

MAS3902

NEWCASTLE UNIVERSITY

SCHOOL OF MATHEMATICS, STATISTICS & PHYSICS

SEMESTER 1 2017/18

MAS3902

Bayesian Inference

(This document contains solutions)

Time allowed: 2 hours

Candidates should attempt all questions. Marks for each question are indicated. However you are advised that marks indicate the relative weight of individual questions, they do not correspond directly to marks on the University scale.

There are TWO questions in Section A and TWO questions in Section B.

Information about standard distributions are given at the end of the paper. Calculators may not be used.

SECTION A

A1. A random sample of n people are each given k scratch cards and the number of winning cards (for each person) is recorded as $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$. We can model these data as a random sample from a binomial $Bin(k, \theta)$ distribution, where k is known.

(a) Show that the likelihood function is

$$f(\mathbf{x}|\theta) \propto \theta^{n\bar{x}}(1 - \theta)^{n(k-\bar{x})}$$

where $\bar{x} = \sum x_i/n$ is the sample mean number of winning cards per person.

Answer:

The likelihood function is

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n \binom{k}{x_i} \theta^{x_i} (1 - \theta)^{k-x_i} \propto \theta^{n\bar{x}} (1 - \theta)^{n(k-\bar{x})}$$

where $\bar{x} = \sum x_i/n$ is the sample mean number of winning cards per person.

[4 marks]

(b) Suppose that your prior beliefs about θ are described by a $Beta(g, h)$ distribution. Determine and name your posterior distribution for θ given the data $\mathbf{x}$.

Answer:

The posterior density is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto \pi(\theta) f(\mathbf{x}|\theta) \\ &\propto \theta^{g-1} (1 - \theta)^{h-1} \times \theta^{n\bar{x}} (1 - \theta)^{n(k-\bar{x})}, \quad 0 < \theta < 1 \\ &\propto \theta^{g+n\bar{x}-1} (1 - \theta)^{h+nk-n\bar{x}-1}, \quad 0 < \theta < 1 \end{aligned}$$

$$i.e. \theta|\mathbf{x} \sim Beta(G = g + n\bar{x}, H = h + nk - n\bar{x}).$$

[5 marks]

(c) Show that vague prior knowledge can be represented as $g \rightarrow 0$ and $h \rightarrow 0$. Hint: re-parameterise the distribution using its mean $m = g/(g + h)$ and $s = g + h + 1$.

Answer:

We represent vague prior information by taking a conjugate prior distribution with large variance. Here the conjugate prior is a beta distribution. The prior variance is

$$\text{Var}(\theta) = \frac{gh}{(g+h)^2(g+h+1)} = \frac{m(1-m)}{s},$$

where $m = g/(g+h)$ and $s = g+h+1$. This is a 1-1 transformation and so we have a free choice of values for m and s subject to the restrictions $0 < m < 1$ and $s > 1$. Clearly, $\text{Var}(\theta)$ is maximised when $m(1-m)$ is maximised (at $m = 1/2$) and $1/s$ is maximised (in the limit as $s \rightarrow 1$). In terms of g and h , these conditions are $g = h$ and $g+h+1 \rightarrow 1$, that is, $g = h \rightarrow 0$.

[6 marks]

- (d) What is the posterior distribution for θ when taking vague prior knowledge? Show that the posterior mean is the sample proportion of successful scratch cards.

Answer:

The posterior distribution under vague prior information is

$$\theta|\mathbf{x} \sim \text{Beta}(n\bar{x}, nk - n\bar{x}).$$

The posterior mean is

$$E(\theta|\mathbf{x}) = \frac{n\bar{x}}{n\bar{x} + nk - n\bar{x}} = \frac{\bar{x}}{k},$$

the sample proportion of successful scratch cards.

[5 marks]

[Total: 20 marks]

A2. A random sample $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ is taken from a log-normal $LN(\mu, \sigma^2)$ distribution.

