Guia 2

soln1_et.pdf

Question 1

1. Name three reasons why a Markov chain may fail to have an equilibrium distribution.
   The three reasons why a Markov chain may not have an equilibrium distribution are reducibility, periodicity and transience.

2. Suppose we are interested in sampling the distribution π on the state space $\mathcal{S}$ using Markov chain Monte Carlo (MCMC).

   a) Let $q(i,j)$, $i,j \in \mathcal{S}$, be the transition probabilities of a Markov chain, where $q(i,j) > 0$ for all $i,j \in \mathcal{S}$. Suppose we use these transition probabilities to define a Metropolis-Hastings Sampler for π. Assume the current state of the Metropolis-Hastings chain is $X_n = i$. The Sampler first proposes a new state accordingto the transition probabilities $q(i,\cdot)$. Then this state is accepted or rejected. Suppose the proposed state is j. What is the acceptance probability for this new state? Depending on acceptance or rejection, which value does X_{n+1} take?
   We accept the new state with probability

$$\alpha(i,j) = \min\left\{1, \frac{\pi(j)\, q(j,i)}{\pi(i)\, q(i,j)}\right\}.$$

   If the state j is accepted then the Markov chain moves to j, that is $X_{n+1} = j$. If, on the other hand the proposed state is rejected the chain remains in its current state, that is $X_{n+1} = i$.

   b) Determine the transition probabilities of the Metropolis-Hastings chain and show that they satisfy detailed balance with respect to the target distribution π.
   The transition probabilities for the Metropolis-Hastings chain are given by

$$p(i,j) = q(i,j)\,\alpha(i,j) + \mathbf{1}_{[i=j]}\, r(i)$$

   for $i,j \in S$ where

$$r(i) = \sum_{j \in S} q(i,j)\,(1 - \alpha(i,j)).$$

Then

$$\begin{aligned}
\pi(i)\, p(i,j) &= \pi(i)q(i,j)\alpha(i,j) + \mathbf{1}_{[i=j]}\pi(i)r(i) \\
&= \pi(i)q(i,j) \min\left\{1, \frac{\pi(j)q(j,i)}{\pi(i)q(i,j)}\right\} + \mathbf{1}_{[i=j]}\pi(j)r(j) \\
&= \min\Big\{\pi(i)q(i,j), \pi(j)q(j,i)\Big\} + \mathbf{1}_{[i=j]}\pi(j)r(j) \\
&= \pi(j)q(j,i) \min\left\{\frac{\pi(i)q(i,j)}{\pi(j)q(j,i)}, 1\right\} + \mathbf{1}_{[i=j]}\pi(j)r(j) \\
&= \pi(j)p(j,i).
\end{aligned}$$

3. Suppose we would like to sample the following joint distribution π of X and Y using Gibbs Sampling.

	$X=0$	$X=1$	$X=2$
$Y=0$	$1/6$	$1/4$	$1/12$
$Y=1$	$1/8$	$1/4$	$1/8$

a) Compute the full conditional distributions

$$\begin{aligned}
\pi_1(x|y) &= \mathbb{P}(X=x|Y=y) \quad \text{and} \\
\pi_2(y|x) &= \mathbb{P}(Y=y|X=x) \quad \text{for } x \in \{0,1,2\} \text{ and } y \in \{0,1\}.
\end{aligned}$$

The first conditional distribution is given by

$$\begin{aligned}
\pi_1(0|0) &= \frac{1/6}{1/6+1/4+1/12} = 1/3 \quad \pi_1(1|0) = \frac{1/4}{1/6+1/4+1/12} = 1/2 \\
\pi_1(2|0) &= \frac{1/12}{1/6+1/4+1/12} = 1/6 \quad \pi_1(0|1) = \frac{1/8}{1/8+1/4+1/8} = 1/4 \\
\pi_1(1|1) &= \frac{1/4}{1/8+1/4+1/8} = 1/2 \quad \pi_1(2|1) = \frac{1/8}{1/8+1/4+1/8} = 1/4
\end{aligned}$$

