MAST30027: Modern Applied Statistics

Exam Solutions

Problem 2

(a) Show that the gamma distribution is an exponential family.
(b) Obtain the canonical link
(c) Obtain the variance function.
Solution: The gamma distribution with shape $\nu > 0$ and rate $\lambda > 0$ has log density

$$\begin{aligned}\log f(x;\nu,\lambda) &= (\nu-1)\log(x) - \lambda x + \nu\log(\lambda) - \log(\Gamma(\nu)) \\ &= \frac{x(-\lambda/\nu)+\log(\lambda/\nu)}{1/\nu} - \nu\log(1/\nu) + (\nu-1)\log(x) - \log(\Gamma(\nu))\end{aligned}$$

Put $\theta = -\lambda/\nu$ and $\phi = 1/\nu$ then we have

$$\log f(x;\nu,\lambda) = \frac{x\theta - \log(-1/\theta)}{\phi} - \frac{\log(\phi)}{\phi} + \left(\frac{1}{\phi}-1\right)\log(x) - \log(\Gamma(1/\phi))$$

This is in the form of an exponential family, with

$$\begin{aligned}b(\theta) &= \log(-1/\theta) \\ a(\phi) &= \phi \\ c(x,\phi) &= \frac{-\log(\phi) + (1-\phi)\log(x) - \phi\log(\Gamma(1/\phi))}{\phi}\end{aligned}$$

Note that with this parameterisation we have $\theta < 0$ and $\phi > 0$.

For the canonical link g we have $g(\mu) = \theta$. Here $\mu = \nu/\lambda = -1/\theta$, so $g(x) = -1/x$. (Note that in practice people tend to use the inverse link $x \mapsto 1/x$ rather than $x \mapsto -1/x$, because it is convenient to keep things positive.) The variance is $\nu/\lambda^2 = \phi\mu^2 = a(\phi)v(\mu)$. That is, the variance function is $v(\mu) = \mu^2$.

Problem 4

The following three-way table refers to results of a case-control study about effects of cigarette smoking and coffee drinking on myocardial infarction (MI) or heart attack for a sample of men under 55 years of age.

Cups Coffee	Cigarettes per Day: 0		1-24		25-34		≥ 35	
per Day	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
0	66	123	30	52	15	12	36	13
1-2	141	179	59	45	53	22	69	25
3-4	113	106	63	65	55	16	119	30
≥ 5	129	80	102	58	118	44	373	85

Eight log-linear models with Poisson error have been fitted, with the residual deviances given in the following table.

Model	Residual deviance
`coffee + cigar + MI`	607.25
`coffee + cigar*MI`	394.43
`cigar + coffee*MI`	484.70
`MI + coffee*cigar`	271.40
`coffee*cigar + coffee*MI`	148.81
`coffee*cigar + cigar*MI`	58.55
`coffee*MI + cigar*MI`	271.88
`coffee*cigar + coffee*MI + cigar*MI`	11.17

You will find the following chi-squared percentage points useful for problems (c) and (d).

```
> qchisq(0.95, df=5:10)
[1] 11.07050 12.59159 14.06714 15.50731 16.91898 18.30704
> qchisq(0.95, df=11:15)
[1] 19.67514 21.02607 22.36203 23.68479 24.99579
> qchisq(0.95, df=16:20)
[1] 26.29623 27.58711 28.86930 30.14353 31.41043
```

(a) What are the residual degrees of freedom (d.f.) for each of the three models: `coffee + cigar + MI`, `cigar + coffee*MI`, and `coffee*cigar + cigar*MI`?
Solution: `coffee + cigar + MI`: 24 `cigar + coffee*MI`: 21 `coffee*cigar + cigar*MI`: 12

(b) Give an interpretation of each of the following models.

(a) `coffee + cigar + MI`

(b) `MI + coffee*cigar`

(c) `coffee*cigar + coffee*MI`

Solution:

i. The three factors `coffee`, `cigar` and `MI` are mutually independent;

ii. `MI` is independent of `coffee` and `cigar` together;

iii. Given `coffee`, `cigar` and MI are independent (unrelated).

(c) Test the hypothesis that there is no association between `coffee` and `MI` when `cigar` level is given (at the 95% level).
Solution: This is to test the adequacy of `coffee*cigar+cigar*MI`. The residual deviance is 58.55 on 12 df. $\chi^2_{12,0.95} = 21.026$ so `coffee*cigar+cigar*MI` is not adequate, and we will reject the hypothesis.

(d) Test the hypothesis that the association between `MI` and `cigar` is the same for all `coffee` levels. That is, test that there is no three-way interaction (at the 95% level).

Solution: We should test the adequacy of `coffee*cigar + coffee*MI + cigar*MI`. The deviance is 11.17 on 9 df and $\chi^2_{9,0.95} = 16.919$, so `coffee*cigar + coffee*MI + cigar*MI` is adequate, and we cannot reject the hypothesis.