MATH 3421: Formula Sheet

Beta(α, β) distribution

If $X|\alpha, \beta \sim \text{Beta}(\alpha, \beta)$ then it has density

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < x < 1, \quad \alpha > 0, \beta > 0.$$

Also, $\text{E}(X|\alpha, \beta) = \alpha/(\alpha+\beta)$ and $\text{Var}(X|\alpha, \beta) = \alpha\beta/\{(\alpha+\beta)^2(\alpha+\beta+1)\}$.

Binomial Bin(k, θ) distribution

If $X|\theta \sim \text{Bin}(k, \theta)$ then it has probability mass function

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

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

Exponential Exp(λ) distribution

If $X|\lambda \sim \text{Exp}(\lambda)$, then it has density

$$f(x|\lambda) = \lambda e^{-\lambda x}, \quad x > 0, \quad \lambda > 0.$$

Also, $\text{E}(X|\lambda) = 1/\lambda$ and $\text{Var}(X|\lambda) = 1/\lambda^2$.

Gamma Gamma(α, β) distribution

If $X|\alpha, \beta \sim \text{Ga}(\alpha, \beta)$, then it has density

$$f(x|\alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \quad x > 0, \quad \alpha > 0, \beta > 0.$$

Also, $\text{E}(X|\alpha, \beta) = \alpha/\beta$ and $\text{Var}(X|\alpha, \beta) = \alpha/\beta^2$.

Geometric Geo(θ) distribution

If $X|\theta \sim \text{Geo}(\theta)$, then it has probability mass function

$$\Pr(x|\theta) = (1-\theta)^x \theta, \quad x \geq 0, \quad 0 < \theta < 1.$$

Also, $\text{E}(X|\theta) = (1-\theta)/\theta$ and $\text{Var}(X|\theta) = (1-\theta)/\theta^2$.

Poisson Pois(λ) distribution

If $X|\theta \sim \text{Pois}(\lambda)$, then it has probability mass function

$$\Pr(X = x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots, \quad \lambda > 0.$$

Also, $\text{E}(X|\lambda) = \lambda$ and $\text{Var}(X|\lambda) = \lambda$.

SECTION A

Q1 Consider a posterior density $\pi(\theta|\boldsymbol{x}) = k\pi(\theta)f(\boldsymbol{x}|\theta)$ where $\theta \in \mathcal{S} \subseteq \mathbb{R}$, $\pi(\cdot)$ is the prior density ascribed to θ, $f(\boldsymbol{x}|\cdot)$ is the likelihood function and the constant k can be evaluated analytically. Suppose that interest lies in estimating an expectation of the form

$$\mu_h = \mathbb{E}_{\pi(\theta|\boldsymbol{x})}[h(\theta)] = \int_{\mathcal{S}} h(\theta)\pi(\theta|\boldsymbol{x})d\theta$$

for some function $h(\cdot)$.

1.1 Write down the form of k in terms of the prior and likelihood.

1.2 Suppose that $\pi(\theta|\boldsymbol{x})$ is difficult to simulate from, but there exists a density $g(\theta)$ which is easy to simulate from, and $g(\theta) > 0$ for all θ with $\pi(\theta|\boldsymbol{x}) \geq 0$.

(a) Give the steps of a Monte Carlo importance sampling algorithm to estimate μ_h using N independent draws from $g(\cdot)$. Write down the form of the estimator explicitly.

(b) Show that the estimator is unbiased.

1.3 Suppose now that k is unknown.

(a) Write down a self-normalised estimator of μ_h that uses N independent draws from the prior $\pi(\cdot)$. Give the weight function explicitly.

(b) Using your answer to **1.3**(a), write down a discrete approximation of $\pi(\theta|\boldsymbol{x})$.

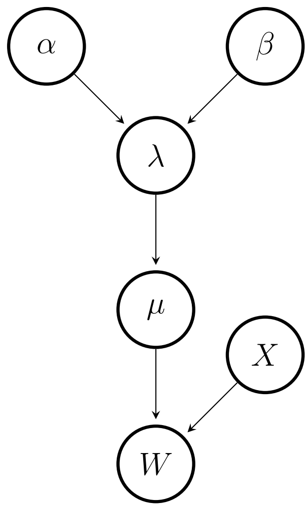
Q2 A faulty machine is known to produce the number of widgets specified by the operator, but it also produces an extra number of widgets due to a fault in the wiring. The number of extra widgets is randomly drawn from a Geometric distribution with parameter 0.8.

2.1 The factory manager is about to find out how many widgets the machine has produced overnight, but they do not know how many the operator specified. They believe that the operator was equally likely to specify 5, 10 or 15, and they know that the machine cannot handle any other number. Suggest a suitable probability distribution that encapsulates his beliefs.

2.2 Derive the factory manager's preposterior distribution for the number of widgets the machine has produced overnight. What does he believe the most likely number is?

2.3 They find that 11 widgets have been produced overnight. Write out the factory manager's posterior distribution for the number of widgets that the operator specified.

2.4 For subsequent analyses, an extended model with six random variables will be used. The proposed dependence structure is represented by the following directed acyclic graph:



Give all independence and conditional independence statements that can be derived from this directed acyclic graph.

