THE UNIVERSITY OF MELBOURNE

Semester 2 Assessment, 2015

Department of Mathematics and Statistics

MATH30027 Modern Applied Statistics

Writing time: 3 hours

Reading time: 15 minutes

This is NOT an open book exam

This paper consists of 10 pages (including this page)

Authorised materials:

- A single two-sided hand-written A4 sheet of notes.
- Scientific calculators. Graphical calculators are *not* allowed.

Instructions to Students

- You may remove this question paper at the conclusion of the examination
- You should attempt all questions. Marks for individual questions are shown.

Instructions to Invigilators

- Students may remove this question paper at the conclusion of the examination

This paper may be held in the Baillieu Library

Question 1 (11 marks) Suppose that $X_1, \ldots, X_n$ are i.i.d. $\Gamma(m, \lambda)$ r.v.s. That is, each X_i has density

$$f(x) = \lambda^m x^{m-1} e^{-\lambda x} / \Gamma(m).$$

(a) Write down the log-likelihood for the sample, and hence show that the maximum likelihood estimator (MLE) for λ is the same as the method of moments estimator.

Note: your estimate will depend on m.

(b) Substitute your estimate for λ back into the log-likelihood, and hence obtain an equation for the MLE of m that does not involve λ.

Hint: your equation should include the digamma function $\Psi(m) = \Gamma'(m)/\Gamma(m)$.

(c) Use Newton's method to obtain an iterative algorithm for estimating m.

What condition would you use to stop the algorithm?

(d) How (and why) would your algorithm change if you used Fisher scoring instead of Newton's method?

Question 2 (14 marks)

(a) What is a generalised linear model (glm)? Give a precise mathematical definition.

(b) What methodology is used to fit a glm? Just the name will suffice.

(c) What is Pearson's χ^2? Give a precise mathematical definition.

How is it distributed, and what is it used for?

(d) What is the deviance? Give a precise mathematical definition.

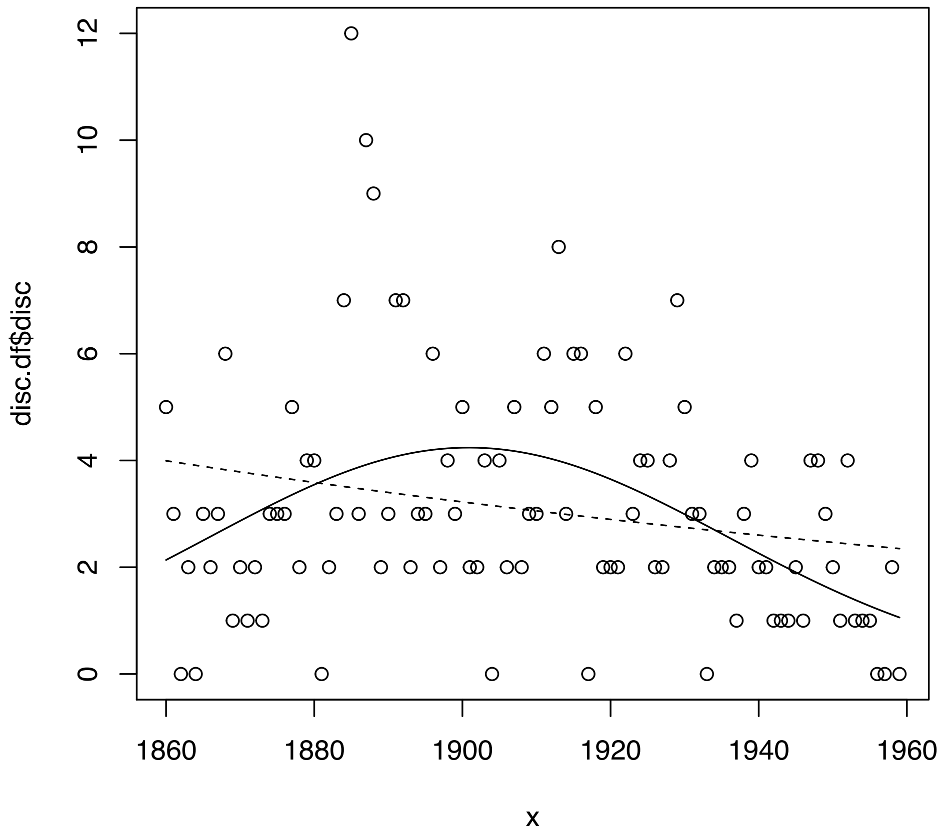
Explain how the deviance can be used to test for model adequacy and to compare nested models.

