Bayesian Statistics
Note 2
Selection of Priors: noninformative, conjugate, and non-conjugate priors

Keunbaik Lee

Sungkyunkwan University

Prior Distributions I

- Suppose we require a prior distribution for

$$\theta = \text{true proportion of U.S. men who are HIV-positive.}$$

- We cannot appeal to the usual long-term frequency notion of probability - it is not possible to even imagine "running the HIV epidemic over again" and reobserving θ. Here θ is random only because it is unknown to us.
- Bayesian analysis is predicated on such a belief in *subjective probability* and its quantification in a prior distribution $p(\theta)$. But:
  - How to create such a prior?
  - Are "objective" choices available?

Prior Distributions II

Elicited Priors

- **Histogram approach**: Assign probability masses to the possible values in such a way that their sum is 1, and their relative contributions reflect the experimenters prior beliefs as closely as possible.
  - BUT: Awkward for continuous or unbounded θ.
- **Matching a functional form**: Assume that the prior belongs to a parametric distributional family $p(\theta|\eta)$, choosing η so that the result matches the elicitee's true prior beliefs as nearly as possible.
  - This approach limits the effort required of the elicitee, and also overcomes the finite support problem inherent in the histogram approach...

Prior Distributions III

- BUT: it may not be possible for the elicitee to “shoehorn” his or her prior beliefs into any of the standard parametric forms.
- **Several priors**: Conjugate priors, Noninformative priors, etc.

Prior Distributions IV

Definition: Let $\mathcal{F}$ be a class of sampling distributions with pdf $P(y|\theta)$ and $\mathcal{P}$ be a class of prior distributions for θ. Then the class $\mathcal{P}$ is **conjugate** for $\mathcal{F}$ if $P(\theta|y) \in \mathcal{P}$ for every $P(y|\theta) \in \mathcal{F}$ and every $P(\theta) \in \mathcal{P}$.

Estimating a probability from Binomial Data I

Let θ denote the proportion of female births and y denote the number of girls in n recorded births. Then

$$P(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}, \quad y = 0, \cdots, n.$$

- Laplace prior: choose the uniform $[0,1]$ prior for θ. Then

$$P(\theta|y) \propto \theta^y(1-\theta)^{n-y},$$

which is Beta$(y+1, n-y+1)$.

Estimating a probability from Binomial Data II

- Let $\tilde{y}$ denote the result of a new trial.

$$
\begin{aligned}
P(\tilde{y}=1|y) &= \int_0^1 P(\tilde{y}=1|\theta,y)P(\theta|y)d\theta \\
&= \int_0^1 P(\tilde{y}=1|\theta)P(\theta|y)d\theta \quad \tilde{y}|\theta \sim \text{Bernoulli}(\theta) \\
&= \int_0^1 \theta P(\theta|y)d\theta \\
&= E(\theta|y) = \frac{y+1}{n+2}.
\end{aligned}
$$

Estimating a probability from Binomial Data III

- Conjugate prior: choose the Beta(α,β) distribution. Then

$$
\begin{aligned}
& P(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} : \text{ kernel of Beta}(\alpha, \beta) \\
=> \quad & P(\theta|y) \propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}
\end{aligned}
$$

  which is Beta($y+\alpha$,$n-y+\beta$). Thus, the prior $P(\theta)$ is a conjugate prior.

Estimating a probability from Binomial Data IV

- The posterior mean is given by

$$E(\theta|y) = \frac{y+\alpha}{n+\alpha+\beta} = \frac{n}{n+\alpha+\beta}\frac{y}{n} + \frac{\alpha+\beta}{n+\alpha+\beta}\frac{\alpha}{\alpha+\beta}$$

which is a weighted average of the sample mean (here, the MLE) and the prior mean.
If n is much larger compared to $\alpha+\beta$, then the posterior mean is leaning towards the sample mean (i.e., $E(\theta|y) \approx \frac{y}{n}$), which if $\alpha+\beta$ is much larger compared to n, the posterior mean is leaning towards the prior mean (i.e., $E(\theta|y) \approx E(\theta)$).

