Problem 1 Suppose that X has cumulative distribution

$$F(x) = \begin{cases} 0, & x < 0 \\ \sqrt{x}, & 0 \le x \le 1 \\ 1, & x > 1. \end{cases}$$

a) Find the expected value and variance of X. Suppose that we were to estimate the expected value of X using Monte-Carlo integration. How would you simulate realisations of X? Using the known variance of X, calculate the true variance of the Monte-Carlo estimator of EX if the estimator is based on $n = 1000$ realizations $X_1, X_2, \ldots, X_n$. Find the same variance if the Monte-Carlo estimator is additionally based on 1000 antithetic realisations $X_1^*, X_2^*, \ldots, X_n^*$.

Problem 2

The ratio-of-uniforms method involves simulating X_1 and X_2 uniformly from a set $C = \{(x_1, x_2) : 0 \le x_1 \le \sqrt{f^*(x_2/x_1)}\}$. It follows that $C \subset [0, a] \times [b_-, b_+]$ where $a = \sup_x \sqrt{f^*(x)}$, $b_- = -\sqrt{\sup_{x\le 0} x^2 f^*(x)}$ and $b_+ = \sqrt{\sup_{x\ge 0} x^2 f^*(x)}$.

a) Derive a ratio-of-uniforms method for simulating a random variable Y having density proportional to

$$f^*(y) = \begin{cases} \frac{1}{(y+1)^2}, & y \ge 0 \\ 0, & y < 0 \end{cases}$$

including an efficient method for simulating X_1, X_2 uniformly from the set C.

Problem 3

Assume that x_i, $i = 1, 2, \ldots, n$, are iid uniformly distributed on the interval from 0 to b and that y_i, $i = 1, 2, \ldots, n$, conditional on $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are independently $N(x_i, \sigma^2)$ distributed. We only observe $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ as shown in Fig. 1. Given these observation, using improper scale prior $\pi(b) \propto \frac{1}{b}$ and $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$ on $b > 0$ and $\sigma^2 > 0$, we want to construct a Gibbs sampler to sample from the joint posterior distribution of $b, \sigma^2, x_1, x_2, \ldots, x_n$ conditional on the observed data $\mathbf{y}$, updating each of these unknown quantities one at the time. Asumme that the posterior is proper.

a) What is the full conditional of σ^2, that is the the distribution of σ^2 conditional on the data $\mathbf{y}$, $\mathbf{x}$ and b?

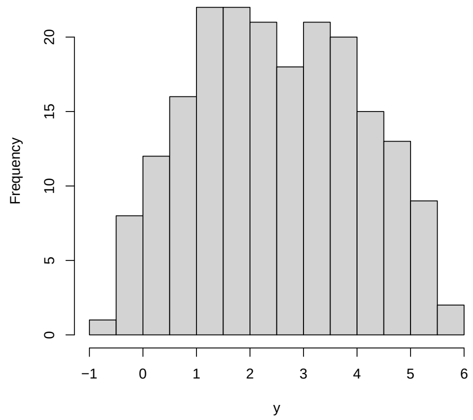

Figure 1: Histogram of the $n = 200$ observations $y_1, y_2, \ldots, y_n$.

b) Similarly, what is the conditional distribution of each x_i, conditional on all other quantities in the model? Derive a method for simulating from this distribution.

c) What is the conditional density of b, again conditioning on the remaining quantities? Find a method to simulate from this distribution.

d) Briefly discuss what you see in Fig. 2 and reasons why the Gibbs sampler based on the update rules in point a) to c) may have a slow rate of convergence.

Conditional on only σ^2 and b, explain why $y_1, y_2, \ldots, y_n$ are independent and show that each y_i has marginal densities given by

$$\pi(y_i|b, \sigma^2) = \frac{1}{b}\Big(\Phi\big(\frac{y_i}{\sigma}\big) - \Phi\big(\frac{y_i - b}{\sigma}\big)\Big)$$

where Φ is the cdf of the standard normal distribution. Briefly discuss how you may exploit this to construct a more efficient way of sampling from the posterior distribution of b and σ.

Finally, what is the limiting value of the likelihood for b and σ^2 as $b \to 0$? Given the above choice of prior, is the posterior proper?

