Biostatistics 682: Applied Bayesian Inference
Lecture 3: More on single parameter models

Jian Kang

Department of Biostatistics
University of Michigan, Ann Arbor

Poisson Distribution

- Some measurements, such as a person's number of children or number of friends, have values that are whole numbers. Our sample space is $\{0, 1, 2, \ldots, \}$ on which the simplest probability model is the Poisson model.
- A random variable X has a Poisson distribution with mean θ if

$$\Pr(X = k \mid \theta) = \frac{\theta^k}{k!} \exp(-\theta), \quad \text{for } k \in \{0, 1, 2, \ldots, \}$$

- People sometimes say that the Poisson family of distributions has a "mean-variance relationship" because if one Poisson distribution has a larger mean than another, it will have a larger variance as well.

Posterior Inference

- Suppose there are n i.i.d. Poisson observation $y = (y_1, \ldots, y_n)$ with mean θ, then the joint probability mass function of our sample data is as follows:

$$\pi(y \mid \theta) \propto \exp\{a(y)\theta + b(y)\log(\theta)\}$$

$$a(y) = -n$$

$$b(y) = \sum_{i=1}^{n} y_i$$

- What is the natural conjugate prior?

Posterior Inference

- Suppose there are n i.i.d. Poisson observation $y = (y_1, \ldots, y_n)$ with mean θ, then the joint probability mass function of our sample data is as follows:

$$\pi(y \mid \theta) \propto \exp\{a(y)\theta + b(y)\log(\theta)\}$$

$$a(y) = -n$$

$$b(y) = \sum_{i=1}^{n} y_i$$

- What is the natural conjugate prior?
  Gamma distribution

Gamma Distribution

- A random variable θ follows a Gamma distribution with shape a and rate b if its probability density function is given by

$$\pi(\theta \mid a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta),$$

where

$$E(\theta) = \frac{a}{b},$$

$$\mathrm{Var}(\theta) = \frac{a}{b^2},$$

$$\mathrm{mode}(\theta) = \frac{a-1}{b} I(a > 1).$$

Posterior Distribution

- If $y_i \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$ for $i = 1, \ldots, n$, and $\theta \sim G(\alpha, \beta)$, then the posterior distribution is

$$\theta \mid y \sim G\left(\alpha + n\bar{y}, \beta + n\right),$$

  where $y = (y_1, \ldots, y_n)$ and $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$.

- What are the posterior mean, variance and mode?

Posterior Distribution

- If $y_i \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$ for $i = 1, \ldots, n$, and $\theta \sim G(\alpha, \beta)$, then the posterior distribution is
$$\theta \mid y \sim G\left(\alpha + n\bar{y}, \beta + n\right),$$
where $y = (y_1, \ldots, y_n)$ and $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$.
- What are the posterior mean, variance and mode?

$$E(\theta \mid y) = (\alpha + n\bar{y})(\beta + n)^{-1}$$

$$\text{Var}(\theta \mid y) = (\alpha + n\bar{y})(\beta + n)^{-2}$$

$$\text{Mode}(\theta \mid y) = (\alpha - n\bar{y} - 1)(\beta + n)^{-1}$$

Posterior Distribution

- If $y_i \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$ for $i = 1, \ldots, n$, and $\theta \sim G(\alpha, \beta)$, then the posterior distribution is
$$\theta \mid y \sim G\left(\alpha + n\bar{y}, \beta + n\right),$$
where $y = (y_1, \ldots, y_n)$ and $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$.
- What are the posterior mean, variance and mode?
$$E(\theta \mid y) = (\alpha + n\bar{y})(\beta + n)^{-1}$$
$$\text{Var}(\theta \mid y) = (\alpha + n\bar{y})(\beta + n)^{-2}$$
$$\text{Mode}(\theta \mid y) = (\alpha - n\bar{y} - 1)(\beta + n)^{-1}$$
- What is the relationship between posterior mean and prior mean?