Estimating a probability from Binomial Data V

- The posterior variance is

$$
\begin{aligned}
var(\theta|y) &= \frac{(y+\alpha)(n-y+\beta)}{(n+\alpha+\beta)^2(n+\alpha+\beta+1)} \\
&= \frac{E(\theta|y)(1-E(\theta|y))}{n+\alpha+\beta+1}
\end{aligned}
$$

  When y and $n-y$ are both very large, the posterior variance $\approx \frac{\frac{y}{n}(1-\frac{y}{n})}{n}$ which is also the classical estimate of $var(\frac{y}{n}|\theta) = \theta(1-\theta)/n$.

Estimating a probability from Binomial Data VI

- For example, if $n = 5$, $y = 3$, and $\alpha = \beta = 1$, then

$$\begin{aligned} &\text{MLE of } \theta = 3/5 = .60 \\ &\text{Posterior mean} = 4/7 = .57. \end{aligned}$$

- Predicted Proabilities

$$\begin{aligned} P(\tilde{y} = 1|y) &= \int_0^1 P(\tilde{y} = 1|\theta, y) P(\theta|y) d\theta \\ &= E(\theta|y) = \frac{y + \alpha}{n + \alpha + \beta}. \end{aligned}$$

Estimating a probability from Binomial Data VII

- $y|\theta \sim B(n, \theta)$, $\theta \sim \text{Beta}(\alpha, \beta) => \theta|y \sim \text{Beta}(\alpha + y, n + \beta - y)$.
  One can find the percentiles of this posterior distribution directly. Alternatively, one can draw a large sample from this posterior distribution, and read off the percentiles from the sample histogram. They should be fairly close. The beta prior is a conjugate prior for the binominal distribution.

Estimating a probability from Binomial Data VIII

Example 1: Consider the maternal condition, placenta previa, an unusual condition of pregnancy in which the placenta implanted very low in the uterus, obstructing the fetus from a normal vaginal delivery. Sex of placenta previa births in Germany: 437 females and 543 males.
Take uniform$[0, 1]$ (beta($\alpha = 1$,$\beta = 1$)) prior. Then posterior distribution of the proportion of females is Beta$(438, 544)$. Then

$$\text{Posterior mean} = .446; \;\; \text{Posterior s.d.} = 0.016$$

Central 95% posterior interval is $[.415, .477]$
Instead, if one uses the interval $.446 \pm (1.96)(.016)$, one gets the same answer. So, a normal approximation of the posterior seems okay.

Estimating a probability from Binomial Data IX

Example 2: 241295 girls and 251527 boys were born in Paris between 1745 and 1770. Estimate the proportion of female births as well as the predictive probability that a future birth is a female. Symbolically, let θ =proportion of female births; y =number of girls in n recorded births; n =total number of recorded births. Then

$$P(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}, \quad y = 0, \cdots, n.$$

Prior: $P(\theta) = 1$, $0 \leq \theta \leq 1$ (Laplace prior). Then

$$\begin{aligned} p(\theta|y) &\propto P(y|\theta)P(\theta) \\ &\propto \theta^y(1-\theta)^{n-y} : Beta(y+1, n-y+1) \end{aligned}$$

Estimating a probability from Binomial Data X

Laplace's estimate of θ is $\frac{y+1}{n+2}$.

$$E(\theta|y) = .4903, \quad \text{MLE of } \theta = \frac{y}{n} = .4903.$$

Why this agreement? n is very very large.
Predicted Proabilities

$$P(\tilde{y} = 1|y) = \frac{y+\alpha}{n+\alpha+\beta} = 0.489619.$$

Estimating a probability from Binomial Data XI

Prior: $\text{logit}(\theta) \sim N(\mu, \tau^2)$
Let $\eta = \text{logit}(\theta) = \log \frac{\theta}{1-\theta}$. Then

$$
\begin{aligned}
& P(y|\eta) = \binom{n}{y} \exp(\eta y)(1 + \exp(\eta))^{-n} \\
=> \quad & P(\eta|y) \propto \exp(\eta y)(1 + \exp(\eta))^{-n} \exp\left\{-\frac{1}{2\tau^2}(\eta - \mu)^2\right\}
\end{aligned}
$$

Closed form expressions for the moments or quantiles of this distribution are not available.

