STA 250: STATISTICS

Notes 9. Conjugate Analysis of Standard Models

Book chapters: 7.3

1 Conjugate prior family

Once we get a posterior pdf/pmf $\xi(\theta|x)$ by combining a model $X \sim f(x|\theta)$ with a prior pdf/pmf $\xi(\theta)$ on $\theta \in \Theta$, a report can be made by summarizing the posterior. It helps to have the posterior pdf/pmf in a recognizable form so that we can easily compute its mean, spread, quantiles etc (by hand or by using `R` functions). This is not guaranteed to happen in general. For example, the model $X \sim \mathsf{Binomial}(n, p)$ and the prior pdf $\xi(p) = e^p/(e-1)$, $p \in [0, 1]$ lead to the posterior $\xi(p|x) = \text{const} \times p^x(1-p)^{n-x}e^p$, $p \in [0, 1]$, with a constant term that is fairly difficult to compute, making it difficult to get summaries of this pdf.

However, for certain models, certain prior pdfs do lead to posterior pdfs that are analytically tractable. We already saw one in the female birth rate analysis, a uniform prior pdf for the binomial model gives a beta posterior pdf. In fact more is true for this model: any beta prior pdf leads to a beta posterior pdf! This phenomenon is called conjugacy. A formal definition is given below.

DEFINITION 1 (Conjugacy). A collection of pdfs (or pmfs) is called a conjugate prior family for a model $X \sim f(x|\theta), \theta \in \Theta$, if whenever a prior $\xi(\theta)$ is chosen from the collection, it leads to a posterior $\xi(\theta|x)$ that is also a member of the collection, for every observation $X = x$.

Conjugacy in itself is not a very useful property. For example the collection of all pdfs on Θ is surely conjugate to the model. It becomes useful when a small collection of pdfs exhibit conjugacy to a certain statistical model. By a small collection we usually mean a collection of pdfs/pmfs $\mathcal{G} = \{g(\theta|a) : a \in A\}$ indexed by a low-dimensional vector a. $\mathcal{G}$ is conjugate to a statistical model $X \sim f(x|\theta)$ if $\xi(\theta) = g(\theta|a)$ for some $a \in A$ means for every x, $\xi(\theta|x) = g(\theta|\tilde{a})$ for some $\tilde{a} \in A$. As mentioned before, the collection of beta pdfs $\{\mathsf{Beta}(a, b) : a > 0, b > 0\}$ is a (2-dimensional) conjugate family to the binomial model $X \sim \mathsf{Binomial}(n, p)$, $p \in [0, 1]$. Table 1 below gives a list of other common models with known, low-dimensional conjugate families. We will establish conjugacy for three of the listed models; you're required to do the maths for the some of the remaining ones in HW5.

2 The binomial-beta conjugacy

The pdf of $\mathsf{Beta}(a, b)$ distribution, for $a > 0, b > 0$, equals $g(p) = p^{a-1}(1-p)^{b-1}/B(a, b)$, $p \in [0, 1]$, where $B(a, b) = \int_0^1 q^{a-1}(1-q)^{b-1}dq$ is known as the Beta function. If we take

