UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Wednesday 8 June, 2005 1:30 to 3:30

PAPER 44

BIOSTATISTICS

*Attempt **THREE** questions.*

*There are **FIVE** questions in total.*

The questions carry equal weight.

STATIONERY REQUIREMENTS
Cover sheet
Treasury Tag
Script paper

SPECIAL REQUIREMENTS
None

You may not start to read the questions printed on the subsequent pages until instructed to do so by the Invigilator.

1 Survival Data Analysis

(a) Define the survivor function F, the hazard h and the integrated hazard H. Set out the relationships between F, h and H for a continuous time-to-event variable T.

A time-to-event dataset consists of n pairs of observations x_i, v_i where x_i is the time to the event or time to censoring and $v_i = 1$ if x_i is an observed event, $v_i = 0$ if x_i is a censored observation.

(b) Write down the log-likelihood in terms of the density f and the survivor function F.

(c) Taking the time-to-event distribution to be exponential with rate parameter θ, find $\hat{\theta}$, the maximum likelihood estimator of θ.

(d) Show that the second derivative of the log-likelihood is proportional to the total number of observed events, and comment.

2 Survival Data Analysis

(a) What is meant by *frailty* in the context of survival analysis? The hazard function for the ith individual in group z ($z = 0$ or 1) is given by

$$h_i^{(z)}(t) = U_i^{(z)} e^{\beta z} h_0(t)$$

where $h_0(t)$ is the baseline hazard, β is a constant and $U_i^{(z)}$, the frailty multiplier, is an independent realization of a non-negative random variable with mean 1 and density g.

(b) Show that the population survivor function $\bar{F}^{(z)}(t)$ for the zth group is given by:

$$\bar{F}^{(z)}(t) = \tilde{g}(e^{\beta z} H_0(t))$$

where H_0 is the integral of h_0 and $\tilde{g}$ is the Laplace transform of g.

(c) Suppose the frailty multiplier has a gamma distribution with unit mean and variance σ^2. Show that the ratio r of the population hazards $\bar{h}^{(z)}$, $z = 0, 1$, is given by:

$$r(t) = \frac{\bar{h}^{(1)}(t)}{\bar{h}^{(0}(t)} = e^{\beta} \frac{1 + \sigma^2 H_0(t)}{1 + \sigma^2 e^{\beta} H_0(t)}$$

Comment on the behaviour of $r(t)$ for small and large t and for small and large σ^2.

[*Hint: The Laplace transform of the density of a gamma distributed random variable with unit mean and variance* $1/\psi$ *is given by*

$$\tilde{g}(s) = \left(\frac{\psi}{\psi + s}\right)^{\psi} \quad]$$

3 Survival Data Analysis

Explain how to set up the *empirical likelihood* for the survivor function, indicating how you would account for right, left and interval censoring. Show how the likelihood can usually be simplified before formal maximisation.

Obtain the Kaplan–Meier estimate $\hat{F}(t)$ of the survivor function $F(t)$ by maximising the empirical likelihood when the only censored observations are right censored.

Indicate how the Kaplan–Meier estimate can be obtained under a constraint $\hat{F}(t) = p$ for some t, p such that $0 < t$ and $0 < p < 1$.

4 Case Studies in Medical Statistics

Persistent young offenders may be sentenced to a community service order of 180 days (CSO180) or prison for 30 days (P30). The Sentencing Commission wants to know which option is safer for the young offenders (lower death rate within 360 days of sentencing), safer for society (lower re-conviction rate within 360 days of sentencing).

(a) Briefly explain to the Sentencing Commission the merits of court-based randomisation of eligible persistent young offenders between the two sentencing options.

40% of persistent young offenders breach their CSO180 at 60 days, which means that the order is revoked and the offender is sent to prison for 60 days. Historically, the re-conviction rate within 330 days of release is 60% for persistent young offenders sent to prison for 30 days. Based on the CSO180 cohort studies, the re-conviction rate within 240 days of release is 80% for those who breached their CSO180 at day 60 and were sent to prison for 60 days; but only 40% of those who complete their CSO180 are re-convicted in the next 180 days.

(b) Assuming no deaths, calculate the expected number re-convicted within 360 days among 10,000 persistent young offenders originally sentenced to CSO180.

(c) Show how to calculate the sample size of persistent young offenders to be randomised between P30 and CSO180 if the Sentencing Commission wants to have 80% power for a test of size α to differentiate between 360-day re-conviction rates of 60% and 56%.

The death rate for the persistent young offenders is 1 death per 30,000 days unless the young offender is in prison, when the rate halves, or is within 30 days of release from prison when it quadruples to 4 per 30,000.

(d) Show how to calculate the expected number of deaths within 360 days for 10,000 persistent young offenders sentenced to CS180.

(e) How feasible is it to investigate differential mortality if England and Wales together sentence about 30,000 persistent young offenders per annum?

5 Case Studies in Medical Statistics

Table 1

Patient	Pre-drug PVCs	Post-drug PVCs
1	6	5
2	9	2
3	17	0
4	22	0
5	7	2
6	5	1
7	5	0
8	14	0
9	9	0
10	7	0
11	9	13
12	51	0

Table 1 presents data from 12 patients who were treated for premature ventricular contractions (PVCs). These abnormal heart-related counts were measured for 1-minute before and after treatment with a drug.

Denote the pre- and post-drug PVC counts for patient i as x_i and y_i respectively. Assume that x_i is a Poisson variate with mean λ_i and that y_i is an independent Poisson variate with mean $\beta\lambda_i$.

(a) Derive a conditional likelihood for the estimation of β that will eliminate the parameters $\lambda_i, i = 1, \ldots, 12$. Express the conditional likelihood as a function of $p = \beta/(1+\beta)$.

(b) Assume now that, after treatment, there is a probability θ of a patient being cured and that, for cured patients, $y_i = 0$. For patients not cured, assume y_i is a Poisson variate with mean $\beta\lambda_i$ as before. Take $L(\theta, p)$ as the approximate conditional likelihood function, where

$$L(\theta, p) = \prod_1 [\theta + (1-\theta)(1-p)^{t_i}] \prod_2 (1-\theta) p^{y_i} (1-p)^{t_i - y_i} \,,$$

where $t_i = x_i + y_i, 1 \leqslant i \leqslant 12, \prod_1$ is the product over i for which $y_i = 0$, and $\prod_2$ is the product over i for which $y_i > 0$.

Hence derive equations for $(\hat{\theta}, \hat{p})$, the values of (θ, p) which maximise $L(\theta, p)$.

(c) Based on the observed data, you may assume that $\hat{\theta}$ and $\hat{p}$ are 0.575 and 0.386 respectively with associated 95% asymptotic confidence intervals of (0.30, 0.81) and (0.27, 0.52) respectively. How do you interpret these results?

END OF PAPER