PUBH 7440: Intro to Bayesian Analysis
Homework from Week 2 — Due Feb 6

For the unknown parameters in each of the following scenarios, *find/show* that something is a conjugate prior for the unknown parameter in each model (i.e., write $p(\theta \mid Y) \propto p(Y \mid \theta) \times p(\theta)$, show a couple of steps, then show the posterior). If you can't find a conjugate prior, explain what you might look for in a prior (e.g., τ^2 must be greater than 0, so a gamma distribution might be appropriate; $\theta \in (0,1)$, so a beta distribution might be appropriate; etc.). For each prior, provide an interpretation of the prior parameters. Vaguely describe what might make for a relatively noninformative prior.

1. $x \sim \text{Bin}(n, \theta)$, n known
2. $x \sim \text{NegBin}(r, \theta)$, r known
3. $\mathbf{x} \sim \text{Mult}(n, \boldsymbol{\theta})$, n known
4. $x \sim \text{Gam}(\alpha, \beta)$, α known
5. $x \sim \text{Gam}(\alpha, \beta)$, β known
6. **OPTIONAL**: $\mathbf{x} \sim \text{Norm}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}$ known
7. **OPTIONAL**: $\mathbf{x} \sim \text{Norm}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\mu}$ known (this one is fun)

Hint: You can use the table on the next page to refresh your memory what different mass/density functions look like.

GLORIOUS (YET INCOMPLETE) TABLE OF DENSITY/MASS FUNCTIONS

Distribution	Density/Mass Function
$x \sim \text{Bern}(\pi)$	$p(x \mid \pi) = \pi^x (1-\pi)^{1-x}$
$x \sim \text{Geometric}(\pi)$	$p(x \mid \pi) = (1-\pi)^{x-1} \pi$
$x \sim \text{Binomial}(n, \pi)$	$p(x \mid \pi) = \binom{n}{x} \pi^x (1-\pi)^{n-x}$
$x \sim \text{NegBin}(r, \pi)$	$p(x \mid \pi) = \binom{x+r-1}{x} (1-\pi)^{x} \pi^{r}$
$x \sim \text{Pois}(\lambda)$	$p(x \mid \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$
$x \sim \text{Laplace}(\mu, b)$	$p(x \mid \mu, b) = \frac{1}{2b} e^{-\lvert x-\mu \rvert / b}$
$x \sim \text{Norm}(\mu, \sigma^2)$	$p(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
$x \sim \text{Gam}(\alpha, \beta)$	$p(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$
$x \sim \text{Exp}(\lambda)$	$p(x \mid \lambda) = \lambda e^{-\lambda x}$
$x \sim \text{IG}(a, b)$	$p(x \mid a, b) = \frac{b^a}{\Gamma(a)} x^{-a-1} e^{-b/x}$
$\mathbf{x} \sim \text{Mult}(n, \boldsymbol{\theta})$	$p(\mathbf{x} \mid n, \boldsymbol{\theta}) = \frac{n!}{\prod_i x_i!} \prod_i \theta_i^{x_i}$
$\mathbf{x} \sim \text{Dir}(\boldsymbol{\alpha})$	$p(\mathbf{x} \mid \boldsymbol{\alpha}) \propto \prod_i x_i^{\alpha_i - 1}$
$\mathbf{x} \sim \text{Norm}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \lvert\boldsymbol{\Sigma}\rvert^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right]$
$\mathbf{X} \sim \text{IW}(\nu, \mathbf{G})$	$p(\mathbf{X} \mid \nu, \mathbf{G}) \propto \lvert\mathbf{X}\rvert^{-(\nu+p+1)/2} \exp\left[-\frac{1}{2}\text{tr}(\mathbf{G}\mathbf{X}^{-1})\right]$

Table 1: Table of density/mass functions for selected distributions. Note that these parameterizations are consistent with those on Wikipedia but may differ from those used in Casella & Berger and other textbooks. I made a table like this for the Biostat PhD exam, but I figured since you might not have a textbook for referencing, it might be good to provide you some guidance here...

1 $X \sim Bin(n, \theta)$

$$p(X|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

since θ represents probability of one succesful Bernoulli outcome, then it makes sense to use Beta as a prior for θ.

- Consider $\theta \sim Beta(\alpha, \beta)$ and

$$f(\theta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \cdot \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1}$$

- $f(\theta/x) \propto f(\theta) \cdot p(x/\theta) =$

$$= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \times \binom{n}{x} \times \theta^{\alpha-1} (1-\theta)^{\beta-1} \times$$

$$\times \theta^x \times (1-\theta)^{n-x}$$

Drop $\Gamma(\alpha), \Gamma(\beta)$ and Binomial coefficients since none of them depend on θ

$$\propto \theta^{\alpha+x-1} \times (1-\theta)^{n+\beta-x-1}$$

one have a kernel of a Beta Distribution with parameters $\alpha+x-1$ and $n+\beta-x-1$

- So, prior of θ is $Beta(\alpha, \beta)$ and posterior of θ/x is $Beta(\alpha+x-1, n+\beta-x-1)$

- Interpretation of parameters:

  for a Beta distribution mean is $\frac{\alpha}{\alpha+\beta}$.