Posterior Distribution

- If $y_i \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$ for $i = 1, \ldots, n$, and $\theta \sim G(\alpha, \beta)$, then the posterior distribution is

$$\theta \mid y \sim G\left(\alpha + n\bar{y}, \beta + n\right),$$

  where $y = (y_1, \ldots, y_n)$ and $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$.

- What are the posterior mean, variance and mode?

$$E(\theta \mid y) = (\alpha + n\bar{y})(\beta + n)^{-1}$$

$$\text{Var}(\theta \mid y) = (\alpha + n\bar{y})(\beta + n)^{-2}$$

$$\text{Mode}(\theta \mid y) = (\alpha - n\bar{y} - 1)(\beta + n)^{-1}$$

- What is the relationship between posterior mean and prior mean?

$$E(\theta \mid y) = wE(\theta) + (1 - w)\bar{y}$$

  where $w = \frac{\beta}{\beta + n}$

Posterior Distribution

- If $y_i \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$ for $i = 1, \ldots, n$, and $\theta \sim G(\alpha, \beta)$, then the posterior distribution is

$$\theta \mid y \sim G\left(\alpha + n\bar{y}, \beta + n\right),$$

where $y = (y_1, \ldots, y_n)$ and $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$.

- What are the posterior mean, variance and mode?

$$E(\theta \mid y) = (\alpha + n\bar{y})(\beta + n)^{-1}$$

$$\text{Var}(\theta \mid y) = (\alpha + n\bar{y})(\beta + n)^{-2}$$

$$\text{Mode}(\theta \mid y) = (\alpha - n\bar{y} - 1)(\beta + n)^{-1}$$

- What is the relationship between posterior mean and prior mean?

$$E(\theta \mid y) = wE(\theta) + (1 - w)\bar{y}$$

where $w = \frac{\beta}{\beta + n}$

- How about the limiting case?

Posterior Distribution

- If $y_i \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$ for $i = 1, \ldots, n$, and $\theta \sim G(\alpha, \beta)$, then the posterior distribution is

$$\theta \mid y \sim G\left(\alpha + n\bar{y}, \beta + n\right),$$

  where $y = (y_1, \ldots, y_n)$ and $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$.

- What are the posterior mean, variance and mode?

$$E(\theta \mid y) = (\alpha + n\bar{y})(\beta + n)^{-1}$$

$$\text{Var}(\theta \mid y) = (\alpha + n\bar{y})(\beta + n)^{-2}$$

$$\text{Mode}(\theta \mid y) = (\alpha - n\bar{y} - 1)(\beta + n)^{-1}$$

- What is the relationship between posterior mean and prior mean?

$$E(\theta \mid y) = wE(\theta) + (1 - w)\bar{y}$$

  where $w = \frac{\beta}{\beta + n}$

- How about the limiting case? $\{E(\theta \mid y) - \bar{y}\} \to 0$ and $\{\text{Var}(\theta \mid y) - n^{-1}\bar{y}\} \to 0$ as $n \to \infty$

- How to interpret hyperparameters α and β in the prior specifications?

Posterior Distribution

- If $y_i \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$ for $i = 1, \ldots, n$, and $\theta \sim G(\alpha, \beta)$, then the posterior distribution is
$$\theta \mid y \sim G\left(\alpha + n\bar{y}, \beta + n\right),$$
where $y = (y_1, \ldots, y_n)$ and $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$.
- What are the posterior mean, variance and mode?
$$E(\theta \mid y) = (\alpha + n\bar{y})(\beta + n)^{-1}$$
$$\text{Var}(\theta \mid y) = (\alpha + n\bar{y})(\beta + n)^{-2}$$
$$\text{Mode}(\theta \mid y) = (\alpha - n\bar{y} - 1)(\beta + n)^{-1}$$
- What is the relationship between posterior mean and prior mean?
$$E(\theta \mid y) = wE(\theta) + (1 - w)\bar{y}$$
where $w = \frac{\beta}{\beta + n}$
- How about the limiting case? $\{E(\theta \mid y) - \bar{y}\} \to 0$ and $\{\text{Var}(\theta \mid y) - n^{-1}\bar{y}\} \to 0$ as $n \to \infty$
- How to interpret hyperparameters α and β in the prior specifications?
  - β: the number of prior observations
  - α: sum of counts from β prior observations

