Subject CS1

Corrections to 2021 study material

0 Introduction

This document contains details of any errors and ambiguities that have been brought to our attention in the Subject CS1 study materials for the 2021 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 CS1@bpp.com

You may also find it useful to refer to the Subject CS1 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 **16 September 2021**.

1 Paper A Course Notes

Chapter 3

Page 16 ***(added on 14 April 2021)***

The second line of the Core Reading formula for $C_X'''(0)$ is incorrect. It should read:

$$= \frac{E[X^3](1)^3 - 3(1)^2 E[X] E[X^2] + 2(1)(E[X])^3}{1^4}$$

Chapter 4

Page 53 ***(added on 9 February 2021)***

The solutions to parts (i) and (ii) of Question 4.5 should include the input values for which they are valid. The solution should read:

(i) ***Marginal density***

$$f_X(x) = \int_{y=0}^{1} \frac{4}{5}\left(3x^2 + xy\right) dy = \left[\frac{4}{5}\left(3x^2 y + \frac{1}{2}xy^2\right)\right]_{y=0}^{1} = \frac{4}{5}\left(3x^2 + \frac{1}{2}x\right) \quad 0 < x < 1 \qquad [2]$$

(ii) ***Conditional density***

$$f_{Y|X=x}(x,y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\frac{4}{5}\left(3x^2 + xy\right)}{\frac{4}{5}\left(3x^2 + \frac{1}{2}x\right)} = \frac{3x^2 + xy}{3x^2 + \frac{1}{2}x} = \frac{3x + y}{3x + \frac{1}{2}} \quad 0 < y < 1 \qquad [1]$$

Similarly, when obtaining the marginal distribution for Y in part (iii) the solution include the input values for which it is valid. It should read:

Obtaining the marginal density function of Y:

$$f_Y(y) = \int_{x=0}^{1} \frac{4}{5}\left(3x^2 + xy\right) dx = \frac{4}{5}\left[x^3 + \frac{1}{2}x^2 y\right]_{x=0}^{1} = \frac{4}{5}\left(1 + \frac{1}{2}y\right) \quad 0 < y < 1$$

Chapter 6

Page 24 ***(added 8 December 2020)***

The solution to Question 6.9, the mean and variance of the approximate normal distribution (200 and 300) are the wrong way around in the standardisations (although the resulting numbers are correct). It should read:

$$
\begin{aligned}
P(279.5 < S < 320.5) &= P(S < 320.5) - P(S < 279.5) \\
&\approx P\left(Z < \tfrac{320.5-300}{\sqrt{200}}\right) - P\left(Z < \tfrac{279.5-300}{\sqrt{200}}\right) \\
&= P(Z < 1.44957) - P(Z < -1.44957) \\
&= P(Z < 1.44957) - [1 - P(Z < 1.44957)] \\
&= 2P(Z < 1.44957) - 1 \\
&= 2 \times 0.92641 - 1 \\
&= 0.85282
\end{aligned}
$$

Chapter 11

Page 22 ***(added 9 April 2021)***

The final sentence on this page is missing the word 'evidence'. It should read:

So we have very strong evidence to reject H_0, and we conclude that the mock scores in CS1 and CS2 are positively correlated.

Chapter 12

Page 20 ***(added 9 April 2021)***

The value of $\hat{\alpha}$ stated at the end of the question is negative when it should be positive. The sentence should say:

Recall that $\hat{\alpha} = 0.164, \hat{\beta} = 0.88231, \hat{\sigma}^2 = 0.0732$ and $S_{xx} = 8.444$.

Chapter 13

Page 42 ***(added 9 April 2021)***

The following paragraph should be included at the end of section 5.4:

The residual deviance outputted by the `glm()` function is a measure of fit, similar to the scaled deviance and deviance defined earlier. However, this output won't necessarily match the scaled deviance or deviance calculated from first principles using the formulae in this section.

Page 42 ***(added 9 April 2021)***

The R code in the box at the bottom of the page uses the wrong type of quotes. The box should read:

The code for comparing two normally distributed models, `model1` **and** `model2`**, in R is:**

```
anova(model1, model2, test = "F")
```

Page 47 ***(added 16 September 2021)***

The first line of R code should be updating `modelA` instead of `model1`. It should read:

We remove the interaction term `wt:disp`**, as this is the least significant.**

```
modelB <- update(modelA, ~.-wt:disp)
```

Page 50 ***(added 9 April 2021)***

There should be an additional paragraph in the R code box. It should read as follows:

The R code for obtaining the Pearson residuals is:

```
residuals(model, type = "pearson")
```

The Pearson residuals returned by R are calculated slightly differently to the definition given in this section. Therefore, this output won't necessarily match the Pearson residuals calculated from first principles using $\frac{y-\hat{\mu}}{\sqrt{\text{var}(\hat{\mu})}}$.

Page 51 ***(added 9 April 2021)***

There should be an additional paragraph in the R code box. It should read as follows:

The R code for obtaining the deviance residuals is:

```
residuals(model)
```

The deviance residuals returned by R are calculated slightly differently to the definition given in this section. Therefore, this output won't necessarily match the deviance residuals calculated from first principles using the formulae in this section.

