Problem 1

a) It follows that X has cdf $F(x) = \int_0^x 2tdt = x^2$. Hence, X can be simulated by setting $X = F^{-1}(U) = \sqrt{U}$ where $U \sim \text{unif}(0,1)$.

The expected value of X can be estimated by

$$\widehat{EX} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where $x_1, x_2, \ldots, x_n$ are simulated realizations of X. This Monte-Carlo estimator has variance

$$\text{Var}(\widehat{EX}) = \frac{1}{n}\text{Var}\,X = 1/18000 \approx 0.0001667$$

since $EX = E\sqrt{U} = \int_0^1 u^{1/2}du = 2/3$, $E(X^2) = EU = 1/2$ so that $\text{Var}\,X = 1/2 - (2/3)^2 = 1/18$.

b) Antithetic realizations X^* can be simulated by setting $X^* = \sqrt{1-U}$.

We then have

$$\begin{aligned} E(XX^*) &= E(\sqrt{U(1-U)}) \\ &= \int_0^1 u(1-u)du \\ &= \int_0^1 \sqrt{-(u^2 - u + 1/4) + 1/4} \\ &= \int_0^1 \sqrt{1/4 - (u-1/2)^2} \\ &= \frac{\pi}{8} \end{aligned}$$

since the last integral is the area of a half disk with radius 1/2 located at $(1/2, 1)$.

Hence,

$$\text{corr}(X, X^*) = \frac{E(XX^*) - (EX)^2}{\text{Var}\,X} = \frac{\pi/8 - 4/9}{1/6} \approx -0.9314165$$

It follows that the improved estimator

$$\widehat{EX} = \frac{1}{2n}(\sum_{i=1}^{n} X_i + \sum_{i=1}^{n} X_i^*)$$

has variance

$$\text{Var}\,\widehat{EX} = \frac{1}{n}\text{Var}(X)(1+\rho) = 3.8101928 \times 10^{-6}$$

that is, the variance is reduced by a factor of $1 + \rho = 0.069$ when using antithetic sampling.

Problem 2

a) When $s = 0$,

$$f(x) = \frac{1}{\sqrt{2\pi}x} e^{-(\ln x)^2/2}$$

which is the density of the standard lognormal distribution.

Alternative 1: A possible method for simulating X is rejection sampling using

$$g(x) = \frac{1}{\sqrt{2\pi}x} e^{-(\ln x)^2/2}$$

as proposal and accepting a sample $X = x$ with probability

$$\frac{f(x)}{cg(x)} = \frac{1 + s\sin(2k\pi \ln x)}{c} \leq 1$$

To ensure that this probability is at most 1 for all x, we must have $c \geq 1+s$, the optimal value being $c = 1 + s$. Thus the acceptance probability becomes

$$\frac{1 + s\sin(2k\pi \ln x)}{1 + s}.$$

The acceptance rate becomes

$$E\frac{f(X)}{cg(X)} = \int_0^\infty \frac{f(x)}{cg(x)} g(x)dx = \frac{1}{c}\int_0^\infty f(x)dx = \frac{1}{c} = \frac{1}{1+s}$$

which can thus be as low as 50% when $s = 1$.

Algorithm 1 Alternative method for simulating from $f(x)$ in problem 2.

Generate $x \sim g$
Generate $u \sim \text{unif}(0, 1)$
$\alpha \leftarrow \min(1, 1 + s\sin(2k\pi \ln x))$
if $u < \alpha$ **then**
  $y \leftarrow x$
else
  $y \leftarrow 1/x$
end if
Return y

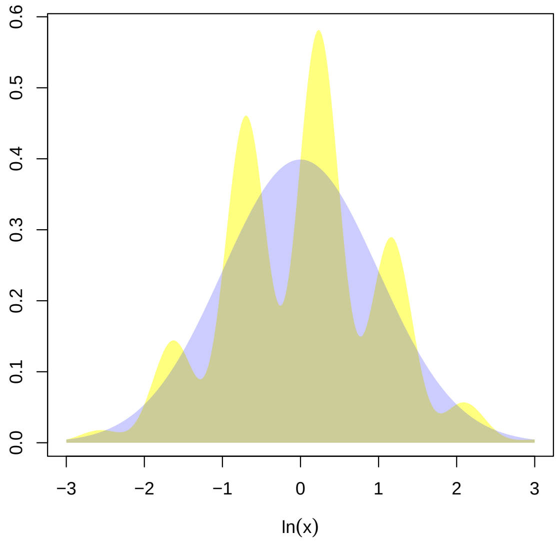

Figure 1: Illustration of algorithm 1 based on moving some of the mass of the proposal (blue areas) to a different locations (yellow areas). Note that the graph shows the densities of log-transformed values.