The second conditional distribution is given by

$$\begin{aligned}
\pi_2(0|0) &= \frac{1/6}{1/6+1/8} = 4/7 \quad \pi_2(1|0) = \frac{1/8}{1/6+1/8} = 3/7 \\
\pi_2(0|1) &= \frac{1/4}{1/4+1/4} = 1/2 \quad \pi_2(1|1) = \frac{1/4}{1/4+1/4} = 1/2 \\
\pi_2(0|2) &= \frac{1/12}{1/12+1/8} = 2/5 \quad \pi_2(1|2) = \frac{1/8}{1/12+1/8} = 3/5
\end{aligned}$$

b) The full conditional distributions from a) can be used to define a Gibbs sampler chain $(X_n, Y_n)_{n\geq 0}$ whose stationary distribution is π. Suppose the states of the Markov chain $(X_n, Y_n)_{n\geq 0}$ are labelled

$$1 = (0,0),\ 2 = (1,0),\ 3 = (2,0),\ 4 = (0,1),\ 5 = (1,1),\ 6 = (2,1).$$

Compute the transition matrix for the update of the first component and the transition matrix for the update of the second component.
The update of the first component is according to the transition matrix

$$P_1 = \begin{pmatrix} 1/3 & 1/2 & 1/6 & 0 & 0 & 0 \\ 1/3 & 1/2 & 1/6 & 0 & 0 & 0 \\ 1/3 & 1/2 & 1/6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1/2 & 1/4 \\ 0 & 0 & 0 & 1/4 & 1/2 & 1/4 \\ 0 & 0 & 0 & 1/4 & 1/2 & 1/4 \end{pmatrix}$$

and the update of the second component according to

$$P_2 = \begin{pmatrix} 4/7 & 0 & 0 & 3/7 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 2/5 & 0 & 0 & 3/5 \\ 4/7 & 0 & 0 & 3/7 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 2/5 & 0 & 0 & 3/5 \end{pmatrix}$$

Empirical Bayes

http://varianceexplained.org/r/empirical_bayes_baseball/

loglinpp.pdf loglinsol.pdf

2. What is the difference between a log-linear model and a linear model after the log transformation of the response?

Respuesta

2. In a log-linear model, the mean of Y is μ and the model is $\log(\mu) = \beta_0 + \beta_1 X_1$. Y is not transformed. If a simple linear regression is used after a log transformation, the model is expressed in terms of the mean of the logarithm of Y. Moreover, the model assumptions are not the same.

3. Why are ordinary residuals $(y_i - \hat{\mu}_i)$ not particularly useful for Poisson regression?

Respuesta

3. The residuals with larger means will have larger variances. So if an observation has a large residual it is difficult to know whether it is an outlier or an observation from a distribution with larger variance than the others. Residuals that are studentized so that they have the same variance are more useful for identifying outliers.

4. Consider the deviance goodness-of-fit test.

   (a) Under what conditions is it valid for Poisson regression?

   (b) When it is valid, what possibilities are suggested by a small p-value?

   (c) When it is valid, what possibilities are suggested by a large p-value?

Respuesta

4. (a) Since it is an asymptotic test (only approximate except in the limit where the sample size goes to infinity), we need large Poisson counts (expected cell counts at least 5 for contingency tables is one rule-of-thumb for "large").

   (b) The Poisson distribution is an inadequate model (for example, there may be extra-Poisson variation), the explanatory variables are inadequate (need more explanatory variables or a different form of the explanatory variables than you have in the model), or there are some outliers.

(c) Either the model is correct, or there is insufficient data to detect any inadequacies.

5. Poisson regression fits the model

$$\log(\mu_i) = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p}$$

where the μ_i's are the means of the Poisson distributions with observed counts y_i, $i = 1, \ldots, n$. Write down the log likelihood function used for maximum likelihood estimation of the β's.

Respuesta

5. The likelihood function is

$$\prod_{i=1}^{n} \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} = \frac{e^{-\sum \mu_i} \prod \mu_i^{y_i}}{\prod y_i!}$$

and the log likelihood function is

$$\log\left(L(\beta_0, \ldots, \beta_p)\right) = -\sum_{i=1}^{n} \mu_i + \sum_{i=1}^{n} y_i \log(\mu_i) - \sum_{i=1}^{n} log(y_i!)$$

where $\mu_i = \exp\left(\beta_0 + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p}\right)$.

