STA 114: Statistics

Notes 10. Conjugate Priors

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. 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 (recall Lab 4).

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, $\xi(\theta|x) = g(\theta|a')$ for some $a' \in A$. 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 **??** 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 remaining ones in HW5.

The binomial-beta conjugacy

The pdf of $bet(a,b)$ 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 $\xi(p) = \mathsf{Beta}(a,b)$ (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 $x \in \{0,1,\cdots,n\}$,

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

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

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

But the pdf of $\mathsf{Beta}(a+x, b+n-x)$ (note: $a+x>0$, $b+n-x>0$) 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 $\xi(p|x) = \mathsf{Beta}(a+x, b+n-x)$.

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 $\xi(\mu|x) = \mathsf{Normal}(a', b'^2)$ for some a' and b' [which we shall identify]. It suffices to show that $\xi(\mu|x)$ is a constant multiple of the $\mathsf{Normal}(a', b'^2)$ density. This is equivalent to showing

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

by going in the log-scale. Now, by definition

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

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!):

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

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, denotes $\mathsf{N}\chi^{-2}(m, k, r, s)$ with parameters $m \in (-\infty, \infty)$, $k > 0$, $r > 0$ and $s > 0$. The pdf of this distributionis given by:

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

where the constant equals $\frac{(rs/2)^{r/2}}{\sqrt{2\pi}\Gamma(r/2)}$. Here are two important results.

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

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

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

The first result can be proved by direct calculations (you're welcome to try it). The second result follows once you note that if $(W, V) \sim \mathsf{N}\chi^{-2}(m, k, r, s)$ then by Result 1, we can write $W = m + \sqrt{V/k}Z$ where $Z \sim \mathsf{Normal}(0, 1)$ is independent of V. Also $U = \frac{rs}{V} \sim \chi^2(r)$. Then,

$$\frac{W - m}{\sqrt{s/k}} = \frac{\sqrt{V/k}Z}{\sqrt{s/k}} = \frac{Z}{\sqrt{U/r}} \sim t(r).$$

With this as the prelude, we're now ready to prove that for $\xi(\mu, \sigma^2) = \mathsf{N}\chi^{-2}(m, k, r, s)$ we get $\xi(\mu, \sigma^2|x) = \mathsf{N}\chi^{-2}(m', k'(s), r', s')$ where

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

Working in the log-scale we have

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

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

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

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

Putting it to use

It is nice to have low-dimensional conjugate prior family for a statistical model one is interested in. But we still have the job of deciding upon one member of the family to ultimately use as the prior pdf/pmf. If a singular choice is hard to justify, we can run the analysis for a multitude of reasonable choices and present the analysis side by side. This can be done both graphically, and more important via tables.

To summarize a prior or a posterior pdf, we can report a number representing the center of the pdf, such as the mean or the median. To summarize the spread of the pdf, we can

Prior	Prior summary		Posterior	Posterior summary	
	Mean	$[\mu_{.025}, \mu_{.975}]$		Mean	$[\mu_{.025}(x), \mu_{.975}(x)]$
$\mathsf{Gamma}(2, .2)$	10	(1.21, 27.86)	$\mathsf{Gamma}(152, 10.2)$	14.9	(12.63, 17.36)
$\mathsf{Gamma}(5, .5)$	10	(3.25, 20.48)	$\mathsf{Gamma}(155, 10.5)$	14.76	(12.53, 17.17)
$\mathsf{Gamma}(50, 5)$	10	(7.42, 12.96)	$\mathsf{Gamma}(200, 15)$	13.33	(11.55, 15.24)
$\mathsf{Gamma}(500, 50)$	10	(9.14, 10.9)	$\mathsf{Gamma}(650, 60)$	10.83	(10.02, 11.68)

Table 2: Poisson-gamma example, prior and posterior summaries

report an interval $[a, b]$ such that the pdf packs most of its area inside the interval. If we want an $1 - \alpha$ area inside, for some small α, then we can take $a =$ the $\alpha/2$-th quantile of the pdf and $b =$ the $(1 - \alpha/2)$-th quantile of it. For a prior $\xi(\theta)$, we'll denote its u-th quantile by θ_u and for a posterior $\xi(\theta|x)$ we will use the notation $\theta_u(x)$. We will usually work with $\alpha = 5\%$ and hence report $[\theta_{.025}, \theta_{.975}]$ from the prior and $[\theta_{.025}(x), \theta_{.975}(x)]$ from the posterior.