Estimation of the normal mean with known variance I

Let $y_1, \cdots, y_n$ denote the observations from $N(\theta, \sigma^2)$ where σ^2 is **known**. The prior for θ is $\theta \sim N(\mu_0, \tau_0^2)$.

- The posterior distribution of θ is

$$P(\theta | y_1, \cdots, y_n) \propto e^{-\frac{1}{2\sigma^2}\sum_i (y_i - \theta)^2} e^{-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2}$$

Here,

$$\begin{aligned}
& \frac{1}{2\sigma^2}\sum_i (y_i - \theta)^2 + \frac{1}{2\tau_0^2}(\theta - \mu_0)^2 \\
= \; & \frac{1}{2}\left(\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}\right)\left[\theta - \frac{\frac{n\bar{y}}{\sigma^2} + \frac{\mu_0}{\tau_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}\right]^2 + \frac{1}{2}\left[\frac{1}{\sigma^2}\sum_i y_i^2 + \frac{\mu_0^2}{\tau_0^2} - \frac{\left(\frac{n\bar{y}}{\sigma^2} + \frac{\mu_0}{\tau_0^2}\right)^2}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}\right].
\end{aligned}$$

Estimation of the normal mean with known variance II

The posterior pdf is

$$\theta | y_1, \cdots, y_n \sim N\left(\frac{\frac{n\bar{y}}{\sigma^2} + \frac{\mu_0}{\tau_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}} \right)$$

- The posterior mean is a weighted average of the sample mean and the prior mean, weights being reciprocals of the corresponding variances.
  Precision=reciprocal of the variance.
- The posterior mean is a weighted average of the sample mean and the prior mean, weights being proportional to the respective precisions (sample precision= n/σ^2, prior precision= $1/\tau_0^2$).

Estimation of the normal mean with known variance III

- If the sample precision outweighs the prior precision, the posterior mean leans towards the sample mean, while if prior precision outweighs sample precision, the posterior mean leans towards the prior mean.
  Posterior precision=sample precision+prior precision=$n/\sigma^2 + 1/\tau_0^2$.

WARNING: Such an exact relationship may not always hold.

Estimation of the normal mean with known variance IV

- Suppose now $\tilde{y}$ denote a single future observation. What is the posterior predictive distribution of $\tilde{y}$ given $y_1, \cdots, y_n$? Let $\mu_1 = \frac{n\bar{y}/\sigma^2 + \mu_0/\tau_0^2}{n/\sigma^2 + 1/\tau_0^2}$ and $\tau_1^2 = \frac{1}{n/\sigma^2 + 1/\tau_0^2}$.

$$P(\tilde{y}|y_1, \cdots, y_n) = \int_{-\infty}^{\infty} P(\tilde{y}|\theta) P(\theta|y_1, \cdots, y_n) d\theta,$$

$$\tilde{y}|\theta \sim N(\theta, \sigma^2) \equiv \tilde{y}|\theta, y_1, \cdots, y_n \sim N(\theta, \sigma^2),$$

$$\theta|y_1, \cdots, y_n \sim N(\mu_1, \tau_1^2).$$

Write $\tilde{y} = \theta + e$ where θ and e given $y_1, \cdots, y_n$ are independent with

$$\theta | y_1, \cdots, y_n \sim N(\mu_1, \tau_1^2),$$
$$e | y_1, \cdots, y_n \sim N(0, \sigma^2).$$

Hence, $\tilde{y} | y_1, \cdots, y_n \sim N(\mu_1, \tau_1^2 + \sigma^2)$.

- As $n \longrightarrow \infty$, $\theta | y_1, \cdots, y_n \approx N(\bar{y}, 0)$.
  i.e., posterior of θ is near degenerate at the sample mean.

$$\tilde{y} | y_1, \cdots, y_n \approx N(\bar{y}, \sigma^2).$$

  Also, we can think of $\theta | y_1, \cdots, y_n$ as approximately $N(\bar{y}, \sigma^2/n)$ if $\tau_0^2 \to \infty$.