GLM.pdf

1. Suppose $X_1, \ldots, X_n$ are i.i.d. Poisson random variables with parameter μ. Show that $\hat{\mu} = \Sigma X_i / n$, and $\operatorname{var}(\hat{\mu}) = \mu / n$.
What is

$$\mathbb{E}(-\frac{\partial^2 L}{\partial \mu^2})?$$

What is the exact distribution of $(n\hat{\mu})$? What is the asymptotic distribution of $\hat{\mu}$?

Respuesta

1. The likelihood is

$$f(x|\mu) = \Pi e^{-\mu} \ \mu^{x_i}/x_i!$$

giving the loglikelihood as

$$L = log(f(x|\mu)) = -n\mu + \Sigma x_i \ log(\mu) + constant.$$

Hence

$$\frac{\partial L}{\partial \mu} = -n + \Sigma x_i/\mu$$

and

$$\frac{\partial^2 L}{\partial \mu^2} = -\Sigma \ x_i/\mu^2 \ \ (< 0)$$

so that L is maximised at

$$\hat{\mu} = \Sigma x_i/n.$$

Clearly $\mathbb{E}(X_i) = \mu = var(X_i)$. Thus $\mathbb{E}(\hat{\mu}) = \mu$

and $var(\hat{\mu}) = \mu/n$, and

$$-\mathbb{E}\frac{\partial^2 L}{\partial \mu^2} = n/\mu.$$

The exact distribution of $\hat{\mu}n = \Sigma X_i$ is $Po(n\mu)$.
The asymptotic distribution of $\hat{\mu}$, by the Central Limit Theorem, is $N(\mu, \mu/n)$.

2. Suppose we have n independent trials, and the outcome of each trial is
Red with probability θ_1,
or White with probability θ_2,
or Blue with probability θ_3,
where $\theta_1 + \theta_2 + \theta_3 = 1$.
Let (X, Y, Z) be the total number of (Red, White, Blue) trials in the sequence of n; write $X = \sum_1^n X_i$, $Y = \sum_1^n Y_i$ for suitably defined (X_i, Y_i).
Find $\mathbb{E}(X)$, $\text{var}\,(X)$, and show that

$$\text{cov}\,(X, Y) = -n\theta_1\theta_2.$$

Find $\binom{\hat{\theta}_1}{\hat{\theta}_2}$, and find the mean vector, and covariance matrix, of its asymptotic distribution (which is of course bivariate normal).

Respuesta

2. (This is the 3-cell multinomial distribution). With

$$f(x,y,z|\theta) = n!\frac{\theta_1^x\theta_2^y\theta_3^z}{x!y!z!}$$

for $x,y,z = 0,1,2...$, and $x+y+z = n$, we have $X = \Sigma X_i$ say, where X_i=1 if ith trial results in a Red, and X_i =0, otherwise, $Y = \Sigma Y_i$, and $Y_i = 1$ if ith trial results in a White, Y_i=0 otherwise.
Clearly, $P(X_i = 1) = \theta_1$, and X is Bi(n,θ_1)
so that var(X)= n$\theta_1(1-\theta_1)$, $\mathbb{E}(X) = n\theta_1$.
Further

$$cov(X,Y) = \Sigma cov(X_iY_i) = n(\mathbb{E}(X_1Y_1) - \mathbb{E}(X_1)\mathbb{E}(Y_1)).$$

