UNIVERSITÀ DEGLI STUDI DI TRIESTE

DEAMS Dipartimento di scienze economiche, aziendali, matematiche e statistiche “Bruno de Finetti”

Bayesian Statistics

Multiple parameter models

Francesco Pauli

A.A. 2018/19

Two-parameters models

A model is specified with two real parameters θ_1, θ_2

$$p(y|\theta_1, \theta_2)$$

the prior is then a bivariate distribution $\pi(\theta_1, \theta_2)$ and the posterior is then a bivariate distribution as well

$$\pi(\theta_1, \theta_2|y) \propto p(y|\theta_1, \theta_2)\pi(\theta_1, \theta_2)$$

Two-parameters models: nuisance parameters

Suppose that one of the parameters, say θ_2 is a nuisance parameter, in which case we may be interested in the marginal posterior for θ_1, which is obtained as

marginal of the joint posteriors

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

or as a mixture of conditional posterior

$$\pi(\theta_1|y) = \int \pi(\theta_1|\theta_2, y)\pi(\theta_2|y)d\theta_2$$

In classical inference, nuisance is σ^2, we're usually only interested in μ, so $\frac{\bar{y}-\mu}{S/\sqrt{n}}$; in Bayes we reason by means of MARGINALIZATION

Indice

Likelihood

Let

$$y_1, \ldots, y_n | \mu, \sigma^2 \sim IID(N(\mu, \sigma^2))$$

The likelihood is

$$\begin{aligned}
p(y|\mu,\sigma^2) &\propto (\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_j (y_j - \mu)^2 \right\} \\
&\propto (\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_j (y_j - \bar{y} + \bar{y} - \mu)^2 \right\} \\
&\propto (\sigma^2)^{-n/2} \exp\left\{ -\frac{n}{2\sigma^2} (\hat{\sigma}^2 + (\bar{y} - \mu)^2) \right\} \\
&\propto (\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} ((n-1)s^2 + n(\bar{y} - \mu)^2) \right\}
\end{aligned}$$

sufficient statistics

a function of the sufficient statistics

$$\bar{y} = \frac{1}{n} \sum_j y_j; \quad s^2 = \frac{1}{n-1} \sum_j (y_j - \bar{y})^2 = \frac{n}{n-1} \hat{\sigma}^2.$$

Indice

1. (Univariate) normal model with μ and σ^2 unknown
   - Normal model with noninformative prior
   - Normal model with conjugate prior

2. Multivariate normal model

Noninformative prior specification

Consider the improper prior

PROPTO

$$\pi(\mu, \sigma^2) \propto (\sigma^2)^{-1}$$

that is, μ and σ^2 are independent and

- $\pi(\mu) \propto k$ ⟶ UNIFORM "DISTRIBUTION" (improper prior)
- $\pi(\sigma^2) \propto (\sigma^2)^{-1}$

  Equivalently, we could say that $\pi(\log \sigma^2) \propto k$

Noninformative prior specification

The posterior is

$$
\begin{aligned}
\pi(\mu,\sigma^2|y) &\propto \overbrace{p(y|\mu,\sigma^2)}^{\text{LIK.}}\overbrace{(\sigma^2)^{-1}}^{\text{"PRIOR"}}\\
&\propto (\sigma^2)^{-1}(\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}((n-1)s^2+n(\bar{y}-\mu)^2)\right\}\\
&\propto \underbrace{\underbrace{(\sigma^2)^{-1/2}}_{\frac{1}{\sqrt{\sigma^2}}}\exp\left\{-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right\}}_{\mu|\sigma^2,y\sim\mathcal{N}\left(\bar{y},\frac{\sigma^2}{n}\right)}\underbrace{(\sigma^2)^{-(n+1)/2}\exp\left\{-\frac{1}{2\sigma^2}(n-1)s^2\right\}}_{\sigma^2|y\sim\text{inv-}\chi^2(n-1,s^2)}
\end{aligned}
$$

INVERSE CHISQ.

We already know that the posterior for μ conditional on σ^2 is

$$\pi(\mu|\sigma^2,y)=\mathcal{N}\left(\bar{y},\frac{\sigma^2}{n}\right)$$

(In fact the problem is no different than what we discussed as a single parameter model assuming σ^2 known.)