(a) Show that the likelihood function is

$$f(\mathbf{x}|\mu, \sigma) \propto \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\log x_i - \mu)^2 \right\}.$$

Answer:

The likelihood function is

$$\begin{aligned} f(\mathbf{x}|\mu, \sigma) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma x_i} \exp \left\{ -\frac{1}{2\sigma^2} (\log x_i - \mu)^2 \right\} \\ &\propto \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\log x_i - \mu)^2 \right\}. \end{aligned}$$

[2 marks]

(b) Show that

$$\begin{aligned} \frac{\partial^2}{\partial \mu^2} \log f(\mathbf{x}|\mu, \sigma) &= -\frac{n}{\sigma^2} \\ \frac{\partial^2}{\partial \mu \partial \sigma} \log f(\mathbf{x}|\mu, \sigma) &= -\frac{2}{\sigma^3} \sum_{i=1}^n (\log x_i - \mu) \\ \frac{\partial^2}{\partial \sigma^2} \log f(\mathbf{x}|\mu, \sigma) &= \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^n (\log x_i - \mu)^2. \end{aligned}$$

Answer:

We have

$$\begin{aligned}
 \log f(\mathbf{x}|\mu, \sigma) &= -n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (\log x_i - \mu)^2 \\
 \implies \frac{\partial}{\partial \mu} \log f(\mathbf{x}|\mu, \sigma) &= \frac{1}{\sigma^2} \sum_{i=1}^n (\log x_i - \mu) \\
 \frac{\partial}{\partial \sigma} \log f(\mathbf{x}|\mu, \sigma) &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (\log x_i - \mu)^2 \\
 \implies \frac{\partial^2}{\partial \mu^2} \log f(\mathbf{x}|\mu, \sigma) &= -\frac{n}{\sigma^2} \\
 \frac{\partial^2}{\partial \mu \partial \sigma} \log f(\mathbf{x}|\mu, \sigma) &= -\frac{2}{\sigma^3} \sum_{i=1}^n (\log x_i - \mu) \\
 \frac{\partial^2}{\partial \sigma^2} \log f(\mathbf{x}|\mu, \sigma) &= \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^n (\log x_i - \mu)^2.
 \end{aligned}$$

[6 marks]

- (c) Hence show that, for large n , the (approximate) posterior distribution for $(\mu, \sigma)^T$ is

$$\begin{pmatrix} \mu \\ \sigma \end{pmatrix} \Big| \mathbf{x} \sim N \left(\begin{pmatrix} \bar{x}_{\log} \\ s_{\log} \end{pmatrix}, \begin{pmatrix} \frac{s_{\log}^2}{n} & 0 \\ 0 & \frac{s_{\log}^2}{2n} \end{pmatrix} \right),$$

where $\bar{x}_{\log} = \sum_{i=1}^n \log x_i / n$ and $s_{\log}^2 = \sum_{i=1}^n (\log x_i - \bar{x}_{\log})^2 / n$ are the sample mean and variance of the $\log x_i$.

Answer:

Now

$$\frac{\partial}{\partial \mu} \log f(\mathbf{x}|\mu, \sigma) = \frac{\partial}{\partial \sigma} \log f(\mathbf{x}|\mu, \sigma) = 0 \implies \hat{\mu} = \bar{x}_{\log}, \quad \hat{\sigma} = s_{\log}.$$

Therefore

$$J_{11}(\hat{\mu}, \hat{\sigma}) = \frac{n}{\hat{\sigma}^2} = \frac{n}{s_{\log}^2},$$

$$J_{12}(\hat{\mu}, \hat{\sigma}) = \frac{2}{\hat{\sigma}^3} \sum_{i=1}^n (\log x_i - \hat{\mu}) = 0,$$

$$J_{22}(\hat{\mu}, \hat{\sigma}) = -\frac{n}{\hat{\sigma}^2} + \frac{3}{\hat{\sigma}^4} \sum_{i=1}^n (\log x_i - \hat{\mu})^2 = -\frac{n}{s_{\log}^2} + \frac{3}{s_{\log}^4} n s_{\log}^2 = \frac{2n}{s_{\log}^2},$$

