MATH 523: EXAMPLE - NEGATIVE BINOMIAL MODEL

Recall the tumour data, with two factors

- Tumour type: `type`
  - 1- Hutchinsons melanotic freckle, 2- Superficial spreading melanoma, 3- Nodular, 4- Indeterminate
- Location: `site`
  - 1- Head and Neck, 2- Trunk, 3-Extremities

defining a 4×3 table.

		Site			
		1	2	3	Total
Tumor	1	22	2	10	34
	2	16	54	115	185
	3	19	33	73	125
	4	11	17	28	56
	Total	68	106	226	400

We can compare the parameter estimates for the main effects (independence) model

`type+site`

under the two assumptions

- M_P : Y Poisson distributed
- M_N : Y Negative binomially distributed.

Recall that the negative binomial model can be generated using a Gamma mixture model:

$$Y_{jk}|Z_{jk} = z_{jk} \sim Poisson(z_{jk}\mu_{jk}) \qquad Z_{jk} \sim Gamma(\theta_Z, \theta_Z)$$

with the Z_{jk} random variables mutually independent, and the Y_{jk} conditionally independent given the Z_{jk}. The marginal pmf of Y is easily computed to be

$$f_Y(y; \theta_Z, \mu) = \frac{\Gamma(\theta_Z + y)}{\Gamma(\theta_Z)y!}\left(\frac{\mu}{\theta_Z + \mu}\right)^y \left(\frac{\theta_Z}{\theta_Z + \mu}\right)^{\theta_Z} \qquad y = 0, 1, 2 \ldots$$

for parameters $\mu, \theta_Z > 0$. Then $\mathbb{E}[Y] = \mu$, but $\text{Var}[Y] = \mu + \mu^2/\theta_Z$. If θ_Z is large, then the Poisson and negative binomial model become very similar. A log link function $\eta = \log \mu$ can be used for both models.

The following code can be used to fit the models, including the estimation of θ_Z.

```
site<-gl(3,1,12)
tumor<-gl(4,3)
Y<-c(22,2,10,16,54,115,19,33,73,11,17,28)

cancer<-data.frame(tumor,site,Y)

#Poisson fit
cancer.mp <- glm(Y ~ tumor + site, family=poisson,data=cancer)

#Negative Binomial fit
library(MASS)
cancer.mn<-glm.nb(Y ~ tumor+site, data=cancer)
```

Results (edited) are as follows:

```
1  Call: glm(formula = Y ~ tumor + site, family = poisson, data = cancer)
2
3  Coefficients:
4              Estimate Std. Error z value Pr(>|z|)
5  (Intercept)   1.7544     0.2040   8.600  < 2e-16 ***
6  tumor2        1.6940     0.1866   9.079  < 2e-16 ***
7  tumor3        1.3020     0.1934   6.731 1.68e-11 ***
8  tumor4        0.4990     0.2174   2.295  0.02173 *
9  site2         0.4439     0.1554   2.857  0.00427 **
10 site3         1.2010     0.1383   8.683  < 2e-16 ***
11 ---
12 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
13
14 (Dispersion parameter for poisson family taken to be 1)
15
16     Null deviance: 295.203  on 11  degrees of freedom
17 Residual deviance:  51.795  on  6  degrees of freedom
18 AIC: 122.91
19
20 Call: glm.nb(formula = Y ~ tumor + site, data = cancer, init.theta =
21 5.020142616, link = log)
22
23 Coefficients:
24             Estimate Std. Error z value Pr(>|z|)
25 (Intercept)  2.20203    0.37444   5.881 4.08e-09 ***
26 tumor2       1.48351    0.40991   3.619 0.000296 ***
27 tumor3       1.12423    0.41330   2.720 0.006526 **
28 tumor4       0.37813    0.42544   0.889 0.374123
29 site2        0.05471    0.35731   0.153 0.878308
30 site3        0.80075    0.34751   2.304 0.021210 *
31 ---
32 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
33
34 (Dispersion parameter for Negative Binomial(5.0201) family taken to be 1)
35
36     Null deviance: 41.328  on 11  degrees of freedom
37 Residual deviance: 13.663  on  6  degrees of freedom
38 AIC: 108.12
39
40               Theta:  5.02
41           Std. Err.:  2.75
42
43  2 x log-likelihood:  -94.124
```

The `glm.nb` call (line 20) fits the negative binomial model using an iterative procedure. The resulting estimate (standard error) of θ_Z is 5.02 (2.75) - see lines 40-41; this is not too large, indicating that the negative binomial model is justified over the Poisson model. Note that the parameter estimates and standard errors change, as do the fitted values (below).

The difference between the $Poisson(\mu)$, and $NegativeBinomial(\mu, \theta_Z)$ is shown below in the figure, with $\mu = \exp\{2\}$ (top panel) and $\mu = \exp\{4\}$ (bottom panel) for $\theta_Z = 5$.

