PRINCIPLES OF STATISTICS: Example Sheet 4 (of 4)

1. A Bayesian statistician takes a prior density

\[ \pi(\theta) \propto (\beta\theta)^{\alpha-1} \exp(-\beta\theta)I_{\{\theta>0\}} \]

for the mean of a Poisson distribution, and then observes a sample \(X_1, \ldots, X_n\) taken from that distribution. What is his posterior distribution given those observations?

**Solution**

Multiplying prior and likelihood we get posterior

\[ \pi(\theta|x) \propto (\beta\theta)^{\alpha-1} \exp(-\beta\theta)\theta^{\Sigma X_i} \exp(-n\theta) \]

which, on simplification, you should be able to identify as a \(\Gamma(\alpha+\Sigma X_i, \beta+n)\) distribution.

*Why is this being asked?* More practice manipulating priors and posteriors; compare with previous example sheets.
2. Find the minimal sufficient statistics for $\theta$ based on a sample of size $n$ in each of the following examples.

(i) $f(x|\alpha, \beta) = (\beta x)^{\alpha-1}e^{-\beta x}\beta I_{x>0}/\Gamma(\alpha)$, where $\theta = (\alpha, \beta)$;

(ii) $f(x|\mu, V) = \exp(-\frac{1}{2}(x - \mu)^T V^{-1} (x - \mu))/(2\pi)^{d/2}\sqrt{\det V}$, where $\theta = (\mu, V)$, and $\mu$ is $d \times 1$, $V$ is $d \times d$;

(iii) $U([a, b])$;

(iv) $f(x|a, b) = b/\pi((x - a)^2 + b^2)$, with $\theta = (a, b)$.

**Solution**

(i) 

$$
\frac{f(x|\theta)}{f(x'|\theta)} = \exp(-\beta \sum^n_i (x_i - x'_i) + (\alpha - 1) \sum^n_i \log(x_i) - \log(x'_i))
$$

so, by the likelihood ratio criterion, $(\sum^n_i x_i, \sum^n_i \log(x_i))$ is minimal sufficient.

(ii) Similarly to (i), the likelihood ratio is

$$
\exp\left(-\frac{1}{2} \sum (x_i - \mu)^T V^{-1} (x_i - \mu) + \frac{1}{2} \sum (x'_i - \mu)^T V^{-1} (x'_i - \mu)\right)
$$

$$
= \exp\left(-\frac{1}{2} \text{tr}(V^{-1} S(x)) - \frac{1}{2} n(\bar{x} - \mu)^T V^{-1} (\bar{x} - \mu) + \frac{1}{2} \text{tr}(V^{-1} S(x')) + \frac{1}{2} n(\bar{x}' - \mu)^T V^{-1} (\bar{x}' - \mu)\right)
$$

where $S(x) = \Sigma(x_i - \bar{x})(x_i - \bar{x})^T$. The likelihood ratio is only identical for all $\theta$ if $\bar{x} = \bar{x}'$ and $S(x) = S(x')$, and so $(\bar{x}, S(x))$ is the minimal sufficient statistic.

(iii) Likelihood ratio criterion again gives $(\text{min } x_i, \text{max } x_i)$ as the minimal sufficient statistic.

(iv) The likelihood ratio is

$$
\prod (x_i - a)^2 + b^2
$$

$$
(x'_i - a)^2 + b^2.
$$

Now, if this is to be the same for all $(a, b)$, consider the case where $a = x_1$ and let $b \to 0$. To keep the ratio of polynomials constant we must have $x_1$ as one of the $x'_j$. Repeat this argument for all $x_i$, and we see that the sets $\{x_1, ..., x_n\}$ and $\{x'_1, ..., x'_n\}$ must be identical. Therefore the minimal sufficient statistic is the whole dataset $x_1, ..., x_n$. (Other arguments are possible)

Why is this being asked? Identification of minimal sufficient statistics. As soon as you get outside ‘nice’ examples, the situation in (iv) is actually very common.
3. Let \( X_1, \ldots, X_n \) be IID geometric(\( p \)) random variables:

\[
P(X_i = k) = (1 - p)^k p \quad (k = 0, 1, \ldots).
\]

(i) Prove that \( t_1(X) \equiv I_{(X_1 = 0)} \) is an unbiased estimator of \( p \).