Example. Consider hurricane counts $X_1, \cdots, X_n$ for $n = 10$ consecutive years in the north Atlantic basin, modeled as $X_i \overset{\text{IID}}{\sim} \mathsf{Poisson}(\mu)$. Figure **??** and Table **??** show prior-posterior summaries for 4 choices of the prior pdf from the conjugate family of gamma pdfs. The observed data are $x = (12, 14, 15, 12, 16, 14, 27, 10, 14, 16)$.

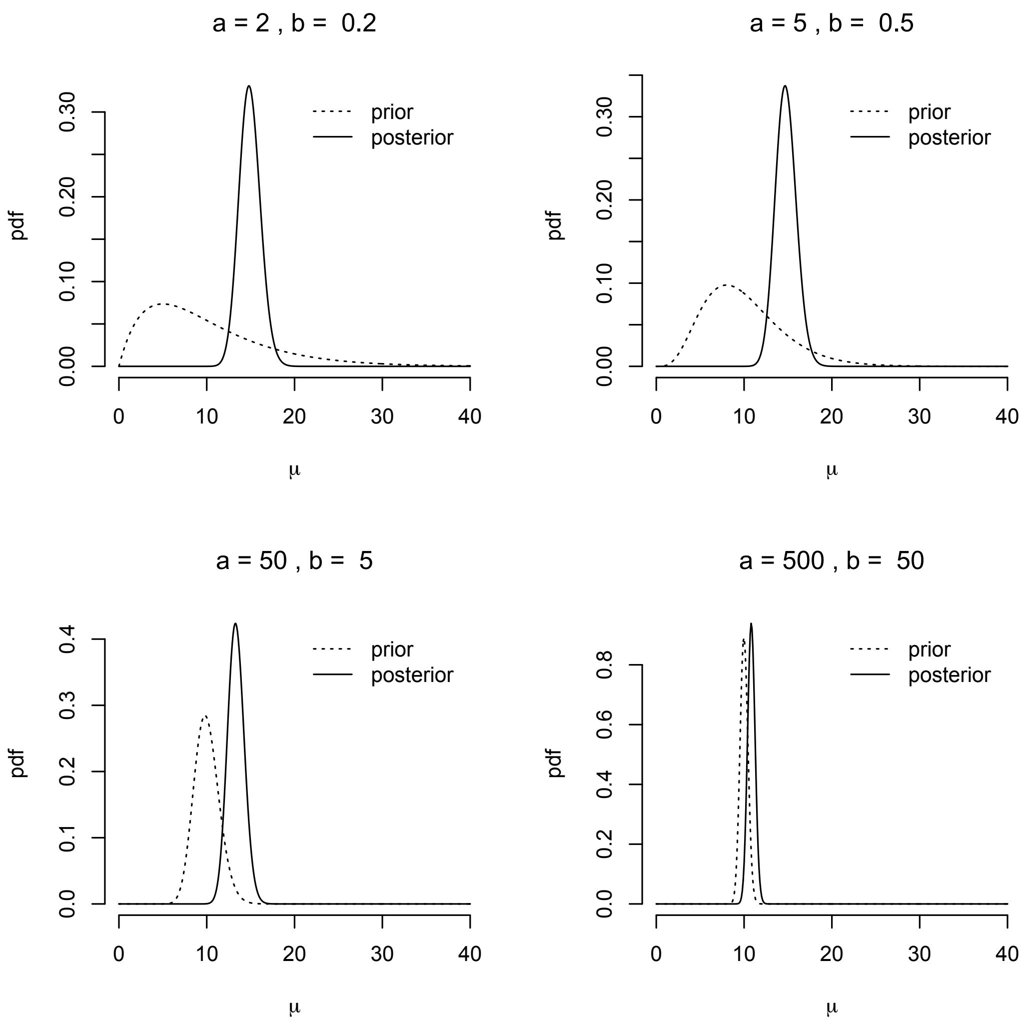

Figure 1: Prior-posterior summary for hurricane count data with $\mathsf{Gamma}(a, b)$ conjugate priors.