STAT30040 Homework 8

Sarah Adilijiang

Problem 1

(a)

For large n, we have approximately: $T_1 = \frac{\hat{\lambda}-\lambda}{\sqrt{\lambda/n}} \sim N(0,1)$, so a $(1-\alpha)100\%$ confidence interval is:

$$P(-z_{\frac{\alpha}{2}} \leq \frac{\hat{\lambda}-\lambda}{\sqrt{\lambda/n}} \leq z_{\frac{\alpha}{2}}) = 1 - \alpha \quad \Rightarrow \quad (\frac{\hat{\lambda}-\lambda}{\sqrt{\lambda/n}})^2 \leq z_{\frac{\alpha}{2}}^2$$

$$\Rightarrow \quad \lambda^2 - (2\hat{\lambda} + z_{\frac{\alpha}{2}}^2/n)\lambda + \hat{\lambda}^2 \leq 0$$

$$\Rightarrow \quad \frac{2\hat{\lambda} + z_{\frac{\alpha}{2}}^2/n - z_{\frac{\alpha}{2}}/n\sqrt{z_{\frac{\alpha}{2}}^2 + 4n\hat{\lambda}}}{2} \leq \lambda \leq \frac{2\hat{\lambda} + z_{\frac{\alpha}{2}}^2/n + z_{\frac{\alpha}{2}}/n\sqrt{z_{\frac{\alpha}{2}}^2 + 4n\hat{\lambda}}}{2}$$

$$\Rightarrow \quad \frac{2\bar{X} + z_{\frac{\alpha}{2}}^2/n - z_{\frac{\alpha}{2}}/n\sqrt{z_{\frac{\alpha}{2}}^2 + 4n\bar{X}}}{2} \leq \lambda \leq \frac{2\bar{X} + z_{\frac{\alpha}{2}}^2/n + z_{\frac{\alpha}{2}}/n\sqrt{z_{\frac{\alpha}{2}}^2 + 4n\bar{X}}}{2}$$

For large n, we also have approximately: $T_2 = \frac{\hat{\lambda}-\lambda}{\sqrt{\hat{\lambda}/n}} \sim N(0,1)$, so a $(1-\alpha)100\%$ confidence interval is:

$$P(-z_{\frac{\alpha}{2}} \leq \frac{\hat{\lambda}-\lambda}{\sqrt{\hat{\lambda}/n}} \leq z_{\frac{\alpha}{2}}) = 1 - \alpha \quad \Rightarrow \quad \hat{\lambda} - z_{\frac{\alpha}{2}}\sqrt{\hat{\lambda}/n} \leq \lambda \leq \hat{\lambda} + z_{\frac{\alpha}{2}}\sqrt{\hat{\lambda}/n}$$

$$\Rightarrow \quad \bar{X} - z_{\frac{\alpha}{2}}\sqrt{\bar{X}/n} \leq \lambda \leq \bar{X} + z_{\frac{\alpha}{2}}\sqrt{\bar{X}/n}$$

(b)

The midpoint of the confidence interval derived from T_1 is: $\frac{2\bar{X}+z_{\frac{\alpha}{2}}^2/n}{2} = \bar{X} + z_{\frac{\alpha}{2}}^2/2n$

The midpoint of the confidence interval derived from T_2 is: $\bar{X}$

(c)

In the confidence interval derived from T_1, for the left endpoint we have:

$$\left(2\bar{X} + z_{\frac{\alpha}{2}}^2/n\right)^2 - \left(z_{\frac{\alpha}{2}}/n\sqrt{z_{\frac{\alpha}{2}}^2 + 4n\bar{X}}\right)^2 = 4\bar{X}^2 \geq 0$$

since both $2\bar{X} + z_{\frac{\alpha}{2}}^2/n$ and $z_{\frac{\alpha}{2}}/n\sqrt{z_{\frac{\alpha}{2}}^2 + 4n\bar{X}}$ are positive, so:

$$\frac{2\bar{X} + z_{\frac{\alpha}{2}}^2/n - z_{\frac{\alpha}{2}}/n\sqrt{z_{\frac{\alpha}{2}}^2 + 4n\bar{X}}}{2} \geq 0$$

Therefore, the confidence interval derived from T_1 is guaranteed to comprise on nonnegative numbers.

However, in the confidence interval derived from T_2, for the left endpoint we have:

$$\bar{X}^2 - \left(z_{\frac{\alpha}{2}}\sqrt{\bar{X}/n}\right)^2 = \bar{X}(\bar{X} - z_{\frac{\alpha}{2}}^2/n) = \frac{\bar{X}}{n}(\sum_{i=1}^{n} X_i - z_{\frac{\alpha}{2}}^2)$$

which could be positive or negative.

