Problem 1

a) X can be generated using the inversion method by setting $X = U^2$ where $U \sim \text{unif}(0, 1)$. Thus, $EX = E(U^2) = 1/3$, $E(X^2) = E(U^4) = 1/5$ and $\text{Var}\, X = 1/5 - 1/9 = 4/45$. The Monte-Carlo estimator of EX is given by $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{1000} X_i$ which has variance $4/45000 \approx 8.9 \cdot 10^{-5}$. Antithetic realizations of X can be obtained by setting $X^* = (1 - U)^2$. This gives $E(X_i X_i^*) = E(U^2(1-U)^2) = \int_0^1 u^2 - 2u^3 + u^4 du = 1/3 - 1/2 + 1/5 = (10 - 15 + 6)/30 = 1/30$, $\text{Cov}(X_i, X_i^*) = 1/30 - 1/9 = (3 - 10)/90 = -7/90$ and $\text{corr}(X_i, X_i^*) = -7/8$. The variance of the Monte-Carlo estimator using antithetic sampling is thus reduced by a factor of 16 to

$$\text{Var}\left(\frac{1}{2n}\sum_{i=1}^{1000}(X_i + X_i^*)\right) = \cdots = \frac{\text{Var}(X_i)}{2 \cdot 1000}(1 + \text{corr}(X_i, X_i^*) = 5.5 \cdot 10^{-6}.$$

Problem 2

a) It follows that $b^- = -\sqrt{\sup_{x \le 0} x^2 f^*(x)} = 0$. In addition, $x_1 \le \sqrt{f^*(x_2/x_1)} = \frac{1}{x_2/x_1 + 1} = \frac{x_1}{x_1 + x_2}$ implies C is given by the inequalities $x_1 + x_2 < 1$, $x_1 \ge 0$ and $x_2 \ge b^- = 0$. An efficient method for simulating uniformly from inside this region is to simulate $x_1, x_2 \overset{\text{iid}}{\sim} \text{unif}(0, 1)$ and then reflect x_1, x_2 about the line $x_1 + x_2 = 1$ if $x_1 + x_2 > 1$ by setting $x_2 = 1 - x_1$ and $x_1 = 1 - x_2$. Finally, set $y = x_2/x_1$.. Alternative, slightly less efficient methods are rejection sampling or first simulating x_1 from its triangular marginal distribution and then $x_2|x_1 \sim \text{unif}(0, 1 - x_1)$.

Problem 3

a) The conditional density of σ^2 becomes

$$\begin{aligned}
\pi(\sigma^2|\mathbf{x}, \mathbf{y}, b) &\propto \pi(\sigma^2)\prod_{i=1}^n \pi(y_i|x_i, \sigma^2) \\
&\propto \frac{1}{\sigma^2}\frac{1}{(\sigma^2)^{n/2}}\exp(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - x_i)^2) \\
&\propto \frac{1}{(\sigma^2)^{n/2+1}}\exp(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - x_i)^2)
\end{aligned}$$

which is an inverse Gamma distribution with shape parameter $\alpha = n/2$ and scale parameter $1/\beta = \frac{1}{2}\sum_{i=1}^n (y_i - x_i)^2$ (or rate parameter $\beta = 2/\sum_{i=1}^n (y_i - x_i)^2)$. Using the inversion method we can thus simulate σ^2 by setting $\sigma^2 = 1/F_\Gamma^{-1}(U; \alpha, \beta)$ where F_Γ^{-1} is the quantile function of the gamma distribution.

