Problem 1

a) The marginal density of X is

$$f_X(x,y) \propto \int_0^\infty x^2 e^{-xy} dy = x^2 \frac{1}{-x} e^{-xy} \Big|_0^\infty = x$$

for $0 < x < 1$ which means that the normalized density is $f_X(x) = 2x$ and so $F_X(x) = x^2$. We can thus generate X by setting $X = F_X^{-1}(U) = \sqrt{U}$ where $U \sim \text{unif}(0,1)$.

b) The conditional density of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = xe^{-yx},$$

that is, $Y|X = x \sim \exp(x)$. We can thus generate a realization of X, Y from its joint density by first using the method in point a) and then set $Y = -\frac{1}{X}\ln(1-V)$ where $V \sim \text{unif}(0,1)$ (again using the inversion method). This also gives us a realization of Y from its marginal density so no additional method is needed to generate Y.

c) Since $Y|X = x \sim \exp(x)$, $E(Y|X) = \frac{1}{X}$. Using the law of total expectation, we thus obtain

$$EY = EE(Y|X) = E\frac{1}{X} = \int_0^1 \frac{1}{x} 2x dx = 2$$

Thus, since Y has finite expectation, the strong law of large numbers implies that $\hat{\mu}_{MC} = \frac{1}{n}\sum_{i=1}^n Y_i$ converges almost surely to EY. In general, convergence almost surely implies convergence in probability.

However, since

$$\begin{aligned}
\operatorname{Var} Y &= E \operatorname{Var}(Y|X) + \operatorname{Var} E(Y|X) \\
&= E\frac{1}{X^2} + \operatorname{Var} \frac{1}{X} \\
&\geq \int_0^1 \frac{1}{x^2} 2x \, dx \\
&= 2 \ln x \Big|_0^1 \\
&= \infty
\end{aligned}$$

the variance of $\hat{\mu}_{MC}$ is also infinite and thus $E((\hat{\mu}_{MC} - EY)^2)$ will be infinite no matter how large we choose n. Hence, $\hat{\mu}_{MC}$ does not converge in mean square to EY. Estimating the standard error of $\hat{\mu}_{MC}$ involves estimating $\operatorname{Var} Y$ by the sample variance but this is clearly not possible when the estimand is infinite.

Problem 2

a) The expected survival time become

$$\begin{aligned}\mu_T &= E(T) \\ &= \int_0^\infty t \cdot 2ate^{-at^2}dt \\ &= \int_0^\infty \sqrt{\frac{u}{a}}e^{-u}du \\ &= \frac{1}{\sqrt{a}}\Gamma(\frac{3}{2}) \\ &= \frac{1}{2}\sqrt{\frac{\pi}{a}}\end{aligned}$$

Since $\ln a \sim N(\mu_0, \sigma_0^2)$ it follows that

$$\ln \mu_T = -\ln 2 - \frac{1}{2}\ln \pi - \frac{1}{2}\ln a \sim N(-\ln 2 - \frac{1}{2}\ln \pi - \frac{1}{2}\mu_0, \frac{1}{4}\sigma_0^2).$$

Accordingly, the prior on μ_T is also log-normal. The lower and upper 5% quantiles of μ_T are $\ln 1 = 0$ and $\ln 10$. Thus, the mean on log-scale

$$-\ln 2 - \frac{1}{2}\ln \pi - \frac{1}{2}\mu_0 = \frac{1}{2}\ln 10$$

which gives $\mu_0 = \ln 10 + \ln \pi + 2\ln 2 = 4.83$. Similarly, the lower 5% quantiles

$$0 = 4.83 - \frac{1}{2}\sigma_0 z_{0.05}$$

which gives $\sigma_0 = 4.93$.