Predictive Distributions

$$\tilde{y}, y_1, \ldots, y_n \mid \theta \overset{\text{iid}}{\sim} \text{Poisson}(\theta), \qquad \theta \sim G(\alpha, \beta).$$

- Prior predictive distribution:

$$\tilde{y} \sim \text{Neg-Bin}(\alpha, \beta), \quad \pi(\tilde{y}) = \frac{\Gamma(\alpha + \tilde{y})}{\Gamma(\tilde{y}+1)\Gamma(\alpha)} \left(\frac{\beta}{\beta+1}\right)^{\alpha} \left(\frac{1}{\beta+1}\right)^{\tilde{y}}.$$

  which represents the negative binomial distribution with parameters α and β

  - $\tilde{y}$: number of failures
  - α: number of sucesses until the experiment stops.
  - β: odds of sucesses
  - $$E(\tilde{y}) = \frac{\alpha}{\beta} \quad \text{and} \quad \text{Var}(\tilde{y}) = \frac{\alpha}{\beta}\left(1 + \frac{1}{\beta}\right)$$

- Posterior predictive distribution:

$$\pi(\tilde{y} \mid y) = \frac{\Gamma(\alpha + n\bar{y} + \tilde{y})}{\Gamma(\tilde{y}+1)\Gamma(\alpha + n\bar{y})} \left(\frac{\beta+n}{\beta+n+1}\right)^{\alpha+n\bar{y}} \left(\frac{1}{\beta+n+1}\right)^{\tilde{y}}.$$

  where $y = (y_1, \ldots, y_n)$. How about mean and variance?

$$\text{E}(\tilde{y} \mid y) = (\alpha + n\bar{y})(\beta + n)^{-1}, \qquad \text{Var}(\tilde{y} \mid y) = (\alpha + n\bar{y})(\beta + n + 1)(\beta + n)^{-2}$$

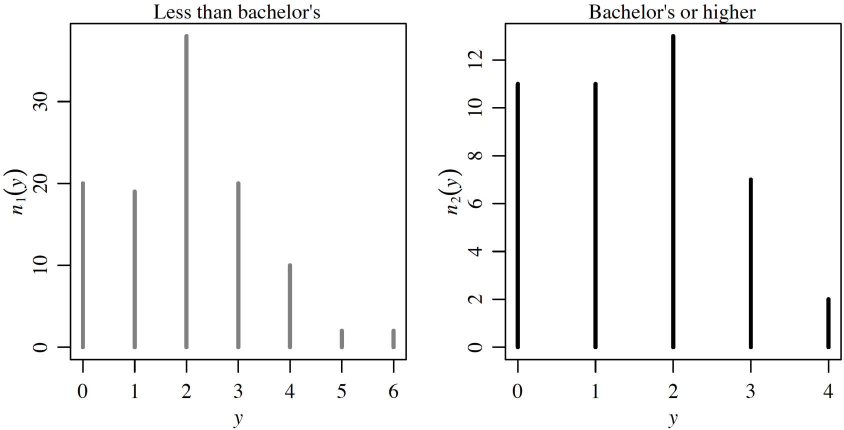
Example: Birth Rate (Hoff, 2009)

- Over the course of the 1990s the General Social Survey gathered data on the educational attainment and number of children of 155 woman who were 40 years of age at the time of their participation in the survey.
- These women were in their 20s during the 1970s, a period of historically low fertility rates in the United States.
- In this example we will examine the difference in the numbers of children between the woman without college degrees (group 1) and those with college degrees (group 2).
- Suppose the number of women without college degrees is n_1 and the number of women with college degrees is n_2. For $k = 1, 2$, $j = 1, \ldots, n_k$, let $y_{k,j}$ be the number of children of women j in group k.
- We assume that

$$y_{k,j} \mid \theta_k \overset{\text{iid}}{\sim} \text{Poisson}(\theta_k),$$

Write $y_k = (y_{k,1}, \ldots, y_{k,n_k})$ for $k = 1, 2$.