and so

$$J(\hat{\mu}, \hat{\sigma}) = \begin{pmatrix} \frac{n}{s_{\log}^2} & 0 \\ 0 & \frac{2n}{s_{\log}^2} \end{pmatrix} \Rightarrow J(\hat{\mu}, \hat{\sigma})^{-1} = \begin{pmatrix} \frac{s_{\log}^2}{n} & 0 \\ 0 & \frac{s_{\log}^2}{2n} \end{pmatrix}.$$

Therefore, for large n , the (approximate) posterior distribution for (μ, σ) is

$$\begin{pmatrix} \mu \\ \sigma \end{pmatrix} \Big| \mathbf{x} \sim N \left(\begin{pmatrix} \bar{x}_{\log} \\ s_{\log} \end{pmatrix}, \begin{pmatrix} \frac{s_{\log}^2}{n} & 0 \\ 0 & \frac{s_{\log}^2}{2n} \end{pmatrix} \right).$$

[8 marks]

- (d) Hence, write down a $100(1 - \alpha)\%$ posterior HDI for σ in terms of the upper percentage points of the standard normal distribution, z_p .

Answer:

From (c), for large n , $\sigma | \mathbf{x} \sim N(s_{\log}, s_{\log}^2/(2n))$. This distribution is symmetric about its mean and so an HDI is an equi-tailed interval. Therefore a $100(1 - \alpha)\%$ posterior HDI for σ is

$$\left(s_{\log} - \frac{z_{\alpha/2} s_{\log}}{\sqrt{2n}}, s_{\log} + \frac{z_{\alpha/2} s_{\log}}{\sqrt{2n}} \right).$$

[4 marks]

[Total: 20 marks]

SECTION B

B3. A random sample $\mathbf{x} = (x_1, \dots, x_n)^T$ is taken from a negative binomial $NegBin(k, \theta)$ distribution, with sample mean $\bar{x}$. Suppose the prior belief about θ follows a mixture beta distribution,

$$\theta \sim p_1 \text{Beta}(g_1, g_1) + p_2 \text{Beta}(g_2, g_2).$$

(a) Show that the mean and variance of the prior distribution are

$$E(\theta) = (p_1 + p_2)/2 \text{ and } Var(\theta) = \sum_{i=1}^2 p_i \frac{g_i + 1}{2(2g_i + 1)} - \frac{(p_1 + p_2)^2}{4}.$$

Answer:

For the prior component i , its mean and variance is

$$E_i(\theta) = \frac{g_i}{g_i + g_i} = \frac{1}{2} \text{ and } Var_i(\theta) = \frac{g_i^2}{(g_i + g_i)^2(g_i + g_i + 1)} = \frac{1}{4(2g_i + 1)}.$$

Then for the mixture prior, its mean is

$$E(\theta) = \sum_{i=1}^2 p_i E_i(\theta) = \frac{p_1 + p_2}{2}.$$

Meanwhile, since

$$\begin{aligned} E(\theta^2) &= \sum_{i=1}^2 p_i E_i(\theta^2) = \sum_{i=1}^2 p_i [Var_i(\theta) + E_i(\theta)^2] \\ &= \sum_{i=1}^2 p_i \left[\frac{1}{4(2g_i + 1)} + \frac{1}{4} \right] \\ &= \sum_{i=1}^2 p_i \frac{g_i + 1}{2(2g_i + 1)}, \end{aligned}$$

the variance of prior distribution is

$$Var(\theta) = E(\theta^2) - E(\theta)^2 = \sum_{i=1}^2 p_i \frac{g_i + 1}{2(2g_i + 1)} - \frac{(p_1 + p_2)^2}{4}.$$

[8 marks]

(b) Show that the posterior distribution is in the form of

$$p_1^* \text{Beta}(G_1, H_1) + p_2^* \text{Beta}(G_2, H_2),$$

and show that $G_i = nk + g_i$ and $H_i = n\bar{x} - nk + g_i$, $i = 1, 2$.