Page 57 ***(added 9 April 2021)***

The summary paragraph for backward selection should say that covariates are removed until the AIC reaches a minimum, not until it reaches a maximum. It should read:

(2) Backward selection. Start by adding all available covariates and interactions. Then remove covariates one by one starting with the least significant until the AIC reaches a minimum or there is no significant improvement in the deviance, and all the remaining covariates have a statistically significant impact on the response.

Chapter 15

Page 20 ***(added 4 January 2021)***

The Core Reading R code to obtain the Monte Carlo credibility premiums has two minor errors. It should read:

```
Z <- n/(n+sigma1^2/sigma2^2)
cp <- rep(0,M)
for (i in 1:M)
{theta <- rnorm(1,mu,sigma2)
x <- rnorm(n,theta,sigma1)
cp[i] <- Z*mean(x)+(1-Z)*mu
}
```


2 Assignments

Assignment X1 solutions

Question 13 ***(added on 9 February 2021)***

The second line of working in the solution for part (iii)(a) is incorrect. The full solution should read:

Since $P(X=1)=\frac{3}{16}$:

$$E[X+Y \mid X=1]=\sum_x \sum_y (x+y)P(X=x,Y=y \mid X=1)$$

$$=\sum_y (1+y)\frac{P(X=1,Y=y)}{P(X=1)}$$

$$=2\times\frac{1/8}{3/16}+3\times\frac{1/16}{3/16} \quad [1]$$

$$=2\times\tfrac{2}{3}+3\times\tfrac{1}{3}=2\tfrac{1}{3} \quad [1]$$

Alternatively, using linearity of expectation:

$$E[X+Y \mid X=1]=E[X \mid X=1]+E[Y \mid X=1]$$

$$=1+E[Y \mid X=1]$$

$$=1+\sum_y y\,P(Y=y \mid X=1)$$

$$=1+\sum_y y\frac{P(X=1,Y=y)}{P(X=1)}$$

$$=1+\left[1\times\tfrac{2}{3}+2\times\tfrac{1}{3}\right]$$

$$=2\tfrac{1}{3}$$

3 Paper B Handouts

Questions

Question 2 ***(added on 9 April 2021)***

The sample size mentioned in part (a) is incorrectly given as 10 instead of 15. It should say:

(a) Simulate 1,000 Bayesian estimates for λ under squared error loss based on samples of 15 values and using a seed of 33.

4 Mock Exam Paper A

Solutions

Solution 12, page 21 ***(added on 16 September 2021)***

The interpolation in the solution to part (ii)(b) is missing the first term. It should be:

$$79.33+\frac{82-80}{90-80}\left(89.33-79.33\right)=81.33$$

5 Mock Exam 2 Paper A

Solutions

Solution 12, page 28 ***(added on 12 April 2021)***

The solution to part (v)(b) gives the wrong model. It should say:

Model 1 has the smallest AIC and so it would be selected using this criterion.

6 Revision Notes

Booklet 4

Solution 5, page 94 ***(added on 16 September 2021)***

The solution to part (ii)(a) shows the incorrect value for the sample variance in the calculation of the observed value of the test statistic. However, the calculated value is correct. It should be:

The observed value of the test statistic is:

$$\frac{19.867-18}{\sqrt{19.432/15}}=1.6403$$

Booklet 5

Solution 1, page 95 ***(added on 16 September 2021)***

The solution to part (i)(b) is missing a square in the numerator for the calculation of R^2.

However, the calculated value is correct. The solution should read:

Hence:

$$R^2=\frac{s_{xy}^2}{s_{xx}s_{yy}}=\frac{84.64^2}{326.93\times 63.536}=0.3449$$

Booklet 6

Solution 3, page 45 ***(added on 16 September 2021)***

In the last line of the solution to part (ii) the second derivative of the CGF should be evaluated at $t=0$, *ie* it should read:

$$\text{var}(Y)=C_Y''(0)=\varphi b''(\theta)$$

Solution 9, page 59 ***(added on 16 September 2021)***

One of the calculation lines is repeated when putting the PDF in exponential family form in part (ii). The penultimate expression is also missing a y. It should read:

The PDF here is:

$$
\begin{aligned}
p(y|\alpha) &= \frac{\alpha^{y-1}}{(1+\alpha)^{y}} \\
&= \exp\left\{(y-1)\ln\alpha - y\ln(1+\alpha)\right\} \\
&= \exp\left\{y\left[\ln\alpha - \ln(1+\alpha)\right] - \ln\alpha\right\} \\
&= \exp\left\{y\ln\left(\frac{\alpha}{1+\alpha}\right) - \ln\alpha\right\} \\
&= \exp\left\{\frac{y\ln\left(\frac{\alpha}{1+\alpha}\right) - \ln\alpha}{1} + 0\right\}
\end{aligned}
$$

Booklet 7

Solution 1, page 86 ***(added on 16 September 2021)***

There are two typos in the fraction in the penultimate calculation. Both occurrences of 0.13829 should be 0.13824. However, the calculated probability is correct. The calculation should be:

$$P(p = 0.4 \mid X = 4) = \frac{0.13824 \times 0.6}{0.13824 \times 0.6 + 0.29663 \times 0.4} = 0.41144$$