Alternative 2: Instead of rejecting samples an alternative idea is to instead move some of the probability mass of a lognormal proposal to new locations based on the factor $1 + s \sin(2\pi k \ln x)$ and thus avoid rejecting any samples (see Algorithm 1 and Fig. 1).

To prove that this works, let I be an indicator variable for the event that we accept the proposed value by setting $Y = X$, the complement of this event being that we set $Y = 1/X$. Conditional on $X = x$, I is thus Bernoulli with parameter ('acceptance probability')

$$\begin{aligned}\alpha(x) &= \min(1, 1 + s \sin(2k\pi \ln x)) \\ &= 1 + \min(0, s \sin(2k\pi \ln x)).\end{aligned}$$

Using the product rule for densities, the joint density of X and I is thus

$$f_{X,I}(x, i) = f_X(x) f_{I|X=x}(i) = \frac{1}{\sqrt{2\pi} x} e^{-\frac{1}{2}(\ln x)^2} \alpha(x)^i (1 - \alpha(x))^{1-i}.$$

The joint density of Y and I follows by applying the transformation formula for densities for the transformation $Y = 1/X$ when $I = 0$ giving

$$\begin{aligned}
f_{Y,I}(y,i) &= \begin{cases} f_{X,I}(y,1) & \text{for } i = 1 \\ f_{X,I}(1/y,1)\left|\frac{dx}{dy}\right| & \text{for } i = 0 \end{cases} \\
&= \begin{cases} \frac{1}{\sqrt{2\pi}y} e^{-\frac{1}{2}(\ln y)^2}\alpha(y) & \text{for } i = 0 \\ \frac{1}{\sqrt{2\pi}(1/y)} e^{-\frac{1}{2}(\ln(1/y))^2}\frac{1}{y^2}(1-\alpha(1/y)) & \text{for } i = 0. \end{cases}
\end{aligned}$$

Hence, using the law of total probability, the marginal density of Y is

$$\begin{aligned}
f_Y(y) &= \sum_{i=0}^{1} f_{Y,I}(y,i) \\
&= f_{Y,I}(y,1) + f_{Y,I}(y,0) \\
&= \frac{1}{\sqrt{2\pi}y} e^{-(\ln y)^2/2}(\alpha(y) + 1 - \alpha(1/y))
\end{aligned}$$

The factor in the parenthesis can be simplified by noting that sin is an odd function and using properties of the min and max functions which yields

$$\begin{aligned}
\alpha(y) + 1 - \alpha(1/y) &= 1 + \min(0, s\sin(2k\pi\ln y)) - \min(0, s\sin(2k\pi\ln(1/y)))) \\
&= 1 + \min(0, s\sin(2k\pi\ln y)) - \min(0, s\sin(-2k\pi\ln y))) \\
&= 1 + \min(0, s\sin(2k\pi\ln y)) - \min(0, -s\sin(2k\pi\ln y))) \\
&= 1 + \min(0, s\sin(2k\pi\ln y)) + \max(0, s\sin(2k\pi\ln y))) \\
&= 1 + s\sin(2k\pi\ln y).
\end{aligned}$$

This completes the proof.

As an aside, the distribution considered in this problem is a standard counterexample to the claim that a distribution is fully determined by its moments, as all the moments of $f(x)$ can be shown to be independent of s and k (Exercise 6.21 in Kendall's Advanced Theory of Statistics (Stuart & Ord, 5th edition)).