Posterior with noninformative prior

The marginal posterior for σ^2 is

$$\begin{aligned}
\pi(\sigma^2|y) &= \int \pi(\mu, \sigma^2|y) d\mu \\
&\propto \int \sigma^{-n-2} \exp\left\{-\frac{1}{2\sigma^2}((n-1)s^2 + n(\bar{y}-\mu)^2)\right\} d\mu \\
&\propto \sigma^{-n-2} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\} \sqrt{\frac{2\pi\sigma^2}{n}} \\
&\propto (\sigma^2)^{-(n+1)/2} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\}
\end{aligned}$$

that is

$$\sigma^2|y \sim \text{inv-}\chi^2(n-1, s^2)$$

Posterior with noninformative prior (cont.)

Reacall that by

$$\sigma^2|y \sim \text{inv-}\chi^2(n-1, s^2)$$

we mean that

$$\sigma^2 =_d \frac{(n-1)s^2}{X}, \quad X \sim \chi^2_{n-1}$$

and compare this with the usual result on the sampling distribution of s^2.

Note also that it is equivalent to write

$$\sigma^2|y \sim \text{inv-Gamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$$

$$(\sigma^2)^{-1}|y \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$$

Marginal posterior for μ

$$
\begin{aligned}
\pi(\mu|y) &= \int_0^\infty \pi(\mu,\sigma^2|y)d\sigma^2 \\
&= \int_0^\infty \sigma^{-n-2}\exp\left\{ \underbrace{-\frac{1}{2\sigma^2}((n-1)s^2+n(\bar{y}-\mu)^2)}_{z=\frac{A}{2\sigma^2}=\frac{(n-1)s^2+n(\bar{y}-\mu)^2}{2\sigma^2}} \right\} d\sigma^2 \\
&= \int_0^\infty \left(\frac{A}{2z}\right)^{-(n+2)/2} \exp\{-z\}\frac{A}{2z}dz \\
&\propto A^{-n/2}\int_0^\infty z^{(n-2)/2}\exp\{-z\}dz \\
&\propto \left(1+\frac{n(\mu-\bar{y})}{(n-1)s^2}\right)^{-n/2} \sim t_{n-1}(\bar{y}, s^2/n)
\end{aligned}
$$

Integral of Kernel of Gamma

Marginal posterior for μ

Hence

$$\mu|y \sim t_{n-1}(\bar{y}, s^2/n)$$

which is equivalent to

$$\left.\frac{\mu - \bar{y}}{s/\sqrt{n}}\right| y \sim t_{n-1}$$

RANDOM VARIABLE

analogous to the usual result for the pivotal quantity

$$\left.\frac{\bar{y} - \mu}{s/\sqrt{n}}\right| \mu, \sigma^2 \sim t_{n-1}$$

RANDOM QUANTITY

Predictive distribution for $\tilde{y}$

In general

$$p(\tilde{y}|y) = \int\int \underbrace{p(\tilde{y}|\mu,\sigma^2,\cancel{y})\overbrace{\pi(\mu,\sigma^2|y)}^{\text{weight by posterior}}}_{\text{GAUSS - INV } \chi^2} d\mu d\sigma^2$$

and it can be shown that

$$\tilde{y}|y \sim t_{n-1}\left(\bar{y}, \left(1+\frac{1}{n}\right)s^2\right)$$

uncertainty of y around its mean (s^2)

uncertainty about μ (s^2/n)

Predictive distribution for $\tilde{y}$, proof

First note that we have proven that

$$\pi(\mu|y) = \int \underbrace{\overset{\text{CONDITIONAL POST.}}{\pi(\mu|\sigma^2, y)}}_{\mathcal{N}(\bar{y},\sigma^2/n)} \overset{\text{MARGINAL POST.}}{\pi(\sigma^2|y)} d\sigma^2 = \left(1 + \frac{n(\mu - \bar{y})}{(n-1)s^2}\right)^{-n/2} \sim t_{n-1}(\bar{y}, s^2/n)$$