SECTION B

Q3 **3.1** Consider the simulation of values from a Gamma$(2, b)$ distribution, where $b > 0$ is fixed and known, using a Metropolis-Hastings independence sampler based on Exp(b) proposals.

(a) Write down and simplify the acceptance probability for a move from θ to θ^*.

(b) If the Markov chain is currently at θ, show that the probability the chain will move (marginalised over the distribution of proposed values) is

$$\frac{1 - e^{-b\theta}}{b\theta}.$$

(c) Find the overall acceptance rate of the chain once it has reached equilibrium, and confirm that it is a constant independent of b.

3.2 Suppose that you wish to generate samples from a posterior density $\pi(\theta|\boldsymbol{x}) \propto \pi(\theta)f(\boldsymbol{x}|\theta)$ with support $\mathcal{S} \subseteq \mathbb{R}$. Consider a Markov chain defined by a Metropolis-Hastings algorithm which performs the following steps for $j = 1, 2, \ldots$

Algorithm. At state $\theta^{(j-1)}$:

- **Step 1.** Draw $\theta^* \sim \pi(\theta^*)$.
- **Step 2.** With probability

$$\alpha_B(\theta^*|\theta^{(j-1)}) = \frac{f(\boldsymbol{x}|\theta^*)}{f(\boldsymbol{x}|\theta^{(j-1)}) + f(\boldsymbol{x}|\theta^*)}$$

  put $\theta^{(j)} = \theta^*$, otherwise put $\theta^{(j)} = \theta^{(j-1)}$.
- **Step 3.** Increment j and go to **Step 1**.

(a) Prove that the Markov chain defined by the algorithm is reversible with stationary density $\pi(\theta|\boldsymbol{x})$.
Hint: show that the detailed balance equation $\pi(\theta|\boldsymbol{x})p(\phi|\theta) = \pi(\phi|\boldsymbol{x})p(\theta|\phi)$ holds.

(b) The Metropolis-Hastings algorithm uses

$$\alpha_{MH}(\theta^*|\theta^{(j-1)}) = \min\left\{1, \frac{f(\boldsymbol{x}|\theta^*)}{f(\boldsymbol{x}|\theta^{(j-1)})}\right\}$$

in Step 2 (but all other steps are unchanged). By comparing $\alpha_B(\theta^*|\theta^{(j-1)})$ and $\alpha_{MH}(\theta^*|\theta^{(j-1)})$, briefly explain why the Metropolis-Hastings algorithm is preferred.

Q4 Consider the model

$$\begin{aligned}\mathbf{x}|\boldsymbol{\theta} &\sim \text{Multinomial}\left[n = 50, (\theta_1, \theta_2, \theta_3)^T\right],\\ \boldsymbol{\theta} &\sim \text{Dirichlet}\left[(\alpha_1, \alpha_2, \alpha_3)^T\right],\end{aligned}$$

where

$$\begin{aligned}\pi(\mathbf{x}|\boldsymbol{\theta}) &= \frac{50!}{x_1!x_2!x_3!}\theta_1^{x_1}\theta_2^{x_2}\theta_3^{x_3},\\ \pi(\boldsymbol{\theta}) &= \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)}\theta_1^{\alpha_1-1}\theta_2^{\alpha_2-1}\theta_3^{\alpha_3-1}.\end{aligned}$$

4.1 Show that $\theta_1 \sim \text{Beta}(\alpha_1, \alpha_2+\alpha_3)$ and $x_1|\boldsymbol{\theta} \sim \text{Bin}(n = 50, \theta_1)$.

4.2 Let $x_1 = 27$, $x_2 + x_3 = 23$ and $\alpha_1 = \alpha_2 = \alpha_3 = 2$; derive the posterior distribution for θ_1.

A second model is proposed that utilises the additive-log-ratio transformation:

$$\begin{aligned}\boldsymbol{\kappa} &= \text{alr}(\boldsymbol{\theta})\\ &= \left[\log\left(\frac{\theta_1}{\theta_3}\right), \log\left(\frac{\theta_2}{\theta_3}\right)\right]^T.\end{aligned}$$

4.3 We plan to use MCMC to derive a sample of $\boldsymbol{\kappa}$; give formulae for θ_1, θ_2 and θ_3 in terms of κ_1 and κ_2 so that we could transform our MCMC sample of $\boldsymbol{\kappa}$ to a sample of $\boldsymbol{\theta}$.

4.4 Weights for the two models are being calculated using the pseudo-Bayesian-model-averaging scheme after collecting 20 observations of $\mathbf{x}$. We have calculated the unnormalised weights that are given by

$$\exp\left(\sum_{i=1}^{20}\log\left[\pi(\mathbf{x}_i|\underline{\mathbf{x}}_{-i}, \mathcal{M}_p)\right]\right), \quad p = 1, 2.$$

For the Dirichlet-based model, we get an unnormalised weight of 0.000181, and, for transformation-based model, we get an unnormalised weight of 0.000154. Calculate the pseudo-Bayesian-model-averaging weights for the two models.