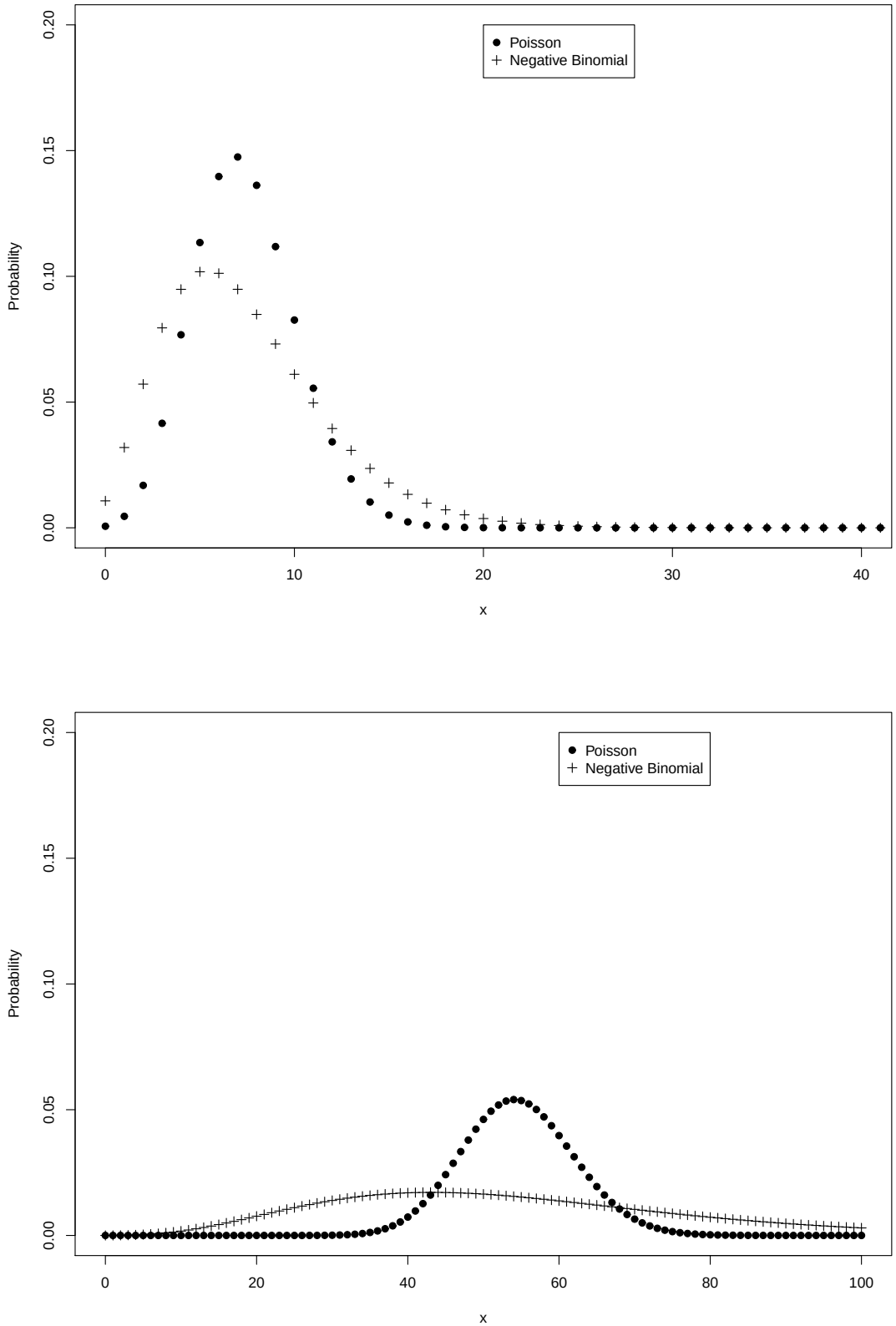


Figure 1: Comparison of Poisson and negative binomial models for different values of μ: $\mu = \exp\{2\}$ (top) and $\mu = \exp\{4\}$ (bottom) for $\theta_Z = 5$.

		Poisson Fit			Neg. Binomial Fit		
		1	2	3	1	2	3
Tumor	1	5.78	9.01	19.21	9.04	9.55	20.14
	2	31.45	49.02	104.52	39.87	42.11	88.79
	3	21.25	33.12	70.62	27.83	29.40	61.99
	4	9.52	14.84	31.64	13.20	13.94	29.40

Table 1: Table of Fitted Values

Model comparison for the Poisson and Negative Binomial models **can** be carried out using a likelihood ratio test as the Poisson model is nested inside the Negative Binomial: if $\varphi = 1/\theta_Z$, then we have that $\varphi = 0$ yields the Poisson, so the likelihood ratio test is

$$\begin{aligned} H_0 &: \varphi = 0 \\ H_1 &: \varphi > 0 \end{aligned}$$

Therefore, the likelihood ratio statistic is

$$-2\log\lambda(\boldsymbol{y}) = -2(l_n^{(P)}(\widehat{\mu}) - l_n^{(N)}(\widehat{\mu},\widehat{\varphi}))$$

where $l_n^{(P)}$ and $l_n^{(N)}$ are the log likelihoods for Poisson and binomial models.

This model, however, is **non-regular**: the nesting structure is obtained by setting φ equal to a value on the boundary of the parameter space, and hence the regularity assumptions required for the usual asymptotic behaviour are not met. It can be shown that in this case

$$-2\log\lambda(\boldsymbol{y}) \xrightarrow{d} V$$

where V is a random variable with a **mixture density**

$$V \stackrel{d}{=} \frac{1}{2}\delta_{\{v=0\}} + \frac{1}{2}\chi_1^2$$

that is with probability mass $1/2$ at $v = 0$, and $1/2$ on the χ_1^2 density. Under this distribution

$$\Pr[V > v] = \frac{1}{2}S_1(v)$$

where $S_1(v)$ is the tail probability of exceeding v for a χ_1^2 density. $\Pr[V > v]$ is the p-value in the likelihood ratio test.

For these data, we can obtain the optimized log-likelihood values via the `logLik` function in `R`:

```
> logLik(cancer.mp);logLik(cancer.mn)
'log Lik.'  -55.45321 (df=6)
'log Lik.'  -47.06196 (df=7)
```

have $v = 16.78249$. Therefore the probability of observing $V > v$ is

```
> 0.5*(1-pchisq(16.78249,1))
[1] 2.095916e-05
```

hence a highly significant result. This implies that the Negative Binomial Model fits significantly better than the Poisson.