(ii) Find the minimal sufficient statistic \( T \) for \( p \) based on the sample of \( n \) observations. What is its distribution? What is the distribution of \( t_1(X) \) given \( T \)?

(iii) Use the Rao-Blackwell Theorem to find an unbiased estimator \( \hat{p}_n \) of \( p \) with smaller variance than \( t_1 \).

**Solution**

(i) \( \mathbb{E}_1(X) = \mathbb{E}[I_{(X_1 = 0)}] = \mathbb{P}(X_1 = 0) = p \)

(ii) \( T = \sum^n_i X_i \) is minimal sufficient by the likelihood ratio criterion. Its distribution (the sum of \( n \) independent geometric distributions) is Negative Binomial \((n, p)\), with

\[
\mathbb{P}(T = k) = p^n (1 - p)^k \binom{n + k - 1}{k}.
\]

Also,

\[
\mathbb{P}(X_1 = 0| T = k) = \frac{\mathbb{P}(X_1 = 0, T = k)}{\mathbb{P}(T = k)}
= \frac{\mathbb{P}(X_1 = 0)\mathbb{P}(\sum^n_2 X_1 = k)}{\mathbb{P}(T = k)}
= \frac{pp^{n-1}(1 - p)^k \binom{n+k-2}{k}}{p^n (1 - p)^k \binom{n+k-1}{k}}
= \frac{n - 1}{n + k - 1}
\]

(iii) Rao-Blackwell tells you that conditioning \( t_1 \) on \( T \) gives you an unbiased estimator with smaller (in fact optimal) variance. In this case the estimator is \( \frac{n-1}{n+k-1} \).

*Why is this being asked?* Without Rao-Blackwell it’s doubtful one would come up with the estimator given - its certainly more difficult to prove its unbiasedness. However, in some situations, insisting on unbiasedness can lead to potentially odd behaviour when we minimise variance in this way.
4. Let $X_1, \ldots, X_n$ be an independent sample from a normal distribution with mean 0 and variance $\sigma^2$. Explain in as much detail as you can how to construct a UMPU test of $H_0 : \sigma = \sigma_0$ against $H_1 : \sigma \neq \sigma_0$.

Solution

The likelihood,

$$f(x|\sigma^2) = (2\pi)^{-n/2} \sigma^{-n} e^{-\frac{1}{2\sigma^2} \sum x_i^2}$$

is in exponential family form, and is routinely manipulated to show that $T = \sum x_i^2$ is a minimal sufficient statistic for $\sigma$. Note that $T/\sigma^2$ has a $\chi^2_n$ distribution. From results in your notes, you can therefore assert that the UMPU test of $H_0$ vs $H_1$ has acceptance region

$$\{c_1 \leq T \leq c_2\}$$

Writing $g$ as the density of a $\chi^2_n$ variable, to satisfy the size requirement we need

$$\int_{c_1/\sigma_0^2}^{c_2/\sigma_0^2} g(y)dy = 1 - \alpha$$

The unbiased requirement is that

$$\int_{c_1/\sigma^2}^{c_2/\sigma^2} g(y)dy < 1 - \alpha$$

for all $\sigma \neq \sigma_0$. Differentiating with respect to $\sigma$, and satisfying the inequality in the ‘worst case’, leads to

$$g(c_1/\sigma_0^2) = g(c_2/\sigma_0^2).$$

You can solve the two requirements (numerically) to get $c_1$ and $c_2$ for any given $\sigma_0, n$.

Why is this being asked? Practice constructing two-sided UMPU tests in regular situations.
5. Let $X_1, \ldots, X_n$ be an independent sample from $N(\mu, \mu^2)$. Let $T_1 = \bar{X}$ and $T_2 = \sqrt{\frac{1}{n}} \sum X_i$. Show that $Z = T_1/T_2$ is ancillary. Explain why the Conditionality Principle would lead to inference about $\mu$ being drawn from the conditional distribution of $T_2$ given $Z$.

