67

1

Alvaro J. Castro R. 11-Oct 2022

Biostats 730 Exam #1

Question 1

(a)

Bayes' rule refers to a relationship between conditional probabilities, that's given by the relationship:

$$(1) \quad P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

This means that the probability of A given B is proportional to the probability of B given A times the probability of A (and then divided by the probability of B) such that

$$(2) \quad P(A|B) \propto P(B|A)P(A)$$

This can be used to estimate conditional probabilities.

For our purposes, it can be used to estimate the posterior distribution of a random variable. To be more clear, we can say that the posterior probability distribution is proportional to the prior distribution times the likelihood, or, analogous to equation (2)

Posterior $\propto$ (Prior) (Likelihood)

⇒

In this case, the posterior distribution refers to the probability distribution (PDF) of a random variable (A in the notation above) given our data (B in the notation above). This can be estimated using the likelihood function of our data times the prior distribution, which is how we think our data is distributed given our prior knowledge. In this case, we are using both prior knowledge AND the likelihood function to estimate the PDF of our random variable. If we have good knowledge of the distribution of our variable, we can use an informative prior that ~~gives us more~~ is weighted with our current data to provide a posterior distribution. But, if we don't have much information, we can use a vague prior, so that the likelihood function carries almost all the weight in informing the posterior. In the most extreme case, the posterior would be essentially equal to the likelihood & would provide the same ~~information as a +~~ estimate as a traditional maximum likelihood estimation.

47

3

Going back to the initial comments, & to be clear, a prior density reflects our prior knowledge about how a random variable (μ in this example) is distributed (as a PDF). If we don't know anything about this, we can use a vague prior & the posterior density would depend almost entirely on the likelihood function. ✓

The posterior density ends up being something of a weighted average between the prior density & the likelihood. The weight given to each of these depends both on how informative our prior is, & how much data we have (& what it says). So, by using Bayesian inference we can combine prior knowledge with our data to get a better estimate of the distribution of a random variable, μ.

5

Finally, a quick note is that we can ignore the denominator since we consider it fixed & can rescale the PDF so that it's integral equals one (1)

$\Rightarrow$

(4)

(b) The effective sample size is ~~the~~ an estimate of the number of independent Monte Carlo samples that we would need to get the results we got. The issue is that we don't get independent samples, but rather use ~~a~~ what is known as a Markov Chain Monte Carlo approximation (MCMC). These Markov Chains explore the parameter space using information from the prior sample ~~observation~~ only. This means that every observation is ~~somewhat~~ correlated with the prior observation, leading to a phenomenon known as autocorrelation. This autocorrelation has to be factored into the estimation & means that you need more MCMC samples to explore the parameter space. ~~A~~ After all MCMC samples are collected, an estimate of the autocorrelation of samples is made, ~~to~~ & from this, ~~we can estimate~~ the computer estimates how many Monte Carlo samples that were drawn independently would have ~~the same sample size~~ given the same results as our MCMC.

Because ~~B~~ there is always autocorrelation ⇒

10 *2 = 20

5

in MCMC, ~~the~~ you need fewer ~~MC~~ independent MC draws to determine the distribution of your random variable. That is why S_{eff} is always smaller than the total MCMC samples drawn

So, to summarize:
Autocorrelation refers to how samples using MCMC depend on the prior observat. & so are not independent, & do not explore the parameter space as quickly as independent MC samples

MCMC - is the Markov-Chain MC algorith that explores the parameter space based on the prior observation

ESS refers to the number of independent MC samples that would give the same results as MCMC. This always less than MCMC because MCMC is autocorrelated & so explores the parameter space more slowly

5

$\Rightarrow$

$\boxed{7} * 2 = \boxed{14}$

6

Question 2

(a) Using `brm` & MCMC, I use the following equation to draw my samples

$y \sim 1 + \beta(a-30)$?

X

This gives me samples for my intercept $\alpha^{(s)}$ & the coefficient $\beta^{(s)}$

2 ~~based on my data for $i=1,2,\ldots n$, & my~~ likelihood function given above.

The ~~probability~~ posterior probability that $\alpha > 0 \Rightarrow P(\alpha > 0)$ is ~~given~~

approximated by the proportion of samples $\alpha^{(s)}$ that are greater than zero, which in R notation can be written as

see sol mean$(\alpha > 0)$, where every true value is given a 1 & false a zero so that the probability ~~is~~ $P(\alpha > 0)$ is estimated from $\alpha^{(s)}$

(b) To obtain the posterior predictive probability that $P(\tilde{y}_x > 0)$ where age $=25$

5 we first obtain our sampling ~~for $\alpha^{(s)}, \beta^{(s)}, \sigma^{2(s)}$~~ distribution as

see sol outlined above & obtain sample parameters $\alpha^{(s)}, \beta^{(s)}, \sigma^{2(s)}$ ~ ⟹

