

# MATH401/501 spec Exam - solutions

Chris Sherlock

## 1

```
## Quick confirmation that the density integrates to 1
ymax=10; ny=10^6; ymin=0.5/ny;
ys=seq(ymin,ymax,len=ny)
theta=2
fs=ys^5/theta^6*exp(-ys^2/theta^2)
dy=(ymax-ymin)/(ny-1)
dy*sum(fs)-0.5*(fs[1]+fs[ny])*dy ## Trapezium rule
```

```
## [1] 1
```

**Note:** If  $X \sim \text{Gam}(3, \lambda)$  then  $Y = X^{1/2}$  and  $\theta = 1/\lambda^{1/2}$ .

```
## Generate data
set.seed(12379)
thetastar=2; lambda=1/thetastar^2
n=100
xs=rgamma(n,shape=3,rate=lambda); ys=sqrt(xs)
S=round(sum(ys^2),1)
round(c(sum(ys),S,sum(exp(ys))),1)
```

```
## [1] 341.2 1248.0 4533.7
```

**(a)**

$$L(\theta; \mathbf{y}) \propto \theta^{-6n} \exp \left\{ -\theta^{-2} \sum_{i=1}^n y_i^2 \right\}$$

$$\ell(\theta; \mathbf{y}) = -6n \log \theta - \theta^{-2} \sum_{i=1}^n y_i^2 + \text{cst}$$

(4 marks)

$$u(\theta; \mathbf{y}) = \frac{-6n}{\theta} + \frac{2}{\theta^3} S,$$

where  $S = \sum_{i=1}^n y_i^2$ . (2 marks)

So,  $\hat{\theta} = (S/(3n))^{1/2}$ . (2 marks).

**(b)**

**(i)**

$$I_O(\theta; \mathbf{y}) = -\frac{6n}{\theta^2} + \frac{6}{\theta^4} S.$$

(4 marks)

**(ii)**

$\mathbb{E} \left[ \sum_{i=1}^n Y_i^2 \right] = n\theta^2 \times 3!/2 = 3n\theta^2$  (2 marks) so

$$I_E(\theta) = -\frac{6n}{\theta^2} + \frac{6}{\theta^4} \times 3n\theta^2 = \frac{12n}{\theta^2},$$

as required (2 marks).

**(c)**

**(i)**

```
thetahat=(S/(3*n))^0.5
IE=12*n/thetahat^2
SE=1/sqrt(IE)
c(thetahat,SE) ## for partial marks
```

```
## [1] 2.03960781 0.05887841
```

```
CI=thetahat+qnorm(0.975)*c(-1,1)*SE
CI
```

```
## [1] 1.924208 2.155007
```

(4 marks)

**(ii)**

$E[Y^2] = 0.5\theta^2\Gamma(4) = 3\theta^2$ , which is *monotonic* in  $\theta$  since  $\theta$  is positive, so the CI is

```
3*CI^2
```

```
## [1] 11.10773 13.93217
```

(4 marks)

**(d)**

**(i)**

```
theta0=2.2
siglevel=0.01
TWald=(thetahat-theta0)/SE
Tcrit=qnorm(1-siglevel/2)
pWald=2*pnorm(-abs(TWald)) ## only possible if online
c(TWald,Tcrit,pWald)
```

```
## [1] -2.724125975 2.575829304 0.006447191
```

Reject  $H_0$  as  $|T_{Wald}| \geq T_{crit}$ . (4 marks)

(ii)

```
u=-6*n/theta0+2*S/(theta0^3)
IE0=12*n/theta0^2
Tscore=u^2/IE0
Tcrit=qchisq(1-siglevel,df=1)
pscore=pchisq(Tscore,df=1,lower.tail=FALSE) ## only possible if online
c(Tscore,Tcrit,pscore)
```

```
## [1] 5.9217267 6.6348966 0.0149553
```

Fail to reject  $H_0$  as  $T_{score} < T_{crit}$ . (4 marks)

(iii)

```
u=-6*n/theta0+2*S/(theta0^3)
Tscore=u^2/IE
Tcrit=qchisq(1-siglevel,df=1)
pscore=pchisq(Tscore,df=1,lower.tail=FALSE) ## only possible if online
c(Tscore,Tcrit,pscore)
```

```
## [1] 5.08974853 6.63489660 0.02406768
```

Fail to reject  $H_0$  (4 marks)

(e)

The score test using  $I_E(\theta_0)$  does not require calculation of the MLE. (2 marks)