Problem 3

a) The joint posterior

$$
\begin{aligned}
\pi(\lambda, a \mid \mathbf{y}) &\propto \pi(a,\lambda)\prod_{i=1}^{n}\pi(y_i \mid \lambda, a) \\
&= \frac{e^{-a}}{\lambda} I(\lambda > 0, a > 0)\prod_{i=1}^{n}\frac{\lambda^a}{\Gamma(a)} y_i^{a-1} e^{-\lambda y_i} \\
&= \lambda^{na-1}\Gamma(a)^{-n}(\prod_{i=1}^{n} y_i)^a e^{-a-\lambda\sum_{i=1}^{n} y_i} I(\lambda > 0, a > 0)
\end{aligned}
$$

and the full conditional of λ is thus

$$
\pi(\lambda \mid a, \mathbf{y}) \propto \lambda^{na-1} e^{-\lambda\sum_{i=1}^{n} y_i} I(\lambda > 0)
$$

which (up to a normalizing constant) is the pdf of a Gamma distribution with shape parameter na and rate parameter $\sum_{i=1}^{n} y_i$.

b) Similarly, the full conditional of a is

$$
\pi(a \mid \lambda, \mathbf{y}) \propto (\lambda^n e^{-1}\prod_{i=1}^{n} y_i)^a \Gamma(a)^{-n} I(a > 0)
$$

which is not the density of any familiar distribution. Using $Q(a'|a) = \frac{1}{2d} I(a - d < a' < a + d)$ as proposal, the log of the acceptance probability of a Metropolis-Hastings within Gibbs step becomes

$$
\begin{aligned}
\ln\alpha &= \ln\min(1, \frac{\pi(a' \mid \lambda, \mathbf{y})Q(a|a')}{\pi(a \mid \lambda, \mathbf{y})Q(a'|a)}) \\
&= \min(0, \ln\frac{(\lambda^n e^{-1}\prod_{i=1}^{n} y_i)^{a'}\Gamma(a')^{-n} I(a' > 0)}{(\lambda^n e^{-1}\prod_{i=1}^{n} y_i)^{a}\Gamma(a)^{-n} I(a > 0)}) \\
&= \min(0, (a' - a)(n\ln\lambda - 1 - \sum_{i=1}^{n}\ln y_i) - n(\ln\Gamma(a') - \ln\Gamma(a)) + \ln I(a' > 0)).
\end{aligned}
$$

While $\ln I(a' > 0))$ in the above expression is undefined for non-positive proposals a', a corresponding R expression would produce a log acceptance probability evaluating to

```
log(aprime > 0)

## [1] -Inf
```

and thus a correct acceptance probability of zero since

```
exp(-Inf)

## [1] 0
```

thus ensuring that the parameters stay inside the support of the posterior distribution.

c) We see that a and λ have a strong posterior correlation and the change in each parameter when updating these in separate Gibbs step will therefore be small leading to a slow rate of convergence.

A joint proposal $Q(a', \lambda'|a, \lambda)$ can be constructed by using a similar random walk proposal for a (but with a different d) combined with a proposal λ' according to its full conditional, conditional on the proposed value a' of a. From the product rule for densities we find that the joint proposal density then is

$$Q(a', \lambda'|a, \lambda) \propto \frac{1}{2d} I(a - d < a' < a + d) \frac{(\sum_{i=1}^n y_i)^{na'}}{\Gamma(na')} (\lambda')^{na'-1} e^{-\lambda' \sum_{i=1}^n y_i} I(\lambda' > 0)$$

This single-block proposal has the advantage that we can increase the step size d while still maintaining a reasonable high acceptance rate since the proposal matches the target density $\pi(a, \lambda|\mathbf{y})$ to a much larger extent. After tuning the step size d we can thus expect a much higher rate of convergence.

Problem 4

a) The bootstrap estimate of $E\hat{\lambda}$ is given by $\frac{1}{B}\sum_{b=1}^B \hat{\lambda}^{b*} = 2.24$ and the bootstrap estimate of the bias $E\hat{\lambda} - \lambda$ is similarly given by $\frac{1}{B}\sum_{b=1}^B \hat{\lambda}^{b*} - \hat{\lambda} = 2.24 - 2.0 = 0.24$.

Subtracting the estimated bias from the original estimate $\hat{\lambda}$ we obtain a bias corrected estimate of $2.0 - 0.24 = 1.76$.