Then, we use those parameters to obtain a sampling distribution for individuals with age 25 given by

$$\tilde{y}_k \mid \tilde{\alpha}, \tilde{\sigma}^2, \tilde{\beta} \overset{iid}{\sim} N\left((\tilde{\alpha} - 5\tilde{\beta}), \tilde{\sigma}^2\right) \quad \checkmark$$

where I use my ~~estimates~~ for $\tilde{\alpha}^{(s)}$ & $\tilde{\sigma}^{2(s)}$ from my samples. In other words, I ~~[illegible]~~ obtain random samples ~~[illegible]~~ from a distribution with mean $(\alpha^{(s)} - 5\beta^{(s)})$ & variance $\sigma^{2(s)}$ such that

$$\tilde{y} = \text{rnorm}\left(\alpha^{(s)} - 5\beta^{(s)}, \sigma^{(s)}\right). \quad \checkmark$$

If I draw a couple thousand samples I can get the posterior distribution for $\tilde{y}$. Again, I just get the proportion of these samples that are greater than zero $\Rightarrow$ mean$(\tilde{y} > 0)$ ✓ & that is the post. pred. prob that the yet to be sampled individual w/ age 25 has a health outcome > 0.

$\Rightarrow$

Question 3

(a) $\hat{\sigma}_\alpha$ refers to the estimated standard deviation of the means around the mean of means. ~~the~~ In other words, it means how much variance there is in health outcomes between counties of residence $\Rightarrow$ that is the variance between groups. In our case, $\hat{\sigma}_\alpha = 1.59$, which is the standard deviation in health outcome units. If this is normally distributed, we can say that 95% of ~~health outcomes are between~~ the means of health outcomes in each county are between 0.44 ± 3.18 health units

10

(b)(i) For dataset B, the variance between counties will be MUCH, MUCH larger so that even if within counties the variance looks the same, & the final point estimate for α_j may be the same, the spread of data will be MUCH larger in B, & this spread will be due to between county variance.

5

9

(ii) There would be much more shrinkage in dataset A toward $\bar{y}$, because the variance $\hat{\sigma}_\alpha$ is much smaller & thus deviations from $\bar{y}$ are more extreme ✓ so that $\bar{y}$ has more weight in determining the pooled mean $\bar{y}_j^*$. On the other hand, for dataset B, since the variance is MUCH larger, the weight of $\bar{y}$ is less in determining the pooled mean $\bar{y}_j^*$

5

Question 4

(a)

double?

$$y_i \mid \alpha_{j[i]}, \sigma_y, \beta_{j[i]} \overset{iid}{\sim} N(\alpha_{j[i]} + \beta_{j[i]}(a_i - 30), \sigma^2)$$

✗

$$\alpha_j \mid \mu_\alpha, \sigma_\alpha, \beta_j \overset{iid}{\sim} N(\mu_\alpha + \beta_j(a_i - 30), \sigma^2)$$

8

✓ $$\beta_j \mid \theta_\beta, \sigma_\beta \sim N(\theta_\beta, \sigma_\beta^2)$$

(I will assume a normal distribution for β_j)

13

~~If $z_i = y_i - \beta(a_i - 30)$~~

~~$y_i \mid \alpha_{j[i]}, \sigma_y, \beta \sim N($~~

~~$\mu \mid y_i, \sigma, \delta, \sigma_\alpha$~~

$$\boxed{\begin{aligned} z_i &= y_i - \beta(a_i - 30) \\ \bar{z} &= \bar{y} - \bar{a}\beta + 30\beta \end{aligned}}$$

~~$\mu \mid y_i,$~~

$$\alpha_j \mid y_i, \mu_\alpha, \beta, \sigma_y, \sigma_\alpha \propto \alpha_j \mid z_i, \mu_\alpha, \sigma_y, \sigma_\alpha$$

$$\sim N\left(\frac{\frac{\mu_\alpha}{\sigma_\alpha^2} + \frac{n \cdot \bar{z}}{\sigma_y^2}}{\frac{1}{\sigma_\alpha^2} + \frac{n}{\sigma_y^2}}, \frac{1}{\frac{1}{\sigma_\alpha^2} + \frac{n}{\sigma_y^2}} \right)$$

5

$$\sim N\left(\frac{\frac{\mu_\alpha}{\sigma_\alpha^2} + \frac{n(\bar{y} - \bar{a}\beta + 30\beta)}{\sigma_y^2}}{\frac{1}{\sigma_\alpha^2} + \frac{n}{\sigma_y^2}}; \frac{\sigma_\alpha^2 \sigma_y^2}{(\sigma_y^2 + n\sigma_\alpha^2)} \right)$$

need y_j