)$ around $\tilde y$ gives*

$$g(y) = \tilde g + \tfrac{1}{2}(y-\tilde y)\tilde g'' + \tfrac{1}{6}(y-\tilde y)^3\tilde g''' + \frac{1}{24}\,(y-\tilde y)^4\tilde g^{(4)} + O\big((y-\tilde y)^5\big),$$

where $\tilde g = g(\tilde y)$, $\tilde g'' = g''(\tilde y)$ etc. Then

$$g_n = e^{-n\tilde g}\int_a^b e^{-\frac{n}{2}(y-\tilde y)^2\tilde g''} e^{-\frac{n}{6}(y-\tilde y)^3\tilde g''' - \frac{n}{24}(y-\tilde y)^4\tilde g^{(4)} + nO(y-\tilde y)^5}\,dy.$$

Multiply and divide by $\sqrt{n\tilde g''/(2\pi)}$ and change the variable of integration to $z = (y - \tilde y)\sqrt{n\tilde g''}$, to obtain

$$\begin{aligned} g_n &\doteq \frac{e^{-n\tilde g}\sqrt{2\pi}}{\sqrt{n\tilde g''}}\int_{\mathbb{R}} \exp\left\{-\frac{z^3\tilde g'''}{6\sqrt{n}(\tilde g'')^{3/2}} - \frac{z^4\tilde g^{(4)}}{24n(\tilde g'')^2} + O(n^{-3/2})\right\}\phi(z)dz \\ &= \frac{e^{-n\tilde g}\sqrt{2\pi}}{\sqrt{n\tilde g''}}\int_{\mathbb{R}}\left(1 - \frac{1}{6\sqrt{n}}\,\frac{\tilde g'''}{(\tilde g'')^{3/2}}\,z^3\right. \\ &\qquad \left. -\frac{1}{24n}\,\frac{\tilde g^{(4)}}{(\tilde g'')^2}\,z^4 + \frac{1}{72n}\,\frac{(\tilde g''')^2}{(\tilde g'')^3}\,z^6 + O(n^{-3/2})\right)\phi(z)dz. \end{aligned}$$

If $Z \sim N(0,1)$, $E(Z^k) = 0$ if k is odd, $E(Z^k) = (k-1)(k-3)\ldots 3.1$ if k is even. Using this we obtain (3.11). The error is of order $O(n^{-2})$ and not $O(n^{-3/2})$, because the term of order $O(n^{-3/2})$ only involves expectations of odd powers of Z.]

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.

[**A22.** *If $Y_1, \ldots, Y_n$ are IID exponential with mean μ, we have $\hat\mu = \bar Y$ and we can write $l(\mu; \hat\mu) = -n\log\mu - n\hat\mu/\mu$. Then*

$$j(\hat\mu) = -\left.\frac{\partial^2 l(\mu;\hat\mu)}{\partial\mu^2}\right|_{\mu=\hat\mu} = n/\hat\mu^2.$$

Then the p^* formula gives

$$\begin{aligned} p^*(\hat\mu;\mu) &= c|j(\hat\mu)|^{\frac12}\exp\{l(\mu)-l(\hat\mu)\} \\ &= c\frac{n^{\frac12}}{\hat\mu}\exp\left\{n\log\frac{\hat\mu}{\mu} - n\left(\frac{\hat\mu}{\mu}-1\right)\right\} \\ &= cn^{\frac12}\frac{1}{\hat\mu}\left(\frac{\hat\mu}{\mu}\right)^n \exp\left(-n\frac{\hat\mu}{\mu}\right)e^n, \end{aligned}$$

where c is a normalising constant. So,

$$p^*(\hat\mu;\mu) \propto \hat\mu^{n-1}\exp\left(-\frac{n}{\mu}\hat\mu\right),$$

which is exact, since the true density of $\hat\mu$ is gamma.]

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.