(e) What is overdispersion? How does one detect it and what effect does it have on a glm?

Question 3 (19 marks) The dataset `discoveries` lists the number of great scientific discoveries for the years 1860 to 1959, as chosen by "The World Almanac and Book of Facts", 1975 Edition. Using the following R output, answer the questions below.

```
> data(discoveries)
> disc.df <- data.frame(year=1860:1959, disc=discoveries)
> model1 <- glm(disc ~ year, family=poisson(link = "log"), disc.df)
> summary(model1)

Call:
glm(formula = disc ~ year, family = poisson(link = "log"), data = disc.df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8112  -0.9482  -0.3533   0.6637   3.5504

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) 11.354807   3.775677   3.007  0.00264 **
```

```
year        -0.005360   0.001982  -2.705  0.00683 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 164.68  on 99  degrees of freedom
Residual deviance: 157.32  on 98  degrees of freedom
AIC: 430.32

Number of Fisher Scoring iterations: 5

> model2 <- glm(disc ~ year + I(year^2), family=poisson(link = "log"), disc.df)
> summary(model2)

Call:
glm(formula = disc ~ year + I(year^2), family = poisson(link = "log"),
    data = disc.df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9066  -0.8397  -0.2544   0.4776   3.3303

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.482e+03  3.163e+02  -4.685 2.79e-06 ***
year         1.561e+00  3.318e-01   4.705 2.54e-06 ***
I(year^2)   -4.106e-04  8.699e-05  -4.720 2.35e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 164.68  on 99  degrees of freedom
Residual deviance: 132.84  on 97  degrees of freedom
AIC: 407.85

Number of Fisher Scoring iterations: 5

> x <- disc.df$year
> plot(x, disc.df$disc)
> beta1 <- model1$coefficients
> lines(x, exp(beta1[1] + beta1[2]*x), lty=2)
> beta2 <- model2$coefficients
> lines(x, exp(beta2[1] + beta2[2]*x + beta2[3]*x^2))
```



(a) Write down the model fitted as `model1`, including the fitted parameter values and any distributional assumptions.

(b) Using `model1`, what is the fitted rate of discoveries for the year 1900?

(c) For `model2`, what is the value of the maximised log-likelihood?

(d) For `model1`, what is the diagonal of the inverse of the Fisher information matrix (evaluated at the MLE)?

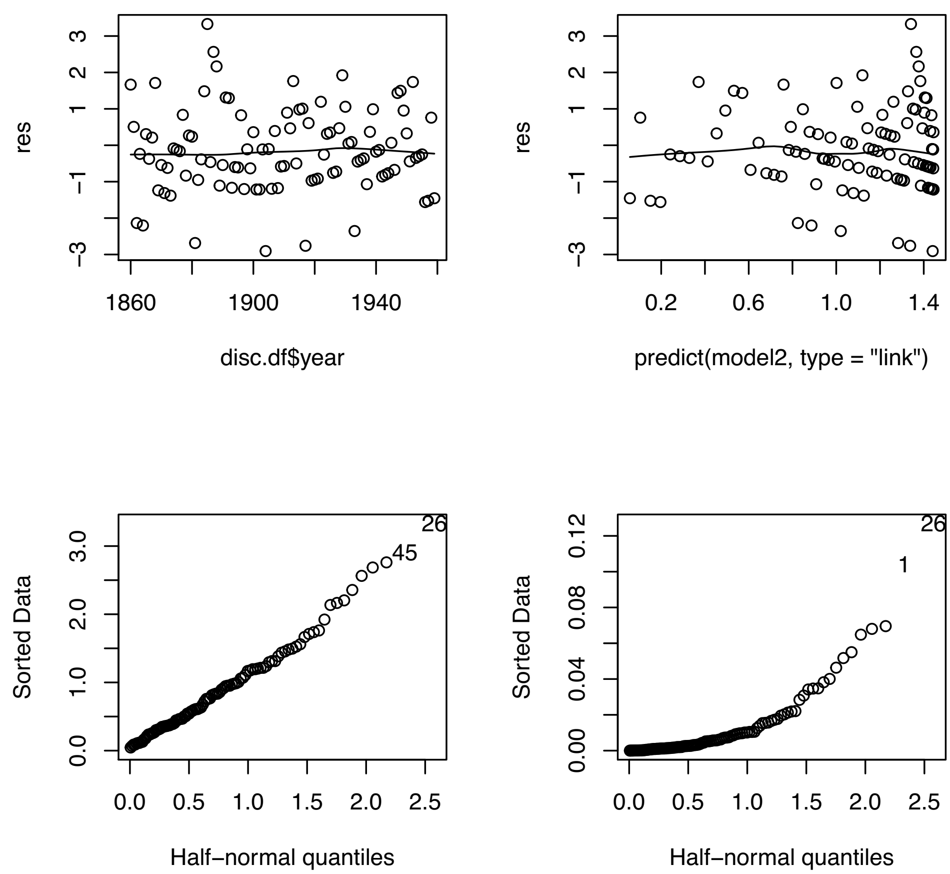
(e) Is `year^2` significant when added to `model1`? Test significance at the 95% and 99% levels. The following percentage points will help

```
> qchisq(.95, c(1,2,3,97,98,99,100))

[1]   3.841459   5.991465   7.814728 120.989644 122.107735 123.225221 124.342113

> qchisq(.99, c(1,2,3,97,98,99,100))

[1]   6.634897   9.210340  11.344867 132.308877 133.475672 134.641617 135.806723
```