Therefore, the confidence interval derived from T_2 is NOT guaranteed to comprise on nonnegative numbers.

(d)

```r
set.seed(1)
z = qnorm(0.975,0,1); n=30
T1_in_CI = T2_in_CI = NULL

for (i in 1:1000) {
    x = rpois(n=30, lambda=1)
    x_bar = mean(x)

    # CI derived from T_1
    T1_l = (2*x_bar + z^2/n - z/n * sqrt(z^2+4*n*x_bar)) /2
    T1_r = (2*x_bar + z^2/n + z/n * sqrt(z^2+4*n*x_bar)) /2
    T1_in_CI[i] = (T1_l<=1 & 1<=T1_r)

    # CI derived from T_2
    T2_l = x_bar - z * sqrt(x_bar/n)
    T2_r = x_bar + z * sqrt(x_bar/n)
    T2_in_CI[i] = (T2_l<=1 & 1<=T2_r)
}

# proportion of 1000 CIs actually contain true lambda=1 for CI derived from T_1
mean(T1_in_CI)
```

```
## [1] 0.947
```

```r
# proportion of 1000 CIs actually contain true lambda=1 for CI derived from T_2
mean(T2_in_CI)
```

```
## [1] 0.934
```

Since we are constructing 95% confidence intervals for λ, the theoretically expected proportion of 1000 confidence intervals that contains the true parameter $\lambda = 1$ should be 95%.

The results of simulations show that 94.7% of confidence intervals derived from T_1 contains the true parameter $\lambda = 1$, and 93.4% of confidence intervals derived from T_2 contains the true parameter $\lambda = 1$. Thus the the confidence interval derived from T_1, i.e. $\dfrac{2\bar{X}+z^2_{\frac{\alpha}{2}}/n \pm z_{\frac{\alpha}{2}}/n\sqrt{z^2_{\frac{\alpha}{2}}+4n\bar{X}}}{2}$, gives a more accurate coverage probability.

Problem 2

(a)

Suppose the number of particles emitted from the radioactive source over the course of 10-second intervals for 180 such 10-second intervals are $X_1, ..., X_n$, where $n = 180$, so we have: $X_1, ..., X_n$ i.i.d. $\sim Poisson(10\lambda)$

Suppose the number of particles emitted from the radioactive source over the course of 10-second intervals for 180 such 10-second intervals are $Y_1, ..., Y_m$, where $m = 20$, so we have: $Y_1, ..., Y_m$ i.i.d. $\sim Poisson(20\lambda)$

Therefore, the likelihood function is:

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-10\lambda}(10\lambda)^{x_i}}{x_i!} \prod_{j=1}^{m} \frac{e^{-20\lambda}(20\lambda)^{y_j}}{y_j!}$$

$$\Rightarrow \quad l(\lambda) = \log L(\lambda) = -10n\lambda + \sum_{i=1}^{n} x_i \log 10\lambda - \sum_{i=1}^{n} \log x_i! - 20m\lambda + \sum_{j=1}^{m} y_j \log 20\lambda - \sum_{j=1}^{m} \log y_j!$$

2

$$= -(10n + 20m)\lambda + (\sum_{i=1}^{n} x_i + \sum_{j=1}^{m} y_j)\log\lambda + \left(\sum_{i=1}^{n} x_i \log 10 - \sum_{i=1}^{n} \log x_i! + \sum_{j=1}^{m} y_j \log 20 - \sum_{j=1}^{m} \log y_j!\right)$$

$$\Rightarrow \quad \frac{\partial l(\lambda)}{\partial \lambda} = -(10n + 20m) + (\sum_{i=1}^{n} x_i + \sum_{j=1}^{m} y_j)/\lambda = 0$$

$$\Rightarrow \quad \hat{\lambda} = \frac{\sum\limits_{i=1}^{n} x_i + \sum\limits_{j=1}^{m} y_j}{10n + 20m} = \frac{310 + 37}{2200} \approx 0.1577$$

Since $\frac{\partial^2 l(\lambda)}{\partial \lambda^2} = -(\sum_{i=1}^{n} x_i + \sum_{j=1}^{m} y_j)/\lambda^2 < 0$, so the $\hat{\lambda}$ here is the maximum likelihood estimator.