[**A23.** *Let $q_0(y) = -\log(p_0(y))$. The log-likelihood $l(\mu,\sigma;y)$ can be written in the form*

$$\begin{aligned} l(\mu,\sigma;\hat\mu,\hat\sigma,a) &= -n\log\sigma - \sum_{i=1}^n q_0\left(\frac{y_i-\mu}{\sigma}\right) \\ &= -n\log\sigma - \sum_{i=1}^n q_0\left(\frac{\hat\sigma}{\sigma}a_i + \frac{\hat\mu-\mu}{\sigma}\right), \end{aligned}$$

on writing $y_i = \hat{\mu} + \hat{\sigma} a_i$*. Then we have*

$$
\begin{aligned}
&l(\mu, \sigma; \hat{\mu}, \hat{\sigma}, a) - l(\hat{\mu}, \hat{\sigma}; \hat{\mu}, \hat{\sigma}, a) \\
&= -n \log \sigma - \sum_{i=1}^{n} q_0 \left(\frac{\hat{\sigma}}{\sigma} a_i + \frac{\hat{\mu} - \mu}{\sigma} \right) + n \log \hat{\sigma} + \sum_{i=1}^{n} q_0(a_i).
\end{aligned}
$$

It is easily checked that

$$|j(\hat{\mu}, \hat{\sigma}; \hat{\mu}, \hat{\sigma}, a)|^{\frac{1}{2}} = (\hat{\sigma})^{-2} \sqrt{D(a)},$$

where

$$D(a) = \left\{ \left(\sum_i q_0''(a_i) \right) \left(n + \sum_i a_i^2 q_0''(a_i) \right) - \left(\sum_i a_i q_0''(a_i) \right)^2 \right\}.$$

The form of the p^**-formula follows immediately.*

Note that of the quantities q_i *only* $n - 2$ *are functionally independent: the likelihood equations* $l_\mu = \frac{\partial}{\partial \mu} l(\mu, \sigma) = 0$ *and* $l_\sigma = \frac{\partial}{\partial \sigma} l(\mu, \sigma) = 0$ *give two constraints:*

$$
\begin{aligned}
\sum_i q_0'(a_i) &= 0 \\
\sum_i a_i q_0'(a_i) &= n.
\end{aligned}
$$

Consider the one-to-one function

$$(y_1, \ldots, y_n) \leftrightarrow (\hat{\mu}, \hat{\sigma}, a_1, \ldots, a_{n-2}).$$

We have

$$p_Y(y_1, \ldots, y_n; \mu, \sigma) = \frac{1}{\sigma^n} \prod_{i=1}^{n} p_0 \left(\frac{y_i - \mu}{\sigma} \right)$$

and because $y_i = \hat{\mu} + \hat{\sigma} a_i$ *the joint density of* $(\hat{\mu}, \hat{\sigma}, a_1, \ldots, a_{n-2})$ *is*

$$p(\hat{\mu}, \hat{\sigma}, a_1, \ldots, a_{n-2}; \mu, \sigma) = \frac{1}{\sigma^n} \prod_{i=1}^{n} p_0 \left(\frac{\hat{\mu} - \mu}{\sigma} + \frac{\hat{\sigma}}{\sigma} a_i \right) |J|$$

where $|J|$ *is the Jacobian determinant of* $(y_1, \ldots, y_n)$ *expressed as a function of* $(\hat{\mu}, \hat{\sigma}, a_1, \ldots, a_{n-2})$*. From the representation*

$$
\begin{aligned}
y_1 &= \hat{\mu} + \hat{\sigma} a_1 \\
&\vdots \\
y_{n-2} &= \hat{\mu} + \hat{\sigma} a_{n-2} \\
y_{n-1} &= \hat{\mu} + \hat{\sigma} f_1(a_1, \ldots, a_{n-2}) \\
y_n &= \hat{\mu} + \hat{\sigma} f_2(a_1, \ldots, a_{n-2})
\end{aligned}
$$