b) The conditional density each x_i,

$$\begin{aligned}\pi(x_i|\mathbf{y},\mathbf{x}_{-i},\sigma,b) &\propto \pi(x_i|b)\pi(y_i|x_i,\sigma)\\ &\propto e^{-\frac{1}{2\sigma^2}(y_i-x_i)^2} I(0<x_i<b),\end{aligned}$$

that is, x_i has normal distribution with mean y_i, standard deviation σ truncated below 0 and above b. Introducing $Z \sim N(y_i,\sigma^2)$, the cdf of the full conditional of x_i can be expressed as

$$\begin{aligned}F_{x_i|\ldots}(x) &= P(X_i \le x \mid \ldots)\\ &= P(Z \le x \mid 0 < Z \le b)\\ &= \frac{P(Z \le x \cap 0 < Z \le b)}{P(0<Z\le b)}\\ &= \frac{P(0<Z\le x)}{P(0<Z\le b)}\\ &= \frac{\Phi(\frac{x-y_i}{\sigma}) - \Phi(-\frac{y_i}{\sigma})}{\Phi(\frac{b-y_i}{\sigma}) - \Phi(-\frac{y_i}{\sigma})}\end{aligned}$$

for $0 \le x \le b$. Using the inversion method, solving $U = F_{x_i|\ldots}(X_i)$ for X_i, gives

$$X_i = y_i + \sigma\Phi^{-1}\left(\Phi(-\frac{y_i}{\sigma}) + \Big(\Phi(\frac{b-y_i}{\sigma}) - \Phi(-\frac{y_i}{\sigma})\Big)U\right).$$

An alternative method of simulation is Box-Muller (or any other methods for simulating from the normal distribution) followed by rejection of samples outside $(0,b)$. In cases where $\sigma \ll b$ so that the rejection probability is mostly small, this method may be just as efficient as the above inversion method as it avoids somewhat costly evaluation of both Φ^{-1} and Φ.

c) The conditional density of b becomes

$$\begin{aligned}\pi(b|\mathbf{x},\mathbf{y},\sigma^2) &\propto \pi(b)\prod_{i=1}^n \pi(x_i|b)\\ &= \frac{1}{b}I(b>0)\prod_{i=1}^n \frac{1}{b}I(0<x_i\le b)\\ &= b^{-(n+1)}I(b \ge x_{(n)})\end{aligned}$$

where $x_{(n)}$ is the maximum of the current values of $x_1, x_2, \ldots, x_n$. This may be recognised as the Pareto distribution with scale parameter $x_{(n)}$ and shape parameter n.

The cdf is

$$
\begin{aligned}
F_{b|\dots}(b) &= \frac{\int_{x_{(n)}}^{b} t^{-(n+1)}dt}{\int_{x_{(n)}}^{\infty} b^{-(n+1)}db} \\
&= \frac{b^{-n} - x_{(n)}^{-n}}{0 - x_{(n)}^{-n}} \\
&= 1 - \left(\frac{x_{(n)}}{b}\right)^n.
\end{aligned}
$$

Equating this to U and solving for b gives the inversion method $b = x_{(n)}(1-U)^{-1/n}$.

d) The trace plots of b and $x_{(n)}$ shows that both of these parameters changes slowly with autocorrelation time of about 100 iterations. It is also clear that b and $x_{(n)}$ have a strong posterior correlation. This explains why the Gibbs sampler will have a slow rate of convergence since b and $x_{(n)}$ can make small steps when b and $\mathbf{x}$ are updated separately using their current conditional distribution.

Using the law of total probability, and making the substitution $u = \frac{x_i - y_i}{\sigma}$,

$$
\begin{aligned}
\pi(y_i|b,\sigma^2) &= \int_0^b \pi(y_i|x_i,\sigma^2)\pi(x_i|b)dx_i \\
&= \frac{1}{b}\int_0^b \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{1}{2\sigma^2}(y_i-x_i)^2)dx_i \\
&= \frac{1}{b}\int_{-\frac{y_i}{\sigma}}^{\frac{b-y_i}{\sigma}} \frac{1}{\sqrt{2\pi}}e^{-u^2}du \\
&= \frac{1}{b}\left(\Phi\left(\frac{b-y_i}{\sigma}\right) - \Phi\left(-\frac{y_i}{\sigma}\right)\right) \\
&= \frac{1}{b}\left(1 - \Phi(\frac{y_i-b}{\sigma}) - 1 + \Phi(\frac{y_i}{\sigma})\right) \\
&= \frac{1}{b}\left(\Phi(\frac{y_i}{\sigma}) - \Phi(\frac{y_i-b}{\sigma})\right)
\end{aligned}
$$

