Multiparameter models (Chapter 3)

- Marginalisation
- Conditional and marginal distribution
- Gaussian with mean and variance unkwown
  - non-informative prior
  - conjugate prior
  - semi-conjugate priori
- Multivariate Gaussian
- Multinomial
  - “multivariate binomial”
- Logistic regression (bioassay data)

Slide 1

Marginalisation - marginal distribution

- Joint distribution

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

- Marginalisation

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

$p(\theta_1|y)$ is marginal distribution
where θ_1 is more interesting
  - one of the model parametesrs
  - some other unobserved quantity like future observation

Slide 2

Example - predictive distribution

- Joint distribution

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

- Marginalisation

Slide 3

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

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

Example - predictive distribution

- Often joint distribution is factored

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

and intergal can be approximated with simulation

- E.g. we can get samples from

Slide 4

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

by

1) sampling $\dot{\theta}_t$ from $p(\theta|y)$
2) sampling $\dot{\tilde{y}}_t$ from $p(\tilde{y}|\dot{\theta}_t)$

- then $\dot{\tilde{y}}_t$ are from $p(\tilde{y}|y)$

Example: marginalisation and Gaussian

- Model $y|\mu, \sigma^2 \sim \mathrm{N}(\mu, \sigma^2)$
  $\mu = \theta_1$ and $\sigma^2 = \theta_2$
- Often μ more interesting
- Or if both are interesting marginals can be used to illustrate the joint distribution
- Matlab-demo (esim4_1.m)
  $y = [93, 112, 122, 135, 122, 150, 118, 90, 124, 114]$

Slide 5

Gaussian - last time

- Gaussian with known variance and non-informative prior $p(\theta) \propto 1$

$$p(\theta|y) \approx \mathrm{N}(\theta|\bar{y}, \sigma^2/n)$$

- Gaussian with known mean and non-informative prior $p(\sigma^2) \propto 1/\sigma^2$

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

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

Slide 6

Gaussian - non-informative prior

- Combining single-parameter cases non-informative prior is

$$p(\mu, \sigma^2) \propto 1/\sigma^2$$

Slide 7

Gaussian - non-informative prior

- Joint posterior

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

Slide 8

where

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

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

Gaussian - non-informative prior

Rearrange terms to get sufficient statistics

Slide 9

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

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

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

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

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

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

Gaussian - non-informative prior

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

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

sama as for Gaussian with known variance

Slide 10

Gaussian - non-informative prior

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

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

Slide 11

Gaussian - non-informative prior

- Compare
  - mean known

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

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

  - mean unknown

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

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

Slide 12

Gaussian - non-informative prior

- Factored joint distribution

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

- It is easy to sample from the joint distribution
  1) sample $\dot{\sigma}_t^2$ from $p(\sigma^2|y)$
  2) sample $\dot{\mu}_t$ from $p(\mu|\dot{\sigma}_t^2)$

Slide 13

  - then $(\dot{\mu}_t, \dot{\sigma}_t^2)$ are from $p(\mu, \sigma^2|y)$

- esim4_2.m

Gaussian - non-informative prior

- If μ interesting, then marginal posterior $p(\mu|y)$ is interesting

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

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

Slide 14

- t-distribution has thicker tails than Gaussian
  - thicker tails are due to uncertainty in σ^2

Gaussian - non-informative prior

- Marginal $p(\mu|y)$ can be illustrated by factoring

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

and by sampling $\dot{\sigma}_t^2$ from $p(\sigma^2|y)$ and plotting conditional distributions $p(\mu|\dot{\sigma}_t^2, y)$, whose average approximates the marginalization over σ^2

Slide 15

- esim4_3.m
- $p(\mu|y)$ is *mixture of normal distributions* where mixing is over $p(\sigma^2|y)$

Gaussian - posterior predictive distribution

- Posterior predictive distribution can be obtained in analytic form; first with known variance (last week)