we obtain

$$J = \begin{bmatrix} 1 & a_1 & \hat\sigma & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ 1 & a_{n-2} & 0 & & \hat\sigma \\ 1 & f_1(a^*) & \hat\sigma f_{1,1}(a^*) & \dots & \hat\sigma f_{1,n-2}(a^*) \\ 1 & f_2(a^*) & \hat\sigma f_{2,1}(a^*) & \dots & \hat\sigma f_{2,n-2}(a^*) \end{bmatrix}$$

say, where $a^* = (a_1, \dots, a_{n-2})$. *We may write this as*

$$J = \begin{bmatrix} \underline{1}_{n-2} & a^* & \hat\sigma I_{n-2} \\ \underline{1}_2 & f(a) & \hat\sigma F(a) \end{bmatrix}.$$

Recall that

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |B||C - DB^{-1}A|.$$

With $B = \hat\sigma I_{n-2}$ *we obtain*

$$|J| = \hat\sigma^{n-2} h_n(a),$$

say.

Denote by $p(a)$ *the marginal density of* $(a_1, \dots, a_{n-2})$. *Then the conditional density of* $(\hat\mu, \hat\sigma)$ *given* a *is*

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

where $c(a) = h_n(a)/p(a)$ *can be interpreted as a normalising constant.*

This **is** *the* p^**-formula, which is therefore exact. Note that this analysis shows that the normalising constant in the* p^**-formula does not depend on* (μ, σ).

24. Let $X_1, \dots, X_n$ be independent exponential random variables with mean $1/\lambda$ and let $Y_1, \dots, Y_n$ 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.

[**A24.** *Simple calculations give* $\hat\psi = \bar x/\bar y, \hat\lambda = 1/\bar x$, *where* $\bar x, \bar y$ *denote the respective sample means. The log-likelihood may be written*

$$l(\theta) = 2n \log\lambda + n \log\psi - n\left(\frac{\lambda\psi}{\hat\lambda\hat\psi} + \frac{\lambda}{\hat\lambda}\right),$$

and

$$j(\hat{\theta}) = n \begin{pmatrix} \hat{\psi}^{-2} & (\hat{\psi}\hat{\lambda})^{-1} \\ (\hat{\lambda}\hat{\psi})^{-1} & 2\hat{\lambda}^{-2} \end{pmatrix}.$$

Noting that no ancillary is required, the approximation to the joint density of $(\hat{\psi}, \hat{\lambda})$ *is*

$$p^*(\hat{\psi}, \hat{\lambda}; \psi, \lambda) = c\frac{n}{\hat{\psi}\hat{\lambda}}\big(\frac{\lambda}{\hat{\lambda}}\big)^{2n}\big(\frac{\psi}{\hat{\psi}}\big)^n \exp\{-n\big(\frac{\lambda\psi}{\hat{\psi}\hat{\lambda}} + \frac{\lambda}{\hat{\lambda}} - 2\big)\}.$$

The approximation to the marginal density of $\hat{\psi}$ *is*

$$\int p^*(\hat{\psi}, \hat{\lambda}; \psi, \lambda)d\hat{\lambda} = c\big(\frac{\psi}{\hat{\psi}}\big)^n \frac{n}{\hat{\psi}} \exp(2n) \int_0^\infty t^{2n-1} \exp\{-n\big(\frac{\psi}{\hat{\psi}} + 1\big)t\}dt,$$

on writing $t = \lambda/\hat{\lambda}$.

Evaluation of the (gamma) integral gives the density approximation

$$c\frac{\Gamma(2n)\exp(2n)}{\hat{\psi}}\big(\frac{\psi}{\hat{\psi}}\big)^n\big(\frac{\psi}{\hat{\psi}} + 1\big)^{-2n},$$

which equals

$$c\Gamma(2n)\exp(2n)\frac{1}{\psi}\big(\frac{\hat{\psi}}{\psi}\big)^{n-1}\big(1 + \frac{\hat{\psi}}{\psi}\big)^{-2n}$$

on simplification, so that the approximation is seen to be exact, apart from the normalising factor.]