Clearly, $\mathbb{E}(X_1Y_1) = 0$, so $cov(X,Y) = -n\theta_1\theta_2$.
Now

$$L = logf(x,y|\theta) = xlog(\theta_1) + ylog(\theta_2) + zlog(\theta_3) + \text{constant}$$

which is maximised subject to $\theta_1+\theta_2+\theta_3$ =1 (use a Lagrange multiplier) by $\hat\theta_1 = x/n, \hat\theta_2 = y/n, \hat\theta_3 = z/n$.
Hence $\mathbb{E}(\hat\theta_i) = \theta_i$ for $i = 1,2,3$. Now

$$\frac{\partial L(\theta)}{\partial\theta_1} = (x/\theta_1) - (z/\theta_3)$$

$$\frac{\partial L(\theta)}{\partial\theta_2} = (y/\theta_2) - (z/\theta_3).$$

Hence minus the matrix of 2nd derivatives of L is

$$\begin{pmatrix} x/\theta_1^2 + z/\theta_3^2 & z/\theta_3^2 \\ z/\theta_3^2 & y/\theta_2^2 + z/\theta_3^2 \end{pmatrix}$$

Substituting for $\mathbb{E}(x), \mathbb{E}(y), \mathbb{E}(z)$, we see that the expectation of the above matrix is

$$\begin{pmatrix} n(1-\theta_2)/\theta_1\theta_3 & n/\theta_3 \\ n/\theta_3 & n(1-\theta_1)/\theta_2\theta_3 \end{pmatrix}.$$

It now remains for you to check that the inverse of this 2×2 matrix is

$$\begin{pmatrix} \theta_1(1-\theta_1)/n & -\theta_1\theta_2/n \\ -\theta_1\theta_2/n & \theta_2(1-\theta_2)/n \end{pmatrix}.$$

This is what the general formula for the **asymptotic** covariance matrix gives us. In this case, it agrees exactly with the **exact** covariance matrix.

3. Suppose Y_i independent Poisson, mean μ_i, and our model is

$$H : \log(\mu_i) = \alpha + \beta x_i$$

where (x_i) are given.
Write down the log likelihood $\log f(y|\alpha, \beta)$ and hence find

(i) the sufficient statistics for (α, β);
(ii) equations for $(\hat{\alpha}, \hat{\beta})$, the maximum likelihood estimator (mle), and
(iii) an expression for

$$\max_{\beta=0} f(y|\alpha, \beta).$$

Respuesta

3.(i)
With $f(y_i|\mu_i)$ proportional to $e^{-\mu_i} {\mu_i}^{y_i}$
and $\mu_i = exp(\alpha + \beta x_i)$ we see that the likelihood for (α, β) is proportional to

$$[exp - \Sigma e^{\alpha + \beta x_i}] exp\ [\alpha\ t_1 + \beta\ t_2]$$

where t_1 is defined as Σy_i , and t_2 as $\Sigma x_i y_i$.
Hence, by the factorisation theorem, (t_1, t_2) are sufficient for (α, β).
The log likelihood is

$$L(\alpha, \beta) = -\Sigma e^{\alpha+\beta x_i} + \alpha\ t_1 + \beta\ t_2 + constant.$$

$ii)$Thus

$$\frac{\partial L}{\partial \alpha} = 0 \text{ for } t_1 = \Sigma e^{\alpha+\beta x_i}$$

$$\frac{\partial L}{\partial \beta} = 0 \text{ for } t_2 = \Sigma x_i e^{\alpha+\beta x_i}.$$

These are the equations for $(\hat{\alpha}, \hat{\beta})$. To verify that this is indeed **the maximum**, we should check that **(minus the matrix of 2nd derivatives)** is positive- definite at $(\hat{\alpha}, \hat{\beta})$.
The equations for $(\hat{\alpha}, \hat{\beta})$ do not have an explicit solution, but we could solve them iteratively to find $(\hat{\alpha}, \hat{\beta})$, and hence we could evaluate the maximum of L.
$iii)$ Now, if $\beta = 0, L(\alpha, \beta) = -\Sigma e^{\alpha} + \alpha\ t_1$. It is easily seen that this is maximised with respect to α by α^* say, where $\alpha^* = log(t_1/n)$.
We know, by Wilks' theorem,that to test $H_0 : \beta = 0$ against $H_1 : \beta$ arbitrary, we should refer

$$2[L(\hat{\alpha}, \hat{\beta}) - L(\alpha^*, 0)] \text{ to } {\chi^2}_1.$$