Solution 1 [10, seen] See parts of slides 22–39.

The essay should cover the repeated sampling, Bayesian and randomisation bases of inference, touching for example on

- repeating sampling/frequentist inference: observed data y^{o} are regarded as sampled from a hypothetical population with density $f(y;\theta)$; role of a relevant subset $\mathcal{A}$ from which y^{o} is supposed to be drawn; properties of estimator $\widehat{\theta}$ due to sampling variation; role of pivots in inference on θ. The probability comes from the notional repeated sampling, and θ is regarded as a fixed unknown constant. One main difficulty is in plausible specification of the sampling framework, and another is that the relevance of hypothetical datasets to the data y^{o} actually observed may be questioned;

- Bayesian inference: the unknown θ is regarded as random with prior density $\pi(\theta)$, which is updated to a posterior density $\pi(\theta \mid y^{\text{o}})$ when y^{o} is obtained. The probability comes from the prior density, so θ is regarded as random and no comparison is made with other hypothetical datasets. The main difficulty is the specification of the prior: Jeffreys priors $\pi(\theta) \propto \imath(\theta)^{1/2}$ are seen as objective but have difficulties in higher dimensions (Stein paradox), and 'non-informative' priors are not transformation-invariant, so weak forms of proper priors are commonly used;

- randomisation inference: the experimenter imposes a scheme whereby treatments are allocated to experimental units, and the basis of the inference is the corresponding randomisation distribution. Similar ideas are applied in sample surveys. Here there is a physical basis for the randomisation inference, but this can only be used in specific settings when the experiment is essentially under the control of the experimenter, and unless the units have been sampled randomly from a population, the inference strictly applies only to the units themselves.

Solution 2

(a) [2, seen] See slides 44–46.

(b) [4, seen] Slide 46.

(c) [4, unseen] The joint density of the two variables is

$$f(y_1, y_2; \theta) = f(y_1;\theta) f(y_2;\theta) = \frac{(m\theta)^{y_1}}{y_1!} e^{-m\theta} \times \frac{\{m(1-\theta)\}^{y_2}}{y_2!} e^{-m(1-\theta)} = \theta^{y_1}(1-\theta)^{y_2} \times h(y,y_2),$$

where the second factor does not depend on θ. Hence $S = s(Y) = (Y_1, Y_2)$ is sufficient for θ, by the factorisation theorem. It is also minimal sufficient, because the ratio

$$\frac{f(y_1';\theta) f(y_2';\theta)}{f(y_1;\theta) f(y_2;\theta)} \propto \frac{\theta^{y_1'}(1-\theta)^{y_2'}}{\theta^{y_1}(1-\theta)^{y_2}}$$

is free of θ iff $y_1 = y_1'$ and $y_2 = y_2'$, i.e., iff $s(Y) = s(Y')$.

Now $A = Y_1 + Y_2$ has a Poisson distribution with mean $m\theta + m(1-\theta) = m$, which is known, so A is ancillary (slide 52) and in principle inferences should be conditioned on the observed value a of A. This distribution is

$$\begin{aligned} f(y_1, y_2 \mid a; \theta) &= \frac{f(y_1;\theta) f(y_2;\theta)}{f(a)} \\ &= \frac{\frac{(m\theta)^{y_1}}{y_1!} e^{-m\theta} \times \frac{\{m(1-\theta)\}^{y_2}}{y_2!} e^{-m(1-\theta)}}{\frac{m^a}{a!} e^{-m}} \\ &= \frac{a!}{y_1!(a-y_1)!} \theta^{y_1}(1-\theta)^{a-y_1}, \end{aligned}$$

where we have set $y_2 = a - y_1$. Hence conditional on A, $Y_1 \sim B(a, \theta)$, and inference for θ would be based on this density.

Alternatively here one state Lemma 17 of the course and argue directly from that.

Solution 3

(a) [4, seen] Slide 63 and its note. In particular the observed P-value may be written as

$$p_{\text{obs}} = \mathrm{P}_0(T \geq t_{\text{obs}}) = 1 - F(t_{\text{obs}})$$

in terms of the observed value t_{obs} of T, so the corresponding random variables satisfy $P_{\text{obs}} = 1 - F(T_{\text{obs}})$, and for $x \in (0, 1)$ this gives

$$\mathrm{P}_0(P \leq x) = \mathrm{P}_0\{1 - F(T_{\text{obs}}) \leq x\} = \mathrm{P}_0\{F^{-1}(1-x) \leq T_{\text{obs}}\} = 1 - F\{F^{-1}(1-x)\} = x,$$

as required.

(b) [4, unseen but related to Problem 5 of week 5].

If all the H_j are true, then $P_1, \ldots, P_m \overset{\text{iid}}{\sim} U(0, 1)$, giving for $0 < x < 1$ that

$$\begin{aligned}
\mathrm{P}(P^* \leq x) &= \mathrm{P}\left(1 - \{\max_j (1 - P_j)\}^m \leq x\right) \\
&= \mathrm{P}\left\{(1-x)^{1/m} \leq \max_j (1 - P_j)\right\} \\
&= 1 - \mathrm{P}\left\{\max_j (1 - P_j) < (1-x)^{1/m}\right\} \\
&= 1 - \mathrm{P}\left\{1 - P_j < (1-x)^{1/m}\right\}^m \\
&= 1 - \{(1-x)^{1/m}\}^m \\
&= x,
\end{aligned}$$

since P_j and $1 - P_j$ have $U(0, 1)$ distributions under H_0. Hence $P^* \sim U(0, 1)$ under H_0.

If P^* is small compared to its null distribution, then at least one of the P_j must have been small compared to its distribution, so P^* can be expected to be unusually small if at least one (but perhaps only one) of the P_j was exceptionally small, casting doubt on the truth of the corresponding H_j.