Solution

Write $X_i = \mu Y_i$, where the $Y_i$ are iid $N(1,1)$ variables. Then, when calculating $T_1/T_2$ in terms of the $Y_i$, the $\mu$ terms all cancel and we are left with a (complicated) function of iid $N(1,1)$ variables. Therefore $Z$ has no dependence on $\mu$ and is ancillary. The Conditionality Principle states that if you can condition on ancillary statistics then you should, leading to inference based on $T_1|Z$.

Why is this being asked? Conditionality is an important and powerful idea; intuitively ignoring irrelevant parts of the likelihood should simplify (and improve) inference in the bits that are of interest.
6. Let $X_1, ..., X_n$ be independent random variables with a common density function

$$f(x; \theta) = \theta e^{-\theta x}, \quad x \geq 0,$$

where $\theta \in (0, \infty)$ is an unknown parameter. Consider testing the null hypothesis $H_0 : \theta \leq 1$ against the alternative $H_1 : \theta > 1$. Construct a uniformly most powerful test of size $\alpha$.

**Solution**

This $Exp(\theta)$ density has MLR with respect to $-\sum X_i$, so the test which rejects when $\sum X_i < c$ is UMP, from theoretical results in your notes. Similarly to Q5, define $Z_i = \theta X_i$, so that the $Z_i$ are iid $Exp(1)$. Then, to get the size requirement right, we need

$$\alpha = \mathbb{P}(\sum X_i \leq c) = \mathbb{P}(\sum Z_i \leq \alpha \theta) = G(\alpha \theta), \quad \forall \theta \leq 1,$$

where $G$ is the cumulative distribution function of a $\Gamma(n, 1)$ distribution. If we solve this for the boundary case, so that $\alpha = G(c)$, then the other values also follow, and so the test is defined up to an (numeric) integral.

*Why is this being asked?* Practice in constructing one-sided UMP tests in regular situations.
7. Let $X_1, \ldots, X_n$ be an independent sample of size $n$ from the uniform distribution on $(0, \theta)$.

Show that there exists a uniformly most powerful size $\alpha$ test of $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$, and find its form.

Let $T = \max(X_1, \ldots, X_n)$. Show that the test

$$
\phi(x) = \begin{cases} 
1, & \text{if } t > \theta_0 \text{ or } t \leq b \\
0, & \text{if } b < t \leq \theta_0,
\end{cases}
$$

where $b = \theta_0 \alpha^{1/n}$, is a uniformly most powerful test of size $\alpha$ for testing $H_0$ against $H'_1 : \theta \neq \theta_0$.

Solution

The first part follows from straightforward Neyman-Pearson theory; consider testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, for $\theta_1 > \theta_0$. Then the usual likelihood ratio test rejects $H_0$ if $T > c$, where $T = \max X_i$. To find $c$, we require

$$
\alpha = \mathbb{P}(T > c) = 1 - \left( \frac{c}{\theta_0} \right)^n,
$$

which is solved when $c = \theta_0(1 - \alpha)^{1/n}$. Therefore the test doesn’t depend on $\theta_1$ and is UMP for testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$. An almost identical argument shows that the test which rejects $H_0$ when $T \leq \theta_0 \alpha^{1/n}$ is UMP for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta < \theta_0$.

The respective powers of these two UMP tests are calculated as

$$
1 - \left( \frac{\theta_0}{\theta_1} \right)^n (1 - \alpha), \quad \theta_1 > \theta_0
$$

$$
\left( \frac{\theta_0}{\theta_1} \right)^n \alpha, \quad \theta_1 < \theta_0
$$

Calculating the power of the given two-sided test, we see it coincides exactly with this, and we can similarly calculate that the size, $\alpha$, is correct. Therefore the given test is UMP for testing $H_0$ against $H_1 : \theta \neq \theta_0$.