(f) Is `model2` adequate? Support your answer with a suitable hypothesis test, if you can.

Here are some diagnostic plots for `model2`. Further questions follow. Note that the function `lowess` produces a local smoothing of the data it is applied to.

```
> library(faraway)
> res <- residuals(model2, type="deviance")
> par(mfrow=c(2,2))
> plot(disc.df$year, res)
> lines(lowess(disc.df$year, res))
> plot(predict(model2, type="link"), res)
> lines(lowess(predict(model2, type="link"), res))
> halfnorm(res)
> halfnorm(cooks.distance(model2))
```



(g) Give a definition of the deviance residuals for this model.

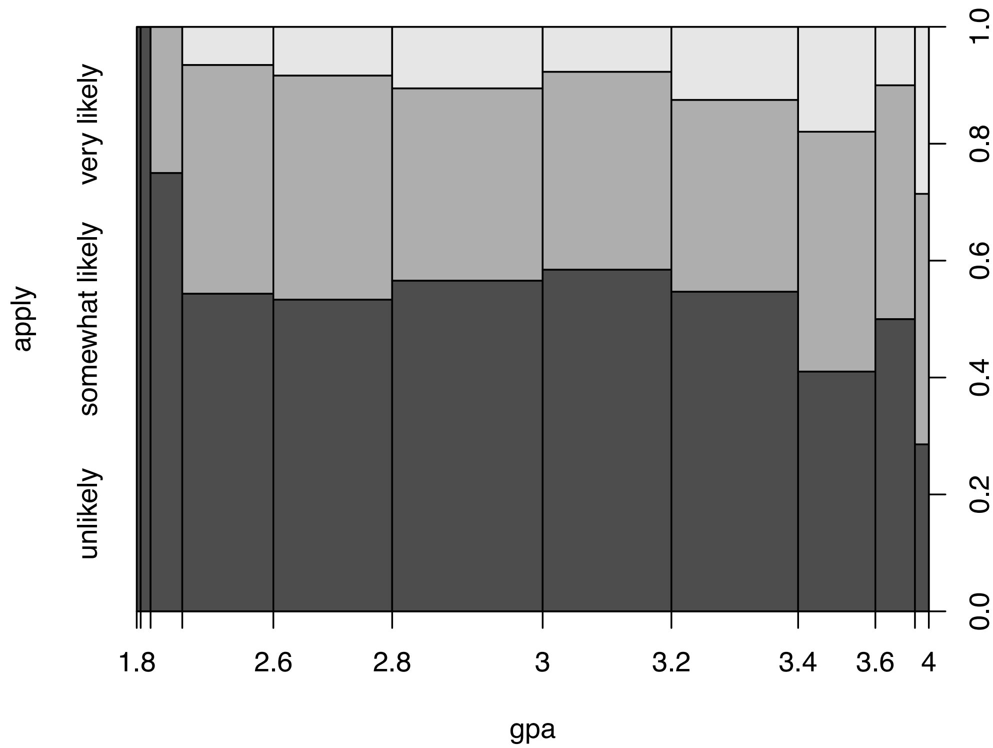
(h) Why do we see bands in the top two residual plots? Do they indicate a problem with the model?

