StAN Exercise Sheet 6

Richard D. Gill

Mathematical Institute, University of Leiden, Netherlands

http://www.math.leidenuniv.nl/~gill

22 November, 2012

1 Bayes estimators; Generalized likelihood ratio test

1.1 Bayes.

Cf. Feigelson and Babu, Section 3.8.

Suppose $\theta \in (0, 1)$ is an unknown probability. Suppose we are prepared to summarize our prior knowledge about θ by the statement that a priori, θ has a beta distribution with parameters α and β, for some $\alpha > 0$ and $\beta > 0$.

Suppose that a priori, we think that the following three possibilities have equal probability 1/3: $\theta < 0.25$, $0.25 \leq \theta \leq 0.5$, and $\theta > 0.5$. Use `R` to determine α and β such that the corresponding beta distribution reproduces these three prior probabilities.

We now observe a single realisation $X = x$ of a binomial random variable X: the number of successes in n independent trials each with success probability θ. In particular, suppose $n = 1000$ and $x = 645$. In the light of this information, what should be now believe about θ?

Use `R` to compute the shortest possible interval of values of the unknown parameter θ having posterior probability 95% given our data x.

1.2 Generalized likelihood ratio test

Cf. Feigelson and Babu, Section 3.7.2.

Suppose that we have data X (possibly a random vector or matrix) and two competing theories about the probability law which generated X. Suppose that according to one theory $\mathcal{P}$, $X \sim p(x, \theta)$ while according to the other theory $\mathcal{Q}$, $X \sim q(x, \phi)$. In the first theory, θ is a vector of unknown parameters and in the second theory ϕ is another vector of unknown parameters.

The two models are called *nested* if one is a special case of the other. In the example above, if for every ϕ there is a θ such that $q(x, \phi) = p(x, \theta)$ for all x, then the second model is nested in the first model; we also say then: the second model is a sub-model of the other.

Such a situation arises when, for instance, the model $\mathcal{Q}$ arises from the model $\mathcal{P}$ by imposing some functional relationships between the components of the parameter vector θ. In such a situation, the dimension of the parameter ϕ is smaller than that of θ.

A powerful method of testing the null-hypothesis that $\mathcal{Q}$ is true, against the hypothesis that $\mathcal{Q}$ is not true but $\mathcal{P}$ is, is based on the generalised log likelihood ratio test: compute the maximum likelihood estimates of θ under model $\mathcal{P}$ and of ϕ under model $\mathcal{Q}$. Define the generalised likelihood ratio $\Lambda = \text{lik}_\mathcal{Q}(\widehat{\phi}_{\text{MLE}})/\text{lik}_\mathcal{Q}(\widehat{\phi}_{\text{MLE}})$. Then compare $-2\log(\Lambda)$ to a chi-square distribution with degrees of freedom equal to the difference in the dimensions of θ and ϕ (or equivalently, the number of independent constraints on the components of θ which are needed to force the distribution of X to be a member of model $\mathcal{Q}$. To be more precise, we reject the null hypothesis that $\mathcal{Q}$ is true if $-2\log(\Lambda)$ is larger than the $(1 - \alpha)$-quantile of that chi-squared distribution, in order to obtain a test of approximate size α (= 0.05 for instance).

According to a theorem of Wilks, this procedure is approximately correct if we have a large amount of data, smooth models (implying "expected" large sample behaviour of maximum likelihood estimates), and the "small model" is actually true with parameter ϕ in the interior of the corresponding parameter space. The corresponding approximate test based on $-2\log\Lambda$ is called the **Wilks generalised likelihood ratio test**.

Explain why the following statement is true: the quantity Λ is always smaller than 1 and, intuitively, the smaller it is, the less plausible is the model $\mathcal{Q}$ within the "background" assumption that $\mathcal{P}$ is true.

Suppose the data is a single observation X from a normal distribution. Suppose that according to model $\mathcal{P}$, X has variance 1 and unknown mean μ, while according to $\mathcal{Q}$, the same is true with $\mu = 0$. Show that under the null-hypothesis that $\mu = 1$, $-2\log\Lambda$ is exactly chi-squared distributed with 1 degree of freedom.

Suppose I take a sample of size n from a gamma distribution with rate parameter λ and shape parameter ν. Let $\widehat{\lambda}, \widehat{\nu}$ denote the maximum likelihood estimators of λ and ν. Explain why $\{(\lambda, \nu) : \log\text{lik}(\lambda, \nu) > \log\text{lik}(\widehat{\lambda}, \widehat{\nu}) - 0.5 \times (\chi^2_2)^{-1}(0.95)\}$ is an approximately 95% confidence region for (λ, ν) by arguing that the region so defined contains the true value (λ_0, ν_0) if and only if the level $\alpha = 5\%$ Wilks' test of the null hypothesis $(\lambda, \nu) = (\lambda_0, \nu_0)$ against the alternative $(\lambda, \nu) \neq (\lambda_0, \nu_0)$ accepts the null.

Use `R` to graphically represent this confidence region given a sample of size 100 from your favourite gamma distribution. Hint: use the function `contour()` for the graphics, and `fitdistr()` (from library `MASS`) for the estimation.