(b)

Since $X_1, ..., X_n$ i.i.d. $\sim Poisson(10\lambda)$, and $Y_1, ..., Y_m$ i.i.d. $\sim Poisson(20\lambda)$, also X_i and Y_j are independent, thus the sum of all 200 counts has a Poisson distribution:

$$S_N = \sum_{k=1}^{N} z_i = \sum_{i=1}^{n} x_i + \sum_{j=1}^{m} y_j \sim Poisson((10n + 20m)\lambda) = Poisson(2200\lambda)$$

where $n = 180, m = 20, N = 200$

By Central Limit Theorem, for large N=200, we have:

$$\frac{S_N - E(S_N)}{\sqrt{Var(S_N)}} = \frac{S_N - N\lambda}{\sqrt{N\lambda}} = \frac{S_N/N - \lambda}{\sqrt{\lambda/N}} = \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/N}} \sim N(0, 1)$$

$$\Rightarrow \quad \hat{\lambda} \sim N(\lambda, \ \lambda/N)$$

The random variable here $\frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/N}}$ is exactly the $T_1 = \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/n}}$ in problem 1, so the CLT suggests the approximation:

$\frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/N}} \sim N(0, 1)$, i.e. $\hat{\lambda} \sim N(\lambda, \ \lambda/N)$, where $N = 200$ and $\hat{\lambda} = S_N/N = \frac{\sum\limits_{i=1}^{n} x_i + \sum\limits_{j=1}^{m} y_j}{10n + 20m} = \frac{310 + 37}{2200} \approx 0.1577$.

Problem 3

(a)

Prior density of θ is: $f(\theta) = 1/\theta$ $(\theta > 0)$

Likelihood is: $f(X_1, ..., X_n | \theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum\limits_{i=1}^{n} x_i}$ $(x_i > 0 \ \forall i)$

So the posterior density of θ is:

$$f(\theta | X_1, ..., X_n) \propto f(\theta) \times f(X_1, ..., X_n | \theta) = 1/\theta \times \theta^n e^{-\theta \sum\limits_{i=1}^{n} x_i} = \theta^{n-1} e^{-\theta \sum\limits_{i=1}^{n} x_i}$$

Since $\theta > 0$, $\sum_{i=1}^{n} x_i > 0$, $n > 0$, thus we can set $\alpha = n$, $\lambda = \sum_{i=1}^{n} x_i$, and get the posterior distribution of θ:

$$\theta | X_1, ..., X_n \sim Gamma(\alpha = n, \ \lambda = \sum_{i=1}^{n} x_i)$$

3

so the posterior density of θ is:

$$f(\theta|X_1,...,X_n) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\lambda\theta} \qquad (\theta > 0, \ \lambda = \sum_{i=1}^{n} x_i > 0, \ \alpha = n > 0)$$

(b)

$n = 64, \ \sum_{i=1}^{n} x_i = 1769.61$ (days), so the posterior distribution of θ is:

$$\theta|X_1,...,X_n \sim Gamma(\alpha = n = 64, \ \lambda = \sum_{i=1}^{n} x_i = 1769.61)$$

Thus the posterior mean of θ is: $E(\theta)_{post} = \frac{\alpha}{\lambda} = \frac{64}{1769.61} \approx 0.0362$

And the posterior variance of θ is : $Var(\theta)_{post} = \frac{\alpha}{\lambda^2} = \frac{64}{1769.61^2} \approx 2.0437 \times 10^{-5}$

By CLT, $\theta_{post} \sim Gamma(\alpha, \lambda) \approx N(E(\theta)_{post}, Var(\theta)_{post})$, thus the approximate 95% highest posterior density credible interval (HPD) for θ is:

$$E(\theta)_{post} \ \pm z_{0.025}\sqrt{Var(\theta)_{post}} \ \approx \ (0.0273, \ 0.0450)$$

OR

Use R to find the exact 95% HPD credible interval of θ:

```
x = seq(0,0.05,0.0001)
lens = NULL
for (i in 1:length(x)) {
    prob = x[i]
    lens[i] = qgamma(prob+0.95, shape=64, rate=1769.61) - qgamma(prob, shape=64, rate=1769.61)
}

# the shortest 95% HPD interval
qgamma(x[which.min(lens)], shape=64, rate=1769.61)
```

[1] 0.02750595

```
qgamma(x[which.min(lens)]+0.95, shape=64, rate=1769.61)
```

[1] 0.0451412

Thus the exact 95% highest posterior density credible interval (HPD) for θ is:

$$(0.0275, \ 0.0451)$$

which is very close to the appoximate HPD interval above, because the posterior density of distribution $\theta|X_1,...,X_n \sim Gamma(\alpha = 64, \ \lambda = 1769.61)$ (shown in the figure below) is nearly symmetric and unimodal, thus the HPD interval will approximately coincide with the interval between the percentiles.

```
x = seq(0.01,0.07,0.00001)
plot(dgamma(x, shape=64, rate=1769.61) ~ x, cex=0.2, main=expression(paste("Gamma Distribution of ", th
```