b) Conditional on a each right censored survivaltime T_i follow left truncated Rayleigh distributions with densities

$$\pi(t_i|a, y_i, \delta_i) = \frac{f_T(t_i)}{P(T_i > y_i)} = \frac{2at_ie^{-at_i^2}}{e^{-ay_i^2}} = 2at_ie^{-a(t_i^2 - y_i^2)}I(t_i > y_i)$$

and since the survival times are independent conditional on a and the censoring indicators, their joint full conditional is

$$\pi(t_1, t_2, \ldots, t_m|a, y_i, \delta_i) = \prod_{i=1}^m 2at_ie^{-a(t_i^2 - y_i^2)}I(t_i > y_i).$$

Using the inversion method, we can simulate from the joint density by setting $u_i \sim \text{unif}(F_T(y_i|a), 1)$ and $t_i = F_T^{-1}(u_i) = \sqrt{-\frac{1}{a}\ln(1-u_i)}$ for each component $i = 1, 2, \ldots, m$.

c) Conditional on $\mathbf{t}$, the full conditional of a can be written

$$\begin{aligned}\pi(a|\mathbf{t}) &\propto \frac{1}{a} e^{-\frac{(\ln a-\mu_0)^2}{2\sigma_0^2}} \prod_{i=1}^{n} 2at_i e^{-at_i^2} \\ &\propto e^{-\frac{(\ln a-\mu_0)^2}{2\sigma_0^2}} a^{n-1} e^{-a\sum_{i=1}^n t_i^2}\end{aligned} \tag{1}$$

This suggest using a Gamma density

$$Q(a'|a,\mathbf{t}) = \frac{(\sum_{i=1}^n t_i^2)^n}{\Gamma(n)} a^{n-1} e^{-a\sum_{i=1}^n t_i^2} \tag{2}$$

as proposal. This leads to

$$\frac{\pi(a'|\mathbf{t})Q(a|a',\mathbf{t})}{\pi(a|\mathbf{t})Q(a'|a,\mathbf{t})} = e^{-\frac{1}{2\sigma_0^2}((\ln a'-\mu_0)^2-(\ln a-\mu_0)^2}$$

after cancellation of terms in normalizing constants and in targets against the proposals and log acceptance probability

$$\ln\alpha = \ln\min(1, \frac{\pi(a'|\mathbf{t})Q(a|a',\mathbf{t})}{\pi(a|\mathbf{t})Q(a'|a,\mathbf{t})}) = \min(0, -\frac{1}{2\sigma_0^2}((\ln a'-\mu_0)^2-(\ln a-\mu_0)^2). \tag{3}$$

It is evident from (3) that the acceptance probability will be high if the prior variance σ_0^2 is high (the prior is vague) and the posterior distribution of $\ln a$ is not located too far out in the tails of the prior.

It is tempting to rewrite the Gaussian term in (1) as

$$e^{-\frac{(\ln a-\mu_0)^2}{2\sigma_0^2}} = e^{-\frac{1}{2\sigma_0^2}((\ln a)^2-2\mu_0\ln a+\mu_0^2)} \propto e^{-\frac{1}{2\sigma_0^2}(\ln a)^2} a^{\frac{\mu_0}{\sigma_0^2}}$$

which suggests including the additional term μ_0/σ_0^2 in the shape parameter of the Gamma proposal. This leads to further simplifications in (3). We can expect a higher overall acceptance probability, however, only if the posterior of $\ln a$ happen to be located in the vicinity of 0 rather than μ_0. So in a real applied setting, this modification of (2) would perhaps in most cases not be a good idea given that μ_0 is already supposed to reflect our prior beliefs about $\ln a$.

d) In Fig. 3 we notice a quite strong posterior dependency between a and and the summary statistic $\sum_{i=1}^n t_i^2$. This is expected since a small value of a will tend to make $\sum_{i=1}^n t_i^2$ large and visa versa. It not unsurprising that the convergence of the Markov chain is slow when a and $\mathbf{t}$ are updated in separate Gibbs step.