Then note that

$$\begin{aligned}
p(\tilde{y}|y) &= \int\int p(\tilde{y}|\mu, \sigma^2, y) \overbrace{\pi(\mu, \sigma^2|y)}^{\pi(\mu|\sigma^2, y)\, \pi(\sigma^2/y)} d\mu d\sigma^2 \\
&= \int\int p(\tilde{y}|\mu, \sigma^2)\pi(\mu, \sigma^2|y) d\mu d\sigma^2 \\
&= \int \underbrace{\overset{\text{mixture of gaussians}}{\left(\int p(\tilde{y}|\mu, \sigma^2)\pi(\mu|\sigma^2, y) d\mu\right)}}_{\mathcal{N}(\bar{y},\sigma^2(1+1/n))} \overbrace{\pi(\sigma^2|y)}^{\text{INV-}\chi^2} d\sigma^2 \\
&\sim t_{n-1}\left(\bar{y}, s^2\left(1 + \frac{1}{n}\right)\right)
\end{aligned}$$

Result analogous to page before

Indice

1. (Univariate) normal model with μ and σ^2 unknown
   - Normal model with noninformative prior
   - Normal model with conjugate prior
2. Multivariate normal model

Conjugate prior specification

Let the prior be

Equivalent # of information conveying the same info as the prior

- $\mu|\sigma^2 \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right)$
- $\sigma^2 \sim \text{inv-}\chi^2(\nu_0, \sigma_0^2)$

Remember that this means

$$\sigma^2 =_d \frac{\nu_0 \sigma_0^2}{\chi^2_{\nu_0}}, \quad \text{i.e. } \sigma^2 \sim \text{inv-Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right)$$

so that

$$E(\sigma^2) = \frac{\nu_0 \sigma_0^2}{\nu_0 - 2}; \quad \text{Mode}(\sigma^2) = \frac{\nu_0 \sigma_0^2}{\nu_0 + 2}$$

and

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

Conjugate prior specification: joint and marginal

Since

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

the joint prior density is

$$\pi(\mu,\sigma^2) = \pi(\mu|\sigma^2)\pi(\sigma^2)$$

Same struct. as likelihood

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

label this as the

$$\text{N-inv-}\chi^2(\mu_0,\sigma_0^2/\kappa_0,\nu_0,\sigma_0^2)$$

JUST A LABEL...

Note also that marginally

Again a mixture of Gaussian and Inv-χ^2

$$\pi(\mu) \propto \left(1+\frac{\kappa_0(\mu-\mu_0)^2}{\nu_0\sigma_0^2}\right)^{-(\nu_0+1)/2} \sim t_{\nu_0}(\mu_0,\sigma_0^2/\kappa_0)$$

Posterior with conjugate prior

$$
\begin{aligned}
\pi(\mu,\sigma^2|y) &\propto p(y|\mu,\sigma^2)\pi(\mu,\sigma^2)\\
&\propto (\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}((n-1)s^2+n(\bar{y}-\mu)^2)\right\}\times\\
&\qquad\times\sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)}\exp\left\{-\frac{1}{2\sigma^2}\left(\nu_0\sigma_0^2+\kappa_0(\mu_0-\mu)^2\right)\right\}\\
&\propto \sigma^{-1}(\sigma^2)^{-((\nu_0+n)/2+1)}\exp\left\{-\frac{1}{2\sigma^2}\left(\nu_0\sigma_0^2+\kappa_0(\mu_0-\mu)^2+(n-1)s^2+n(\bar{y}-\mu)^2\right)\right\}\\
&\propto \sigma^{-1}(\sigma^2)^{-(\nu_n/2+1)}\exp\left\{-\frac{1}{2\sigma^2}\left(\nu_n\sigma_n^2+\kappa_n(\mu_n-\mu)^2\right)\right\}
\end{aligned}
$$

where

$$
\begin{aligned}
\mu_n &= \frac{\kappa_0\mu_0+n\bar{y}}{\kappa_0+n}\\
\kappa_n &= \kappa_0+n\\
\nu_n &= \nu_0+n\\
\nu_n\sigma_n^2 &= \nu_0\sigma_0^2+(n-1)s^2+\frac{\kappa_0 n}{\kappa_0+n}(\bar{y}-\mu_0)^2
\end{aligned}
$$