Gamma Distribution of θ
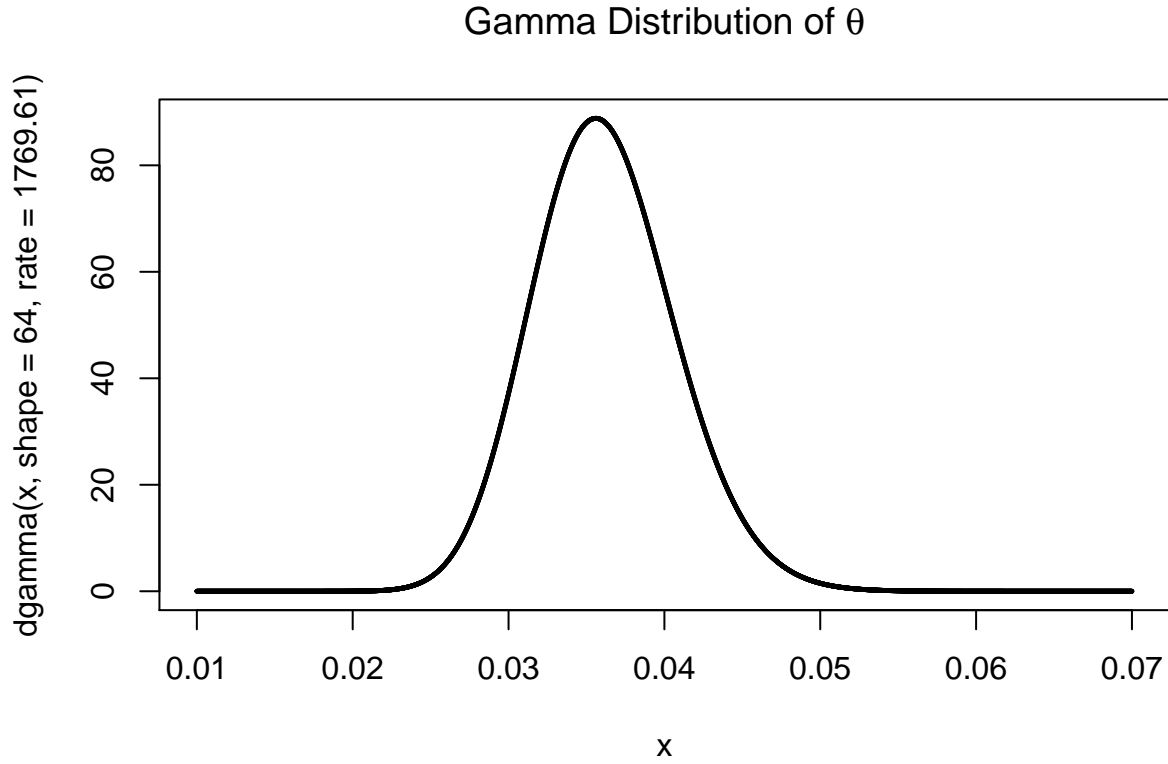
(c)

From question (b), we get the posterior distribution of θ is:

$$\theta | X_1, ..., X_n \sim Gamma(\alpha = n = 64, \ \lambda = \sum_{i=1}^{n} x_i = 1769.61)$$

Since $\mu = 1/\theta$, thus the posterior mean of μ is: $E(\mu)_{post} = E(\frac{1}{\theta})_{post} = \frac{\lambda}{\alpha - 1} = \frac{1769.61}{64 - 1} \approx 28.0891$

And the posterior variance of μ is : $Var(\mu)_{post} = Var(\frac{1}{\theta})_{post} = E(\frac{1}{\theta^2})_{post} - (E(\frac{1}{\theta})_{post})^2 = \frac{\lambda^2}{(\alpha - 1)(\alpha - 2)} - (\frac{\lambda}{\alpha - 1})^2 = \frac{\lambda^2}{(\alpha - 1)^2 (\alpha - 2)} = \frac{1769.61^2}{(64 - 1)^2 (64 - 2)} \approx 12.7257$

By CLT, approximately we have: $\mu_{post} \sim N(E(\mu)_{post}, Var(\mu)_{post})$, thus the 95% highest posterior density credible interval (HPD) for μ is:

$$E(\mu)_{post} \ \pm z_{0.025} \sqrt{Var(\mu)_{post}} \ \approx \ (21.0972, \ 35.0809)$$