We can thus omit $x_1, x_2, \dots, x_n$ from the MCMC sampler since the marginal posterior density of b and σ^2 can be written in closed form as

$$
\pi(b,\sigma^2|\mathbf{y}) \propto \pi(\sigma^2)\pi(b)\prod_{i=1}^{n}\pi(y_i|b,\sigma^2)
$$

To sample from this density, standard random-walk Metropolis-Hastings with e.g. a bivariate normal proposal may be used. Fig. 2 suggests using proposal standard deviations for both b and σ^2 somewhere around 0.5.

In the limit as $b \to 0$, each factor $\pi(y_i|b, \sigma^2)$ tends to a Gaussian density with mean zero and standard deviation σ, that is, $\phi(y_i/\sigma)/\sigma$, since the underlying x_i are then just constants equal to zero. More formalistically, we have

$$\begin{aligned}\lim_{b\to 0} \frac{1}{b}\left(\Phi\Big(\frac{b-y_i}{\sigma}\Big) - \Phi\Big(-\frac{y_i}{\sigma}\Big)\right) &= \lim_{b\to 0} \frac{1}{b}\left(\Phi\Big(\frac{t+b-y_i}{\sigma}\Big) - \Phi\Big(\frac{t-y_i}{\sigma}\Big)\right)\Big|_{t=0} \\ &= \frac{d}{dt}\Phi\Big(\frac{t-y_i}{\sigma}\Big)\Big|_{t=0} = \frac{1}{\sigma}\phi(-y_i/\sigma)\end{aligned}$$

by definition of the derivative and by the chain rule. The whole likelihood thus tends to a positive limit, say a, that is, for any $\epsilon > 0$ we can find a $b > 0$ such that

$$\left|\prod_{i=1}^{n} \pi(y_i|b, \sigma^2) - a\right| < \epsilon.$$

Thus, $\prod_{i=1}^{n} \pi(y_i|b, \sigma^2) > a - \epsilon = c$ where $c > 0$ is a constant that may depend on σ^2 and

$$\int_0^b \left(\pi(\sigma^2)\pi(b)\prod_{i=1}^{n} \pi(y_i|b, \sigma^2)\right) db > \int_0^b \frac{1}{\sigma^2 b} c\, db = \infty$$

for some constants $b > 0$ and $c > 0$. This holds for a continuum of σ^2 values. The target density does thus not have a finite normalizing constant and, although not evident from trace plot, the posterior is therefore improper. See Hobert and Casella (1996) for a similar example.

Problem 4

a) The full data likelihood is

$$\begin{aligned} L(p,\mu_0,\mu_1;\mathbf{x},\mathbf{z}) &= \prod_{i=1}^n f(x_i,z_i) \\ &= \prod_{i=1}^n f(x_i|z_i)f(z_i) \\ &= \prod_{i=1}^n \frac{e^{-\mu_{z_i}}\mu_{z_i}^{x_i}}{x_i!}p^{z_i}(1-p)^{1-z_i} \end{aligned}$$

and its log

$$l(p,\mu_0,\mu_1;\mathbf{x},\mathbf{z}) = \sum_{i=1}^n \Big(x_i \ln \mu_{z_i} - \mu_{z_i} - \ln x_i! + z_i \ln p + (1-z_i)\ln(1-p)\Big)$$

b) Conditional on x_i and the current values of the model parameters, it follows from Bayes theorem that