Why is this being asked? This question produces a UMP test of a two-sided hypothesis, in a non-regular situation. In regular situations such things don’t exist.
8. A local councillor suspects that traffic conditions have become more hazardous in Ambridge than in Borchester, so she records the numbers $A$ and $B$ of accidents occurring in each place in the course of a month. Assuming that $A$ and $B$ are independent Poisson random variables with parameters $\lambda$ and $\mu$, it is desired to construct an unbiased test of size $\alpha = 1/16$ of $H_0 : \lambda \geq \mu$ against $H_1 : \lambda < \mu$.

(i) Prove that $A + B \sim \text{Poi}(\lambda + \mu)$, and that conditional on $A + B = n$, $A \sim \text{Bin}(n, p)$, where $p = \lambda/(\lambda + \mu)$.

(ii) Show that if $X \sim \text{Bin}(n, p)$ then the UMPU test of $H_0 : p \geq 1/2$ against $H_1 : p < 1/2$ is of the form

$$\varphi(k, n) = \begin{cases} 1 & (0 \leq k < \kappa_n(\alpha)) \\ \pi_n(\alpha) & (k = \kappa_n(\alpha)) \\ 0 & (\kappa_n(\alpha) < k \leq n) \end{cases}$$

where $\kappa_n(\alpha)$ and $\pi_n(\alpha)$ are chosen so that $E_{1/2} \varphi(X, n) = \alpha$.

(iii) Show that the test of the original hypotheses defined by choosing $H_1$ with probability $\varphi(A, A + B)$ is unbiased with size $\alpha$.

(iv) Carry out the test when $A = 2$ and $B = 5$, and also when $A = 0$ and $B = 3$.

**Solution**

(i) First result should be familiar from IA Probability. For the conditional distribution,

$$\Pr(A = a | A + B = n) = \frac{\Pr(A = a) \Pr(B = n - a)}{\Pr(A + B = n)} = \frac{\lambda^a e^{-\lambda} \mu^{n-a} e^{-\mu}}{a! (n-a)! e^{-\lambda - \mu}} \frac{n!}{n}$$

$$= \binom{n}{a} p^a (1-p)^{n-a},$$

where $p$ is as specified in the question.

(ii) Follows from direct application of theory from the course. (Binomial distribution has MLR with respect to $X$)

(iii) Reparameterising as $\gamma = \lambda + \mu$, $p = \lambda/(\lambda + \mu)$, then from the construction of the test we see that it is unaffected by varying $\gamma$. $\mathbb{E} \varphi$ can also be seen to be decreasing in $p$, and is equal to $\alpha$ at $p = 1/2$ by construction. Hence the test is unbiased.

(iv) In the first instance, we compare $A = 2$ to a Binomial$(7, 1/2)$ distribution; The probability of $A = 0, 1, 2, \ldots$ are therefore $1/128, 7/128, 21/128, \ldots$. Therefore to get a size $1/16$ (=8/128) test reject when $A \leq 1$. For the second dataset, reject with probability $1/2$ when $A = 0$.

*Why is this being asked?* Conditional tests are useful when you have nuisance parameters to deal with; make sure you understand the procedure. In fact the test you end up with here is UMPU, and this is a general result. However, as in Q9, with such a tiny dataset you should expect very, very little power, and so even this ‘optimal’ analysis probably won’t tell you much.
9. A local councillor suspects that traffic conditions have become more hazardous in Ambridge than in Borchester, so she records the numbers $A$ and $B$ of accidents occurring in each place in the course of a month. Assuming that $A$ and $B$ are independent Poisson random variables with parameters $\lambda$ and $\mu$, which have independent exponential(1) prior distributions, Compute the posterior probability that $\lambda \geq \mu$ if $A = 2, B = 5$, and again when $A = 0, B = 3$.

**Solution**

The posterior for $\lambda$ is

$$f(\lambda|A) \propto \lambda^A e^{-\lambda}$$

and so we see that it must have a $\Gamma(A + 1, 2)$ distribution. The posterior for $\mu$ follows with a similar dependence on $B$. Integrate these (independent) posteriors over the range $0 \leq \mu \leq \lambda \leq \infty$ to get the required probability; $37/256$ in the first instance and $1/16$ in the second.