Detail

$$
\begin{aligned}
&\kappa_0(\mu_0-\mu)^2+n(\bar{y}-\mu)^2= \\
&\quad=\kappa_0(\mu_0^2-2\mu\mu_0+\mu^2)+n(\bar{y}^2-2\mu\bar{y}+\mu^2) \\
&\quad=(\kappa_0+n)\mu^2-2\mu(\kappa_0\mu_0+n\bar{y})+(\kappa_0\mu_0^2+n\bar{y}^2) \\
&\quad=(\kappa_0+n)\left(\mu-\frac{\kappa_0\mu_0+n\bar{y}}{\kappa_0+n}\right)^2-(\kappa_0+n)\left(\frac{\kappa_0\mu_0+n\bar{y}}{\kappa_0+n}\right)^2+\kappa_0\mu_0^2+n\bar{y}^2 \\
&\quad=\kappa_n(\mu-\mu_n)^2-\frac{1}{\kappa_0+n}\left(\kappa_0^2\mu_0^2+2\kappa_0\mu_0 n\bar{y}+n^2\bar{y}^2\right)+\kappa_0\mu_0^2+n\bar{y}^2 \\
&\quad=\kappa_n(\mu-\mu_n)^2-\frac{1}{\kappa_0+n}\left(\kappa_0^2\mu_0^2+2\kappa_0\mu_0 n\bar{y}+n^2\bar{y}^2-(\kappa_0+n)(\kappa_0\mu_0^2+n\bar{y}^2)\right) \\
&\quad=\kappa_n(\mu-\mu_n)^2-\frac{1}{\kappa_0+n}\left(\kappa_0^2\mu_0^2+2\kappa_0\mu_0 n\bar{y}+n^2\bar{y}^2-\kappa_0^2\mu_0^2-n\kappa_0\bar{y}^2-\kappa_0 n\mu_0^2-n^2\bar{y}^2\right) \\
&\quad=\kappa_n(\mu-\mu_n)^2-\frac{1}{\kappa_0+n}\left(2\kappa_0\mu_0 n\bar{y}-n\kappa_0\bar{y}^2-\kappa_0 n\mu_0^2\right) \\
&\quad=\kappa_n(\mu-\mu_n)^2+\frac{n\kappa_0}{\kappa_0+n}\left(\bar{y}^2+\mu_0^2-2\mu_0\bar{y}\right) \\
&\quad=\kappa_n(\mu-\mu_n)^2+\frac{n\kappa_0}{\kappa_0+n}(\mu_0-\bar{y})^2
\end{aligned}
$$

Detail (cont.)

$$
\begin{aligned}
&\nu_0\sigma_0^2 + \kappa_0(\mu_0 - \mu)^2 + (n-1)s^2 + n(\bar{y} - \mu)^2) = \\
&\quad = \nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{\kappa_0 + n}\left(\mu_0 - \bar{y}\right)^2 + \kappa_n\left(\mu - \mu_n\right)^2 \\
&\quad = \nu_n\sigma_n^2 + \kappa_n\left(\mu - \mu_n\right)^2
\end{aligned}
$$

Posterior with conjugate prior

Then

$$\mu, \sigma^2 | y \sim \text{N-inv-}\chi^2(\mu_n, \sigma_n^2/\kappa_n, \nu_n, \sigma_n^2)$$

with

weighted average of prior means and evidence mean (weights are the respective uncertainties)

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

$$\kappa_n = \kappa_0 + n \rightarrow \text{\# obs. implied by the prior + posterior}$$

$$\nu_n = \nu_0 + n$$

$$\nu_n \sigma_n^2 = \nu_0 \sigma_0^2 + (n-1)s^2 + \overbrace{\frac{\kappa_0 n}{\kappa_0 + n}\underbrace{(\bar{y} - \mu_0)^2}_{\text{difference btw. parameter and its estimator}}}^{\text{TERM TO ACCOUNT FOR USAGE OF ESTIMATOR}}$$

Note that

$$E(\sigma^2 | y) = \frac{\nu_n \sigma_n^2}{\nu_n - 2} = \frac{\nu_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)^2}{\nu_0 + n - 2}$$