The biases corrected estimator can be expressed as

$$\hat{\lambda}_c = \hat{\lambda} - \widehat{\text{Bias}(\hat{\lambda})} = \hat{\lambda} - (\frac{1}{B}\sum_{b=1}^B \hat{\lambda}^{b*} - \hat{\lambda}) = 2\hat{\lambda} - \frac{1}{B}\sum_{b=1}^B \hat{\lambda}^{b*}.$$

b) Since we know that $E(\hat{\lambda}) = \frac{n}{n-1}\lambda$, the plug-in principle implies that the same hold analogously for for the bootstrapped model, that is, each bootstrap replicate of $\hat{\lambda}$ has conditional expected value $E(\hat{\lambda}^{b*} \mid \hat{\lambda}) = \frac{n}{n-1}\hat{\lambda}$.

Hence,

$$\begin{aligned}
E(\hat{\lambda}_c \mid \hat{\lambda}) &= E\Big(2\hat{\lambda} - \frac{1}{B}\sum_{b=1}^{B} \hat{\lambda}^{b*} \mid \hat{\lambda}\Big) \\
&= 2\hat{\lambda} - \frac{1}{B}\sum_{b=1}^{B} E(\hat{\lambda}^{b*} \mid \hat{\lambda}) \\
&= 2\hat{\lambda} - \frac{n}{n-1}\hat{\lambda} \\
&= \frac{2n-2-n}{n-1}\hat{\lambda} \\
&= \frac{n-2}{n-1}\hat{\lambda}.
\end{aligned}$$

Using the law of total expectation

$$\begin{aligned}
E\hat{\lambda}_c &= EE(\hat{\lambda}_c|\hat{\lambda}) \\
&= E\left(\frac{n-2}{n-1}\hat{\lambda}\right) \\
&= \frac{n-2}{n-1}E\hat{\lambda} \\
&= \frac{n-2}{n-1}\cdot\frac{n}{n-1}\lambda \\
&= \frac{n(n-2)}{(n-1)^2}\lambda.
\end{aligned}$$

The bias corrected estimator $\hat{\lambda}_c$ is thus still biased. However, for $n = 10$, $\frac{n(n-2)}{(n-1)^2} = 0.988$ so $\hat{\lambda}_c$ underestimates λ by only about 1.2%. In comparison, the original estimator overestimates λ by a factor of $\frac{n}{n-1} = 1.11$ or about 11%.

Problem 5

a) Since the counts have a multinomial distribution, the likelihood is

$$\begin{aligned}
L(p; Z_{AA}, Z_{Aa}, Z_{aa}) &= \frac{n!}{Z_{AA}!Z_{Aa}!Z_{aa}!}(p^2)^{Z_{AA}}(2p(1-p))^{Z_{Aa}}((1-p)^2)^{Z_{aa}} \qquad (1)\\
&= \frac{n!}{Z_{AA}!Z_{Aa}!Z_{aa}!}2^{Z_{Aa}}p^{2Z_{AA}+Z_{Aa}}(1-p)^{Z_{Aa}+2Z_{aa}} \qquad (2)
\end{aligned}$$

Taking the log the result follows.

b) Conditional on $X_{A-}, X_{aa}, p_{(t)}$, Z_{AA} has a binomial distribution with parameters X_{A-} and $p_{(t)}^2/(p_{(t)}^2 + 2p_{(t)}(1-p_{(t)}))$. Thus

$$\begin{aligned}
Z_{AA}^* &= E(Z_{AA}|X_{A-}, X_{aa}, p^{(t)}) \\
&= X_{A-}\frac{p_{(t)}^2}{p_{(t)}^2 + 2p(t)(1-p(t))}, \\
Z_{Aa}^* &= X_{A-}\Big(1 - \frac{p_{(t)}^2}{p_{(t)}^2 + 2p(t)(1-p(t))}\Big)
\end{aligned}$$

Since $Z_{aa} = X_{aa}$,

$$Z_{aa}^* = X_{aa}.$$

Taking conditional expectation of the log likelihood in point a) we find that

$$Q(p|p_{(t)}) = C + (2Z_{AA}^* + Z_{Aa}^*)\ln p + (2Z_{aa}^* + Z_{Aa}^*)\ln(1-p)$$

where C is a constant that depends on $p_{(t)}$ but not p.

Maximizing Q w.r.t. p, we find that that

$$p_{(t+1)} = \frac{2Z_{AA}^* + Z_{Aa}^*}{2Z_{AA}^* + 2Z_{aa}^* + 2Z_{AA}^*} = \frac{2Z_{AA}^* + Z_{Aa}^*}{2n}$$