Translate this interval back into the interval for θ is:

$$(\frac{1}{35.0809}, \ \frac{1}{21.0972}) \approx (0.0285, \ 0.0474)$$

OR

Use R to find the exact 95% HPD credible interval of μ, where $\mu \sim invGamma(\alpha = n = 64, \ \lambda = \sum_{i=1}^{n} x_i = 1769.61)$:

```r
library(invgamma)
```

```
## Warning: package 'invgamma' was built under R version 3.5.2
```

```r
x = seq(0,0.05,0.0001)
lens = NULL
for (i in 1:length(x)) {
    prob = x[i]
    lens[i] = qinvgamma(prob+0.95, shape=64, rate=1769.61) - qinvgamma(prob, shape=64, rate=1769.61)
}

# the shortest 95% HPD interval
qinvgamma(x[which.min(lens)], shape=64, rate=1769.61)
```

```
## [1] 21.47405
```

```r
qinvgamma(x[which.min(lens)]+0.95, shape=64, rate=1769.61)
```

```
## [1] 35.23283
```

Thus the exact 95% highest posterior density credible interval (HPD) for μ is:

$$(21.4741, 35.2328)$$

which is a little different from the appoximate HPD interval above, because the posterior Inverse Gammma distribution of μ is not symmetric (shown in the figure below), thus the HPD interval will not coincide with the interval between the percentiles.

Translate this interval back into the interval for θ is:

$$(\frac{1}{35.2328}, \frac{1}{21.4741}) \approx (0.0284, 0.0466)$$

```r
x = seq(10,50,0.001)
plot(dinvgamma(x, shape=64, rate=1769.61) ~ x, cex=0.2, main=expression(paste("Inverse Gamma Distributio
```

Inverse Gamma Distribution of μ
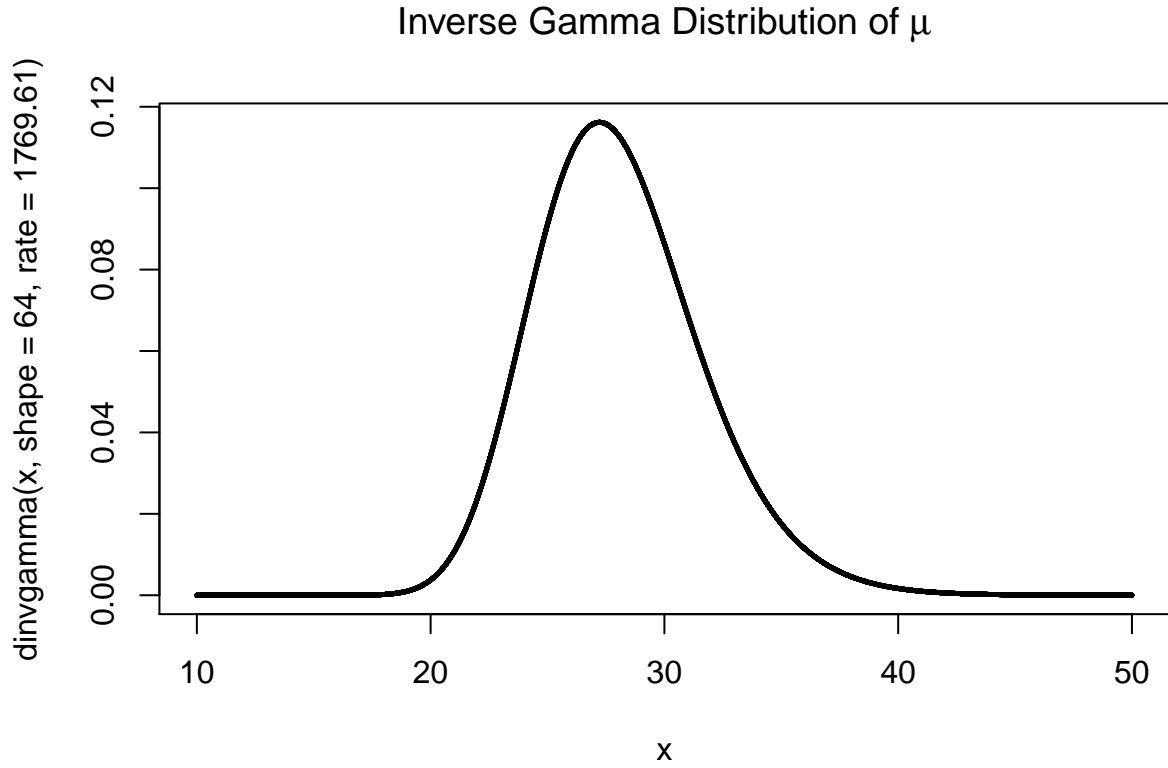
(d)

The intervals in (b) and (c) are different, though a little similar.