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

Slide 16

this is up to scaling term same as $p(\mu|\sigma^2, y)$, so based on earlier slide

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

- It is easy to sample from the predictive distribution by
  1) sample $(\dot{\mu}_t, \dot{\sigma}_t^2)$ from the posterior
  2) and then sample $\dot{\tilde{y}}_t$ from $\mathrm{N}(\dot{\mu}_t, \dot{\sigma}_t^2)$
- esim4_4.m

Normaalijakauma - esimerkki

- Simon Newcomb measured in 1882 speed of light
  - measured 66 times the time for light to travel 7442m
  - (direct measurement is not time or speed)
- Matlab-demo (esim4_5.m)
- Note that the accuracy of the posterior inference is limited both by the model and the observations

Slide 17

Gaussian - conjugate prior

- Conjugate prior has to be in product form $p(\sigma^2)p(\mu|\sigma^2)$
- Convenient parametrisation is

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

  which can be written as

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

- Note that μ and σ^2 dependent a priori
  - e.g. if σ^2 big, then μ has wide prior distribution

Slide 18

Gaussian - conjugate prior

- Joint posterior (ex 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

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

Slide 19

Gaussian - conjugate prior

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

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

Slide 20

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

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

- Marginal $p(\mu|y)$

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

Semiconjugacy

- For each parameter alone prior is conjugate, but joint prior is not conjugate
  - semi-term is often dropped
- Usage
  - avoids prior dependence
  - if there is no joint-conjugate
  - convenient in hierarchical models when using Gibbs-sampling or Rao-Blackwellization

Slide 21

Gaussian - semiconjugate prior

- Often used semiconjugate prior

$$\mu|\sigma^2 \sim \mathrm{N}(\mu_0, \tau_0^2)$$
$$\sigma^2 \sim \mathrm{Inv}\text{-}\chi^2(\nu_0, \sigma_0^2)$$

where μ and σ^2 a priori independent

Slide 22

Gaussian - semiconjugate prior

- Factored joint distribution

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

- Conditional $p(\mu|\sigma^2, y)$ same as if σ^2 known
- Marginal $p(\sigma^2|y)$

Slide 23

$$p(\sigma^2|y) \propto \tau_n \mathrm{N}(\mu_n|\mu_0, \tau_0^2)\, \mathrm{Inv}\text{-}\chi^2(\sigma^2|\nu_0, \sigma_0^2) \prod_{i=1}^{n} \mathrm{N}(y_i|\mu_n, \sigma^2)$$

Not easy form, but since unidimensional easy to sample from

Multivariate Gaussian

- Multidimensional observations
- Likelihood

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

where μ is d dimensional vector and Σ is $d \times d$ symmetric and positive definite covariance matrix

Slide 24

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

where

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

Multivariate Gaussian

- Covariance matrix Σ
  - diagonal describes the marginal variance
  - non-diagonal elements describe covariance between dimensions (ie linear dependency)
  - scaled covariance is called correlation

Slide 25

Multivariate Gaussian

- Posterior, conditional and marginal distribution generalize from corresponding univariate cases
  - additionally marginal and conditional distributions for subsets of parameters

Slide 26

Multivariate Gaussian - known Σ

- Posterior

$$
\begin{aligned}
\mu|y,\Sigma &\sim \mathrm{N}(\mu_n,\Lambda_n)\\
\mu_n &= (\Lambda_0^{-1}+n\Sigma^{-1})^{-1}(\Lambda_0^{-1}\mu_0+n\Sigma^{-1}\bar{y})\\
\Lambda_n^{-1} &= \Lambda_0^{-1}+n\Sigma^{-1}
\end{aligned}
$$