Number of Children in Two Groups



Posterior Inference

- The group sums and means are as follows:
  - Less than bachelor's: $n_1 = 111$, $\sum_{j=1}^{n_1} y_{1,j} = 217$, $\bar{y}_1 = 1.95$.
  - Bachelor's or higher: $n_2 = 44$, $\sum_{j=1}^{n_2} y_{2,j} = 66$, $\bar{y}_2 = 1.50$.
- Suppose we assign the prior

$$\theta_1, \theta_2 \overset{\text{iid}}{\sim} G(2, 1).$$

- The posterior distributions:

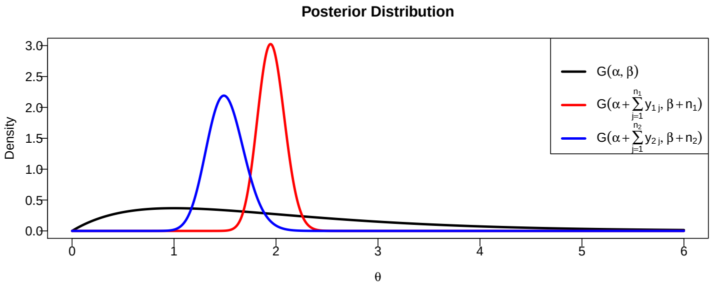
$$\theta_1 \mid y_1 \sim G(219, 112) \qquad \theta_2 \mid y_2 \sim G(68, 45)$$

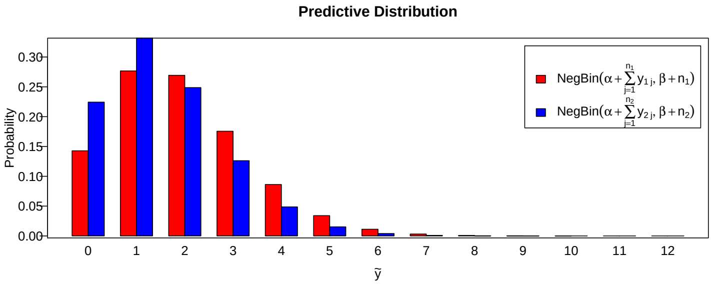
- The posterior predictive distributions:

$$\tilde{y}_{1,*} \mid y_1 \sim \text{Neg-Bin}(219, 112) \qquad \tilde{y}_{2,*} \mid y_2 \sim \text{Neg-Bin}(68, 45).$$

- Conditional on data, are θ_1 and θ_2 independent? Yes
- Conditional on data, are $\tilde{y}_{1,*}$ and $\tilde{y}_{2,*}$ independent? Yes

Posterior Distribution and Predictive Distribuiton





More on Posterior Inference

To answer the following questions, what should we compute and how to do it?

- How likely the average number of children for women without college degree is large than the average number of children for women with college degree given the data?

- How likely the number of children of a woman with college degree is strictly larger than that of a woman without college degree given the data?

- How likely the number of children of a woman with college degree is exactly the same as that of a woman without college degree given the data?