Answer:

Since the prior distribution is a mixture, the posterior distribution is in the form of $p_1^* \pi_1(\theta | \mathbf{x}) + p_2^* \pi_2(\theta | \mathbf{x})$ where $\pi_i(\theta | \mathbf{x})$ is the posterior distribution

$$\pi_i(\theta | \mathbf{x}) = \frac{\pi_i(\theta) f(\mathbf{x} | \theta)}{f_i(\mathbf{x})},$$

where $\pi_i(\theta)$ is the density of $\text{Beta}(g_i, g_i)$, $i = 1, 2$. The likelihood function is

$$f(\mathbf{x} | \theta) = \prod_{j=1}^n \binom{x_j - 1}{k - 1} \theta^k (1 - \theta)^{x_j - k}.$$

Then for $i = 1, 2$,

$$\begin{aligned} \pi_i(\theta | \mathbf{x}) &\propto \pi_i(\theta) f(\mathbf{x} | \theta) \\ &\propto \theta^{g_i-1} (1 - \theta)^{g_i-1} \times \theta^{nk} (1 - \theta)^{\sum_{j=1}^n x_j - nk}, \quad 0 < \theta < 1, \\ &\propto \theta^{nk+g_i-1} (1 - \theta)^{n\bar{x}-nk+g_i-1}, \quad 0 < \theta < 1, \end{aligned}$$

and $\pi_i(\theta | \mathbf{x})$ must be the density of $\text{Beta}(G_i = nk + g_i, H_i = n\bar{x} - nk + g_i)$. Therefore the posterior distribution is in the form of $p_1^* \text{Beta}(G_1, H_1) + p_2^* \text{Beta}(G_2, H_2)$.

[8 marks]

(c) Show that $p_1^* + p_2^* = 1$. Not a good question.

Answer:

Because the integral of the posterior distribution needs to be 1, we have

$$\begin{aligned} &\int_0^1 [p_1^* \pi_1(\theta | \mathbf{x}) + p_2^* \pi_2(\theta | \mathbf{x})] d\theta = 1 \\ \implies &p_1^* \int_0^1 \pi_1(\theta | \mathbf{x}) d\theta + p_2^* \int_0^1 \pi_2(\theta | \mathbf{x}) d\theta = 1 \\ \implies &p_1^* + p_2^* = 1. \end{aligned}$$

[4 marks]

- (d) Find expressions for p_1^* and p_2^* .

Answer:

From the results of (b),

$$f_i(\mathbf{x}) = \frac{\pi(\theta)f(\mathbf{x} | \theta)}{\pi_i(\theta | \mathbf{x})} = \frac{\prod_{j=1}^n \binom{x_j - 1}{k - 1} B(G_i, H_i)}{B(g_i, g_i)},$$

where $B(g, h)$ is the beta function. Then

$$p_i^* = \frac{p_i f_i(\mathbf{x})}{\sum_{i=1}^2 p_i f_i(\mathbf{x})} = \frac{p_i B(G_i, H_i) / B(g_i, g_i)}{\sum_{i=1}^2 p_i B(G_i, H_i) / B(g_i, g_i)}.$$

[4 marks]

[Total: 24 marks]

- B4.** Suppose you have a random sample $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ from an Exponential-gamma $EG(\alpha, \lambda)$ distribution, where α and λ are unknown. Your prior distribution has $\alpha \sim Ga(a, b)$ and $\lambda \sim Ga(c, d)$, independently, for known values a, b, c and d .

- (a) Show that the likelihood function is

$$f(\mathbf{x} | \alpha, \lambda) = \lambda^{n\alpha} \exp(\alpha n \bar{x} - \lambda n \bar{x}_{\text{exp}}) / \Gamma(\alpha)^n,$$

where $\bar{x} = \sum_{i=1}^n x_i / n$ and $\bar{x}_{\text{exp}} = \sum_{i=1}^n \exp(x_i) / n$ are the sample means of x and $\exp(x)$.