Model	Parameter	Prior	Posterior
$X \sim \mathsf{Binomial}(n,p)$	$0 \le p \le 1$	$\mathsf{Beta}(a,b)$ $a>0, b>0$	$\mathsf{Beta}(\tilde{a},\tilde{b})$ $\tilde{a} = a + x$ $\tilde{b} = b + n - x$
$X = (X_1, \cdots, X_n)$ $X_i \overset{\text{IID}}{\sim} \mathsf{Poisson}(\lambda)$	$\lambda > 0$	$\mathsf{Gamma}(a,b)$ $a>0, b>0$	$\mathsf{Gamma}(\tilde{a},\tilde{b})$ $\tilde{a} = a + n\bar{x}$ $\tilde{b} = b + n$
$X = (X_1, \cdots, X_n)$ $X_i \overset{\text{IID}}{\sim} \mathsf{Exponential}(\lambda)$	$\lambda > 0$	$\mathsf{Gamma}(a,b)$ $a>0, b>0$	$\mathsf{Gamma}(\tilde{a},\tilde{b})$ $\tilde{a} = a + n$ $\tilde{b} = b + n\bar{x}$
$X = (X_1, \cdots, X_n)$ $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu,\sigma^2)$ σ^2 known	$-\infty < \mu < \infty$	$\mathsf{Normal}(a,b^2)$ $-\infty < a < \infty$ $b > 0$	$\mathsf{Normal}(\tilde{a},\tilde{b}^2)$ $\tilde{a} = \frac{nb^2\bar{x}+\sigma^2 a}{nb^2+\sigma^2}$ $\tilde{b}^2 = \frac{\sigma^2 b^2}{nb^2+\sigma^2}$
$X = (X_1, \cdots, X_n)$ $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu,\sigma^2)$	$-\infty < \mu < \infty$ $\sigma^2 > 0$	$\mathsf{N}\chi^{-2}(m,k,r,s^2)$ $-\infty < m < \infty$ $k>0, r>0, s>0$	$\mathsf{N}\chi^{-2}(\tilde{m},\tilde{k},\tilde{r},\tilde{s}^2)$ $\tilde{m} = \frac{km+n\bar{x}}{k+n}$ $\tilde{k} = k+n$ $\tilde{r} = r+n$ $\tilde{s}^2 = \frac{rs^2+\frac{kn}{k+n}(\bar{x}-m)^2+(n-1)s_x^2}{r+n}$

Table 1: Conjugate prior and posterior for some common models.

$\xi(p) = \mathsf{Beta}(a,b)$ (for some $a > 0, b > 0$) as the prior pdf for a binomial model $X \sim \mathsf{Binomial}(n,p)$, $p \in [0,1]$, then for any observations $x \in \{0, 1, \cdots, n\}$,

$$\begin{aligned}\xi(p|x) &= \text{const} \times f(x|p)\xi(p) = \text{const} \times p^x(1-p)^{n-x}p^{a-1}(1-p)^{b-1} \\ &= \text{const} \times p^{a+x-1}(1-p)^{b+n-x-1}.\end{aligned}$$

But the pdf of $\mathsf{Beta}(a+x, b+n-x)$ is $p^{a+x-1}(1-p)^{b+n-x-1}/B(a+x,b+n-x)$. Therefore $\xi(p|x)$ is a constant multiple of the $\mathsf{Beta}(a+x,b+n-x)$ pdf. But if two pdfs are constant multiples of each other, they must be identical (and the constant must be 1). So $\xi(p|x) = \mathsf{Beta}(a+x, b+n-x)$.

3 The normal-normal conjugacy

Next we show that for the model $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $\mu \in (-\infty, \infty)$, σ fixed, the prior pdf $\xi(\mu) = \mathsf{Normal}(a, b^2)$ gives a posterior pdf $\xi(\mu|x) = \mathsf{Normal}(\tilde{a}, \tilde{b}^2)$ for some $\tilde{a}$ and $\tilde{b}$ [which we shall identify]. It suffices to show that $\xi(\mu|x)$ is a constant multiple of the $\mathsf{Normal}(\tilde{a}, \tilde{b}^2)$ density. This is equivalent to showing

$$\log \xi(\mu|x) = \text{const} + \frac{(\mu - \tilde{a})^2}{2\tilde{b}^2}$$

by working on the log-scale. Earlier we worked out that $\ell_x(\mu) = n(\bar{x}-\mu)^2/(2\sigma^2)$. Therefore,

$$\begin{aligned}\log \xi(\mu|x) &= \text{const} + \ell_x(\mu) + \log \xi(\mu) \\ &= \text{const} - \frac{1}{2}\left[\frac{n(\bar{x}-\mu)^2}{\sigma^2} + \frac{(\mu-a)^2}{b^2}\right] \\ &= \text{const} - \frac{1}{2}\frac{(\mu - \frac{nb^2\bar{x}+\sigma^2 a}{nb^2+\sigma^2})^2}{\frac{b^2\sigma^2}{nb^2+\sigma^2}}\end{aligned}$$

and therefore, $\xi(\mu|x) = \mathsf{Normal}(\frac{nb^2\bar{x}+\sigma^2 a}{nb^2+\sigma^2}, \frac{b^2\sigma^2}{nb^2+\sigma^2})$. The last equality above follows from a "completion of squares" identity (give it a try!):

$$\frac{n(\bar{x}-\mu)^2}{\sigma^2} + \frac{(\mu-a)^2}{b^2} = \frac{(nb^2+\sigma^2)(\mu - \frac{nb^2\bar{x}+\sigma^2 a}{nb^2+\sigma^2})^2}{b^2\sigma^2} + \frac{n(\bar{x}-a)^2}{nb^2+\sigma^2}.$$