Exponential Family

- Binomial, Poisson and Normal models in the exponential family.
- A one-parameter exponential family model is any model whose densities can be expressed as

$$\pi(y \mid \phi) = h(y)c(\phi)\exp\{\phi t(y)\},$$

  where ϕ is the unknown parameter and $t(y)$ is the sufficient statistic.
- Conjugate prior distributions of the form

$$\pi(\phi \mid n_0, t_0) = \kappa(n_0, t_0)c(\phi)^{n_0}\exp\{n_0 t_0 \phi\}.$$

  Interpretations:
  - n_0: prior sample size (measure of how informative the prior is)
  - t_0: prior expected value of $t(Y)$.
- Suppose $y_i \overset{\text{iid}}{\sim} \pi(y \mid \phi)$, for $i = 1, \ldots, n$, then

$$\pi(\phi \mid y) \propto \pi\left(\phi \mid n_0 + n, \frac{n_0 t_0 + n\bar{t}(y)}{n_0 + n}\right),$$

  where $y = (y_1, \ldots, y_n)$ and $\bar{t}(y) = n^{-1}\sum_{i=1}^{n} t(y_i)$.

Binomial Model Representation

- Representation of the binomial model:

$$\pi(y \mid \theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}.$$

Let $\phi = \log\{\theta/(1-\theta)\}$. Then

$$\pi(y \mid \phi) = \binom{n}{y}\{1+\exp(\phi)\}^{-n}\exp(\phi y).$$

- The conjugate prior for ϕ:

$$\pi(\phi \mid n_0, t_0) \propto \{1+\exp(\phi)\}^{-n_0}\exp(n_0 t_0 \phi).$$

where t_0 represents the prior expectation of $t(y) = y$.

- What is the prior distribution in terms of θ?

$$\pi(\theta \mid n_0, t_0) \propto \theta^{n_0 t_0 - 1}(1-\theta)^{n_0(1-t_0)-1}$$

- What is the posterior distribution?

$$\pi(\theta \mid n_0, t_0, y) \propto \theta^{n_0 t_0 + n\bar{y} - 1}(1-\theta)^{n_0(1-t_0)+n(1-\bar{y})-1}$$

Poisson Model Representation

- The $\text{Poisson}(\theta)$ model can be shown to be an exponential family model with

$$\begin{aligned} t(y) &= y \\ \phi &= \log(\theta) \\ c(\phi) &= \exp\{\exp(-\phi)\} \end{aligned}$$

- The conjugate prior for ϕ is thus

$$\pi(\phi \mid n_0, t_0) = \exp\{n_0 \exp(-\phi)\} \exp(n_0 t_0 y).$$

- What is the prior distribution in terms of θ?

Poisson Model Representation

- The $\mathrm{Poisson}(\theta)$ model can be shown to be an exponential family model with

$$\begin{aligned} t(y) &= y \\ \phi &= \log(\theta) \\ c(\phi) &= \exp\{\exp(-\phi)\} \end{aligned}$$

- The conjugate prior for ϕ is thus

$$\pi(\phi \mid n_0, t_0) = \exp\{n_0 \exp(-\phi)\} \exp(n_0 t_0 y).$$

- What is the prior distribution in terms of θ?

  Since $\theta = \exp(\phi)$ and $d\phi = d\theta/\theta$,

$$\pi(\theta \mid n_0, t_0) \;\propto\; \theta^{\,n_0 t_0 - 1} \exp\{-n_0\, \theta\}, \quad \theta > 0,$$

  i.e.

$$\boxed{\theta \sim \mathrm{Gamma}\big(\text{shape } \alpha = n_0 t_0,\ \text{rate } \beta = n_0\big)}$$

  Its density is

$$\pi(\theta \mid n_0, t_0) = \frac{n_0^{\,n_0 t_0}}{\Gamma(n_0 t_0)}\, \theta^{\,n_0 t_0 - 1} e^{-n_0 \theta}, \quad \theta > 0.$$

Noninformative pirors

- Let the data speak for themselves
- How to guarantee to prior distributions to play a minimal role in the posterior distribution?
- Flat or diffuse improper priors that lead to proper posterior distributions
  - $y \sim \mathrm{N}(\mu, 1)$ with $\pi(\mu) \propto 1$
  - $y \sim \mathrm{N}(0, \sigma^2)$ with $\pi(\sigma^2) \propto 1/\sigma^2$
- Other approaches?