Since we can both simulate from and evaluate the joint density of the entire block $t_1, t_2, \ldots, t_m$, this suggest constructing a block proposal where we first generate a proposal a' for a and then simulate $t_1, t_2, \ldots, t_m$ from its full conditional given a' using the method in point b). The optimal proposal for a here would be to use the posterior marginal of a but this is clearly not feasible is is reasonable to resort to some form of random walk proposal, e.g. a Gaussian. This leads to the block proposal

$$Q(a', \mathbf{t}'_{1:m} | a, \mathbf{t}_{1:m}) \propto e^{-\frac{(a'-a)^2}{2v}} \prod_{i=1}^{m} 2a' t_i e^{-a'(t_i^2 - y_i^2)} I(t_i > y_i).$$

The step length v will need tuning, and we can expect a quite large step length to work well since the joint proposal will align well with the overall shape of the joint target posterior.

Problem 3

a) The cdf of $\hat{\theta} = X_{(1)}$ becomes

$$\begin{aligned} F_{X_{(1)}}(x) &= P(X_{(1)} \leq x) \\ &= 1 - P(X_{(1)} > x) \\ &= 1 - P(X_1 > x \cap \cdots \cap X_n > x) \\ &= 1 - \prod_{i=1}^{n} P(X_i > x) \\ &= 1 - (e^{-\lambda(x-\theta)})^n \\ &= 1 - e^{-n\lambda(x-\theta)} \end{aligned}$$

Conditional on $\hat{\lambda}, \hat{\theta}$, the bootstrap replicates $\hat{\theta}^b$ would be distributed the same way but with $\hat{\lambda}, \hat{\theta}$ as the parameter values, that is,

$$F_{\hat{\theta}^b}(x | \hat{\lambda}, \hat{\theta}) = 1 - e^{-n\hat{\lambda}(x-\hat{\theta})}$$

Quantiles of the bootstrap replicates are given by the inverse of this cdf and so the percentile confidence interval would be given by

$$(\hat{\theta} - \frac{1}{n\hat{\lambda}} \ln(1 - \alpha/2), \hat{\theta} - \frac{1}{n\hat{\lambda}} \ln(\alpha/2))$$

if we neglect Monte-Carlo error.

The percentile method is valid if a transformation exist such that $\phi(\hat{\theta}) - \phi(\theta)$ has cdf satisfying $H(z) = 1 - H(-z)$.

Given that $\hat{\theta} = X_{(1)} > \theta$ with probability 1 and that the lower limit of the percentile interval is larger than $\hat{\theta}$, the percentile interval will contain θ with probability zero so the method fails very badly.

That the true coverage is different from the nominal $(1-\alpha)$-level implies that no such transformation H exists in the present case.

b) Since θ is a location parameter and $1/\lambda$ is a scale parameter, this suggest using

$$T = \frac{\hat{\theta} - \theta}{1/\hat{\lambda}}$$

as a pivotal quantity. Alternatively, since $\text{Var}(\hat{\theta}) = \frac{1}{n^2\lambda^2}$, we could use the estimator $1/(n\hat{\lambda})$ of standard error of $\hat{\theta}$ in the denominator, which would clearly lead to an identical interval.

The lower and upper $T_{\alpha/2}$ and $T_{1-\alpha/2}$ quantiles of T can be estimated from bootstrap replicates

$$T^b = \frac{\hat{\theta}^b - \hat{\theta}}{1/\hat{\lambda}^b}$$

of T from the fitted model $F(x; \hat{\lambda}, \hat{\theta})$. Since T and T^b have identical distributions, we then know that

$$P(T_{\alpha/2} < \frac{\hat{\theta} - \theta}{1/\hat{\lambda}} < T_{1-\alpha/2}) = 1 - \alpha$$

and so

$$(\hat{\theta} - \frac{1}{\hat{\lambda}} T_{1-\alpha/2}, \hat{\theta} - \frac{1}{\hat{\lambda}} T_{\alpha/2})$$

is a $(1-\alpha)$ confidence interval for θ.

