Deber Seminario Investigación. Ejercicios pares Capítulo 4 "Bayesian Data Analysis"

Fausto Fabian Crespo Fernandez

Junio 2016

0.1. Ejercicio 2

2. Normal approximation: derive the analytic form of the information matrix and the normal approximation variance for the bioassay example.

Solución Tenemos que
$p(y_i|\theta) \propto (logit^{-1}(\alpha+\beta x_i))^{y_i}(1-logit^{-1}(\alpha+\beta x_i))^{n_i-y_i}$
$l_i = lnp(y_i|\theta) = cte + y_i ln(logit^{-1}(\alpha+\beta x_i)) + (n_i - y_i) ln(1-logit^{-1}(\alpha+\beta x_i))$
$\frac{d^2 l_i}{d\alpha^2} = -\frac{n_i exp(\alpha+\beta x_i)}{(1+exp(\alpha+\beta x_i))^2}$
$\frac{d^2 l_i}{dd\beta} = -\frac{n_i x_i exp(\alpha+\beta x_i)}{(1+exp(\alpha+\beta x_i))^2}$
$\frac{d^2 l_i}{d\beta^2} = -\frac{n_i x_i^2 exp(\alpha+\beta x_i)}{(1+exp(\alpha+\beta x_i))^2}$
La densidad a priori de (α,β)es uniforme y entonces $lnp(\theta|y) = cte + \sum_{i=1}^4 l_i$ y
$I(\hat{\theta}) = \begin{pmatrix} \sum_{i=1}^4 \frac{n_i exp(\alpha+\beta x_i)}{(1+exp(\alpha+\beta x_i))^2} & \sum_{i=1}^4 \frac{n_i x_i exp(\alpha+\beta x_i)}{(1+exp(\alpha+\beta x_i))^2} \\ \sum_{i=1}^4 \frac{n_i x_i exp(\alpha+\beta x_i)}{(1+exp(\alpha+\beta x_i))^2} & \sum_{i=1}^4 \frac{n_i x_i^2 exp(\alpha+\beta x_i)}{(1+exp(\alpha+\beta x_i))^2} \end{pmatrix}\Bigg|_{(\alpha,\beta)=(\hat{\alpha},\hat{\beta})}$ donde $\hat{\theta} = (\hat{\alpha},\hat{\beta})$ es la moda a posteriori. Denotando I como $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ entonces las aproximaciones normales a las varianzas son los elementos de la diagonal principal de I^{-1}:$\frac{b}{ab-c^2}$ y $\frac{a}{ab-c^2}$ respectivamente

0.2. Ejercicio 4

4. Asymptotic normality: assuming the regularity conditions hold, we know that $p(\theta|y)$ approaches normality as $n \to \infty$. In addition, if $\phi = f(\theta)$ is any one-to-one continuous transformation of θ, we can express the Bayesian inference in terms of ϕ and find that $p(\phi|y)$ also approaches normality. But a nonlinear transformation of a normal distribution is no longer normal. How can both limiting normal distributions be valid?

Solución
En el límite cuando $n \to \infty$, la varianza a posteriori tiende a 0 o sea la distribución a posterior se concentra en un solo punto y cualquier transformación uno-a-uno continua sobre los números reales es localmente lineal es una vecindad de ese punto.

0.3. Ejercicio 6

6. Statistical decision theory: a decision-theoretic approach to the estimation of an unknown parameter θ introduces the loss function $L(\theta, a)$ which, loosely speaking, gives the cost of deciding that the parameter has the value a, when it is in fact equal to θ. The estimate a can be chosen to minimize the *posterior expected loss*,

$$\mathrm{E}(L(a|y)) = \int L(\theta, a)p(\theta|y)d\theta.$$

This optimal choice of a is called a *Bayes estimate* for the loss function L. Show that:

(a) If $L(\theta, a) = (\theta - a)^2$ (squared error loss), then the posterior mean, $\mathrm{E}(\theta|y)$, if it exists, is the unique Bayes estimate of θ.

(b) If $L(\theta, a) = |\theta - a|$, then any posterior median of θ is a Bayes estimate of θ.