## 2

**Note:**  $Y \sim \text{Gamma}(\alpha, \theta)$  and  $Z \sim \text{Weibull}(2, \theta)$ .

```
## Simulate some data
set.seed(123579)
n=150; m=50; alpha=3; theta=0.5
ys=rgamma(n,alpha,rate=theta);
zs=sqrt(rexp(m,rate=theta^2))
```

```
S0=round(sum(log(ys)),2); S1=round(sum(ys),2); S2=round(sum(zs^2),2)
c(S0,S1,S2)
```

```
## [1] 240.28 925.64 190.05
```

```
Q2log.like<-function(pars,ys,zs) {
  n=length(ys); m=length(zs); alpha=pars[1]; theta=pars[2]
  if ((alpha>0) && (theta>0)) {
    S0=sum(log(ys)); S1=sum(ys); S2=sum(zs^2)
    l=(n*alpha+2*m)*log(theta)-n*lgamma(alpha)+alpha*S0-theta*S1-theta^2*S2
  }
  else {
    l=-Inf
  }
  return(l)
}
```

**(a)**

$$L(\alpha, \theta; \mathbf{y}) \propto \frac{\theta^{n\alpha}}{\Gamma(\alpha)^n} \left( \prod_{i=1}^n y_i \right)^\alpha \exp \left\{ -\theta \sum_{i=1}^n y_i \right\} \times \theta^{2m} \exp \left\{ -\theta^2 \sum_{i=1}^m z_i^2 \right\}.$$

$$\Rightarrow \ell(\alpha, \theta; \mathbf{y}, \mathbf{z}) = -n \log \Gamma(\alpha) + (n\alpha + 2m) \log \theta + \alpha \sum_{i=1}^n \log y_i - \theta \sum_{i=1}^n y_i - \theta^2 \sum_{i=1}^m z_i^2 + \text{cst.}$$

as required (2 marks)

$$\frac{\partial \ell}{\partial \theta} = \frac{n\alpha + 2m}{\theta} - \sum_{i=1}^n y_i - 2\theta \sum_{i=1}^m z_i^2.$$

So

$$2\hat{\theta}_\alpha^2 S_z + \hat{\theta}_\alpha S_y - (n\alpha + 2m) = 0,$$

where  $S_z = \sum_{i=1}^m z_i^2$  and  $S_y = \sum_{i=1}^n y_i$ . Thus

$$\hat{\theta}_\alpha = \frac{-S_y + \sqrt{S_y^2 + 8S_z(n\alpha + 2m)}}{4S_z},$$

where we have taken the positive root because  $\theta$  is positive. (6 marks)

**(b)**

```
ctl=list(fnscale=-1)
a=optim(c(1,1),Q2log.like,ys=ys,zs=zs,hessian=TRUE,control=ctl); a

## $par
## [1] 2.8447060 0.4760389
##
## $value
## [1] -274.3389
##
## $counts
## function gradient
##      73      NA
##
## $convergence
## [1] 0
##
## $message
## NULL
##
## $hessian
##      [,1]      [,2]
## [1,] -63.05874  315.1008
## [2,] 315.10076 -2704.3820
```

**(i)**

$$\frac{\partial^2 \ell}{\partial \theta^2} = -\frac{\alpha n + 2m}{\theta^2} - 2 \sum_{i=1}^m z_i^2.$$

Substituting  $\hat{\alpha}$  for  $\alpha$ ,  $\hat{\theta}$  for  $\theta$  etc, we obtain:

```
alphahat=a$par[1]; thetahat=a$par[2]
-(n*alphahat+2*m)/(thetahat^2)-2*S2
```

```
## [1] -2704.352
```

(4 marks)

(ii)

$\text{Var}(\hat{\alpha})$  is the (1,1) entry of  $I_O^{-1}$ :

```
I0=-a$hessian
V=solve(I0)[1,1]; SE=sqrt(V) ## in-person exam: use (2x2) mx inverse formula
c(V,SE)
```

```
## [1] 0.03795814 0.19482848
```

```
CI=alphahat+c(-1,1)*qnorm(0.995)*SE
CI
```

```
## [1] 2.342861 3.346551
```

(6 marks)

(iii)

```
coeffs=c(1,-4)
MLElc=sum(a$par*coeffs)
Vlc=(t(coeffs) %*% solve(I0) %*% coeffs)[1,1] ## turn from 1x1 mx to scalar
SElc=sqrt(Vlc)
c(MLElc,SElc)
```

```
## [1] 0.9405504 0.1293750
```

```
CIlc=MLElc+c(-1,1)*qnorm(0.995)*SElc
CIlc
```

```
## [1] 0.6073024 1.2737984
```