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.

[**A25.** *We have* $\hat{\mu}_{\sigma^2} = \bar{Y}$, *so that the profile log-likelihood is*

$$l_p(\sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\Sigma(Y_i - \bar{Y})^2.$$

The profile score is therefore

$$\frac{\partial l_p(\sigma^2)}{\partial\sigma^2} = -\frac{n}{2}(\sigma^2)^{-1}(1 - \hat{\sigma}^2/\sigma^2),$$

where $\hat{\sigma}^2 = \frac{1}{n}\Sigma(Y_i - \bar{Y})^2$ *has expectation*

$$E(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)\sigma^2 = \left(1 - \frac{1}{n}\right)\sigma^2.$$

So,

$$E\left(\frac{\partial l_p(\sigma^2)}{\partial\sigma^2}\right) = -\frac{n}{2\sigma^2}\left\{1-\frac{1}{\sigma^2}\left(1-\frac{1}{n}\right)\sigma^2\right\} = -\frac{1}{2\sigma^2} \neq 0.$$

In general, the expectation of the profile score is of order $O(1)$. *We have*

$$\begin{aligned} l(\mu,\sigma^2) &= l(\mu,\sigma^2;\hat\mu,\hat\sigma^2) \\ &= -\frac{n}{2}\log\sigma^2 - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum_i (Y_i-\mu)^2 \\ &= -\frac{n}{2}\log\sigma^2 - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\{n\hat\sigma^2 + n(\hat\mu-\mu)^2\}. \end{aligned}$$

Then

$$l_{\mu;\hat\mu}(\mu,\sigma^2;\hat\mu,\hat\sigma^2) = \frac{\partial}{\partial\mu}\left(-\frac{n}{\sigma^2}(\hat\mu-\mu)\right) = \frac{n}{\sigma^2}.$$

Also,

$$j_{\mu\mu}(\mu,\sigma^2;\hat\mu,\hat\sigma^2) = \frac{n}{\sigma^2},$$

so the modifying factor $M(\sigma^2) = \left(\frac{\sigma^2}{n}\right)^{\frac{1}{2}}$ *and, ignoring constants, the modified profile log-likelihood is*

$$\begin{aligned} \tilde l_p &= -\frac{(n-1)}{2}\log\sigma^2 - \frac{1}{2\sigma^2}n\hat\sigma^2 \\ \frac{\partial\tilde l_p(\sigma^2)}{\partial\sigma^2} &= -\frac{(n-1)}{2\sigma^2} + \frac{n\hat\sigma^2}{2\sigma^4} = -\frac{n}{2\sigma^2}\left(1-\frac{\hat\sigma^2}{\sigma^2}\right) + \frac{1}{2\sigma^2}. \end{aligned}$$

Then, $E\left(\frac{\partial\tilde l_p(\sigma^2)}{\partial\sigma^2}\right) = 0$.

In general, the expectation of the modified profile score is of order $O(n^{-1})$.]

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.

[**A26.** *The log-likelihood is*

$$l(\theta) = -n\log\lambda - \psi\sum x_j - \frac{1}{\lambda}\sum\exp(-\psi x_j)y_j,$$

which is equivalent to

$$-n\log\lambda - \psi\sum x_j - \frac{\hat\lambda}{\lambda}\sum\exp\{(\hat\psi-\psi)x_j + a_j)\},$$

using the definition of the ancillary.