Since the posterior distribution of θ is a Gamma distribution that is nearly symmetric and unimodal, thus the HPD interval will approximately coincide with the interval between the percentiles.

However, the posterior distribution of μ is an Inverse Gamma distribution that is not symmetric though being unimodal, thus the HPD interval will not coincide with the interval between the percentiles. So when transferring it back to an interval for θ, it will be different from the HPD interval that is directly computed from the posterior Gamma distribution of θ. But these two HPD intervals are a little similar with each other, because the Inverse Gamma distribution here is not that skewed.

Therefore, these two HPD intervals should be different but a little similar with each other.

(e)

Suppose the waiting time for the next major earthquakes is Y, and $Y \sim Exp(\theta)$, i.e. $f(y|\theta) = \theta e^{-\theta y}$ $(y > 0)$

In question (b), we get the posterior density of θ is:

$$f(\theta|X_1, ..., X_n) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta} \qquad (\theta > 0, \ \lambda = \sum_{i=1}^{n} x_i = 1769.61, \ \alpha = n = 64)$$

Thus the joint density of (Y, θ) is:

$$f(y, \theta|X_1, ..., X_n) = f(y|\theta) \times f(\theta|X_1, ..., X_n) = \theta e^{-\theta y} \times \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta} = \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^\alpha e^{-(\lambda+y)\theta} \quad (y > 0, \theta > 0)$$

Then the marginal density of Y is:

$$f(y|X_1, ..., X_n) = \int_0^\infty f(y, \theta|X_1, ..., X_n)d\theta = \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)}\theta^\alpha e^{-(\lambda+y)\theta}d\theta$$

$$= \frac{\alpha\lambda^\alpha}{(\lambda+y)^{\alpha+1}}\int_0^\infty \frac{(\lambda+y)^{\alpha+1}}{\Gamma(\alpha+1)}\theta^{(\alpha+1)-1}e^{-(\lambda+y)\theta}d\theta = \frac{\alpha\lambda^\alpha}{(\lambda+y)^{\alpha+1}} \quad (y > 0)$$

Therefore, the (marginal) probability that the waiting time for the next major quake will be at least 60 days is:

$$P(Y \geq 60|X_1, ..., X_n) = \int_{60}^\infty f(y|X_1, ..., X_n)dy = \int_{60}^\infty \frac{\alpha\lambda^\alpha}{(\lambda+y)^{\alpha+1}}dy = \left(\frac{\lambda}{\lambda+60}\right)^\alpha = \left(\frac{1769.61}{1769.61+60}\right)^{64} \approx 0.1184$$

Problem 4

(a)

Prior density of (θ_X, θ_Y) is: $f(\theta_X, \theta_Y) = \frac{1}{\theta_X \theta_Y} \quad (\theta_X, \theta_Y > 0)$

Likelihood is: $f(X_1, ..., X_n, Y_1, ..., Y_m|\theta_X, \theta_Y) = \prod_{i=1}^n \theta_X e^{-\theta_X x_i} \prod_{j=1}^m \theta_Y e^{-\theta_Y y_j} = \theta_X^n e^{-\theta_X \sum_{i=1}^n x_i} \theta_Y^m e^{-\theta_Y \sum_{j=1}^m y_j}$

So the posterior density of (θ_X, θ_Y) is:

$$f(\theta_X, \theta_Y|X_1, ..., X_n, Y_1, ..., Y_m) \propto f(\theta_X, \theta_Y) \times f(X_1, ..., X_n, Y_1, ..., Y_m|\theta_X, \theta_Y) = \theta_X^{n-1} e^{-\theta_X \sum_{i=1}^n x_i} \theta_Y^{m-1} e^{-\theta_Y \sum_{j=1}^m y_j}$$