The MLE is  $\hat{\alpha} - 4\hat{\theta} = 0.9405504$ . (4 marks) The variance of the linear combination is  $V_{1,1} + 2 \times 1 \times (-4)V_{1,2} + 16V_{2,2} = 0.0167379$ . (4 marks)

Which gives a CI of 0.6073024, 1.2737984. (2 marks)

(c)

```
alpha0=3.3
thetahatalpha0=sqrt(S1^2+8*S2*(alpha0*n+2*m))-S1
thetahatalpha0=thetahatalpha0/(4*S2)
```

From (a)(ii)

$$\hat{\theta}_\alpha = \frac{-S_y + \sqrt{S_y^2 + 8S_z(\alpha n + 2m)}}{4S_z} = 0.5282233.$$

(4 marks)

So we need  $\ell(3.3, 0.5282233)$ :

```
l0=q2log.likelihood(c(alpha0,thetahatalpha0),ys,zs)
TWilks=2*(a$val-l0)
Tcrit=qchisq(0.975,df=1)
pWilks=pchisq(TWilks,df=1,lower.tail=FALSE) ## only if online (& is unnecessary)
c(l0,TWilks,Tcrit,pWilks)
```

```
## [1] -276.86373341    5.04970686    5.02388619    0.02463011
```

Reject (just)  $H_0$  as  $T_{Wilks} \geq T_{crit}$ .

(6 marks)

### 3

#### (a)

##### (i)

$$\mathbf{Y} \sim N_n(X\boldsymbol{\beta}, \sigma^2 I_n),$$

where  $I_n$  is the  $n \times n$  identity matrix.

##### (ii)

$$\begin{aligned} \text{Var}[(X^\top X)^{-1}X^\top \mathbf{Y}] &= (X^\top X)^{-1}X^\top \text{Var}[\mathbf{Y}] \{(X^\top X)^{-1}X^\top\}^\top \\ &= \sigma^2 (X^\top X)^{-1}X^\top I_n X (X^\top X)^{-1} \\ &= \sigma^2 (X^\top X)^{-1} \end{aligned}$$

(6 marks) because  $(X^\top X)^{-1}$  is symmetric (2 marks).

