Normal distribution with known mean, unknown variance

Consider y_i iid $N(\theta, \sigma^2)$, $i=1,\dots,n$

The likelihood function as a function of σ^2 is

$$p(y|\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right\}$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right\}$$

The natural conjugate prior distribution for this likelihood for σ^2 is a scaled Inverse-χ^2 distribution with hyperparameters: scale σ_0^2 and ν_0 degrees of freedom (which equals the inverse-gamma with $\alpha=\frac{\nu_0}{2}$, $\beta=\frac{\nu_0}{2}\cdot\sigma_0^2$)

hence $p(\sigma^2) \propto \left(\frac{\sigma_0^2}{\sigma^2}\right)^{\nu_0/2+1} \exp\left\{-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right\}$

The posterior for σ^2 is hence

$$p(\sigma^2|y) \propto p(\sigma^2)\cdot p(y|\sigma^2) \propto (\sigma^2)^{-((n+\nu_0)/2+1)} \exp\left\{-\frac{1}{2\sigma^2}\left(\nu_0\sigma_0^2 + n\sum_{i=1}^{n}(y_i-\theta)^2\right)\right\}$$

$$\propto \text{Inv-}\chi^2\left(\sigma^2 \,\middle|\, n+\nu_0, \frac{\nu_0\sigma_0^2 + n\sum_{i=1}^{n}(y_i-\theta)^2}{\nu_0+n}\right)$$

sep 22-12:12

Introduction to multiparameter models

- In most practical cases, we have more than one parameter $\theta=(\theta_1,\dots,\theta_m)$, for which we have the joint posterior distribution $p(\theta|y)$

↳ Some parameters of interest, maybe one, or a few

Marginal posterior distribution of interest

↳ Some/many parameters are not of interest, but important for constructing the model

→ Called "nuisance parameters"

Several ways of finding the marginal posterior distribution for one parameter θ_i, two ways:

1. Exact:
   - Find the joint posterior $p(\theta_1,\dots,\theta_n|y)$ analytically
   - Integrate out all the other parameters. For example, if $\theta=(\theta_1,\theta_2)$, θ_1 is of interest, θ_2 is a nuisance parameter, then

$$p(\theta_1|y) = \int_{\theta_2} p(\theta_1,\theta_2|y)\, d\theta_2$$

sep 22-12:31

2. By simulation:
   - Draw samples from the joint posterior distribution
   - Consider the posterior samples of the parameter(s) of interest

Two-parameter application

Data: Lung cancer death rates per 100.000 inhabitants in 1960 for each of 44 American states

Death rate for state i : y_i , $i=1,\dots,44$

Sampling distribution:

We assume that the y_i's are iid

$$y_i \sim N(\mu, \sigma^2) \quad , \; i=1,\dots,n \quad , \; n=44$$

We need to find a prior distribution for $\theta = (\mu, \sigma^2)$

The likelihood as a function of both μ and σ^2

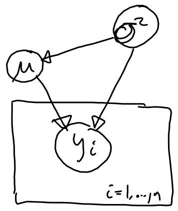
$$p(y \mid \mu, \sigma^2) \underset{\text{Prop. w.r.t. } \mu \text{ and } \sigma^2}{\propto} \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right\} = \sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\bar{y})^2 + n(\bar{y}-\mu)^2\right]\right\}$$

sep 22-12:39

It turns out that a conjugate prior for this likelihood is

$$\mu \mid \sigma^2 \sim N(\mu_0, \sigma^2/\kappa_0)$$

$$\sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$$



Where $\mu_0, \kappa_0, \nu_0, \sigma_0^2$ are hyperparameters, assumed known

The joint prior density is

$$p(\mu, \sigma^2) \propto \sigma^{-1}\exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma^2/\kappa_0}\right\} \cdot (\sigma^2)^{-(\nu_0/2+1)} \cdot \exp\left\{-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right\}$$

$$= \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \cdot \exp\left\{-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + \kappa_0(\mu_0-\mu)^2\right\}\right\}$$

We call this the $N\text{-Inv-}\chi^2(\mu_0, \sigma_0^2/\kappa_0 ; \nu_0, \sigma_0^2)$ density

The joint posterior density is hence

$$p(\mu, \sigma^2 \mid y) \propto p(\mu, \sigma^2) \cdot p(y \mid \mu, \sigma^2) \propto \sigma^{-1}(\sigma^2)^{-\left(\frac{\nu_0+n}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2}\left[\nu_0\sigma_0^2 + \sum_{i=1}^{n}(y_i-\bar{y})^2 + \kappa_0(\mu_0-\mu)^2 + n(\bar{y}-\mu)^2\right]\right\}$$