4 A conjugate family for the full normal model

For the full normal model $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $(\mu, \sigma^2) \in (-\infty, \infty) \times (0, \infty)$, we need a bivariate prior pdf $\xi(\mu, \sigma^2)$ on $(-\infty, \infty) \times (0, \infty)$. There are several choices here. For example we could take $\xi(\mu, \sigma^2) = g(\mu)h(\sigma^2)$ where $g(\mu)$ is a pdf on $(-\infty, \infty)$ and $h(\sigma^2)$ is a pdf on $(0, \infty)$. This is in fact widely used, with $g(\mu)$ usually taken to be a normal pdf and $h(\sigma^2)$ taken to be an inverse-gamma pdf (i.e., the prior pdf of $1/\sigma^2$ is a gamma pdf). However, the family of such pdfs are not conjugate to the model. In particular, the posterior pdf $\xi(\mu, \sigma^2|x)$ does not factor into a product $g(\mu|x)h(\sigma^2|x)$.

There is however, a conjugate family of pdfs, known as the normal-inverse-chi-square pdfs, denoted $\mathsf{N}\chi^{-2}(m, k, r, s^2)$ with parameters $m \in (-\infty, \infty)$, $k > 0$, $r > 0$ and $s > 0$. The pdf of this distribution equals

$$g(w, v) = \text{const.} \times v^{-\frac{r+3}{2}} \exp\left(-\frac{k(w-m)^2 + rs^2}{2v}\right), \quad (w, v) \in (-\infty, \infty) \times (0, \infty)$$

where the constant equals $\frac{(rs^2/2)^{r/2}}{\sqrt{2\pi}\Gamma(r/2)}$. Suppose for the normal model, we take $\xi(\mu, \sigma^2) = \mathsf{N}\chi^{-2}(m, k, r, s^2)$ for some valid choices of m, k, r and s. We will again work in the log-scale: $\log \xi(\mu, \sigma^2) = \text{const} + \ell_x(\mu, \sigma^2) + \log \xi(\mu, \sigma^2)$. We have worked out before that

$$\ell_x(\mu, \sigma^2) = \text{const} - \frac{n}{2}\log \sigma^2 - \frac{(n-1)s_x^2 + n(\bar{x}-\mu)^2}{2\sigma^2}.$$

Therefore,

$$\begin{aligned}\log \xi(\mu, \sigma^2|x) &= \ell_x(\mu, \sigma^2) + \log \xi(\mu, \sigma^2) \\ &= \text{const} - \frac{n+r+3}{2}\log \sigma^2 - \frac{(n-1)s_x^2 + n(\bar{x}-\mu)^2 + k(\mu-m)^2 + rs^2}{2\sigma^2}.\end{aligned}$$

There is another square-completion identity that we need (try this):

$$n(\bar{x}-\mu)^2 + k(\mu-m)^2 = (k+n)\left(\mu - \frac{km+n\bar{x}}{k+n}\right)^2 + \frac{kn}{k+n}(\bar{x}-m)^2.$$

Plugging this in we get

$$\log \xi(\mu, \sigma^2 | x) = \text{const} - \frac{\tilde{r}+3}{2}\log\sigma^2 - \frac{\tilde{k}(\mu-\tilde{m})^2 + \tilde{r}\tilde{s}^2}{2\sigma^2}$$

and hence $\xi(\mu, \sigma^2|x) = \mathsf{N}\chi^{-2}(\tilde{m}, \tilde{k}, \tilde{r}, \tilde{s}^2)$, where

- $\tilde{m} = \frac{km+n\bar{x}}{k+n}$
- $\tilde{k} = n + k$
- $\tilde{r} = n + r$
- $\tilde{s}^2 = \frac{1}{n+r}(rs^2 + \frac{kn}{k+n}(\bar{x}-m)^2 + (n-1)s_x^2)$.