$$\begin{aligned} w_i^{(t)} &= E(z_i \mid x_i, p^{(t)}, \mu_0^{(t)}, \mu_1^{(t)}) \\ &= P(z_i = 1|x_i, \dots) \\ &= \frac{P(x_i \mid z_i = 1, ...)P(z_i = 1 \mid \dots)}{P(x_i \mid z_i = 0, ...)P(z_i = 0 \mid \dots) + P(x_i \mid z_i = 1, \dots)P(z_i = 1 \mid \dots)} \\ &= \frac{(\mu_1^{(t)})^{x_i} e^{-\mu_1^{(t)}} p^{(t)}}{(\mu_0^{(t)})^{x_i} e^{-\mu_0^{(t)}}(1-p^{(t)}) + (\mu_1^{(t)})^{x_i} e^{-\mu_1^{(t)}} p^{(t)}}. \end{aligned}$$

In terms of these weights, keeping in mind that μ_{z_i} is a binary random variable taking values of either μ_0 or μ_1 with probabilites $1 - w_I^{(t)}$ and $w_I^{(t)}$, respectively,

$$\begin{aligned} &Q(p,\mu_0,\mu_1 \mid p^{(t)}, \mu_0^{(t)}, \mu_1^{(t)}) \\ &= \sum_{i=1}^n (1-w_i^{(t)})(x_i \ln \mu_0 - \mu_0) + w_i^{(t)}(x_i \ln \mu_1 - \mu_1) - \ln x_i! + w_i^{(t)} \ln p + (1-w_i^{(t)}) \ln(1-p) \end{aligned}$$

c) Setting the partial derivatives equal to zero gives,

$$\frac{\partial Q}{\partial p} = \frac{1}{p}\sum_{i=1}^n w_i^{(t)} - \frac{1}{1-p}\sum_{i=1}^n (1-w_i^{(t)}) = 0$$

which solved for p yields

$$p^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} w_i^{(t)},$$

and similarly,

$$\frac{\partial Q}{\partial \mu_1} = \frac{1}{\mu_1}\sum_{i=1}^{n} w_i^{(t)} x_i - \sum_{i=1}^{n} w_i^{(t)} = 0$$

which yields,

$$\mu_1^{(t+1)} = \frac{\sum_{i=1}^n w_i^{(t)} x_i}{\sum_{i=1}^n w_i^{(t)}}.$$

Symmetry implies that

$$\mu_0^{(t+1)} = \frac{\sum_{i=1}^n (1 - w_i^{(t)}) x_i}{\sum_{i=1}^n (1 - w_i^{(t)})}.$$

d) Having obtained MLEs $\hat{p}, \hat{\mu}_0, \hat{\mu}_1$ by applying the EM-algorithm to the original data, parametric bootstrapping is done by generating B bootstrap samples $\mathbf{y}^b$, $b = 1, 2, \ldots, B$. To simulate each observation y_i^b we would first simulate $z_i \sim \text{Bernoulli}(\hat{p})$ and then $y_i^b \sim \text{Poisson}(\hat{\mu}_{z_i^b})$. Applying the EM-algorithm to each bootstrap sample we obtain bootstrap replicates $\hat{p}^b, \hat{\mu}_0^b, \hat{\mu}_1^b$ of the MLEs. Standard errors can then be estimated by the sample standard deviations of these bootstrap replicates.

The model is not identifiable since the distribution of the data is the same if the parameters are p, μ_0, μ_1 and

$$p' = 1 - p, \tag{1}$$
$$\mu_0' = \mu_1, \tag{2}$$
$$\mu_1' = \mu_0 \tag{3}$$

respectively. To make the model identifiable, we can impose some additional restrictions on the parameters, for example, the restriction that $\mu_0 \leq \mu_1$. To fit the model with this restriction on the parameters, we could use the above algorithm but apply the transformation (1) in cases where the final MLEs $\hat{\mu}_0 > \hat{\mu}_1$. Alternatively, one could impose the constraint $p \leq 1/2$.