



ST2304 Statistical modelling for biologists and biotechnologists - solution
9. juni, 2011

Problem 1

- a) `dpois(5,0.5*10)`
- b) `ppois(5,0.5*A,lower.tail=FALSE)`
- c) `rpois(5,0.5*A)`

Problem 2

- a) Based on the linear model $y = a + bx$ the predicted proportion \hat{y} of surviving nestlings given a clutch date of $x = 150$ becomes

$$\hat{y} = \hat{a} + \hat{b}x = -1.73 + 0.0193 \cdot 150 = 1.165. \quad (1)$$

This prediction, following from the questionable assumption of linearity between clutch-date and survival probability, is unreasonable since the proportion surviving has to be between 0 and 1. The model also assumes that the response variable (the proportion of nestlings surviving in each nest) is normally distributed with a constant variance which is questionable since we are dealing with a discrete variable with variance which must go to zero as the survival probability tends to either zero or one.

Since the model on which the statistical significance of the effect of clutchdate is wrong, this result is questionable. Over the range of clutchdates considered the relationship to survival probability may be only approximately linear. On the other hand, just like a t -test, a linear model is known to be robust against violations of the assumption of normality. Not taking into account the non-constant variance of the response variable makes the test somewhat inefficient.

- b) The fitted model assumes that the number of nestlings surviving $X_i \sim \text{bin}(p_i, n_i)$ where

$$\text{logit } p_i = \beta_0 + \beta_c \text{clutchdate}_i + \beta_b \text{badgesize}_i \quad (2)$$

Furthermore, the number of nestlings X_i surviving in different nests are independent. The covariates thus have linear (additive) effect on the logit of the survival probability and a multiplicative effect on the odds of survival.

- c) Rewriting (2), the predicted survival for these values of the explanatory variables becomes

$$\hat{p} = \frac{1}{1 + e^{-(\hat{\beta}_0 + \hat{\beta}_c \text{clutchdate} + \hat{\beta}_b \text{badgesize})}} = \frac{1}{1 + e^{-(-9.73 + 0.078 \cdot 150 + 0.0016 \cdot 400)}} = 0.92. \quad (3)$$

- d) An estimate of the scale parameter of the model is

$$\hat{\phi} = D/(n - p) = 15.57/17 = 0.92. \quad (4)$$

This indicate a small amount of under dispersion but this is not significant since D is larger than the lower 0.05/2-quantile of the chi-square distribution with 17 degrees of freedom, 7.56.

Overdispersion may in this case be caused by important covariates not included in the model, or a wrong functional form between the covariate and the survival probability p . Positive covariance between the survival of individuals nestling within each nest would also inflate the variance of each X_i beyond $n_i p_i (1 - p_i)$. This could arise as a result of say, improved thermoregulation.

Similarly, negative covariance as a result of say, competition between individual nestlings, such that the survival probability of on individual is increased if a siblings dies, would cause underdispersion.

Problem 3

- a) `lnL <- function(p, x)
-sum(dgamma(x, shape=p[1], scale=p[2], log=T))`
- b) The MLEs are $\hat{\alpha} = 1.27$ and $\hat{\beta} = 20.40$. The maximum likelihood is -423.90 and standard errors are 0.16 and 3.16.
- c) The exponential model is a special case of the gamma model which can be seen by setting $\alpha = 1$. The density function then becomes

$$f(x) = \frac{1}{\beta^1 \Gamma(1)} x^{1-1} e^{-x/\beta} = \frac{1}{\beta} e^{-x/\beta} \quad (5)$$

which is the density function of an exponentially distributed random variable.

For large samples size n , the change in two times the log likelihood is approximately chi-square with degrees of freedom equal to the change in number of parameters.

The observed value of this test statistic becomes

$$2(\ln L_1 - \ln L_0) = 2(-423.90 - (-425.56)) = 3.32. \quad (6)$$

Since this is smaller than the critical value of the test, the upper 0.05-quantile in a chi-square distribution with $p_1 - p_0 = 2 - 1 = 1$ degrees of freedom equal to 3.84 we can not reject the simpler exponential model in favour of the extended gamma model.