(c) If k_0 and k_1 are nonnegative numbers, not both zero, and

$$L(\theta, a) = \begin{cases} k_0(\theta - a) & \text{if} \quad \theta \geq a \\ k_1(a - \theta) & \text{if} \quad \theta < a, \end{cases}$$

then any $\frac{k_0}{k_0+k_1}$ quantile of the posterior distribution $p(\theta|y)$ is a Bayes estimate of θ.

Solución

Se asume por simplicidad que la distribución a posteriori es continua y que los momentos necesarios existen(la demostración se puede adaptar fácilmente en el caso de distribuciones discontinuas)

- (a) Derivando respecto a a,$E(L(a|y))$ con $L(\theta, a) = (\theta - a)^2$ tenemos $\frac{d}{da}E(L(a|y)) = \frac{d}{da}\int(\theta - a)^2 p(\theta|y)d\theta = -2\int(\theta - a)p(\theta|y)d\theta$ $= -2(E(\theta|y) - a)$ que es 0 si $a = E(\theta|y)$

- (b) Podemos aplicar el inciso c) con $k_0 = k_1 = 1$

- (c) Tenemos que
  $\frac{d}{da}E(L(a|y)) = \frac{d}{da}(\int_{-\infty}^{a} k_1(a - \theta)p(\theta|y)d\theta + \int_{a}^{\infty} k_0(\theta - a)p(\theta|y)d\theta)$
  $= k_1\int_{-\infty}^{a} p(\theta|y)d\theta + k_0\int_{a}^{\infty} p(\theta|y)d\theta$ y como
  $\int_{a}^{\infty} p(\theta|y)d\theta = 1 - \int_{-\infty}^{a} p(\theta|y)d\theta$
  $\frac{d}{da}E(L(a|y)) = (k_1 + k_0)\int_{-\infty}^{a} p(\theta|y)d\theta - k_0$ y es cero si $\int_{-\infty}^{a} p(\theta|y)d\theta = \frac{k_0}{k_0+k_1}$

0.4. Ejercicio 8

8. Regression to the mean: work through the details of the example of mother's and daughter's heights on page 94, illustrating with a sketch of the joint distribution and relevant conditional distributions.