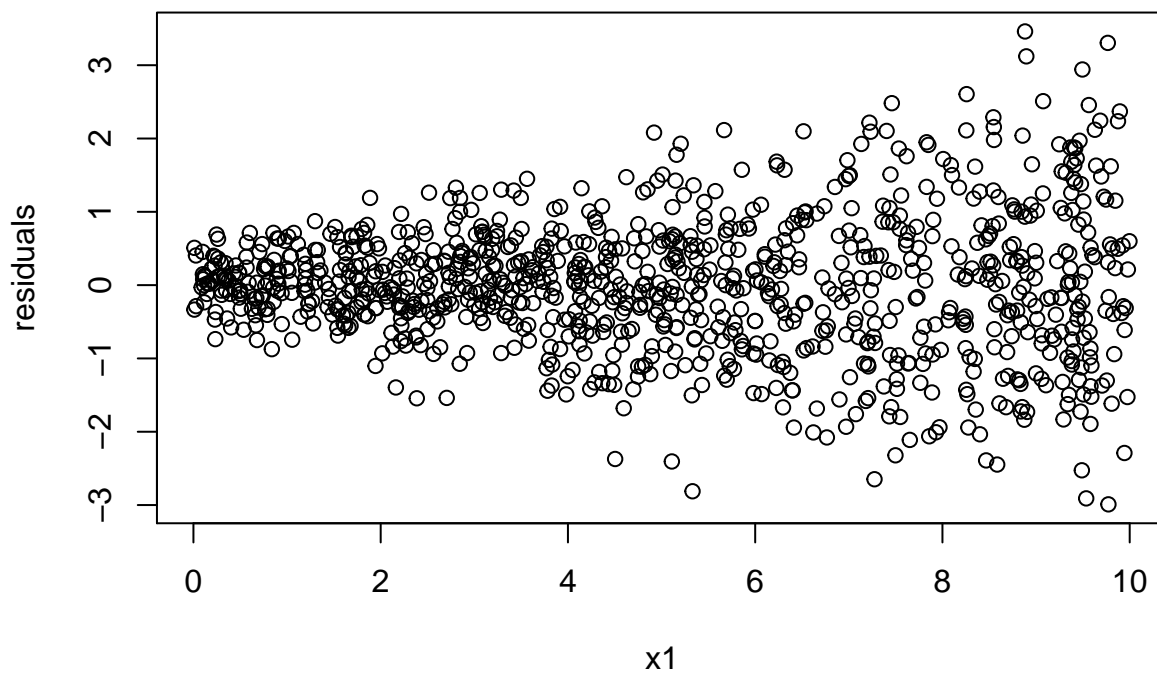


**(b)**

```
## Generate data
n=1000
set.seed(1235713)
x1=runif(n)*10
x3us=runif(n)
x2us=runif(n)/2+x3us/2
T21=x2us<0.5; T22=x2us>0.5
T31=x3us<0.2; T33=x3us>0.7; T32=!T31 & !T33
x2=rep("No",n); x2[T22]="Yes"
x3=rep("R",n); x3[T32]="G"; x3[T33]="B"
X=cbind(rep(1,n), x1, T22, T32, T33, T22*T32)
betas=c(6, 0.7, 1.0, 2, 3, -1.5)
sigma=0.8
etas=X %*% betas
Q3y=etas+sigma*rnorm(n)*(0.375+x1/8)
Mb=lm(Q3y~x1+x2+x3+x2*x3)
Sb=summary(Mb)
Sb$coeff
```

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	8.9584809	0.147068762	60.9135535	0.000000e+00
## x1	0.7133570	0.009844503	72.4624690	0.000000e+00
## x2Yes	0.9055976	0.148985487	6.0784284	1.728214e-09
## x3G	-0.9822243	0.147983942	-6.6373708	5.244480e-11
## x3R	-2.9402937	0.153737248	-19.1254481	1.157324e-69
## x2Yes:x3G	-1.5021881	0.169362111	-8.8696822	3.348565e-18
## x2Yes:x3R	0.1110128	0.273643059	0.4056846	6.850617e-01

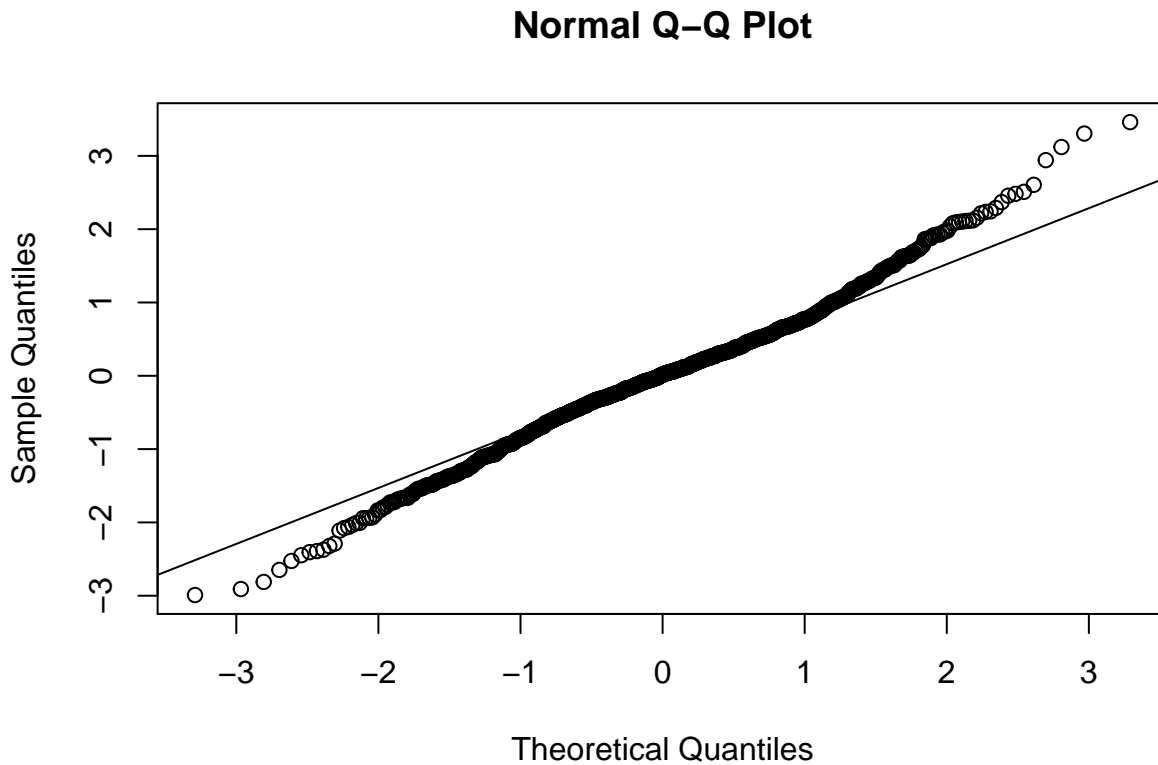
```
resids=residuals(Mb)
plot(x1,resids,ylab="residuals")
```



```
dev.print(pdf, file="specFig1.pdf")
```

```
## pdf  
## 2
```

```
qqnorm(resids)  
qqline(resids)
```



```
dev.print(pdf, file="specFig2.pdf")
```

```
## pdf
## 2
```

**Note to student:** data values here are slightly different to those appearing in the specimen exam. Answers are for the exam values.