Jeffrey's invariance principle

- One approach is based on Jeffrey's invariance principle
- Consider one-to-one transformations of the parameter $\phi = h(\theta)$.
- By transformation of variables, the prior density $\pi(\theta)$ is equivalent, in terms of expressing the same beliefs, to the following prior density on ϕ:

$$\pi(\phi) = \pi(\theta)|h'(\theta)|^{-1}. \tag{1}$$

- The Jeffrey's general principle:
  Any rule for determining the prior density $\pi(\theta)$ should yield an equivalent result if applied to the transformed parameter $\phi = h(\theta)$ for any one-to-one transformation h. That is,
  - According the rule, determine $\pi(\theta)$ using $\pi(y \mid \theta)$.
  - According the rule, determine $\pi(\phi)$ using $\pi(y \mid \phi)$.
  - (1) should be satisfied.

The Jeffreys Prior

- The Jeffreys prior is given by

$$\pi(\theta) \propto \{I(\theta)\}^{1/2},$$

where $I(\theta)$ is the Fisher information for θ:

$$I(\theta) = E\left[\left\{\frac{d \log \pi(y \mid \theta)}{d\theta}\right\}^2 \middle| \theta\right] = -E\left\{\frac{d^2 \log \pi(y \mid \theta)}{d\theta^2}\middle|\theta\right\}$$

- Can we verify the Jeffreys principle?

Verifying Jeffreys Principle

- Consider a smooth, one-to-one reparameterization $\eta = g(\theta)$ with inverse $\theta = h(\eta)$.
- By the chain rule,

$$\frac{\partial}{\partial\eta}\log\pi(Y \mid \eta) = \frac{\partial}{\partial\theta}\log\pi(Y \mid \theta)\,\frac{d\theta}{d\eta}.$$

- Hence the Fisher information transforms as

$$I_\eta(\eta) = E\left[\left(\frac{\partial}{\partial\eta}\log\pi(Y \mid \eta)\right)^2\right] = I_\theta(\theta)\left(\frac{d\theta}{d\eta}\right)^2.$$

Jeffreys Prior: Invariance

- Jeffreys prior under η is

$$\pi_J(\eta) \ \propto \ \sqrt{I_\eta(\eta)} = \sqrt{I_\theta(\theta)} \left| \frac{d\theta}{d\eta} \right|.$$

- But the change-of-variables rule for priors gives

$$\pi_J(\eta) \ = \ \pi_J(\theta) \left| \frac{d\theta}{d\eta} \right|, \qquad \pi_J(\theta) \ \propto \ \sqrt{I_\theta(\theta)}.$$

- Therefore, Jeffreys prior is invariant to reparameterization.

- *Multivariate case:* For $\theta \in \mathbb{R}^d$,

$$\pi_J(\theta) \ \propto \ |I_\theta(\theta)|^{1/2}.$$

  If $\eta = g(\theta)$ with Jacobian $J = \partial\theta/\partial\eta$, then

$$I_\eta(\eta) = J^\top I_\theta(\theta) J \ \Rightarrow \ |I_\eta(\eta)|^{1/2} = |J| \, |I_\theta(\theta)|^{1/2},$$

  again matching the change-of-variables factor.

Example: Normal Model (Jeffreys Prior for Location)

Suppose

$$y \sim \mathrm{N}(\mu, \sigma^2), \quad \sigma^2 \text{ known.}$$

- Log-likelihood:
  $$\log \pi(y \mid \mu) = -\tfrac{1}{2}\log(2\pi\sigma^2) - \tfrac{(y-\mu)^2}{2\sigma^2}.$$
- Score for μ:
  $$\frac{\partial}{\partial \mu}\log \pi(y \mid \mu) = \frac{y-\mu}{\sigma^2}.$$
- Fisher information:
  $$I(\mu) = \mathrm{E}\left[\left(\frac{y-\mu}{\sigma^2}\right)^2\right] = \frac{1}{\sigma^4}\mathrm{E}[(y-\mu)^2] = \frac{1}{\sigma^4}\sigma^2 = \frac{1}{\sigma^2}.$$
- Jeffreys prior:
  $$\pi(\mu) \propto \sqrt{I(\mu)} \propto 1,$$
  improper but posterior is proper.
- How about μ^3?