Answer:

The likelihood function is

$$\begin{aligned} f(\mathbf{x} | \alpha, \theta) &= \prod_{i=1}^n \frac{\lambda^\alpha \exp\{\alpha x_i - \lambda \exp(x_i)\}}{\Gamma(\alpha)} \\ &= \frac{\lambda^{n\alpha} \exp(\alpha n \bar{x} - \lambda n \bar{x}_{\text{exp}})}{\Gamma(\alpha)^n}. \end{aligned}$$

[3 marks]

(b) Show that your posterior density is, for $\alpha > 0$ and $\lambda > 0$,

$$\pi(\alpha, \lambda | \mathbf{x}) \propto \alpha^{a-1} \lambda^{n\alpha+c-1} \exp\{\alpha(n\bar{x} - b) - \lambda(n\bar{x}_{\text{exp}} + d)\} / \Gamma(\alpha)^n.$$

Answer:

Using Bayes' Theorem, the posterior density is, for $\alpha > 0$ and $\lambda > 0$,

$$\begin{aligned} \pi(\alpha, \lambda | \mathbf{x}) &\propto \pi(\alpha, \lambda) f(\mathbf{x} | \alpha, \lambda) \\ &\propto \alpha^{a-1} e^{-b\alpha} \times \lambda^{c-1} e^{-d\lambda} \times \frac{\lambda^{n\alpha} \exp(\alpha n\bar{x} - \lambda n\bar{x}_{\text{exp}})}{\Gamma(\alpha)^n}, \\ &\propto \frac{\alpha^{a-1} \lambda^{n\alpha+c-1} \exp\{\alpha(n\bar{x} - b) - \lambda(n\bar{x}_{\text{exp}} + d)\}}{\Gamma(\alpha)^n}. \end{aligned}$$

[4 marks]

(c) Suppose you want to construct a Metropolis–Hastings algorithm which samples this posterior distribution by using an independent proposal for (α, λ) . Suppose the proposal is made using the prior distribution, that is, $(\alpha^*, \lambda^*) | (\alpha, \lambda) \sim \pi(\alpha^*, \lambda^*)$. What is the acceptance probability for the proposed value (α^*, λ^*) ?

Answer:

If we use the prior distribution as the independent proposal, the proposal (α^*, λ^*) has acceptance probability $\min(1, A)$ where

$$\begin{aligned} A &= \frac{\pi(\alpha^*, \lambda^* | \mathbf{x})}{\pi(\alpha, \lambda | \mathbf{x})} \times \frac{\pi(\alpha, \lambda)}{\pi(\alpha^*, \lambda^*)} \\ &= \frac{f(\mathbf{x} | \alpha^*, \lambda^*)}{f(\mathbf{x} | \alpha, \lambda)} \\ &= \left\{ \frac{\lambda^{*\alpha^*}}{\lambda^\alpha} / \frac{\Gamma(\alpha^*)}{\Gamma(\alpha)} \right\}^n \exp\{(\alpha^* - \alpha)n\bar{x} - (\lambda^* - \lambda)n\bar{x}_{\text{exp}}\}. \end{aligned}$$

[4 marks]

(d) Determine the posterior conditional densities for $\alpha | \lambda$ and $\lambda | \alpha$ up to a multiplicative constant.