Posterior for μ with conjugate prior

One can draw conclusions directly from the bivariate posterior distribution (for instance, a posterior credibility region may be obtained for the pair), it may also be interesting, however, to investigate one parameter only, typically the mean.
it is then relevant to know that

- conditionally to a value for the variance σ^2,

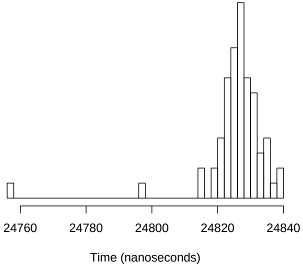
$$(\mu|\sigma^2, y) = \mathcal{N}\left(\frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_0 + n}, \frac{\sigma^2}{\kappa_0 + n}\right)$$

- Marginally

$$\pi(\mu|y) \propto \left(1 + \frac{\kappa_n(\mu - \mu_n)^2}{\nu_n\sigma_n^2}\right)^{-(\nu_n+1)/2} \sim t_{\nu_n}(\mu_n, \sigma_n^2/\kappa_n)$$

Example: Newcomb measurements

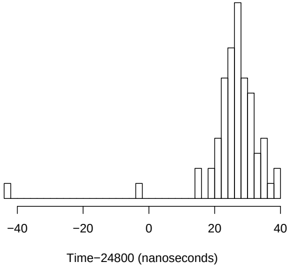
Simon Newcomb set up an experiment in 1882 to measure the speed of light by observing the time required for light to travel 7442 meters.



$n = 66$
$\bar{z} = 24826.2$
$s = 10.8$
True value: 24833.02

Example: Newcomb measurements

Simon Newcomb set up an experiment in 1882 to measure the speed of light by observing the time required for light to travel 7442 meters.



$n = 66$
$\bar{z} = 24826.2$
$s = 10.8$
True value: 24833.02
Trasformation: $y = z - 24800$ for simplicity
$\bar{y} = 26.2$
True value: 33.02

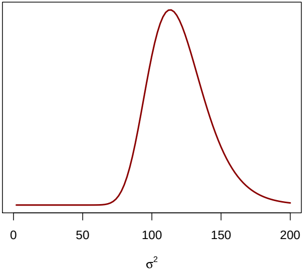
Newcomb measurements, posterior for σ^2

UNIFORM PRIORS

$$\sigma^2|y \sim \text{inv-}\chi^2(65, 10.8^2)$$

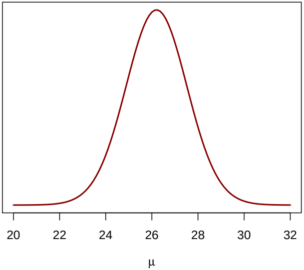
$$E(\sigma^2|y) = \frac{65}{65-2}10.8^2 = 120.34$$

$$\sqrt{E(\sigma^2|y)} = 10.97$$



Newcomb measurements, posterior for μ

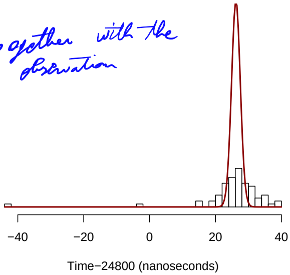
$$\mu|y \sim t_{65}\left(26.2, \frac{10.8^2}{66} = 1.329^2\right)$$



Newcomb measurements, posterior for μ

$$\mu|y \sim t_{65}\left(26.2, \frac{10.8^2}{66} = 1.329^2\right)$$

together with the observation



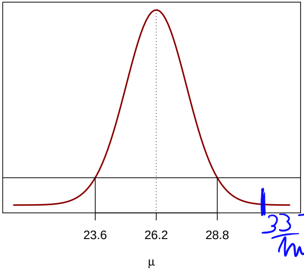
Newcomb measurements, posterior for μ

$$\mu|y \sim t_{65}\left(26.2, \frac{10.8^2}{66} = 1.329^2\right)$$

Posterior interval:

$$\bar{y} \pm t_{66,0.975}\frac{s}{\sqrt{66}} = 26.2 \pm 1.997 \times 1.329$$

$$[23.6, 28.8]$$



33 true value

HPD ≡ CREDIBILITY due to symm.