**(i)**

For an increase of  $x_1$  by 1,  $\mathbb{E}[Y]$  increases by 0.71m (or  $y$  increases by 0.71m on average).

**(ii)**

When  $x_2 = \text{No}$ ,  $y$  is 0.93m lower when  $x_3 = \text{G}$  than when  $x_3 = \text{B}$ , and is 2.87m lower when  $x_3 = \text{R}$  than when  $x_3 = \text{B}$  (2 marks). When  $x_2 = \text{Yes}$ ,  $y$  is 2.46m ( $0.93 + 1.53$ ) lower when  $x_3 = \text{G}$  than when  $x_3 = \text{B}$ , and is 2.94m ( $2.868 + 0.069$ ) lower (OR essentially the same discrepancy as when  $x_2 = \text{No}$ ) when  $x_3 = \text{R}$  than when  $x_3 = \text{B}$  (4 marks). Finally, when  $x_3 = \text{B}$ ,  $y$  is 0.93m higher when  $x_2 = \text{Yes}$

than when  $x_2 = \text{No}$ . (2 marks) (lose two marks for inappropriate number of dp e.g. none or  $\geq 4$ ).

*How to decide on dp?:* look at the standard errors. For example, there is no point giving a precision which is a thousand times smaller than the uncertainty arising from the variability in the data. Depending on the context - the use to which the numbers will be put - it could also be crazy to give a precision which is worse than this uncertainty (or even the same size). Please use common sense.

**(iii)**

The residual variance increases with  $x_1$  (2 marks), this means that the overall distribution has fatter tails than a standard Gaussian, as is shown in the QQplot (2 marks). Specify (something like)  $\epsilon_i \sim N(0, \sigma^2 \exp(\gamma x_{1,i}))$ . Or, more generally, some monotonically increasing variance function  $\sigma^2 f(x_1; \gamma)$ . (2 marks) The log-likelihood is then

$$-n \log \sigma - \frac{1}{2} \sum_{i=1}^n \log f(x_{1,i}; \gamma) - \frac{1}{2\sigma^2} \sum_{i=1}^n \frac{(y_i - \eta_i)^2}{f(x_{1,i}; \gamma)}.$$

(6 marks)

**(c)**

```
## Generate data
set.seed(321)
n=75
x1=runif(n)*3-1
x2=runif(n)*2-1.5
betas=c(-1,.2,-.06)
sigma=1
X=cbind(rep(1,n),x1,x2)
logity=X %*% betas+rnorm(n)*sigma
y=1/(1+exp(-logity))
Mc=lm(log(y/(1-y))~x1+x2)
Sc=summary(Mc)
Sc$coeff
```

##		Estimate	Std. Error	t value	Pr(> t )
##	(Intercept)	-0.7776230	0.1546968	-5.0267546	3.516277e-06
##	x1	0.1140109	0.1287506	0.8855178	3.788254e-01
##	x2	0.1788088	0.1993881	0.8967876	3.728211e-01

(i)

$Y$  must be between 0 and 1 (2 marks), and the logit transform ensures this (2 marks).

(ii)

```
predlin=Sc$coeff[1,1]
pred=1/(1+exp(predlin))
pred
```

```
## [1] 0.6851676
```

(2 marks)

```
CIlin=predlin+c(-1,1)*Sc$coeff[1,2]*qnorm(0.975)
CI=1/(1+exp(-CIlin))
CIlin
```

```
## [1] -1.0808233 -0.4744228
```

```
CI
```

```
## [1] 0.2533503 0.3835700
```

(6 marks)

(iii)

$Y_i \sim \text{Beta}(a_i, b_i)$ , (2 marks). One sensible possibility is:  $a_i = c \exp(\eta_i)$  and  $b_i = c$ , where  $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ . Then  $\mathbb{E}[Y_i] = \exp(\eta_i) / [1 + \exp(\eta_i)]$ . The point is that there is a linear predictor (2 marks), the parameters have the right range - they are positive (2 marks) and the expectation increases monotonically with the linear predictor (2 marks). In the particular example given, the extra parameter  $c$  controls the variance of  $Y_i$ . **NB:** the solution in the original version of this

document, which I started to go through in the revision session, then queried, was indeed wrong.