(i) Is there any evidence of a trend in the residuals?

(j) Are there any potential outliers? Justify your answer.

Question 4 (9 marks) College juniors (undergraduates) at US universities were asked if they were "unlikely", "somewhat likely", or "very likely" to apply to graduate school. Data on whether the parents have graduate education (`pared`), whether the undergraduate institution is public (`public`), and current GPA (`gpa`), were also collected.

```
> load("ologit.Rdata")
> head(dat)

            apply pared public  gpa
1     very likely     0      0 3.26
2 somewhat likely     1      0 3.21
3        unlikely     1      1 3.94
4 somewhat likely     0      0 2.81
5 somewhat likely     0      0 2.53
6        unlikely     0      1 2.59

> ftable(xtabs(~ public + apply + pared, data = dat))

                       pared   0   1
public apply                        
0      unlikely              175  14
       somewhat likely        98  26
       very likely            20  10
1      unlikely               25   6
       somewhat likely        12   4
       very likely             7   3

> plot(apply ~ gpa, data = dat)
```



Using the R output below, answer the questions that follow.

```
> library(MASS)
> mod <- polr(apply ~ pared + public + gpa, data = dat, method="logistic")
> summary(mod)

Call:
polr(formula = apply ~ pared + public + gpa, data = dat, method = "logistic")

Coefficients:
          Value Std. Error t value
pared   1.04769     0.2658  3.9418
public -0.05879     0.2979 -0.1974
gpa     0.61594     0.2606  2.3632

Intercepts:
                            Value   Std. Error t value
unlikely|somewhat likely     2.2039  0.7795     2.8272
somewhat likely|very likely  4.2994  0.8043     5.3453

Residual Deviance: 717.0249
AIC: 727.0249
```

(a) Specify the model that has been fitted, including the fitted parameter values.

Hint: the model assumes that the data is ordinal.

(b) What is the fitted probability that a student whose parents have graduate degrees and who went to a public university, with a gpa of 3.0, is very likely to undertake graduate study?

(c) What is the odds ratio of being unlikely to apply to graduate school for a student at a private college compared to a student at a public college?

(d) Let $\alpha(\mathbf{x})$ be the probability that a student with predictor variable $\mathbf{x}$ is unlikely to apply to graduate school, and let $\beta(\mathbf{x})$ be the probability that a student with predictor variable $\mathbf{x}$ is either unlikely or somewhat likely to apply to graduate school.

Show that for the type of model fitted here, the odds ratio of $\alpha(\mathbf{x}_1)$ and $\alpha(\mathbf{x}_2)$ must be the same as the odds ratio of $\beta(\mathbf{x}_1)$ and $\beta(\mathbf{x}_2)$. We say the model is a "proportional odds" model.

Question 5 (6 marks) A density belongs to an exponential family if it can be written as

$$f(x;\theta,\phi) = \exp\left(\frac{x\theta - b(\theta)}{a(\phi)} + c(x,\phi)\right)$$

(a) Show that if $a(\phi) = 1$ then f can be written as

$$\alpha(x)\exp\left(x\theta + \beta(\theta)\right) \tag{1}$$

(b) Name four distributions that can be put into the form (1)

(c) Show that if X has a density of the form (1) and θ has the prior density

$$p(\theta;\gamma,\delta) \propto \exp\left(\gamma\theta + \delta\beta(\theta)\right)$$

then X and θ are a conjugate pair, and give the parameters of the posterior.

Question 6 (10 marks)

The random variable Y has density

$$h(x) = \frac{2\pi}{3\sqrt{3}}\cos\left(\frac{\pi}{3}x\right)1_{[0,1]}(x).$$

Here 1_A is the indicator function for the set A.

(a) Show how to simulate from h using the inversion method.

You may assume that you have access to an unlimited number of $U(0,1)$ random variables.

The random variable X has a truncated normal density

$$f(x) = ce^{-x^2/2}1_{[0,1]}(x).$$

(b) Give an expression for c. You do not have to solve the expression.