Since:

$$\int_{\theta_X, \theta_Y} f(\theta_X, \theta_Y|X_1, ..., X_n, Y_1, ..., Y_m)d\theta_X d\theta_Y = c \times \int_0^\infty \theta_X^{n-1} e^{-\theta_X \sum_{i=1}^n x_i} d\theta_X \int_0^\infty \theta_Y^{m-1} e^{-\theta_Y \sum_{j=1}^m y_j} d\theta_Y = 1$$

$$\Rightarrow \quad c = \frac{(\sum_{i=1}^n x_i)^n (\sum_{j=1}^m y_j)^m}{\Gamma(n)\,\Gamma(m)}$$

$$\Rightarrow \quad f(\theta_X, \theta_Y|X_1, ..., X_n, Y_1, ..., Y_m) = \frac{(\sum_{i=1}^n x_i)^n (\sum_{j=1}^m y_j)^m}{\Gamma(n)\,\Gamma(m)}\theta_X^{n-1} e^{-\theta_X \sum_{i=1}^n x_i} \theta_Y^{m-1} e^{-\theta_Y \sum_{j=1}^m y_j} \quad (\theta_X, \theta_Y > 0)$$

Set $U = \frac{\theta_Y}{\theta_X}$, $V = \theta_X$, so $\theta_X = V$, $\theta_Y = UV$, and we have:

$$J = det\begin{pmatrix} \frac{\partial \theta_X}{\partial u} & \frac{\partial \theta_X}{\partial v} \\ \frac{\partial \theta_Y}{\partial u} & \frac{\partial \theta_Y}{\partial v} \end{pmatrix} = det\begin{pmatrix} 0 & 1 \\ v & u \end{pmatrix} = -v$$

Thus the joint density of (U, V) is:

$$f_{U,V}(u, v) = f_{\theta_X, \theta_Y}(v, uv)\,|J| = \frac{(\sum_{i=1}^n x_i)^n (\sum_{j=1}^m y_j)^m}{\Gamma(n)\,\Gamma(m)}v^{n-1} e^{-v \sum_{i=1}^n x_i} (uv)^{m-1} e^{-uv \sum_{j=1}^m y_j} v \quad (u, v > 0)$$

8

Therefore, the posterior density of $U = \frac{\theta_Y}{\theta_X}$ is:

$$f_U(u) = \int_0^\infty f_{U,V}(u,v)dv = \int_0^\infty \frac{(\sum\limits_{i=1}^n x_i)^n \ (\sum\limits_{j=1}^m y_j)^m}{\Gamma(n)\ \Gamma(m)} v^{n+m-1}\ u^{m-1}\ e^{-v\ (\sum\limits_{i=1}^n x_i + u \sum\limits_{j=1}^m y_j)} dv$$

$$= \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m)} \frac{(\sum\limits_{i=1}^n x_i)^n \ (u\sum\limits_{j=1}^m y_j)^m}{(\sum\limits_{i=1}^n x_i + u \sum\limits_{j=1}^m y_j)^{n+m}} \frac{1}{u} = \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m)} \left(\frac{\sum\limits_{i=1}^n x_i}{\sum\limits_{i=1}^n x_i + u \sum\limits_{j=1}^m y_j} \right)^n \left(\frac{u\sum\limits_{j=1}^m y_j}{\sum\limits_{i=1}^n x_i + u \sum\limits_{j=1}^m y_j} \right)^m \frac{1}{u}$$

$$= \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m)} \left(\frac{n\bar{X}/m\bar{Y}}{n\bar{X}/m\bar{Y} + u} \right)^n \left(\frac{u}{n\bar{X}/m\bar{Y} + u} \right)^m \frac{1}{u} \qquad (u > 0)$$

When $n = m$, the posterior density of $U = \frac{\theta_Y}{\theta_X}$ is:

$$f_U(u) = \frac{\Gamma(2n)}{\Gamma(n)^2} \left(\frac{\bar{X}/\bar{Y}}{\bar{X}/\bar{Y} + u} \right)^n \left(\frac{u}{\bar{X}/\bar{Y} + u} \right)^n \frac{1}{u} \qquad (u > 0)$$

(b)

In homework 1 problem 5, we got that:

$$\frac{\bar{X}}{\bar{Y}} \ \sim \ \frac{\theta_Y}{\theta_X} \ F(2n, 2m)$$

So by CLT, we can get the approximate distribution:

$$\frac{\frac{\bar{X}}{\bar{Y}} - \frac{\theta_Y}{\theta_X} E[F(2n,2m)]}{\frac{\theta_Y}{\theta_X} \sqrt{Var[F(2n,2m)]}} \ \sim \ N(0,1)$$

from which, we will be able to easily obtain the inferences of $\frac{\theta_Y}{\theta_X}$.

However, for $\theta_Y - \theta_X$, there is no such convenient approximate distributions.

Therefore, it is a good reason to consider the inference of θ_Y/θ_X instead of the inference of $\theta_Y - \theta_X$.