Answer:

The posterior conditional density for α is, for $\alpha > 0$

$$\begin{aligned}
 \pi(\alpha|\lambda, \mathbf{x}) &= \frac{\pi(\alpha, \lambda|\mathbf{x})}{\pi(\lambda|\mathbf{x})} \\
 &\propto \pi(\alpha, \lambda|\mathbf{x}) \\
 &\propto \alpha^{a-1} \lambda^{n\alpha+c-1} \exp\{\alpha(n\bar{x} - b) - \lambda(n\bar{x}_{\text{exp}} + d)\} / \Gamma(\alpha)^n \\
 &\propto \alpha^{a-1} \lambda^{n\alpha} \exp\{\alpha(n\bar{x} - b)\} / \Gamma(\alpha)^n \\
 &\propto \alpha^{a-1} \exp\{\alpha(n\bar{x} - b + n \log \lambda)\} / \Gamma(\alpha)^n.
 \end{aligned}$$

The posterior conditional density for λ is, for $\lambda > 0$

$$\begin{aligned}
 \pi(\lambda|\alpha, \mathbf{x}) &= \frac{\pi(\alpha, \lambda|\mathbf{x})}{\pi(\alpha|\mathbf{x})} \\
 &\propto \pi(\alpha, \lambda|\mathbf{x}) \\
 &\propto \alpha^{a-1} \lambda^{n\alpha+c-1} \exp\{\alpha(n\bar{x} - b) - \lambda(n\bar{x}_{\text{exp}} + d)\} / \Gamma(\alpha)^n \\
 &\propto \lambda^{n\alpha+c-1} \exp\{-(n\bar{x}_{\text{exp}} + d)\lambda\},
 \end{aligned}$$

i.e. $\lambda|\alpha, \mathbf{x} \sim \text{Ga}(n\alpha + c, n\bar{x}_{\text{exp}} + d)$.

[8 marks]

You decide to develop an MCMC algorithm with separate steps for each parameter to simulate realisations from the posterior distribution. You have available a simulator which generates values from a gamma distribution.

- (e) Why might a Metropolis within Gibbs algorithm be appropriate for simulating realisations from the posterior distribution?

Answer:

The distribution of $\lambda|\alpha, \mathbf{x}$ is standard but that for $\alpha|\lambda, \mathbf{x}$ is non-standard. Therefore we can use a Metropolis within Gibbs algorithm which uses a Gibbs step for λ and a M-H step for α .

[3 marks]

- (f) You decide to use a normal $N(\alpha, \Sigma_\alpha)$ random walk proposal for α , that is, make proposals using $\alpha^*|\alpha \sim N(\alpha, \Sigma_\alpha)$. What is the acceptance probability for the proposed value α^* ?

Answer:

If we use a normal random walk proposal with $q(\alpha^*|\alpha) = N(\alpha, \Sigma_\alpha)$,

the proposal α^* has acceptance probability $\min(1, A)$ where

$$\begin{aligned}
 A &= \frac{\pi(\alpha^*|\lambda, \mathbf{x})}{\pi(\alpha|\lambda, \mathbf{x})} \times \frac{q(\alpha|\alpha^*)}{q(\alpha^*|\alpha)} \\
 &= \frac{\pi(\alpha^*|\lambda, \mathbf{x})}{\pi(\alpha|\lambda, \mathbf{x})} \\
 &= \frac{\alpha^{*a-1} \exp\{\alpha^*(n\bar{x} - b + n \log \lambda)\} / \Gamma(\alpha^*)^n}{\alpha^{a-1} \exp\{\alpha(n\bar{x} - b + n \log \lambda)\} / \Gamma(\alpha)^n} \\
 &= \left(\frac{\alpha^*}{\alpha}\right)^{a-1} / \left\{ \frac{\Gamma(\alpha^*)}{\Gamma(\alpha)} \right\}^n \exp\{(\alpha^* - \alpha)(n\bar{x} - b + n \log \lambda)\}.
 \end{aligned}$$

[4 marks]

- (g) Write down the steps in your MCMC algorithm for simulating realisations from the posterior distribution.