Maximising $l(\theta)$ with respect to λ gives

$$\hat{\lambda}_\psi = \frac{\sum \exp\{-\psi x_j\} y_j}{n}.$$

Then, on noting that

$$\hat{\lambda}_\psi = \hat{\lambda}\frac{\sum \exp\{(\hat{\psi} - \psi)x_j + a_j\}}{n},$$

simple manipulation shows that the modifying factor $M(\psi)$ is free of ψ and depends only on the data. Hence $l_p(\psi) \equiv \tilde{l}_p(\psi)$. In detail: $\partial\hat{\lambda}_\psi/\partial\hat{\lambda} = S$, say, with $-\hat{j}_\psi = n/\hat{\lambda}_\psi^2 - 2\hat{\lambda}S/\hat{\lambda}_\psi^3$. Hence $\hat{j}_\psi = n^3/(\hat{\lambda}^2 S^2)$, so that $M(\psi)$ is a function of $n, \hat{\lambda}$, not 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 $w' = w/(1 + 3/2n)$.

[**A27.** *The first part is a simple exercise. Now*

$$\begin{aligned} w &= n \log\{1 + t^2/(n-1)\} \\ &\simeq n\left\{\frac{t^2}{n-1} - \frac{1}{2}\frac{t^4}{(n-1)^2}\right\}, \end{aligned}$$

on using the expansion $\log(1+x) = x - \frac{x^2}{2} + \cdots$. Also $E(t^2) = \mathrm{var(t)} = \frac{\mathrm{n-1}}{\mathrm{n-3}}$, since $t \sim t_{n-1}$. Also $E(t^4) = 3 + O(n^{-1})$ [recall $t_\nu \to N(0,1)$ as $\nu \to \infty$]. So, by the above,

$$\begin{aligned} E(w) &= \frac{n}{n-3} - \frac{3}{2n} + O(n^{-2}) \\ &= \frac{1}{1-3/n} - \frac{3}{2n} + O(n^{-2}) \\ &= 1 + \frac{3}{n} - \frac{3}{2n} + O(n^{-2}) \\ &= 1 + \frac{3}{2n} + O(n^{-2}), \end{aligned}$$

as required.

I performed the following simple simulation exercise. For each of a set of values of n I generated 1 million null values of w and w', generating from $N(0,1)$ and testing $H_0 : \mu = 0$, and recorded the proportion exceeding 3.84, the upper 5% point of $\mathcal{X}_1^2$. It is apparent that the chi-squared approximation to the distribution of w' is adequate for small n, while the distribution of w approaches χ_1^2 more slowly.

n	w	w'
10	0.0702	0.0520
20	0.0593	0.0506
30	0.0558	0.0501
40	0.0548	0.0503

]

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 $n \to \infty$.*]

[**A28.** *It is straightforward to obtain the maximum likelihood estimator as*

$$\hat{\sigma}^2 = \frac{\sum (X_j - Y_j)^2}{4n}.$$

Then note that $\hat{\sigma}^2$ has mean $\sigma^2/2$ and by WLLN converges in probability to $\sigma^2/2$, so that it is inconsistent.

The profile log-likelihood is

$$l_p(\sigma^2) = -\frac{S}{4\sigma^2} - 2n \log \sigma.$$

Then observe that $\hat{\mu}_i \equiv \hat{\mu}_{i_{\sigma^2}}$, and that

$$\frac{\partial^2 l}{\partial \mu_i \partial \mu_j} = \begin{cases} -\frac{2}{\sigma^2}, & \textit{if } i = j, \\ 0, & \textit{otherwise} \end{cases},$$

so that $| j_{\mu\mu}(\sigma^2, \hat{\mu}_{\sigma^2}) | = (\frac{2}{\sigma^2})^n$. *Hence*

$$\tilde{l}_p(\sigma^2) = -\frac{S}{4\sigma^2} - n \log \sigma.$$

It is easily checked that the estimator obtained by maximising $\tilde{l}_p(\sigma^2)$ *is consistent.*

We have $S/(2\sigma^2) \sim \chi^2_n$, *so that the density of* S *is proportional to* $\sigma^{-n} \exp\{-s/(4\sigma^2)\}$, *so that the marginal log-likelihood is equivalent to the modified profile log-likelihood.*]