Normal with known mean and unknown variance I

Conjugate prior for σ^2 is

$$P(\sigma^2) \propto (\sigma^2)^{-\frac{1}{2}a-1} e^{-\frac{b}{2\sigma^2}}.$$

This density is known as inverse gamma density ($\text{IG}(\frac{a}{2}, \frac{b}{2})$). This is because if $Z = \frac{1}{\sigma^2}$ i.e., $\sigma^2 = \frac{1}{z}$ so that $\|\frac{d\sigma^2}{dz}\| = \frac{1}{z^2}$, Z has pdf

$$\begin{aligned} P(Z) &\propto z^{\frac{a}{2}+1} \frac{1}{z^2} e^{-\frac{bz}{2}} \\ &\propto z^{\frac{a}{2}-1} e^{-\frac{bz}{2}} \end{aligned}$$

which is $\text{Gamma}(\frac{a}{2}, \frac{b}{2})$.

Normal with known mean and unknown variance II

We will write $\sigma^2 \sim \mathrm{IG}(\frac{a}{2}, \frac{b}{2})$. Then

$$\begin{aligned} & P(\sigma^2 | y_1, \cdots, y_n) \\ \propto \quad & (\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_i (y_i-\theta)^2} (\sigma^2)^{-\frac{a}{2}-1} e^{-\frac{b}{2\sigma^2}} \\ = \quad & (\sigma^2)^{-\frac{1}{2}(n+a)-1} e^{-\frac{1}{2\sigma^2}\left[\sum_i (y_i-\theta)^2 + b\right]} \end{aligned}$$

which is $IG(\frac{1}{2}(n+a), \frac{1}{2}\left[\sum_i (y_i-\theta)^2 + b\right])$.

- Prior mean:

$$
\begin{aligned}
E(\sigma^2) &= E\left(\frac{1}{z}\right) = \frac{\int_0^\infty \frac{1}{z} e^{-\frac{bz}{2}} z^{\frac{a}{2}-1} dz}{\int_0^\infty e^{-\frac{b}{2}z} z^{\frac{a}{2}-1} dz} \\
&= \frac{b}{a-2}, \text{ if } a > 2.
\end{aligned}
$$

The posterior mean of σ^2 is

$$
E(\sigma^2 | y_1, \cdots, y_n) = \frac{\sum_{i=1}^n (y_i - \theta)^2 + b}{n + a - 2} \text{ if } n + a - 2 > 0.
$$

Normal with unknown mean and unknown variance I

Suppose $y_1, \cdots, y_n \sim N(\theta, \sigma^2)$ where θ is unknown and σ^2 is unknown. Then

$$P(y_1, \cdots, y_n|\theta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\sum_i (y_i-\theta)^2}$$

Let $r = \sigma^{-2}$. Then

$$P(y|\theta, r) \propto r^{\frac{n}{2}} \exp\left\{-\frac{r}{2}\sum_i (y_i - \theta)^2\right\}.$$

Normal with unknown mean and unknown variance II

The following prior is considered for θ and r.

$$P(r) \propto r^{a/2-1} \exp\left\{-\frac{b}{2}r\right\} \quad \left(r \sim \text{Gamma}\left(\frac{a}{2}, \frac{b}{2}\right)\right)$$

$$P(\theta|r) \propto r^{1/2} \exp\left\{-\frac{\lambda r}{2}(\theta-\mu)^2\right\} \quad \left(\theta|r \sim N\left(\mu, \frac{1}{\lambda r}\right)\right)$$

Normal with unknown mean and unknown variance III

where μ and $\lambda(> 0)$ are known. Then

$$
\begin{aligned}
P(\theta, r|y) &\propto \exp\left\{-\frac{r}{2}\left[n(\bar{y}-\theta)^2 + \sum_i (y_i - \bar{y})^2\right]\right\} \exp\left\{-\frac{\lambda r}{2}(\theta-\mu)^2 - \frac{1}{2}br\right\} \\
&\quad \times r^{\frac{1}{2}(n+a-1)} \\
&\propto \exp\left\{-\frac{r(n+\lambda)}{2}\left(\theta - \frac{n\bar{y}+\lambda\mu}{n+\lambda}\right)^2\right\} r^{1/2} \\
&\quad \times \exp\left\{-\frac{r}{2}\left[\sum_i (y_i-\bar{y})^2 + \frac{n\lambda}{n+\lambda}(\bar{y}-\mu)^2 + b\right]\right\} r^{\frac{1}{2}(n+a)-1}
\end{aligned}
$$