Newcomb measurements, predictive distribution

$$\tilde{y}|y \sim t_{65}\left(26.2, 10.8^2\left(1+\frac{1}{66}\right) = 10.88^2\right)$$

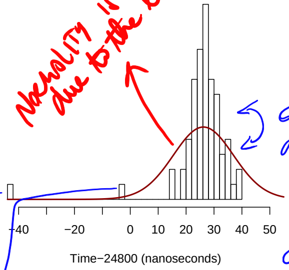
Posterior interval:

$$\bar{y} \pm t_{66,0.975}\, s\sqrt{1+\frac{1}{66}}$$

$$26.2 \pm 1.997 \times 10.88$$

$$[4.47, 47.93]$$

NORMALITY IS INADEQUATE
due to the left outliers



Quite different

These comparison are good to check model adequacy

These 2 outliers change everything...

«POSTERIOR PREDICTIVE CHECKS»

Reparametrization

It is convenient to reparametrize the model writing $\tau = 1/\sigma^2$, so the likelihood is

$$p(y|\mu, \tau) \propto \tau^{n/2} \exp\left\{-\frac{n\tau}{2}(\hat{\sigma}^2 + (\bar{y} - \mu)^2)\right\}$$

the parameter τ is also called precision. Used a lot in ML...

Indice

Likelihood

Let $y \in \mathbb{R}^d$ be a vector of observations and assume

$$y|\mu, \Sigma \sim \mathcal{N}(\mu, \Sigma),$$

then, for one observation,

$$p(y|\mu, \Sigma) = (2\pi)^{-d/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right\}$$

while for n observations

$$\begin{aligned} p(y_1, \ldots, y_n|\mu, \Sigma) &\propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)\right\} \\ &\propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}S_0)\right\} \end{aligned}$$

where $S_0 = \sum_{i=1}^n (y_i-\mu)(y_i-\mu)^T$.

Model with Σ known

A priori, let $\mu \sim \mathcal{N}(\mu_0, \Lambda_0)$, then

$$\begin{aligned} p(\mu|y,\Sigma) &\propto p(y|\mu,\Sigma)\pi(\mu) \\ &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu) - \frac{1}{2}(\mu-\mu_0^T)\Lambda_0^{-1}(\mu-\mu_0)\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\mu-\mu_n)(\Lambda_0^{-1}+n\Sigma^{-1})(\mu-\mu_n)\right\} \end{aligned}$$

where

$$\mu_n = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1}(\Lambda_0^{-1}\mu_0 + n\Sigma^{-1}\bar{y})$$

Note that the result resembles that for the unidimensional normal distribution, the posterior is a $\mathcal{N}(\mu_n, \Lambda_n)$ with $\Lambda_n^{-1} = \Lambda_0^{-1} + n\Sigma^{-1}$.

Model with μ, Σ unknown

We consider the prior defined by

$$\mu|\Sigma \sim \mathcal{N}\left(\mu_0, \Sigma/\kappa_0\right)$$

$$\Sigma \sim \text{Inv-wishart}(\Lambda_0^{-1}, \nu_0)$$

where the latter means that

$$\pi(\Sigma) \propto |\Sigma|^{-\nu_0/2-1} \exp\left\{-\frac{1}{2}\text{tr}(\Lambda_0\Sigma^{-1})\right\}$$

and so the prior is

$$\pi(\mu, \Sigma) \propto |\Sigma|^{-\frac{d+\nu_0}{2}-1} \times \\ \times \exp\left\{-\frac{1}{2}\text{tr}(\Lambda_0\Sigma^{-1}) - \frac{\kappa_0}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)\right\}$$

Model with μ, Σ unknown (cont.)

The posterior distribution belongs to the same family with parameters

$$\begin{aligned}
\mu_n &= \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y} \\
\kappa_n &= \kappa_0 + n \\
\nu_n &= \nu_0 + n \\
\Lambda_n &= \Lambda_0 + S + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)(\bar{y} - \mu_0)^T
\end{aligned}$$

where

$$S = \sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})^T$$