(c) Show that the following algorithm (which does not use c) will simulate from f

1 Simulate X from h

2 Simulate $Y \sim U(0, h(X)\frac{3\sqrt{3}e^{-1/2}}{\pi})$

3 If $Y < e^{-X^2/2}$ then return X, otherwise go back to 1.

You may use the fact that

$$\sup_{0\le x\le 1} \frac{e^{-x^2/2}}{\cos(\pi x/3)} \le 2e^{-1/2}$$

(d) Show that the probability of returning X on any given loop is

$$\frac{\pi e^{1/2}}{3\sqrt{3}c}.$$

Question 7 (6 marks) Consider the data from Question 3.

Give a Bayesian model for these data. Your description should include a directed acyclic graph (DAG) showing the relationships between the stochastic and logical nodes, distributions for all the stochastic nodes, and descriptions of all the constants.

Use appropriate vague priors for all the parameters.

Question 8 (20 marks) We observe independent samples $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$ from two normal populations with the same variance.

Here is a Bayesian model for these data.

$$\begin{aligned} X_i &\sim N(\mu_1, \sigma^2), \quad i = 1, \ldots, n_1 \\ Y_i &\sim N(\mu_2, \sigma^2), \quad i = 1, \ldots, n_2 \\ \mu_1 &\sim N(0, 10^4) \\ \mu_2 &\sim N(0, 10^4) \\ \tau = 1/\sigma^2 &\sim \Gamma(0.001, 0.001) \end{aligned}$$

(a) Are the priors for μ_1, μ_2 and τ vague or informative?

(b) Write down the kernel of the joint likelihood for $X_1, \ldots, X_{n_1}$, $Y_1, \ldots, Y_{n_2}$, μ_1, μ_2 and τ.

That is, you only need give the likelihood up to a constant.

(c) Write down the kernels of

$$\begin{gathered} \mu_1|\mathbf{X}, \mathbf{Y}, \mu_2, \tau \\ \mu_2|\mathbf{X}, \mathbf{Y}, \mu_1, \tau \\ \tau|\mathbf{X}, \mathbf{Y}, \mu_1, \mu_2 \end{gathered}$$

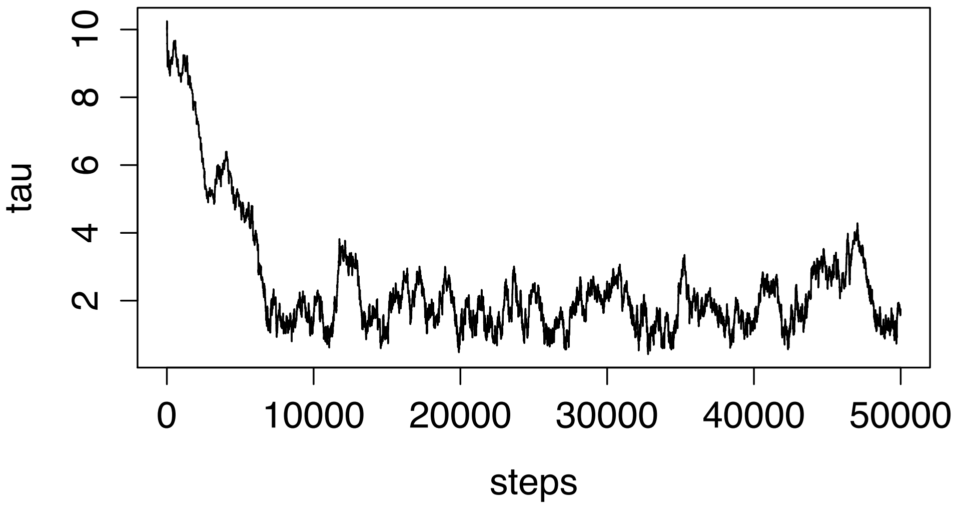
Hence give the (conditional) distributions of these variables, including their parameters.

(d) Give a Gibbs sampling scheme for sampling (μ_1, μ_2, τ).

(e) If you were allowed to use improper (vague) priors for μ_1, μ_2 and τ, what would they be?

How would the Gibbs sampling scheme change if you used improper priors? Would it still produce sensible output (and why)?

(f) Below is a plot of a sample of τ values from a Gibbs sampler for this model. How many samples should we discard as a burn-in?



(g) I wish to use the mean of the sample above (after the burn-in is discarded) as a point estimate for τ. How can I decide if my sample is large enough that the sample mean is a good estimate of the posterior mean?

End of Exam—Total Available Marks = 95