- Marginal for subset $\mu^{(1)}$

Slide 27

$$
\mu^{(1)}|y \sim \mathrm{N}(\mu_n^{(1)},\Lambda_n^{(11)})
$$

- Conditional for subset $\mu^{(1)}$

$$
\mu^{(1)}|\mu^{(2)},y \sim \mathrm{N}(\mu_n^{(1)}+\beta^{1|2}(\mu^{(2)}-\mu_n^{(2)}),\Lambda^{1|2})
$$

missä $\beta^{1|2} = \Lambda_n^{(12)}\left(\Lambda_n^{(22)}\right)^{-1}$

$$
\Lambda_n^{1|2} = \Lambda_n^{(11)} - \Lambda_n^{(12)}\left(\Lambda_n^{(22)}\right)^{-1}\Lambda_n^{(21)}
$$

Multivariate Gaussian - conjugate prior

- Conjugate prior

$$
\begin{aligned}
\Sigma &\sim \text{Inv-Wishart}_{\nu_0}(\Lambda_0^{-1})\\
\mu|\Sigma &\sim \mathrm{N}(\mu_0,\Sigma/\kappa_0)
\end{aligned}
$$

- Marginal $p(\mu|y)$

Slide 28

$$
\mu|y \sim t_{\nu_n-d+1}(\mu_n,\Lambda_n/(\kappa_n(\nu_n-d+1)))
$$

Multivariate Gaussian - non-informative prior

- Multivariate Jeffreys' prior

$$p(\mu, \Sigma) \propto |\Sigma|^{-(d+1)/2}$$

- Marginal $p(\mu|y)$

$$\mu|y \sim t_{n-d}(\bar{\mu}, S/(n(n-d)))$$

Slide 29

Multinomial

- Multinomial is generalization of binomial when more than classes
- Conjugate prior is Dirichlet which is multivariate generalization of Beta

Slide 30

Bioassay

- Example with no closed from posterior
- Generalized linear model
  - refers to linear model with non-Gaussian observation model
  - example of regression model

Slide 31

Bioassay example

Dose, x_i (log g/ml)	Number of animals, n_i	Number of deaths, y_i
-0.86	5	0
-0.30	5	1
-0.05	5	3
0.73	5	5

- LD50?
  - lethal dose 50%
  - half of the animals die with this dose
  - the authorities require classification of chemicals
    - e.g., 1984 EEC directive had 4 toxity levels
- Using Bayesian approach
  - less animals needed
  - allows sequential design-of-experiment

Slide 32

Bioassay example

Dose, x_i (log g/ml)	Number of animals, n_i	Number of deaths, y_i
-0.86	5	0
-0.30	5	1
-0.05	5	3
0.73	5	5

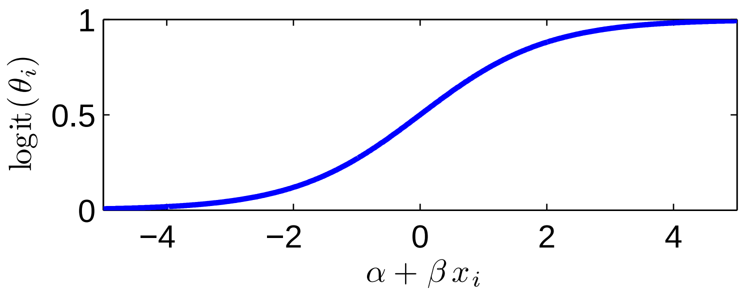
- Binomial model for number of deaths given probability
  $y_i|\theta_i \sim \mathrm{Bin}(n_i, \theta_i)$
- How θ_i is related to dose x_i?

Slide 33

Bioassay example

- Logistic regression describes how probability depends on dose
  $\mathrm{logit}(\theta_i) = \alpha + \beta x_i$
  - $\mathrm{logit}(\theta_i) = \log(\theta_i/(1-\theta_i))$
  - transformation $(-\infty, \infty) \rightarrow (0, 1)$
  - has also information theoretic justification
    - there are other transformations
    - logit is one of the most commonly used

Slide 34



Bioassay example

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

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

Slide 35

- Posterior

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

- esim4_6.m (contains some help for making exercise 3.11)