Chapter 3

- 3.1 Marginalisation
- 3.2 Normal distribution with a noninformative prior (very important)
- 3.3 Normal distribution with a conjugate prior (very important)
- 3.4 Multinomial model (can be skipped)
- 3.5 Multivariate normal with known variance (needed later)
- 3.6 Multivariate normal with unknown variance (glance through)
- 3.7 Bioassay example (very important, related to one of the exercises)
- 3.8 Summary (summary)

R demos

- demo3_1: visualise joint density and marginal densities of posterior of normal distribution with unknown mean and variance
- demo3_2: visualise factored sampling and corresponding marginal and conditional density
- demo3_3: visualise marginal distribution of mu as a mixture of normals
- demo3_4: visualise sampling from the posterior predictive distribution
- demo3_5: visualise Newcomb's data
- demo3_6: visualise posterior in bioassay example

Marginalization

- Joint distribution

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

- Marginalization

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

  $p(\theta_1 | y)$ is marginal distribution
- Goal is to find marginal posterior of an interesting quantity
  - a parameter of the model
  - future event

Marginalization - predictive distribution

- Joint distribution

$$\begin{aligned} p(\tilde{y}, \theta | y) &= p(\tilde{y} | \theta, y) p(\theta | y) \\ &= p(\tilde{y} | \theta) p(\theta | y) \qquad \text{(often)} \end{aligned}$$

- Marginalization

$$p(\tilde{y} | y) = \int p(\tilde{y} | \theta) p(\theta | y) d\theta$$

$p(\tilde{y} | y)$ is a predictive distribution

Marginalization - predictive distribution

- Often joint distribution can be factorized

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

  and the integral can be solved easily by simulation
- For example

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)p(\theta|y)d\theta$$

  can be approximated by
  1) sampling $\theta^{(t)}$ from $p(\theta|y)$
  2) for each $\theta^{(t)}$ sample $\tilde{y}^{(t)}$ from $p(\tilde{y}|\theta^{(t)})$.

  Then $\tilde{y}^{(t)}$ come from $p(\tilde{y}|y)$

Gaussian - non-informative prior

- Joint posterior (-2 is due to prior $1/\sigma^2$)

$$
\begin{aligned}
p(\mu, \sigma^2 | y) &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right) \\
&= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right]\right) \\
&= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[(n-1)s^2 + n(\bar{y} - \mu)^2\right]\right)
\end{aligned}
$$

where

$$
s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2
$$

$\bar{y}$ and s^2 (and n) are sufficient statistics

Gaussian - non-informative prior

$$\sum_{i=1}^{n}(y_i - \mu)^2$$

Gaussian - non-informative prior

$$\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2)$$

Gaussian - non-informative prior

$$\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2)$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2 - \bar{y}^2 + \bar{y}^2 - 2y_i\bar{y} + 2y_i\bar{y})$$

Gaussian - non-informative prior

$$\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2)$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2 - \bar{y}^2 + \bar{y}^2 - 2y_i\bar{y} + 2y_i\bar{y})$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\bar{y} + \bar{y}^2) + \sum_{i=1}^{n}(\mu^2 - 2y_i\mu - \bar{y}^2 + 2y_i\bar{y})$$

Gaussian - non-informative prior

$$\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2)$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2 - \bar{y}^2 + \bar{y}^2 - 2y_i\bar{y} + 2y_i\bar{y})$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\bar{y} + \bar{y}^2) + \sum_{i=1}^{n}(\mu^2 - 2y_i\mu - \bar{y}^2 + 2y_i\bar{y})$$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\mu^2 - 2\bar{y}\mu - \bar{y}^2 + 2\bar{y}\bar{y})$$

Gaussian - non-informative prior

$$\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2)$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2 - \bar{y}^2 + \bar{y}^2 - 2y_i\bar{y} + 2y_i\bar{y})$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\bar{y} + \bar{y}^2) + \sum_{i=1}^{n}(\mu^2 - 2y_i\mu - \bar{y}^2 + 2y_i\bar{y})$$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\mu^2 - 2\bar{y}\mu - \bar{y}^2 + 2\bar{y}\bar{y})$$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\bar{y} - \mu)^2$$

Gaussian - non-informative prior

- Factorize $p(\mu, \sigma^2|y) = p(\mu|\sigma^2, y)p(\sigma^2|y)$
- *conditional* posterior $p(\mu|\sigma^2, y)$

$$\mu|\sigma^2, y \sim \mathrm{N}(\bar{y}, \sigma^2/n)$$

is same as Gaussian with a known variance

Gaussian - non-informative prior

- Factorize $p(\mu, \sigma^2|y) = p(\mu|\sigma^2, y)p(\sigma^2|y)$
- *marginal* posterior $p(\sigma^2|y)$

$$
\begin{aligned}
p(\sigma^2|y) &\propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\mu \\
&\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \\
&\qquad \int \exp\left(-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right) d\mu \\
&\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \sqrt{2\pi\sigma^2/n} \\
&\propto (\sigma^2)^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \\
\sigma^2|y &\sim \text{Inv-}\chi^2(n-1, s^2)
\end{aligned}
$$

- Gaussian integral

$$\int p(y|\theta)dy = \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y-\theta)^2\right) dy = 1$$

Gaussian - non-informative prior

- demo3_1.m

Gaussian - non-informative prior

- Compare
  - known mean

$$\sigma^2 | y \sim \text{Inv-}\chi^2(n, v)$$

$$\text{where} \quad v = \frac{1}{n} \sum_{i=1}^{n} (y_i - \theta)^2$$

  - unknown mean

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

$$\text{where} \quad s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Gaussian - non-informative prior

