**STATISTICS** Part IB

**Example Sheet 2 (of 3)** RJS/Lent 2009

*Comments and corrections to r.samworth@statslab.cam.ac.uk*

**1.** Let $X$ have density function

$$f(x; \theta) = \frac{\theta}{(x + \theta)^2}, \quad x > 0,$$

where $\theta \in (0, \infty)$ is an unknown parameter. Find the likelihood ratio test of $H_0 : \theta = 1$ against $H_1 : \theta = 2$ and show that the probability of Type II error is $19/21$.

**2.** Let $f_0$ and $f_1$ be probability mass functions on a countable set $\mathcal{X}$. State and prove a version of the Neyman–Pearson lemma for a size $\alpha$ test of $H_0 : f = f_0$ against $H_1 : f = f_1$ assuming that $\alpha$ is such that there exists a likelihood ratio test of exact size $\alpha$.

**3.** Let $X_1, \ldots X_n, Y_1, \ldots, Y_n$ be independent, with $X_1, \ldots, X_n \sim \mathrm{Exp}(\theta_1)$ and $Y_1, \ldots, Y_n \sim \mathrm{Exp}(\theta_2)$. Find the likelihood ratio test of size $\alpha$ of $H_0 : \theta_1 = \theta_2$ against $H_1 : \theta_1 \neq \theta_2$, expressing it in terms of the statistic

$$T = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i + \sum_{i=1}^n Y_i},$$

and the quantiles of a standard distribution.

**4.** Let $C = \{x : t(x) < k\}$ denote the critical region of a test of a simple null hypothesis based on a test statistic $T = t(X)$. Show that if the null distribution function $F$ of $T$ is continuous then under this hypothesis the $p$-value $P = F(T)$ has a $U(0, 1)$ distribution.

[*Hint: For $u \in (0, 1]$, let $F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$ and show that $x \geq F^{-1}(u)$ iff $F(x) \geq u$.*]

**5.** Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Exp}(\theta)$. Find the likelihood ratio test of size $\alpha$ of $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1 \ (> \theta_0)$ and write down an expression for the power function. Is the test uniformly most powerful for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$?

Find chi-squared distribution tables either by looking in a textbook or, for example, from `http://www.statsoft.com/textbook/sttable.html#chi`, or (better still) by downloading the statistical programming language **R** from `http://cran.r-project.org/` and using one of the manuals on that site to help you write your own tables. Use these tables to find the smallest sample size yielding a test of size 0.05 of $H_0 : \theta \leq 1$ against $H_1 : \theta > 1$ whose power function is at least 0.9 at $\theta = 3$.

**6.** A machine produces plastic articles in bunches of three articles at a time. The process is rather unreliable, and quite a few defective articles are produced. In an experimental

run of the machine, 512 bunches were produced. Of these, the numbers of bunches with $i = 0, 1, 2, 3$ defective articles were 213 $(i = 0)$, 228 $(i = 1)$, 57 $(i = 2)$ and 14 $(i = 3)$. Test the hypothesis that each article has a constant (but unknown) probability $\theta$ of being defective, independently of all other articles.

**7.** A random sample of 59 people from the planet Krypton yielded the results below.

|  |  | Eye-colour | |
|---|---|---|---|
|  |  | 1 (Blue) | 2 (Brown) |
| Sex | 1 (Male) | 19 | 10 |
|  | 2 (Female) | 9 | 21 |

Carry out a chi-squared test at the 5% level that sex and eye-colour are independent factors on Krypton. Now carry out a chi-squared test at the 5% level of the null hypothesis that each of the cell probabilities is equal to $1/4$. Comment on your results.

**8.** Write down the model and hypotheses for a test of homogeneity in a two-way contingency table. Show that the likelihood ratio and Pearson's chi-squared tests are identical to those for the independence test. Perform the homogeneity test on the data below from a clinical trial for a drug, obtained by randomly allocating 150 patients to three equal groups.

|  | Improved | No difference | Worse |
|---|---|---|---|
| Placebo | 18 | 17 | 15 |
| Half dose | 20 | 10 | 20 |
| Full dose | 25 | 13 | 12 |

**9.** If $X \sim N(0, 1)$ and $Y \sim \chi_n^2$ are independent, we say that $T = \frac{X}{\sqrt{Y/n}}$ has a $t$-distribution with $n$ degrees of freedom and write $T \sim t_n$. Derive the probability density function of $T$.

**10.** Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, where $\sigma$ is unknown, and suppose we are interested in testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. Derive the likelihood ratio test of size $\alpha$.

**11.** Statisticians A and B obtain independent samples $X_1, \ldots, X_{10}$ and $Y_1, \ldots, Y_{17}$, both independent and identically distributed from a $N(\mu, \sigma^2)$ distribution with both $\mu$ and $\sigma$ unknown. They estimate $(\mu, \sigma^2)$ by $(\bar{X}, S_{XX}/9)$ and $(\bar{Y}, S_{YY}/16)$ respectively, where, for example, $\bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i$ and $S_{XX} = \sum_{i=1}^{10} (X_i - \bar{X})^2$. Given that $\bar{X} = 5.5$ and $\bar{Y} = 5.8$, which statistician's estimate of $\sigma^2$ is more probable to have exceeded the true value by more than 50%? Find this probability (approximately) in each case.

**12.*** In **2.**, does there exist a version of the Neyman–Pearson lemma when a likelihood ratio test of exact size $\alpha$ does not exist?