  So, let α be the number of successful outcomes,
  β be the number of unsuccessful outcomes.

  Then, such model is good for modeling ratios of desired outcomes in binomial setting.

- Vaguely non-informative prior:
  small numbers for α and β, both >0.
  This way a small number of "trials" helps us form a proportion θ, and we can update θ with data

2 $X \sim \text{NegBin}(r, \theta)$

$$p(x|\theta) = \binom{x+r-1}{x} \theta^r (1-\theta)^x$$

r is known # successes, x is random number of unsuccessful outcomes.

i) again, let $\theta \sim \text{Beta}(\alpha, \beta)$.

$$f(\theta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

2) $f(\theta|x) \propto f(\theta) \cdot p(x|\theta) =$

$$= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \times \binom{x+r-1}{x} \times \theta^{r} \times \theta^{\alpha-1} \times (1-\theta)^{x} \times (1-\theta)^{\beta-1}$$

$$\propto \theta^{r+\alpha-1} \times (1-\theta)^{x+\beta-1}$$

3) Therefore, $\theta | X \sim \text{Beta}(r+\alpha,\ x+\beta)$.

4) Same interpretation of α and β, as well as non-informative parameters from 1 apply here as well.

we essentially have the same model for X.

3 $\vec{x} \sim \text{Mult}(n, \vec{\theta})$

$$p(\vec{x}|\vec{\theta}) = \frac{n!}{\prod x_i!} \cdot \prod \theta_i^{x_i}$$

1) Since $\vec{\theta}$ is a vector of values θ_i, we use Dirichlet multivariate distribution, where each α_i is a parameter in a prior that corresponds to θ_i

2) let $\vec{\theta} \sim \text{Dir}(\vec{\alpha})$

$$p(\vec{\theta}|\vec{\alpha}) \propto \prod \theta_i^{\alpha_i - 1}$$

3) $p(\vec{\theta}/x) = p(\vec{\theta}) \times p(\vec{x}/\vec{\theta})$

$$\propto \frac{n!}{\prod x_i!} \times \prod \theta_i^{x_i} \times \prod \theta_i^{\alpha_i - 1}$$

$$\propto \prod \theta_i^{x_i + \alpha_i - 1}$$

So, $\vec{\theta}/\vec{x} \sim \text{Dir}(\vec{\alpha} + \vec{x})$

a weakly informative prior is the one where $\vec{\alpha}$ values are relatively low compared to $\vec{x}$, such as $\mathcal{E} = 1$ or something similar

4) $X \sim \text{Gamma}(\alpha, \beta)$

$$f(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad ; \alpha, \beta > 0$$

α is known, so we need to worry about β distribution only.

1) Assume $\beta \sim \text{Gamma}(\alpha_0, \beta_0)$, then

$$f(\beta|x) \propto f(x|\alpha,\beta) \times f(\beta|\alpha_0,\beta_0) =$$

$$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \times \beta^{\alpha_0-1} e^{-\beta_0 \beta}$$

collect terms that contain β

$$\propto \beta^{\alpha+\alpha_0-1} e^{-\beta x - \beta_0 \beta}$$

$$= \beta^{\alpha+\alpha_0-1} e^{-\beta(x+\beta_0)}$$

2) we still have kernel of a gamma distribution.

$\alpha, \alpha_0 > 0$ by assumption that it was a parameter for a Gamma

$X > 0$ because it was sampled from a Gamma distribution

3) So, $\beta \sim \text{Gamma}(\alpha + \alpha_0, X + \beta_0)$

Mean of a Gamma distribution is

$\frac{\alpha}{\beta}$, so we have $\frac{\alpha + \alpha_0}{\beta_0 + x}$

4) we commonly use Gamma distribution when working with rates.

let α be some number of interest in the population, and β be the size of this population.

5) So, population size itself depends on some parameters α_0 and β_0

non-informative prior would be one w/ low values of α_0 and β_0 such that the "size" of population contributing to prior knowledge is small, and affects posterior estimates less as we collect data from more population

⑤ $X \sim \text{Gamma}(\alpha, \beta)$ β known.

1) Again, let $\alpha \sim \text{Gamma}(\alpha_0, \beta_0)$

Then

$$f(\alpha \mid x) \propto \frac{\beta^{\alpha}}{\Gamma(\alpha)} \underline{x^{\alpha-1}}\ \underline{e^{-\beta x}} \times$$

$$\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \underline{\alpha^{\alpha_0-1}}\ \underline{e^{-\beta_0 \alpha}}$$

$$\propto \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \alpha^{\alpha_0-1} e^{-\beta x-\beta_0 \alpha}$$

There is no further way to reduce the equation to the kernel such that we obtain a recognizeable density.

I do not think there is a conjugate prior in this situation