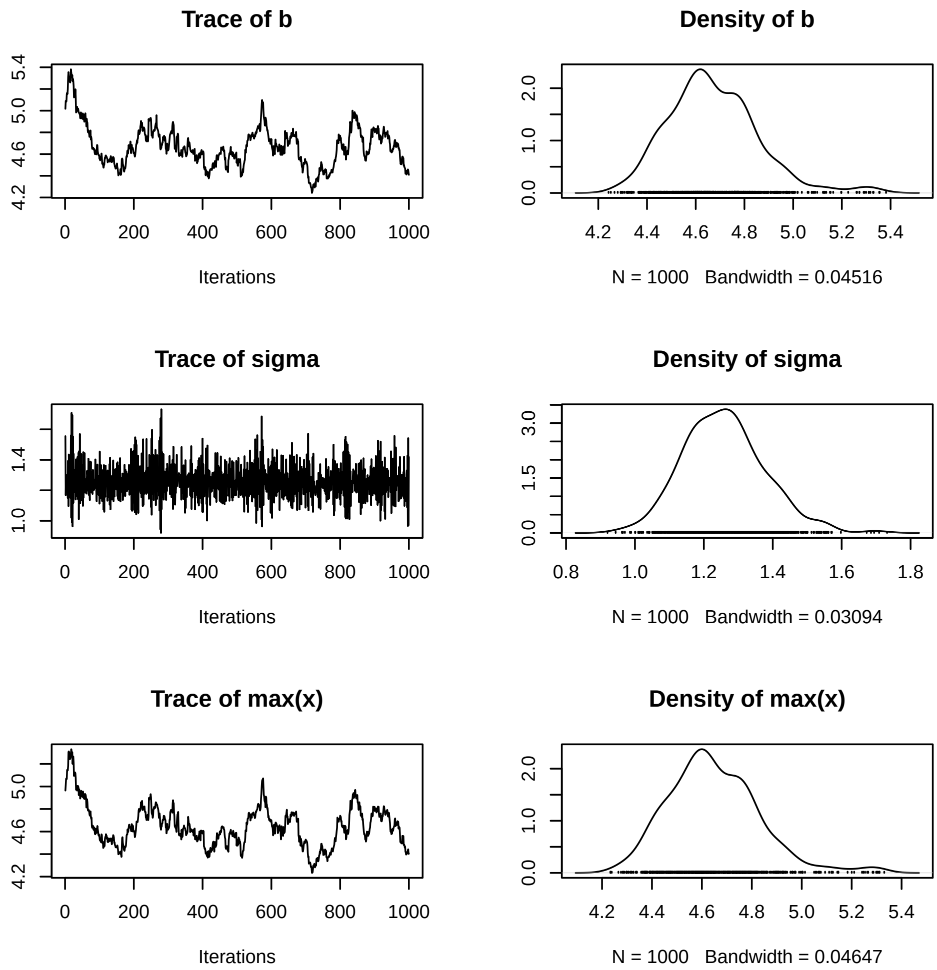

Figure 2: Trace plots and kernel density estimates of the marginal posteriors of b, σ and $x_{(n)} = \max(x_1, x_2, \ldots, x_n)$ when running the Gibbs sampler in problem 2a-c for 1000 iterations for the data shown in Fig. 1.

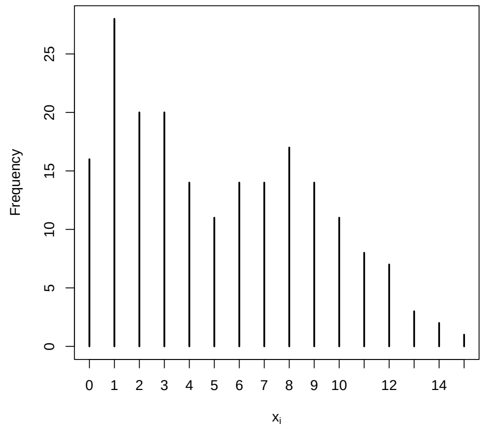

Figure 3: Observations $x_1, x_2, \ldots, x_{200}$ from the Poisson mixture described in the main text.

Problem 4

Suppose that we observe an iid sample $x_1, x_2, \ldots, x_n$ from a Poisson mixture with probability mass function

$$f(x) = (1-p)\frac{\mu_0^x e^{-\mu_0}}{x!} + p\frac{\mu_1^x e^{-\mu_1}}{x!}$$

as shown in Fig. 3. We would like to derive an expectation-maximization algorithm for computing the maximum likelihood estimates of p, μ_1, μ_2.

Note first that the model is equivalent to assuming that $x_i|z_i \sim \text{Pois}(\mu_{z_i})$ where $z_i \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$, $i = 1, 2, \ldots, n$, are "missing" unobserved variables.

a) Write down the full data likelihood and its logarithm $l(p, \mu_0, \mu_1; \mathbf{x}, \mathbf{z})$ had both $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ been observed.

b) Derive expressions for

$$w_i^{(t)} = E(z_i|x_i, p^{(t)}, \mu_0^{(t)}, \mu_1^{(t)})$$

and

$$Q(p, \mu_0, \mu_1 \mid p^{(t)}, \mu_0^{(t)}, \mu_1^{(t)}) = E(l(p, \mu_0, \mu_1; \mathbf{x}, \mathbf{z}) \mid \mathbf{x}, p^{(t)}, \mu_0^{(t)}, \mu_1^{(t)})$$

(the E-step of the algorithm) where superscripts (t) indicates values at the t'th iteration.

c) Find $p^{(t+1)}, \mu_0^{(t+1)}, \mu_1^{(t+1)}$ by maximizing Q with respect to p, μ_0, μ_1 (the M-step of the algorithm).

d) Explain in detail how you would estimate standard errors of the MLEs using parametric bootstrapping. Are there any issues with indentifiability of the above model? If so, how can the model be modified to make it identifiable and how would you modify the EM-algorithm to obtain MLEs for the modified model?