*Why is this being asked?* This Bayesian approach to Q8 doesn’t require the (sophisticated) argument used in the frequentist approach; however you are required to give prior distributions to $\lambda$ and $\mu$, on which the final analysis depends critically.

Keen people may be interested in the WinBUGS code for this question:

```r
a~dpois(lambda)
b~dpois(mu)
lambda~dexp(1)
mu~dexp(1)
I.lambda.bigger <- step(lambda-mu)
```

# data
list(a=2,b=5)

This is all you need to set up Gibbs Sampling (Monte Carlo Markov Chain) to get an approximation to the integrals required for this analysis; note how it’s reasonably straightforward to add extra data, or to change the priors.

See [www.mrc-bsu.cam.ac.uk/bugs/overview/contents.shtml](http://www.mrc-bsu.cam.ac.uk/bugs/overview/contents.shtml) for more details.
10. Let $x_1, x_2, ..., x_n$, with ordered values $x_{(1)} < x_{(2)} < ... < x_{(n)}$, be an independent, identically distributed sample from an unknown distribution $F$. Let $\hat{\theta}(X_1, ..., X_n)$ be a statistic.

Explain what is meant by the ‘bootstrap estimate’ of the standard deviation of $\hat{\theta}$, and describe a Monte Carlo algorithm for its evaluation.

Suppose that $n = 2m - 1$ and that $\hat{\theta}$ is $X_{(m)}$, the middle order statistic. By showing that the bootstrap distribution is concentrated on the observed data points and obtaining an expression for the bootstrap probability of $\hat{\theta}^*$ being equal to $x_{(k)}$, show that in this case the bootstrap estimate of standard deviation can be calculated theoretically, without simulation.

Solution

The bootstrap estimate of the standard deviation is the standard deviation of $\hat{\theta}(X_1^*, ..., X_n^*)$, the same statistic calculated from samples from the empirical distribution $F^*$, which takes value $X_i$ with probability $1/n$. (The bootstrap works because, with enough data, $F^*$ will be a good approximation to $F$, the distribution of interest.)

A Monte Carlo estimate of this quantity can be taken by taking many samples of size $n$ from $F^*$, calculating $\hat{\theta}(X^*)$ for each sample, and then reporting the standard deviation of the various $\hat{\theta}(X^*)$ values.

For the last part, we can calculate the distribution of $\hat{\theta}(X^*)$ exactly, and therefore its standard deviation. For a sample of size $n$ from $F^*$,

$$P(\hat{\theta}(X^*) > x_k) = P(\text{no more than } m-1 \text{ of the } n \text{ observations are } \leq x_{(k)})$$

$$= \sum_{j=0}^{m-1} \binom{n}{j} (k/n)^j (1 - k/n)^{n-j}.$$

From these expressions you can work out the probabilities that $\hat{\theta}(X^*)$ takes any particular value, and therefore its standard deviation.

Why is this being asked? Familiarisation with the bootstrap; the basic idea is simple, as is its practical implementation; see Q11 for why exact answers are more difficult to produce.
11. Given a dataset of \( n \) distinct values, show that the number of distinct bootstrap samples is

\[
\binom{2n-1}{n}
\]

Solution

A combinatorial argument; Think of a sequence of \( n \) balls, laid out in a line interspersed with \( n - 1 \) bars. Assign the quantity of \( X_1 \)'s in the (re)sample \( X^* \) as the number of balls before the first bar. The quantity of \( X_2 \)'s is the number of balls between the first and second bar, and so on for the quantities of all the original data points \( X_i \) in the sample of size \( n \) from \( F^* \).

Each layout of balls and bars corresponds to a particular bootstrap sample, and so there are \( \binom{2n-1}{n} \) of these in total.

Why is this being asked? This number gets very big very quickly; try calculating it for \( n = 100 \). For practical problems it's much easier to use the Monte Carlo algorithm from Q10; in terms of tripos questions some neat solutions are available - as at the end of Q10 - but they tend to involve clever counting arguments.