Subject CS2

Corrections to 2023 study material

0 Introduction

This document contains details of any errors and ambiguities that have been brought to our attention in the Subject CS2 study materials for the 2023 exams. We will incorporate these changes into the study material each year. We are always happy to receive feedback from students, particularly details concerning any errors, contradictions or unclear statements in the courses. If you have any such comments on this course please email them to CS2@bpp.com.

You may also find it useful to refer to the Subject CS2 threads on the ActEd Discussion Forum. (You can reach the Forums by clicking on the 'Discussion Forums' button at the top of the ActEd homepage, or by going to **www.acted.co.uk/forums/**.)

This document was last updated on **3 April 2023**.

1 Course Notes

Chapter 6

Page 35 ***(added on 1 December 2022)***

The equations in question 6.7 are incorrect. The question should read:

6.7 Given that $e_{50} = 30$ and $\mu_{50+t} = 0.005$ for $0 \le t \le 1$, calculate the value of e_{51}.

Chapter 13

Page 65 ***(added on 14 November 2022)***

The values given for the autocovariances in the solution to 13.5 should all be multiplied by σ^2. It should read:

The autocovariance function is:

$$\text{cov}\left(X_n, X_{n+k}\right) = \text{cov}\left(e_n - 5e_{n-1} + 6e_{n-2}, e_{n+k} - 5e_{n+k-1} + 6e_{n+k-2}\right)$$

$$= \begin{cases} 62\sigma^2 & k = 0 \\ -35\sigma^2 & |k| = 1 \\ 6\sigma^2 & |k| = 2 \\ 0\sigma^2 & |k| > 2 \end{cases}$$

For example:

$$\begin{aligned} \gamma_0 &= \text{cov}(X_n, X_n) \\ &= \text{cov}(e_n - 5e_{n-1} + 6e_{n-2}, e_n - 5e_{n-1} + 6e_{n-2}) \\ &= \text{cov}(e_n, e_n) + \text{cov}(-5e_{n-1}, -5e_{n-1}) + \text{cov}(6e_{n-2}, 6e_{n-2}) \\ &= \sigma^2 + (-5)^2\sigma^2 + 6^2\sigma^2 = 62\sigma^2 \end{aligned}$$

and similarly for the other values of k.

Chapter 14

Page 28 ***(added on 6 March 2023)***

There is a mistake in the first equation of the first sentence of the second bullet point at the top of the page. It should read:

Let $\hat{\sigma}_d^2$ denote the sample variance of the process $z^{(d)} = \nabla^d x$, *ie* the sample variance of the data values after they have been differenced d times.

Page 31 ***(added on 6 March 2023)***

There is a typo in the second half of the first paragraph on this page. It should read:

In the case of a more general ARMA process, we encounter the difficulty that the e_t cannot be deduced from the z_t. For example, in the case of an $ARMA(1,1)$ model, we have:

$$e_t = z_t - \alpha_1 z_{t-1} - \beta_1 e_{t-1}$$

an equation which can be solved iteratively for e_t as long as some starting value e_0 is assumed. For an $ARMA(p,q)$ model the list of starting values is $(e_0, \ldots, e_{q-1})$.

2 Assignments

Assignment X1 Solutions

Solution X1.3 ***(added on 14 November 2022)***

There is a typo in the return times to State 1 in the first chain. It should read:

The possible return times to State 1 are $\{2,3,4,5,6...\}$.

Solution X1.7, part (iii) ***(added on 1 December 2022)***

There is a typo in the typed solution on page 14. The calculation for the estimate of the variance of $\tilde{p}$ should be:

(p^)^2 / 26 = (26/112)^2 / 26 = 0.0020727

Assignment X4 Questions

Question X4.9, part (iv)(a) ***(added on 14 November 2022)***

There is a typo in the wording of the question. It should refer to the autocovariance function of X_t, not the autocorrelation function:

Show that the autocovariance function of Y_t, γ_k^Y, can be expressed in terms of the autocovariance function of X_t, γ_k, as follows:

Assignment X4 Solutions

Question X4.9, part (iv)(b) ***(added on 6 March 2023)***

There is a typo in the title of the solution. It should refer to the autocorrelation instead of the autocovariance. It should read:

(iv)(b) ***lowest lag such that autocorrelation is 0***

Solution X4.10, part (i) ***(added on 14 November 2022)***

There is a mistake in the limits of x given when deriving the formula for the CDF of X_M near the bottom of the page. It should read:

Using the formula for the CDF of X_M above, for $-n < x < 0$:

$$P\left(\frac{X_M - 500}{\frac{1}{n}500} \le x\right) = \left(\frac{x\frac{1}{n}500 + 500}{500}\right)^n$$

$$= \left(1 + \frac{x}{n}\right)^n$$

Assignment X5 Questions

Question 10, part (i) ***(added on 14 November 2022)***

The number of marks available is missing from this part. It is worth 2 marks.