Answer:

The MCMC algorithm is

1. Initialise the iteration counter to $j = 1$, and initialise the chain to $\alpha^{(0)}$, a realisations from its $Ga(a, b)$ prior distribution or use the prior mean a/b or moment/mle estimates.
2. Obtain a new value

$$\lambda^{(j)} \sim Ga(n\alpha^{(j-1)} + c, n\bar{x}_{\text{exp}} + d).$$

3. Generate a proposed value $\alpha^* \sim N(\alpha^{(j-1)}, \Sigma_\alpha)$
4. Evaluate the acceptance probability $\min(1, A)$ at α^* , $\alpha^{(j-1)}$ and $\alpha^{(j)}$
5. Set $\alpha^{(j)} = \alpha^*$ probability $\min(1, A)$, and set $\alpha^{(j)} = \alpha^{(j-1)}$ otherwise.
6. Change the counter from j to $j + 1$ and return to step 2.

[10 marks]

[Total: 36 marks]

Distributions for data

Binomial distribution: If $X \sim \text{Bin}(k, \theta)$ then it has probability function

$$f(x|\theta) = \binom{k}{x} \theta^x (1 - \theta)^{k-x}, \quad x = 0, 1, \dots, k, \quad 0 < \theta < 1.$$

Also, $E(X) = k\theta$ and $\text{Var}(X) = k\theta(1 - \theta)$.

Negative binomial distribution: If $X|\theta \sim \text{NegBin}(k, \theta)$ then it has probability function

$$f(x|\theta) = \binom{x-1}{k-1} \theta^k (1 - \theta)^{x-k}, \quad x = k, k+1, \dots,$$

where k is a positive integer and $0 < \theta < 1$. Also $E(X) = k/\theta$ and $\text{Var}(X) = k(1 - \theta)/\theta^2$.

Normal distribution: If $X \sim N(\mu, 1/\tau)$ then it has density

$$f(x|\mu, \tau) = \left(\frac{\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{\tau}{2}(x - \mu)^2\right\}, \quad -\infty < x < \infty$$

where $-\infty < \mu < \infty$ and $\tau > 0$. Also, $E(X) = \mu$ and $\text{Var}(X) = 1/\tau$.

Exponential-gamma distribution: If $X \sim \text{EG}(\alpha, \lambda)$ then it has probability density

$$f(x|\alpha, \lambda) = \lambda^\alpha \exp\{\alpha x - \lambda \exp(x)\} / \Gamma(\alpha),$$

where $\alpha > 0$ and $\lambda > 0$.

Log-normal distribution: If $X \sim \text{LN}(\mu, \sigma^2)$ then and X has density

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{1}{2\sigma^2}(\log x - \mu)^2\right\}, \quad x > 0$$

where $-\infty < \mu < \infty$ and $\sigma > 0$. Also, $E(X) = \exp\{\mu + \sigma^2/2\}$, $\text{Var}(X) = \{\exp(\sigma^2) - 1\} \exp(2\mu + \sigma^2)$. Further $\log X \sim N(\mu, \sigma^2)$ and so $E(\log X) = \mu$ and $\text{Var}(\log X) = \sigma^2$.

Prior distributions

Exponential distribution: If $\theta \sim \text{Exp}(h)$ then it has density

$$\pi(\theta) = h e^{-h\theta}, \quad \theta > 0,$$

where $h > 0$. Also, $E(\theta) = 1/h$ and $\text{Var}(\theta) = 1/h^2$.

Beta distribution: If $\theta \sim \text{Beta}(g, h)$ then it has density

$$\pi(\theta) = \frac{\theta^{g-1}(1-\theta)^{h-1}}{B(g, h)}, \quad 0 < \theta < 1, \quad g > 0, h > 0.$$

Also, $E(\theta) = g/(g+h)$ and $\text{Var}(\theta) = gh/\{(g+h)^2(g+h+1)\}$.

Gamma distribution: If $\theta \sim \text{Ga}(g, h)$ then it has density

$$\pi(\theta) = \frac{h^g \theta^{g-1} e^{-h\theta}}{\Gamma(g)}, \quad \theta > 0, \quad g > 0, h > 0.$$

Also, $E(\theta) = g/h$ and $\text{Var}(\theta) = g/h^2$.