(c) [2, unseen]

The Bonferroni method would take some $\alpha \in (0, 1)$ and reject H_0 when

$$\begin{aligned}
\min_j P_j \leq \alpha/m \quad &\Longleftrightarrow \quad \max_j (1 - P_j) \geq 1 - \alpha/m \\
&\Longleftrightarrow \quad 1 - \{\max_j (1 - P_j)\}^m \leq 1 - (1 - \alpha/m)^m
\end{aligned}$$

and if m is large and α is small, then $1 - (1 - \alpha/m)^m \approx \alpha$. Hence in this case both methods are similar if rejection at a fixed level α is required.

Solution 4

(a) [2, seen] See slide 115 etc.

(b) [2, seen] See slides 112–113.

(c) [4, seen/unseen] This is a simplified version of Problem 4 of Week 9. The overall log likelihood is

$$\ell(\mu_1, \ldots, \mu_m, \sigma^2) \equiv -\tfrac{1}{2}\left\{2m \log \sigma^2 + \frac{1}{\sigma^2}\sum_{j=1}^{m}(y_{j1} - \mu_j)^2 + (y_{j2} - \mu_j)^2\right\},$$

and differentiation with respect to μ_j yields $\widehat{\mu}_j = (y_{j1} + y_{j2})/2$. Inserting these into $\ell(\mu_1, \ldots, \mu_m, \sigma^2)$ and simplifying using the fact that

$$(y_{j1} - \widehat{\mu}_j)^2 = (y_{j2} - \widehat{\mu}_j)^2 = (y_{j1} - y_{j2})^2/4$$

yields the required expression for the profile log likelihood for σ^2.

A further differentiation yields

$$\widehat{\sigma}^2 = \frac{1}{4m}\sum_{j=1}^{m}(y_{j1} - y_{j2})^2,$$

and the hint implies that $(y_{j1} - y_{j2})^2 \stackrel{\mathrm{D}}{=} 2\sigma^2\chi_1^2$, so $\widehat{\sigma}^2 \stackrel{\mathrm{D}}{=} (2m)^{-1}\chi_m^2 \stackrel{P}{\longrightarrow} \sigma^2/2$ as $m \to \infty$. Hence $\widehat{\sigma}^2$ is inconsistent.

(d) [2, unseen] If we maximise the log likelihood of $z_j = y_{j1} - y_{j2} \stackrel{\mathrm{iid}}{\sim} \mathcal{N}(0, 2\sigma^2)$, i.e.,

$$-\tfrac{1}{2}\left\{m \log(2\sigma^2) + \frac{1}{2\sigma^2}\sum_{j=1}^{m} z_j^2\right\},$$

then we obtain an unbiased (and consistent) *marginal likelihood* estimator $\widehat{\sigma}^2 = (2m)^{-1}\sum_{j=1}^{m} z_j^2$.

Solution 5

(a) [2, seen] See slides 224–231.

(b) [5, unseen] The joint density is [1]

$$\prod_{j=1}^{n} f(y_j \mid \lambda_j)\pi(\lambda_j \mid \beta) \times \pi(\beta) = \left\{\prod_{j=1}^{n} \lambda_j e^{-\lambda_j y_j}\frac{\beta^\alpha \lambda_j^{\alpha-1}}{\Gamma(\alpha)}e^{-\beta\lambda_j}\right\} \times \tau e^{-\tau\beta}, \quad y_j, \lambda_j, \beta > 0,$$

and integration over the λ_j gives [1]

$$f(y_1, \ldots, y_n \mid \beta) = \left\{\prod_{j=1}^{n} \frac{\Gamma(\alpha+1)\beta^\alpha}{\Gamma(\alpha)(y_j + \beta)^{\alpha+1}}\right\} \times \tau e^{-\tau\beta}, \quad y_j, \beta > 0,$$

and recalling that $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ for any $\alpha > 0$ and a further integration over β gives the expression above.

For Laplace approximation [3] we write the integrand as $e^{-h(\beta)}$, where

$$h(\beta) = \tau\beta - n\alpha \log \beta - \log \tau + (\alpha+1)\sum_{j=1}^{n} \log(y_j + \beta), \quad \beta > 0,$$

then find $\tilde{\beta}$ as the (positive) solution to the equation $h'(\beta) = 0$, set $h_2 = h''(\tilde{\beta})$, and approximate the integral by $e^{-h(\tilde{\beta})}(2\pi/h_2)^{1/2}$. This has a relative error of order $1/n$.

(c) [3, unseen] The marginal density of λ_1 is

$$f(\lambda_1 \mid y_1, \ldots, y_n) = \frac{f(\lambda_1, y_1, \ldots, y_n)}{f(y_1, \ldots, y_n)} = \frac{\int_0^\infty e^{-h_1(\beta,\lambda_1)}\,\mathrm{d}\beta}{\int_0^\infty e^{-h(\beta)}\,\mathrm{d}\beta},$$

say, where

$$h_1(\beta, \lambda_1) = \lambda_1 y_1 - \log\lambda_1 + \tau\beta - n\alpha\log\beta - \log\tau + (\alpha+1)\sum_{j=2}^{n}\log(y_j+\beta), \quad \beta > 0.$$

Laplace approximation would now involve computing $\tilde{\beta}(\lambda_1)$ by minimising $h_1(\beta, \lambda_1)$ for each value of λ_1 on a grid, and then computing the ratio of the two Laplace approximations (here and for (b)) at the points of the grid. We would expect this to be more accurate than the approximation in (b), because the errors in the numerator and denominator are both $O(n^{-1})$ and might be expected to cancel to some extent.