- Factorize

$$p(\mu, \sigma^2|y) = p(\mu|\sigma^2, y)p(\sigma^2|y)$$

- Sample from the joint distribution
  1) sample $(\sigma^2)^{(t)}$ from $p(\sigma^2|y)$
  2) sample $\mu^{(t)}$ from $p(\mu|(\sigma^2)^{(t)}, y)$
  - then $(\mu^{(t)}, (\sigma^2)^{(t)})$ are samples from $p(\mu, \sigma^2|y)$
- demo3_2.m

Gaussian - non-informative prior

- If interested in μ, then marginal posterior $p(\mu|y)$

$$p(\mu|y) = \int_0^\infty p(\mu, \sigma^2|y) d\sigma^2$$

- Transformation

$$z = \frac{A}{2\sigma^2}, \quad \text{where} \quad A = (n-1)s^2 + n(\mu - \bar{y})^2$$

- Recognize unnormalized gamma integral

$$\begin{aligned} p(\mu|y) &\propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) dz \\ &\propto [(n-1)s^2 + n(\mu - \bar{y})^2]^{-n/2} \\ &\propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2}\right] \end{aligned}$$

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

Gaussian - non-informative prior

- Story of Student's t

Gaussian - non-informative prior

- Marginal posterior $p(\mu|y)$

$$p(\mu|y) = \int_0^\infty p(\mu|\sigma^2, y)p(\sigma^2|y)d\sigma^2$$

- see visualization demo3_3.m
- marginal posterior of μ a mixture of normal distributions where mixing density is the marginal posterior of σ^2

Gaussian - posterior predictive distribution

- Posterior predictive distribution given known variance

$$
\begin{aligned}
p(\tilde{y}|\sigma^2, y) &= \int p(\tilde{y}|\mu, \sigma^2, y) p(\mu|\sigma^2, y) d\mu \\
&= \mathrm{N}(\tilde{y}|\bar{y}, (1 + \tfrac{1}{n})\sigma^2)
\end{aligned}
$$

this is up to scaling factor same as $p(\mu|\sigma^2, y)$, and thus

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

- Sampling from the the posterior predictive distribution
  1) sample $(\mu^{(t)}, (\sigma^2)^{(t)})$ from the posterior
  2) sample $\tilde{y}^{(t)}$ from $\mathrm{N}(\mu^{(t)}, (\sigma^2)^{(t)})$
- see visualization in demo3_4.m

Gaussian - example

- Simon Newcomb's light of speed experiment in 1882
- Matlab demo (demo3_5.m)

Gaussian - conjugate prior

- Conjugate prior has to have a form $p(\sigma^2)p(\mu|\sigma^2)$ (see comments3.pdf)
- Handy parametrization

$$\mu|\sigma^2 \sim \mathrm{N}(\mu_0, \sigma^2/\kappa_0)$$
$$\sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$$

which can be written as

$$p(\mu, \sigma^2) = \text{N-Inv-}\chi^2(\mu_0, \sigma_0^2/\kappa_0; \nu_0, \sigma_0^2)$$

- μ and σ^2 are a priori dependent
  - if σ^2 is large, then μ had wide prior

Gaussian - conjugate prior

- Joint posterior (exercise 3.9)

$$p(\mu, \sigma^2|y) = \text{N-Inv-}\chi^2(\mu_n, \sigma_n^2/\kappa_n; \nu_n, \sigma_n^2)$$

where

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

Gaussian - conjugate prior

- Conditional $p(\mu|\sigma^2, y)$

$$
\begin{aligned}
\mu|\sigma^2, y &\sim \mathrm{N}(\mu_n, \sigma^2/\kappa_n) \\
&= \mathrm{N}\left(\frac{\frac{\kappa_0}{\sigma^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}\right)
\end{aligned}
$$

- Marginal $p(\sigma^2|y)$

$$\sigma^2|y \sim \text{Inv-}\chi^2(\nu_n, \sigma_n^2)$$

- Marginal $p(\mu|y)$

$$\mu|y \sim t_{\nu_n}(\mu|\mu_n, \sigma_n^2/\kappa_n)$$

Bioassay

- Example of bivariate model without closed form posterior
- Generalized linear model
  - y is assume to be distributed as $p(y|\mu_i, \phi)$, where μ_i is a location parameter and possible ϕ denotes other parameters for the distribution
  - expectation μ_i depends on $x_i = (x_{i1}, \ldots, x_{ip})$ as

$$g(\mu_i) = \sum_{j=1}^{p} x_{ij}\beta_j$$

where g is a link function

Bioassay

- $y_i|\theta_i \sim \text{Bin}(n_i, \theta_i)$
- $\text{logit}(\theta_i) = \alpha + \beta x_i$
- Likelihood

$$p(y_i|\alpha, \beta, n_i, x_i) \propto \theta_i^{y_i}[1 - \theta_i]^{n_i - y_i}$$

$$p(y_i|\alpha, \beta, n_i, x_i) \propto [\text{logit}^{-1}(\alpha + \beta x_i)]^{y_i}[1 - \text{logit}^{-1}(\alpha + \beta x_i)]^{n_i - y_i}$$

- Posterior

$$p(\alpha, \beta|y, n, x) \propto p(\alpha, \beta) \prod_{i=1}^{n} p(y_i|\alpha, \beta, n_i, x_i)$$

- demo3_6.m (useful when making exercises)