Example: Normal Model (Jeffreys Prior for Location)

Suppose

$$y \sim \mathrm{N}(\mu, \sigma^2), \quad \sigma^2 \text{ known.}$$

- Log-likelihood:

$$\log \pi(y \mid \mu) = -\tfrac{1}{2}\log(2\pi\sigma^2) - \tfrac{(y-\mu)^2}{2\sigma^2}.$$

- Score for μ:

$$\frac{\partial}{\partial\mu}\log \pi(y \mid \mu) = \frac{y-\mu}{\sigma^2}.$$

- Fisher information:

$$I(\mu) = \mathrm{E}\left[\left(\frac{y-\mu}{\sigma^2}\right)^2\right] = \frac{1}{\sigma^4}\mathrm{E}[(y-\mu)^2] = \frac{1}{\sigma^4}\sigma^2 = \frac{1}{\sigma^2}.$$

- Jeffreys prior:

$$\pi(\mu) \propto \sqrt{I(\mu)} \propto 1,$$

  improper but posterior is proper.

- How about μ^3? Let $\eta = \mu^3$, then $\mu = \eta^{1/3}$, and $\pi_J(\eta) = \pi_J(\mu)\left|\frac{d\mu}{d\eta}\right| \ \propto \ 1 \cdot \frac{1}{3|\eta|^{2/3}}$

$$\boxed{\pi(\mu^3) \ \propto \ |\mu^3|^{-2/3}}$$

  which matches Jeffreys' invariance principle.

Example: Normal Model (Jeffreys Prior for Scale)

Suppose

$$y \sim \mathrm{N}(\mu, \sigma^2), \quad \mu \text{ known.}$$

- Log-likelihood:

$$\log \pi(y \mid \sigma^2) = -\tfrac{1}{2}\log(2\pi\sigma^2) - \tfrac{(y-\mu)^2}{2\sigma^2}.$$

- Score for σ^2:

$$\frac{\partial}{\partial\sigma^2}\log \pi(y \mid \sigma^2) = -\frac{1}{2\sigma^2} + \frac{(y-\mu)^2}{2\sigma^4}.$$

- Fisher information:

$$I(\sigma^2) = \mathrm{E}\left[\left(-\tfrac{1}{2\sigma^2} + \tfrac{(y-\mu)^2}{2\sigma^4}\right)^2\right] = \frac{1}{2\sigma^4}.$$

- Jeffreys prior:

$$\pi(\sigma^2) \propto \sqrt{I(\sigma^2)} \propto \frac{1}{\sigma^2}.$$

- How about $\log(\sigma^2)$?

Example: Normal Model (Jeffreys Prior for Scale)

Suppose

$$y \sim \mathrm{N}(\mu, \sigma^2), \quad \mu \text{ known.}$$

- Log-likelihood:

$$\log \pi(y \mid \sigma^2) = -\tfrac{1}{2}\log(2\pi\sigma^2) - \tfrac{(y-\mu)^2}{2\sigma^2}.$$

- Score for σ^2:

$$\frac{\partial}{\partial\sigma^2}\log \pi(y \mid \sigma^2) = -\frac{1}{2\sigma^2} + \frac{(y-\mu)^2}{2\sigma^4}.$$

- Fisher information:

$$I(\sigma^2) = \mathrm{E}\left[\left(-\tfrac{1}{2\sigma^2} + \tfrac{(y-\mu)^2}{2\sigma^4}\right)^2\right] = \frac{1}{2\sigma^4}.$$

- Jeffreys prior:

$$\pi(\sigma^2) \propto \sqrt{I(\sigma^2)} \propto \frac{1}{\sigma^2}.$$

- How about $\log(\sigma^2)$? Let $\eta = \log(\sigma^2)$, so $\sigma^2 = e^\eta$ and $d\sigma^2/d\eta = e^\eta$.

$$\pi_J(\eta) = \pi_J(\sigma^2)\left|\tfrac{d\sigma^2}{d\eta}\right| \propto \frac{1}{\sigma^2}\cdot\sigma^2 = 1.$$

$$\boxed{\pi(\log\sigma^2) \ \propto \ 1}$$

a flat prior in the log-scale parameterization.

