Prof. Donna Ankerst, Katharina Selig

July 1st & 2nd 2019

Problem 1

Suppose one wants to use a Metropolis-Hastings sampler to generate a sample from a Rayleigh(4) distribution using as proposal density the Gamma(X, 1) density.

- Rayleigh(σ) density: $f(x) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, \quad x \geq 0, \sigma > 0$
- Gamma(α, β) density: $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0, \alpha, \beta > 0$

(a) Develop and write down the algorithm for N iterations.

`In R:`

(b) Implement the algorithm in `R` for 10,000 iterations. Calculate the acceptance rate and compare to the acceptance rate of 60% observed for the χ^2-distribution used in the lecture. Show realizations of the chain in a time series plot, where the first 20% of the iterations are used as a burn-in.

Problem 2

(a) Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, 1)$ and a uniform prior $\pi(\mu) \propto 1$ is specified for $\mu \in \mathbb{R}$. What is the posterior distribution of μ?

(b) Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(0, \sigma^2)$ and $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$ for $\sigma^2 > 0$. What is the posterior distribution of σ^2?
Hint: If $X \sim \text{Gamma}(\alpha, \beta)$ then $Y = \frac{1}{X} \sim \text{InvGamma}(\alpha, \beta)$.

Problem 3

We consider the following density of a 2-component mixture

$$f^*(Z) = pf_1(Z) + (1-p)f_2(Z),$$

with $p = (0, 1)$, the unknown mixing parameter and f_1 and f_2 are known probability distributions. f_1 is the χ^2_2 - distribution and f_2 is the χ^2_{10} - distribution.

(a) Suppose we have observed $z_1, \ldots, z_n$ i.i.d. samples from f^* and assume the prior distribution $\pi(p) \sim \mathcal{U}(0, 1)$ for p. To draw random samples from the posterior distribution of p we will use independence sampling with the $\mathcal{U}(0, 1)$ density as a proposal distribution. Develop and write down the algorithm for the independence sampler.

(b) Write down the acceptance probability for a new candidate p.

`In R:`

(c) For fitting the algorithm later, simulate $n = 1000$ i.i.d. observations from f^* with $p = 0.3$. Plot a histogram of the data and overlay the theoretical density f^* as well as the density of f_1 and f_2 on the plot using different colors.

(d) Given the observed sample $z_1, \ldots, z_{1000} \sim f^*$ simulated in (c), implement the algorithm of (a) in `R` using 10,000 iterations and $p = 0.1, 0.5, 0.9$ as starting values. Plot the three different chains as a time series using different colors. Compute the means of the three chains, after discarding 100 iterations as burn-in. Have the chains converged?

(e) Repeat (d) with a random sample of $n = 10$ observations instead of $n = 1000$. Does anything different happen? Can you find an explanation?

(f) We want to monitor the convergence of the three chains from (d) using the Gelman-Rubin method. Refer to Example 9.8 of the lecture and write an `R` function to compute the Gelman-Rubin statistic $\hat{R} = \hat{V}(\psi)/W$ using the mean of the i^{th} chain up to time j as the scalar summary statistic. Determine $\hat{R}$ for the three chains from (d). Plot the sequence of the $\hat{R}$ statistics and add a horizontal line at $y = 1.2$. Have the chains converged according to the Gelman-Rubin method? How many iterations would you discard as a burn-in?

Problem 4

Suppose $X = (X_1, X_2) \in \mathbb{R}^2$ is bivariate normal distributed with marginal means μ_1, μ_2, marginal variances σ_1^2, σ_2^2, and correlation coefficient ρ. The density of the bivariate normal distribution is given by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{Q}{2(1-\rho^2)}\right),$$

where

$$Q = \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 - 2\rho\frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2}.$$

(a) Derive the conditional density $f(x_1|x_2)$ for $X_1|X_2 = x_2$. Use that $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

(b) Given the formula for $f(x_1|x_2)$, what is the formula for $f(x_2|x_1)$?

(c) Which has smaller variance, X_1 or $X_1|X_2 = x_2$? Argue why and under what conditions.

(d) Write a Gibbs sampler to sample n observations from $(X_1, X_2) \sim$ bivariate Normal with means $(2, 3)$, variances $(1, 1)$ and $\rho = 0.5$, given you have an univariate normal sampler.

(e) Given you have an univariate normal sampler write the most efficient sampler for (d) when $\rho = 0$ and $\rho = 1$.