Normal distribution: If $\mu \sim N(b, 1/d)$ then it has density

$$\pi(\mu) = \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu - b)^2\right\}, \quad -\infty < \mu < \infty$$

where $-\infty < b < \infty$ and $c > 0$. Also, $E(\mu) = b$ and $\text{Var}(\mu) = 1/d$.

Normal-gamma distribution: If $\begin{pmatrix} \mu \\ \tau \end{pmatrix} \sim \text{NGa}(b, c, g, h)$ then it has density

$$\pi(\mu, \tau) \propto \tau^{g-\frac{1}{2}} \exp\left\{-\frac{\tau}{2} [c(\mu - b)^2 + 2h]\right\}, \quad -\infty < \mu < \infty, \tau > 0$$

where $-\infty < b < \infty$ and $c, g, h > 0$. Also, $\mu|\tau \sim N\left(b, \frac{1}{c\tau}\right)$, $\tau \sim \text{Ga}(g, h)$

and $\mu \sim t_{2g}\left(b, \frac{h}{gc}\right)$.

Multivariate normal distribution: If $\boldsymbol{\theta} \sim N_p(\mathbf{b}, c)$ then it has density

$$\pi(\boldsymbol{\theta}) \propto (2\pi)^{-p/2} |c|^{-1/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \mathbf{b})^T c^{-1} (\boldsymbol{\theta} - \mathbf{b}) \right\}, \quad \boldsymbol{\theta} \in \mathbb{R}^p$$

where $\mathbf{b} \in \mathbb{R}^p$ and c is a symmetric positive-definite symmetric matrix. Also, $E(\boldsymbol{\theta}) = \mathbf{b}$, $Var(\boldsymbol{\theta}) = c$ and $\theta_i \sim N(b_i, c_{ii})$, $i = 1, \dots, p$.

Multivariate normal-gamma distribution: If $\begin{pmatrix} \boldsymbol{\beta} \\ \tau \end{pmatrix} \sim N_p Ga(\mathbf{b}, c, g, h)$ then it has density

$$\pi(\boldsymbol{\beta}, \tau) \propto \tau^{g + \frac{p}{2} - 1} \exp \left\{ -\frac{\tau}{2} [(\boldsymbol{\beta} - \mathbf{b})^T c (\boldsymbol{\beta} - \mathbf{b}) + 2h] \right\}, \quad \boldsymbol{\beta} \in \mathbb{R}^p, \tau > 0$$

where $\mathbf{b} \in \mathbb{R}^p$, $g, h > 0$ and c is a symmetric positive-definite symmetric matrix. Also, $\boldsymbol{\beta}|\tau \sim N_p(\mathbf{b}, (c\tau)^{-1})$, $\tau \sim Ga(g, h)$ and $\boldsymbol{\beta} \sim t_{2g} \left(\mathbf{b}, \frac{h}{g} c^{-1} \right)$.

Multivariate t distribution: If $\boldsymbol{\beta} \sim t_a(\mathbf{b}, c)$ then it has density

$$\pi(\boldsymbol{\beta}) = \frac{\Gamma\left(\frac{a+p}{2}\right) |c|^{-1/2}}{(a\pi)^{p/2} \Gamma\left(\frac{a}{2}\right)} \left\{ 1 + \frac{(\boldsymbol{\beta} - \mathbf{b})^T c^{-1} (\boldsymbol{\beta} - \mathbf{b})}{a} \right\}^{-\frac{a+p}{2}}, \quad \boldsymbol{\beta} \in \mathbb{R}^p$$

where $a > 0$, $\mathbf{b} \in \mathbb{R}^p$ and c is a symmetric positive semi-definite matrix. Also, $E(\boldsymbol{\beta}) = \mathbf{b}$, $Var(\boldsymbol{\beta}) = \frac{ac}{a-2}$, if $a \geq 2$ and $\beta_i \sim t_a(b_i, c_{ii})$, $i = 1, \dots, p$.

THE END