Example: Binomial model

Suppose

$$y \sim \mathrm{Binomial}(n, \theta)$$

Then

$$\log \pi(y \mid \theta) = \log \binom{n}{y} + y \log \theta + (n-y) \log(1-\theta)$$

$$\frac{\partial}{\partial \theta} \log \pi(y \mid \theta) = \frac{y}{\theta} - \frac{n-y}{1-\theta}, \qquad \frac{\partial^2}{\partial \theta^2} \log \pi(y \mid \theta) = -\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2}$$

Using $E_\theta[y] = n\theta$,

$$I(\theta) = -E_\theta\left[\frac{\partial^2}{\partial \theta^2} \log \pi(y \mid \theta)\right] = \frac{n}{\theta(1-\theta)}.$$

The Jeffreys prior for θ is

$$\pi(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}.$$

This is

$$\theta \sim \mathrm{beta}(1/2, 1/2)$$

What is the prior for $\mathrm{logit}(\theta)$?

The Prior for $\operatorname{logit}(\theta)$

Let $\eta = \operatorname{logit}(\theta) = \log\{\theta/(1-\theta)\}$.

$$\theta = \frac{e^\eta}{1+e^\eta}, \qquad \frac{d\theta}{d\eta} = \theta(1-\theta).$$

Change of variables:

$$\pi(\eta) \;=\; \pi(\theta)\left|\tfrac{d\theta}{d\eta}\right| \;\propto\; \{\theta(1-\theta)\}^{-1/2}\cdot\theta(1-\theta) \;=\; \{\theta(1-\theta)\}^{1/2}.$$

Express as a function of η:

$$\theta(1-\theta) = \frac{e^\eta}{(1+e^\eta)^2} \quad\Rightarrow\quad \pi(\eta) \;\propto\; \frac{e^{\eta/2}}{1+e^\eta} \;=\; \frac{1}{e^{-\eta/2}+e^{\eta/2}} \;=\; \frac{1}{2\cosh(\eta/2)}.$$

$$\boxed{\pi\big(\operatorname{logit}(\theta)\big) \;\propto\; \operatorname{sech}\left(\frac{\eta}{2}\right)} \quad \text{(proper)}.$$

Difficulties with Noninformative Priors

- Searching for a prior distribution that is always vague seems misguided: If the likelihood is truly dominant in a given problem, then the choice among a range of relatively flat prior densities cannot matter.

- For many problems, there is no clear choice for a vague prior distribution, since a density that is flat or uniform in one parameterization will not be in another. For example, for normal model $y \sim \mathrm{N}(\mu, 1)$, we can assume $\pi(\mu) \propto 1$, i.e., a uniform flat prior. How about $\pi\{\exp(\mu)\}$? still uniform?

- Noninformative priors are often useful when it does not seem to be worth the effort to quantify one's real prior knowledge as a probability distribution, as long as one is willing to perform the mathematical work to check that the posterior density is proper

Weakly Informative Priors

- Prior distribution is weakly informative if it is proper but is set up so that the information it does provide is intentionally weaker than whatever actual prior knowledge.
- Examples: $\mu \sim \mathrm{N}(0, 10^6)$, $\sigma^2 \sim \mathrm{G}^{-1}(0.001, 0.001)$.
- Principles for setting up the weakly informative priors.
- Start with some version of a noninformative prior distribution and then add enough information so that inferences are constrained to be reasonable
- Start with a strong, highly informative prior and broaden it to account for uncertainty in ones' prior beliefs and in the applicability of any historically based prior distribution to new data.