Solución
Tenemos que las alturas de la madre y la hija tienen distribución normal conjunta con medias iguales de 160 cm, varianzas iguales(σ)y coeficiente de correlacion de 0.5.La distribución normal bivariada de θ, y es :
$f(\theta, y) = \frac{1}{2\pi\sigma^2\sqrt{(1-\rho^2)}} exp(-\frac{1}{2(1-\rho^2}[(\frac{\theta-160}{\sigma})^2 - 2\rho\frac{(\theta-160)(y-160)}{\sigma^2} + (\frac{y-160}{\sigma})^2])$
Las distribución condicional $f(\theta|y)$ es $f(\theta|y) = \frac{f(\theta,y)}{fy(y)} = \frac{f(\theta,y)}{\int_{-\infty}^{\infty} f(\theta|y)d\theta}$ y
$\int_{-\infty}^{\infty} f(\theta|y)d\theta = \frac{1}{2\pi\sigma^2\sqrt{(1-\rho^2)}} \int_{-\infty}^{\infty} exp(-\frac{1}{2\sigma^4(1-\rho^2)}[(\theta-160)^2\sigma^2 - 2\rho\sigma^2(\theta-160)(y-160) + (y-160)^2\sigma^2])$ completando el cuadrado en función de θ en el exponente se obtiene la integral de una densidad normal con media $160 + (y-160)(160/160)\rho$ y varianza $(1-\rho^2)\sigma^2$ con lo que queda al final :
$f_y(y) = \frac{1}{\sqrt{(2\pi)}\sigma} exp(-\frac{(y-160)^2}{2\sigma^2})$ y la densidad condicional es
$f(\theta|y) = \frac{1}{\sqrt{(2\pi(1-\rho^2))}\sigma} exp(-\frac{1}{2(1-\rho^2)\sigma^2}[\theta - (\bar{\theta} + (y-\bar{y})\frac{\rho\sigma}{\sigma})]^2)$
$= \frac{1}{\sqrt{(2\pi(1-\rho^2))}\sigma} exp(-\frac{1}{2(1-\rho^2)\sigma^2}[\theta - (160 + (y-160)\rho]^2)$
la media condicional es
$E(\theta|y) = \int_{-\infty}^{\infty} \theta f(\theta|y)dy$ que finalmente se tiene
$E(\theta|y) = \bar{\theta} + \frac{(y-\bar{y})\rho\sigma}{\sigma} = 160 + \rho(y-160) = 160 + 0{,}5(y-160)$
Analogamente $E(y|\theta) = 160 + 0{,}5(\theta - 160)$
El valor esperado de la media a posteriori($E(\theta|y)$)dado θ es
$E(E(\theta|y)) = 160 + 0{,}5(E(y|\theta) - 160) = 160 + 0{,}25(\theta - 160)$ que es sesgado hacia 160 ya que partiendo de θ_1 obtenemos
$\theta_2 = 160 + 0{,}25(\theta_1 - 160) = 0{,}75 * 160 + 0{,}25\theta_1$ y
$\theta_3 = 160 + 0{,}25(\theta_2 - 160) = 0{,}75 * 160(1 + 0{,}25) + (0{,}25)^2\theta_1$
$\theta_4 = 160 + 0{,}25(\theta_3 - 160) = 0{,}75 * 160(1 + 0{,}25 + (0{,}25)^2) + (0{,}25)^3\theta_1$
...
$\theta_n = 160 + 0{,}25(\theta_{n-1} - 160) = 0{,}75 * 160(1 + 0{,}25 + (0{,}25)^2 + \cdots + (0{,}25)^{n-2}) + (0{,}25)^{n-1}\theta_1$
Cuando n es grande $(0{,}25)^{n-1}\theta_1 \to 0$ y $\theta_n \to 0{,}75 * 160\frac{1}{1-0{,}25} = 160$
Sin embargo el estimado $\hat{\theta} = 160 + 2(y-160)$ es insesgado ya que
$E(E(\hat{\theta}|y)) = 160 + 2(E(y|\theta) - 160) - 160) = 160 + 2(160 + 0{,}5(\theta - 160) - 160) = \theta$. Pero este estimador $\hat{\theta}$no tiene sentido para valores de y distintos de 160: si

la madre tiene altura $y = \bar{y} + 10 = 170$ entonces la hija tiene altura de 180

0.5. Ejercicio 10

10. Non-Bayesian inference: replicate the analysis of the bioassay example in Section 3.7 using non-Bayesian inference. This problem does not have a unique answer, so be clear on what methods you are using.
    (a) Construct an 'estimator' of (α, β); that is, a function whose input is a dataset, (x, n, y), and whose output is a point estimate $(\hat{\alpha}, \hat{\beta})$. Compute the value of the estimate for the data given in Table 5.2.
    (b) The bias and variance of this estimate are functions of the true values of the parameters (α, β) and also of the sampling distribution of the data, given α, β. Assuming the binomial model, estimate the bias and variance of your estimator.
    (c) Create approximate 95% confidence intervals for α, β, and the LD50 based on asymptotic theory and the estimated bias and variance.
    (d) Does the inaccuracy of the normal approximation for the posterior distribution (compare Figures 3.3 and 4.1) cast doubt on the coverage properties of your confidence intervals in (c)? If so, why?
    (e) Create approximate 95% confidence intervals for α, β, and the LD50 using the jackknife or bootstrap (see Efron and Tibshirani, 1993).
    (f) Compare your 95% intervals for the LD50 in (c) and (e) to the posterior distribution displayed in Figure 3.4 and the posterior distribution based on the normal approximation, displayed in 4.2b. Comment on the similarities and differences among the four intervals. Which do you prefer as an inferential summary about the LD50? Why?

Solución

- (a)
- (b)

0.6. Ejercicio 12

12. Bayesian interpretation of non-Bayesian estimates: repeat the above problem but with σ replaced by s, the sample standard deviation of $y_1, \dots, y_n$.

Solución

Tenemos

- (a)

- (b)

0.7. Bibliografía

[1]http://www.stat.columbia.edu/ gelman/book/solutions2.pdf
[2]http://www.stat.columbia.edu/ gelman/book/solutions3.pdf