Normal with unknown mean and unknown variance IV

Then

$$\theta | r, y \sim N\left(\frac{n\bar{y} + \lambda\mu}{n+\lambda}, \frac{1}{r(n+\lambda)}\right)$$

$$r | y \sim \text{Gamma}\left(\frac{n+a}{2}, \left[\sum_i (y_i - \bar{y})^2 + \frac{n\lambda}{n+\lambda}(\bar{y}-\mu)^2 + b\right] \diagup 2\right).$$

Poisson Distribution I

- Let $y_1, \cdots, y_n$ denote the observations from $Poisson(\theta)$ where $\theta > 0$. Then the joint distribution of $y_1, \cdots, y_n$ is

$$P(y_1, \cdots, y_n|\theta) = \frac{e^{-n\theta}\theta^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!}.$$

The prior for θ is

$$P(\theta) = \frac{\theta^{a-1}b^a}{\Gamma(a)}e^{-b\theta} \quad \text{Gamma}(a, b).$$

Then the posterior distribution is

$$P(\theta|y_1, \cdots, y_n) \propto e^{-(n+b)\theta}\theta^{\sum_{i=1}^n y_i + a - 1}$$

Poisson Distribution II

which is Gamma($\sum_{i=1}^{n} y_i + a$, $n + b$). The posterior mean is

$$
\begin{aligned}
E(\theta | y_1, \cdots, y_n) &= \frac{\sum_{i=1}^{n} y_i + a}{n + b} \\
&= \frac{n}{n + b} \frac{\sum_{i=1}^{n} y_i}{n} + \frac{b}{n + b} \frac{a}{b}
\end{aligned}
$$

which is a weighted average of the sample mean and the prior mean.

Poisson Distribution III

- Suppose now $\tilde{y}$ denote a single future observation. Then the posterior predictive distribution of $\tilde{y}$ given $y_1, \cdots, y_n$ is

$$
\begin{aligned}
& P(\tilde{y}|y_1,\cdots,y_n) \\
= & \int_0^\infty P(\tilde{y}|\theta)P(\theta|y_1,\cdots,y_n)d\theta \\
= & \frac{\Gamma(\sum_{i=1}^n y_i + \tilde{y} + a)}{\Gamma(\sum_{i=1}^n y_i + a)\tilde{y}!}\left(\frac{n+b}{n+b+1}\right)^{\sum_{i=1}^n y_i + a}\left(\frac{1}{n+b+1}\right)^{\tilde{y}}, \\
& \text{ for } \tilde{y} = 0, 1, 2, \cdots
\end{aligned}
$$

which is a negative binomial distribution.

Noninformative prior distributions I

When prior distributions have no population basis, it is a **noninformative prior** that can be guaranteed to play a minimal role in the posterior distribution.

Noninformative prior distributions II

Example:

1. Let $y_1, \cdots, y_n|\theta \sim^{iid} N(\theta, \sigma^2)$ where σ^2 is known. Then

$$P(y|\theta) \propto \exp\left\{-\frac{n}{2\sigma^2}(\bar{y}-\theta)^2\right\}$$
$$P(\theta) \propto c \text{ (constant)}$$

where σ^2 is known and $-\infty < \theta < \infty$. Then

$$P(\theta|y) \propto \exp\left\{-\frac{n}{2\sigma^2}(\bar{y}-\theta)^2\right\}$$

which is $N\left(\bar{y}, \frac{\sigma^2}{n}\right)$.

Noninformative prior distributions III

2. Let $y_1, \cdots, y_n | \sigma^2 \sim^{iid} N(0, \sigma^2)$. Then

$$P(y|\sigma^2) \propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_i y_i^2\right\}$$
$$P(\sigma^2) \propto (\sigma^2)^{-1}.$$

Then

$$P(\sigma^2|y) \propto (\sigma^2)^{-n/2-1} \exp\left\{-\frac{1}{2\sigma^2}\sum_i y_i^2\right\}$$

which is an inverse gamma distribution.