Now $\kappa_0 \cdot (\mu_0-\mu)^2 + n(\bar{y}-\mu)^2 \underset{\text{Do yourselves}}{=} \kappa_n \cdot (\mu-\mu_n)^2 + \frac{n\kappa_0}{\kappa_0+n} \cdot (\bar{y}-\mu)^2$, where $\kappa_n = \kappa_0 + n$

sep 22-12:47

Hence

$$p(\mu, \sigma^2 | y) \propto \sigma^{-1} (\sigma^2)^{-(\frac{\nu_n}{2}+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[\nu_0 \sigma_n^2 + \kappa_n (\mu_n - \mu)^2\right]\right\}$$

where $\mu_n = \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y}$

$\kappa_n = \kappa_0 + n$

$\nu_n = \nu_0 + n$

$\nu_n \cdot \sigma_n^2 = \nu_0 \sigma_0^2 + \sum_{i=1}^{n} (y_i - \bar{y})^2 + \frac{\kappa_0 n}{\kappa_0 + n} (\bar{y} - \mu_0)^2$

which is $N\text{-}Inv\text{-}\chi^2(\mu_n, \sigma_n^2/\kappa_n ; \nu_n, \sigma_n^2)$

sep 22-12:49

We can also write (for $\theta = (\theta_1, \theta_2)$) ↑ nuisance

$$p(\theta_1 | y) = \int p(\theta_1 | \theta_2, y)\, p(\theta_2 | y)\, d\theta_2 \quad (*)$$

Which shows that the marginal posterior distribution of interest is a mixture of the conditional posterior distributions for θ_1, given θ_2 for all values of θ_2, weighted by $p(\theta_2 | y)$.

In some cases, the marginal posterior distr. of the nuisance parameter θ_2 is available analytically.

We usually don't evaluate $(*)$ explicitly, but it is a motivation/justification for simulations.

Back to the Normal model:

Find $p(\sigma^2 | y) = \int_{-\infty}^{\infty} p(\mu, \sigma^2 | y)\, d\mu$

$$\propto \sigma^{-1} (\sigma^2)^{-(\frac{\nu_n}{2}+1)} \exp\left\{-\frac{1}{2\sigma^2}\nu_n \sigma_n^2\right\} \cdot \overbrace{\int_{-\infty}^{\infty} \underbrace{\exp\left\{-\frac{\kappa_n}{2\sigma^2}(\mu_n - \mu)^2\right\}}_{\text{Kernel of a } N(\mu | \mu_n, \sigma^2/\kappa_n)} d\mu}^{\sqrt{\frac{2\pi\sigma^2}{\kappa_n}}}$$

$$\propto \sigma^{2^{-(\frac{\nu_n}{2}+1)}} \exp\left\{-\frac{1}{2\sigma^2}\nu_n \sigma_n^2\right\} \propto \underline{Inv\text{-}\chi^2(\nu_n, \sigma_n^2)}$$

sep 22-13:17

We can approximate the marginal posterior distr. for μ by simulation:

For i in 1 to S

- Draw $\sigma^{2(i)}$ from $p(\sigma^2|y)$
- Draw $\mu^{(i)}$ from $p(\mu|\sigma^{2(i)}, y) = N\left(\mu \mid \mu_n, \frac{1}{\frac{1}{\kappa_0} + \frac{n}{\sigma^{2(i)}}}\right)$

Then $\mu^{(1)}, \dots, \mu^{(S)}$, for large enough S, can be used to approximate $p(\mu|y)$ (they are samples from this distribution)

However, in this case, we can explicitly find the marginal $p(\mu|y)$:

Let $A = \nu_n \sigma_n^2 + \kappa_n(\mu_n - \mu)^2 = \nu_n \sigma_n^2 \left(1 + \frac{\kappa_n(\mu - \mu_n)^2}{\nu_n \sigma_n^2}\right)$

and $z = \frac{A}{2\sigma^2}$.

Hence

$$p(\mu|y) = \int_0^\infty p(\mu, \sigma^2|y)\, d\sigma^2 \propto A^{-(\nu_n+1)/2} \cdot \underbrace{\int_0^\infty z^{\nu_n - 1/2} \cdot \exp\{-z\}\, dz}_{\text{integrates to a constant w.r.t. } \mu}$$

$$\propto \left(1 + \frac{\kappa_n \cdot (\mu - \mu_n)^2}{\nu_n \sigma_n^2}\right)^{-(\nu_n+1)/2}$$

sep 22-13:28

$$= \left(1 + \frac{1}{\nu_n} \cdot \frac{(\mu - \mu_n)^2}{\sigma_n^2/\kappa_n}\right)^{-(\nu_n+1)/2}$$

which is prop. to the $t_{\nu_n}(\mu_n, \sigma_n^2)$-distribution (see Appendix A)

sep 22-13:41