Expressing the observations $X_i = \theta + Z_i/\lambda$ where Z_i are iid standard exponential (with the same releation applying to the order statistics) it follows that

$$\begin{aligned} T = (\hat{\theta} - \theta)\hat{\lambda} &= \frac{n(X_{(1)} - \theta)}{\sum_{i=1}^n X_i - X_{(1)}} \\ &= \frac{n(\theta + Z_{(1)}/\lambda - \theta)}{\sum_{i=1}^n \theta + Z_i/\lambda - \theta + Z_{(1)}/\lambda} \\ &= \frac{nZ_{(1)}}{\sum_{i=1}^n Z_i - Z_{(1)}}. \end{aligned}$$

Thus, since the distribution of $Z_1, Z_2, \ldots, Z_n$ does not depend on θ and λ, T is pivotal. Thus the interval is exact, at least if we neglect Monte-Carlo error or estimate the quantiles using the same method as in project 3.

As an aside, it can be noted that $2nZ_{(1)} \sim \chi^2_{2n}$. Also, $2\sum_{i=1}^{n}(Z_i - Z_{(1)}) \sim \chi^2_{2n-2}$ and independent of $Z_{(1)}$ since the $n-1$ nonzero terms in the sum, by the memoryless property of the exponential distribution, are all independent of $Z_{(1)}$ and standard exponentially distributed. Thus, T has a Fisher distribution with 2 and $2n-2$ degrees of freedom so in this case, estimating the quantiles via bootstrapping is not really needed.

c) Since $\hat{\theta} = X_{(1)}$ also has a shifted exponential distribution but with rate parameter $n\lambda$, it's clear that $E(\hat{\theta}) = \theta + \frac{1}{n\lambda}$.

Similarly, replacing the parameters with their original estimates and the estimates with their bootstrap replicates,

$$E(\hat{\theta}^b|\hat{\lambda},\hat{\theta}) = \hat{\theta} + \frac{1}{n\hat{\lambda}}.$$

We can express the bias-corrected estimator as

$$\hat{\theta}_c = 2\hat{\theta} - \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^b.$$

Thus,

$$\begin{aligned}
E(\hat{\theta}_c|\hat{\lambda},\hat{\theta}) &= E(2\hat{\theta} - \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^b|\hat{\lambda},\hat{\theta}) \\
&= 2\hat{\theta} - E(\hat{\theta}^b|\hat{\lambda},\hat{\theta}) \\
&= 2\hat{\theta} - E(\hat{\theta}^b|\hat{\lambda},\hat{\theta}) \\
&= 2\hat{\theta} - \hat{\theta} - \frac{1}{n\hat{\lambda}} \\
&= \hat{\theta} - \frac{1}{n\hat{\lambda}}
\end{aligned}$$

Using the law of total expectation,

$$\begin{aligned}
E\hat{\theta}_c &= E(E(\hat{\theta}_c|\hat{\lambda},\hat{\theta})) \\
&= E(\hat{\theta} - \frac{1}{n\hat{\lambda}}) \\
&= \theta + \frac{1}{n\lambda} - \frac{1}{n}E\frac{1}{\hat{\lambda}} \\
&= \theta + \frac{1}{n\lambda} - \frac{1}{n^2}E\sum_{i=1}^{n}(X_i - X_{(1)}) \\
&= \theta + \frac{1}{n\lambda} - \frac{1}{n^2}\frac{n-1}{\lambda} \\
&= \theta + \frac{1}{n^2\lambda}
\end{aligned}$$

In the second last step we have used the memoryless property of the exponential distribution which implies that the $n-1$ nonzero terms $X_i - X_{(1)}$ in the sum ("times passed since the first failure") are all exponentially distributed with parameter λ.

Thus $\hat{\theta}_c$ is not unbiased but the bias is reduced by a factor of n compared to that of $\hat{\theta}$.