Noninformative prior distributions IV

Jeffreys Priors

- Motivating example:

$$\begin{aligned} Y|\theta &\sim B(n,\theta) \\ \theta &\sim \text{uniform}(0,1). \end{aligned}$$

Here the prior dist. of θ above is flat.
Let $\theta^* = \theta^2$. Then the pdf of θ^* is

$$p(\theta^*) = \frac{1}{2}\theta^{*-1/2},$$

which is not flat.

Noninformative prior distributions V

- **Fisher Information Number**
  Suppose y has pdf (or pf) $P(y|\theta)$. Then the Fisher Information number is given by

$$
\begin{aligned}
E\left[\left\{\frac{d\log P(y|\theta)}{d\theta}\right\}^2|\theta\right] &= E\left[-\frac{d^2\log P(y|\theta)}{d\theta^2}|\theta\right] \\
&= I(\theta).
\end{aligned}
$$

  under same regularity conditions.

Noninformative prior distributions VI

- **Jeffreys' Prior**

$$P(\theta) = I^{1/2}(\theta).$$

Theorem: Such a prior satisfies an invariance property in the sense that if ϕ is a one-to-one function of θ, then

$$I^{1/2}(\phi) = I^{1/2}(\theta) \left| \frac{d\theta}{d\phi} \right|.$$

Proof)

Noninformative prior distributions VII

If a prior density $P(\theta) \propto I^{1/2}(\theta)$ is used, then by the above result

$$P(\phi) \propto I^{1/2}(\phi).$$

This rule has the valuable property that the prior is *invariant* in that, whatever scale we choose to measure the unknown parameter in, the same prior results when the scale is transformed to any particular scale.

Noninformative prior distributions VIII

Examples

- Binomial distribution:

$$
\begin{aligned}
P(y|\theta) &= \left(\begin{array}{c} n \\ y \end{array} \right) \theta^y (1-\theta)^{n-y}, \\
\log P(y|\theta) &= \log \left(\begin{array}{c} n \\ y \end{array} \right) + y \log \theta + (n-y) \log(1-\theta), \\
\frac{d \log P(y|\theta)}{d\theta} &= \frac{y}{\theta} - \frac{n-y}{1-\theta}, \\
\frac{d^2 \log P(y|\theta)}{d\theta^2} &= -\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2}, \\
I(\theta) &= E\left[-\frac{d^2 \log P(y|\theta)}{d\theta^2} \right] = \frac{n}{\theta(1-\theta)}. \\
I^{1/2}(\theta) &\propto \theta^{-1/2}(1-\theta)^{-1/2}
\end{aligned}
$$

which is Beta(1/2,1/2).

Noninformative prior distributions IX

- Normal distribution:

$$P(y|\theta) = \frac{e^{-\frac{1}{2\sigma^2}(y-\theta)^2}}{\sqrt{2\pi}\sigma}, \quad \sigma(>0) \text{ known}$$

$$\log P(y|\theta) = -\log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}(y-\theta)^2$$

$$\frac{d \log P(y|\theta)}{d\theta} = \frac{y-\theta}{\sigma^2}$$

$$\frac{d^2 \log P(y|\theta)}{d\theta^2} = -\frac{1}{\sigma^2}$$

$$I(\theta) = \frac{1}{\sigma^2} \propto 1.$$

Hence, Laplace's prior and Jeffreys' prior are identical.

Noninformative prior distributions X

- Noninformative priors for location and scale models
  1. Location Family of Distribution

$$P(y|\theta) = f(y-\theta),$$

where f is a pdf. i.e., $\int_{-\infty}^{\infty} f(z)dz = 1$.

- Laplace's Prior: $P(\theta) \propto 1$.
  The heuristic idea here is that if θ is a location parameter, find a vague prior for θ such that the posterior is proportional to the likelihood. i.e.,

$$P(\theta|y) \propto P(y|\theta) = f(y-\theta).$$

But

$$P(\theta|y) \propto f(y-\theta)P(\theta) \propto f(y-\theta).$$

Noninformative prior distributions XI

- Jeffreys' prior
  Suppose f is differentiable in θ. Then

$$\frac{d \log P(y|\theta)}{d\theta} = \frac{d}{d\theta}\left[\log f(y-\theta)\right] = -\frac{f'(y-\theta)}{f(y-\theta)}.$$