3 PBOR

Chapter 8 Proportional hazards models – Summary

Page 4 ***(added on 6 March 2023)***

Section 2.4 of the summary should include the below information on adding interaction terms and confidence intervals:

To include an interaction between p of the variables we can use:

```
<var 1>:<var 2>: … :<var p>
```

To include main effects and all possible interactions between the n variables we can use:

```
<var 1>*<var 2>* … *<var n>
```

To output a summary of the fit, we can use:

```
summary(<fit>)
```

Confidence intervals

An approximate $100\alpha\%$ confidence interval for the parameters can be calculated using:

```
confint(<fit>, level = <alpha>)
```

For the exponent of the parameters, we can use:

```
exp(confint(<fit>, level = <alpha>))
```

This is also included in the summary output.

Page 4 ***(added on 6 March 2023)***

In the section about testing significance against the null model, the `logtest` object is extracted from the summary rather than the fit itself. It should read:

Testing significance against the null model

Extracting the p-value from the output:

```
summary(<fit>)$logtest["pvalue"]
```

Chapters 10 and 11 Graduation – Course Notes

Page 6 ***(added on 14 November 2022)***

There is a typo in the calculation of the graduated rates at the top of the page. It should read:

```
Grad$GRAD <- B * c ^ Grad$AGE
```

Page 17 ***(added on 14 November 2022)***

There is a typo in the description of the observed value of the test statistic. A square root is missing from the denominator. It should read:

ie:

$$\frac{\text{total observed deaths} - \text{total expected deaths}}{\sqrt{\text{total expected deaths}}}$$

Page 20 ***(added on 14 November 2022)***

The number of groups of positive deviations is incorrect in a few places. The R output of 9 on page 19 is correct. Near the top of page 20, it should read:

So, there are 9 groups of positive deviations. We now need to check if this is 'too few' to complete our test.

Recall from the Course Notes that we have:

$$P(G=t)=\binom{n_1-1}{t-1}\binom{n_2+1}{t}\bigg/\binom{n_1+n_2}{n_1}$$

Since this is a one-sided test, the *p*-value in our case is therefore:

$$P(G\le 9)=\sum_{t=1}^{9}\binom{n_1-1}{t-1}\binom{n_2+1}{t}\bigg/\binom{n_1+n_2}{n_1}$$

Page 20 ***(added on 14 November 2022)***

There is another instance of the incorrect number of positive groups when introducing the `phyper()` function near the bottom of the page. It should read:

So we can use the `phyper()` *function to calculate* $P(G\le 9)$ *as follows:*

Page 21 ***(added on 14 November 2022)***

There is another instance of the incorrect number of positive groups near the top of this page. When describing the continuity correction, it should read:

This is a one-sided test, where we reject values in the lower tail. As we are approximating a discrete distribution using a continuous distribution, we should apply a continuity correction $P(G \leq 9) \rightarrow P(G \leq 9.5)$. *So the p-value is:*

Chapters 19 and 20 Collective risk model - Solutions

Solution 19-20.3, part (iv), page 11 ***(added on 6 March 2023)***

The chi-squared test performed at the top of this page does not use the correct number of degrees of freedom. As we estimated the exponential parameter using the available data, we need to subtract an additional degree of freedom. This can't be done directly with the `chisq.test()` function and so we need to calculate the observed value of the test statistic and compare it to the critical value directly or calculate a *p*-value. This section should read:

Like the histogram, this looks like a reasonable fit at a glance. However, a chi-squared test is needed to formally reach this conclusion. Calculating the observed value of the test statistic:

```
      sum((obs - exp)^2/exp)
[1] 5.031482
```

We lose 1 degree of freedom (DOF) for making the totals match (ie the number of claims) and 1 for estimating the parameter from the data. So, there are 5 DOF. Calculating the critical value:

```
      qchisq(0.95, 5)
[1] 11.0705
```

So, because the test statistic of 5.0315 is less than the critical value of 11.0705, there is insufficient evidence to reject H_0 at the 5% level. So, it is reasonable to conclude that the waiting times are exponentially distributed.

We could also calculate the p -value directly:

```
      pchisq(sum((obs - exp)^2/exp), 5, lower = FALSE)
[1] 0.4120501
```

As this is greater than 5%, we have insufficient evidence to reject the null hypothesis. This is consistent with the conclusion above.

4 Paper A Handouts

Day 5 Handouts

Page 4 ***(added on 3 April 2023)***

There is a type in the hint in question 4 part (vii). It should read:

$$P(X > x, Y > y, Z > z) = 1 - P(X \le x) - P(Y \le y) - P(Z \le z)$$
$$+ P(X \le x, Y \le y) + P(X \le x, Z \le z) + P(Y \le y, Z \le z)$$
$$- P(X \le x, Y \le y, Z \le z)$$

5 Revision notes

Booklet 9

Page 122 ***(added on 3 April 2023)***

There is a typo in the CDF for the threshold exceedance amount about halfway down the page. It should read:

The excess of a random variable X over the threshold u is $X-u \mid X>u$. Its CDF is:

$$F_{X-u|X>u}(x)=\frac{F_X(x+u)-F_X(u)}{1-F_X(u)}$$