5 Computing with normal-inverse-chi-square

The pdf formula of $\mathsf{N}\chi^{-2}(m, k, r, s^2)$ is useful to establish conjugacy, but does not help much in computing prior or posterior summaries for (μ, σ^2). Here are some useful results that will help us with computing:

RESULT 1. A pair of random variables (W, V) has a $\mathsf{N}\chi^{-2}(m, k, r, s^2)$ pdf if and only if

1. $\frac{rs^2}{V} \sim \chi^2(r) = \mathsf{Gamma}(r/2, 1/2)$, and
2. $[W|V = v] \sim \mathsf{Normal}(m, v/k)$.

RESULT 2. If $(W, V) \sim \mathsf{N}\chi^{-2}(m, k, r, s^2)$ then $\sqrt{k}(W - m)/s \sim t(r)$.

Example (NSW)**.** For the NSW study we have $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$ and suppose we assign $\xi(\mu, \sigma^2) = \mathsf{N}\chi^{-2}(0, 4, 2, 9000^2)$. Results 1 says: under the prior $(2 \cdot 9000^2)/\sigma^2 \sim \mathsf{Gamma}(1, 1/2)$. Hence a 95% prior range for $(2 \cdot 9000^2)/\sigma^2$ is

$$[\texttt{qgamma(.025, 1, 1/2)}, \texttt{qgamma(0.975, 1, 1/2)}] = [0.051, 7.378]$$

and so a 95% prior range for σ is:

$$\left[\sqrt{\frac{2\cdot 9000^2}{7.378}}, \sqrt{\frac{2\cdot 9000^2}{0.051}}\right] = [4685.85, 56360.19]$$

Also the prior median of σ is $\sqrt{2 \cdot 9000^2/\texttt{qgamma(0.5, 1, 1/2)}} = \sqrt{2\cdot 9000^2/1.386} = 10811.25$. Result 2 says: under the prior $\sqrt{4}(\mu - 0)/9000 \sim t(2)$ and hence a 95% prior range for μ is $0 \mp \frac{9000}{\sqrt{4}} z_2(0.05) = [-19361.94, 19361.94]$.

The recorded data shows $n = 185$, $\bar{x} = 4253.57$ and $s_x = 8926.985$. So

$$\tilde{m} = \frac{4 \cdot 0 + 185 \cdot 4253.57}{4 + 185} = 4163.547, \quad \tilde{k} = 189, \quad \tilde{r} = 187,$$

$$\tilde{s}^2 = = \frac{1}{187} \left\{ 2 \cdot 2000^2 + \frac{4 \cdot 185}{4 + 185}(4253.57 - 0)^2 + 184 \cdot 8926.985^2 \right\} = 8878.862^2,$$

and hence the posterior pdf of (μ, σ^2) is $\mathsf{N}\chi^{-2}(4163.547, 189, 187, 8878.862^2)$. So to calculate a 95% posterior range for σ, we first get a 95% range for $\tilde{r}\tilde{s}^2/\sigma^2 \sim \mathsf{Gamma}(187/2, 1/2)$ which is $[151.024, 226.761]$ and then transform it to an interval for σ:

$$\left[\sqrt{\frac{187 \cdot 8878.862^2}{226.761}}, \sqrt{\frac{187 \cdot 8878.862^2}{151.024}} \right] = [8062.73, 9879.69].$$

Also, a posterior 95% range for μ is:

$$4163.547 \mp \frac{8878.862}{\sqrt{189}} z_{187}(0.05) = [2884.452, 5442.642].$$

If we were interested in assessing whether μ was larger than 2000, we would calculate:

$$\begin{aligned} P(\mu > 2000 | X = x) &= P\left(\frac{\sqrt{\tilde{k}}(\mu - \tilde{m})}{\tilde{s}} > \frac{\sqrt{\tilde{k}}(2000 - \tilde{m})}{\tilde{s}} \right) \\ &= 1 - \Phi_{\tilde{r}}\left(\frac{\sqrt{\tilde{k}}(2000 - \tilde{m})}{\tilde{s}} \right) \\ &= 1 - \Phi_{187}\left(\frac{\sqrt{189} \cdot (2000 - 4163.547)}{8878.862} \right) \\ &= 1 - \Phi_{187}(-3.35) = 0.9995 \end{aligned}$$