Noninformative prior distributions XII

Thus,

$$
\begin{aligned}
I(\theta) &= E\left[\left\{\frac{d\log P(y|\theta)}{d\theta}\right\}^2 |\theta\right] \\
&= E\left[\left\{-\frac{f^{'}(y-\theta)}{f(y-\theta)}\right\}^2 |\theta\right] \\
&= \int_{-\infty}^{\infty}\left\{\frac{f^{'}(y-\theta)}{f(y-\theta)}\right\}^2 f(y-\theta)dy \\
&= \int_{-\infty}^{\infty}\left(\frac{f^{'}(z)}{f(z)}\right)^2 f(z)dz \quad (z = y-\theta)
\end{aligned}
$$

Therefore,

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

Noninformative prior distributions XIII

2. Scale Family of Distribution

$$P(y|\sigma) = \frac{1}{\sigma} f(\frac{y}{\sigma}), \quad \sigma > 0,$$

where $\int_{-\infty}^{\infty} f(y)dy = 1$.

- Laplace:
  Let $\phi = \log \sigma$ and put $P(\phi) \propto 1$.
  $\frac{d\phi}{d\sigma} = \frac{1}{\sigma}$. So take $P(\sigma) \propto \frac{1}{\sigma}$.

Noninformative prior distributions XIV

- Jeffreys:

$$\log P(y|\sigma) = -\log\sigma + \log f(\frac{y}{\sigma})$$
$$\frac{d\log P(y|\sigma)}{d\sigma} = -\frac{1}{\sigma} - \frac{f^{'}(y/\sigma)}{f(y/\sigma)}\frac{y}{\sigma^2}.$$

Then

$$\begin{aligned} I(\sigma) &= \frac{1}{\sigma^2}\int_{-\infty}^{\infty}\left[1 + z\frac{f^{'}(z)}{f(z)}\right]^2 f(z)dz \\ &\propto \frac{1}{\sigma^2}, \end{aligned}$$

where $z = y/\sigma$. Thus, Jeffreys' prior is

$$P(\sigma) \propto I^{1/2}(\sigma) \propto \frac{1}{\sigma}.$$

Noninformative prior distributions XV

3. Example:

$$y_1, \cdots, y_n|\sigma \sim^{\text{i.i.d.}} N(\mu, \sigma^2).$$

where μ is known and σ is unknown. Then the likelihood function is

$$P(y_1, \cdots, y_n|\sigma) \propto \sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum_1^n (y_i-\mu)^2}.$$

Jeffreys' prior for σ is

$$P(\sigma) \propto \frac{1}{\sigma}.$$

The posterior distribution of σ is

$$P(\sigma|y_1, \cdots, y_n) \propto \sigma^{-n-1} e^{-\frac{1}{2\sigma^2}\sum_1^n (y_i-\mu)^2}.$$

Noninformative prior distributions XVI

Let $r = 1/\sigma^2$. Then $|d\sigma/dr| = 1/2r^{-3/2}$ and

$$\begin{aligned} P(r|y_1, \cdots, y_n) &\propto \left(r^{-1/2}\right)^{-(n+1)} e^{-r/2\sum_1^n (y_i-\mu)^2} \frac{1}{r^{3/2}} \\ &= r^{n/2-1} e^{-\frac{r}{2}\sum_1^n (y_i-\mu)^2} \end{aligned}$$

which is Gamma$(n/2, \sum_1^n (y_i - \mu)^2/2)$.

Noninformative prior distributions XVII

4. Remark:

$$P(\sigma) \propto \frac{1}{\sigma}$$

Let $\sigma^2 = z \ => \ \sigma = \sqrt{z}$.

$$P(z) \propto \frac{1}{\sqrt{z}} \frac{1}{2\sqrt{z}} \propto \frac{1}{z}.$$

i.e., $P(\sigma^2) \propto \frac{1}{\sigma^2}$ which puts most of the mass near $\sigma = 0$ i.e., for small values of σ.
Let $r = \frac{1}{\sigma^2}$, $\sigma^2 = \frac{1}{r}$, $|d\sigma^2/dr| = 1/r^2$. Then

$$P(r) \propto r \frac{1}{r^2} \propto r^{-1}$$

Noninformative prior distributions XVIII

which puts most of the mass near $r = 0$. i.e., for large values of σ.