Problem 5

(a)

The model is $Y = x_1\beta + \varepsilon$, where $Y \in R^d, x_1 \in R^d, \beta \in R, \varepsilon \in R^d$

From the figure, we have: $x_1 \perp Y \ \Rightarrow \ x_1^T Y = x_1 \cdot Y = 0$, so:

$$\hat{\beta} = (x_1^T x_1)^{-1} x_1^T Y = ||x_1||^2 x_1^T Y = 0$$

and

$$RSS = ||\hat{\varepsilon}||^2 = ||Y - x_1\hat{\beta}||^2 = ||Y||^2 = 1$$

(b)

The model is $Y = x_2\beta + \varepsilon$, where $Y \in R^d, x_2 \in R^d, \beta \in R, \varepsilon \in R^d$

From the figure, we can roughly get that the length of x_2 is half the length of Y, and the angle between Y and x_2 is roughly 30 degrees. Thus: $||x_2|| = \frac{||Y||}{2} = \frac{1}{2}$, and:

$$x_2^T Y = x_2 \cdot Y = ||x_2|| \times ||Y|| \times \cos\frac{\pi}{6} = \frac{\sqrt{3}}{4}$$

So:

$$\hat{\beta} = (x_2^T x_2)^{-1} x_2^T Y = \frac{x_2^T Y}{||x_2||^2} = \sqrt{3}$$

and

$$RSS = ||\hat{\varepsilon}||^2 = (||Y|| \times \sin\frac{\pi}{6})^2 = \frac{1}{4}$$

(c)

For model $Y = x_2\beta + \varepsilon$, we get $n = d, p = 1$, so the degrees of freedom is: $df = n - p = d - 1$

Since under model assumptions: $\hat{\beta} \sim N(\beta, \sigma^2(x_2^T x_2)^{-1})$, so:

$$\hat{Var}(\hat{\beta}) = \hat{\sigma}^2(x_2^T x_2)^{-1} = \frac{RSS}{d-1}\frac{1}{||x_2||^2} = \frac{1}{d-1}$$

Under the null hypothesis: $H_0 : \beta = 0$, we have:

$$\frac{\hat{\beta} - 0}{\sqrt{\hat{Var}(\hat{\beta})}} = \frac{\hat{\beta}}{\sqrt{\frac{1}{d-1}}} = \hat{\beta}\sqrt{d-1} \sim t_{d-1}$$

Therefore, since $d \geq 20$, the upper bond for p-value is:

$$p = 2 \times P\left(\hat{\beta}\sqrt{d-1} > \sqrt{3(d-1)} \,|H_0\right) = 2 \times P\left(t_{d-1} > \sqrt{3(d-1)}\right) \leq 2 \times P\left(t_{19} > \sqrt{3 \times 19}\right) = 3.9182 \times 10^{-7}$$

So the p-value is very small, nearly close to zero, thus we reject the null hypothesis that the regression coefficient β is 0.

(d)

The model is $Y = X\beta + \varepsilon$, where $Y \in R^d, \varepsilon \in R^d, X = (x_1\ x_2) \in R^{d \times 2}, \beta = (\beta_1\ \beta_2)^T \in R^2$

Since Y is already in the $span(X)$, so we have the projection of Y to the span of X is: $\hat{Y} = P_X Y = Y$.

Also, from the figure, we can roughly get that the length of x_1 is half the length of x_2, and the angle between x_1 and x_2 is roughly 120 degrees. Thus: $||x_1|| = \frac{||x_2||}{2} = \frac{1}{4}$, and:

$$x_1 \cdot x_2 = ||x_1|| \times ||x_2|| \times \cos\frac{2\pi}{3} = -\frac{1}{16}$$

So:

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{pmatrix} ||x_1||^2 & x_1 \cdot x_2 \\ x_1 \cdot x_2 & ||x_2||^2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \cdot Y \\ x_2 \cdot Y \end{pmatrix} = \begin{pmatrix} \frac{1}{16} & -\frac{1}{16} \\ -\frac{1}{16} & \frac{1}{4} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \frac{\sqrt{3}}{4} \end{pmatrix} = \frac{16}{3}\begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ \frac{\sqrt{3}}{4} \end{pmatrix} = \begin{pmatrix} \frac{4\sqrt{3}}{3} \\ \frac{4\sqrt{3}}{3} \end{pmatrix}$$

which means $\hat{\beta}_1 = \hat{\beta}_2 = \frac{4\sqrt{3}}{3}$

Also:

$$RSS = ||\hat{\varepsilon}||^2 = ||Y - \hat{Y}||^2 = 0$$