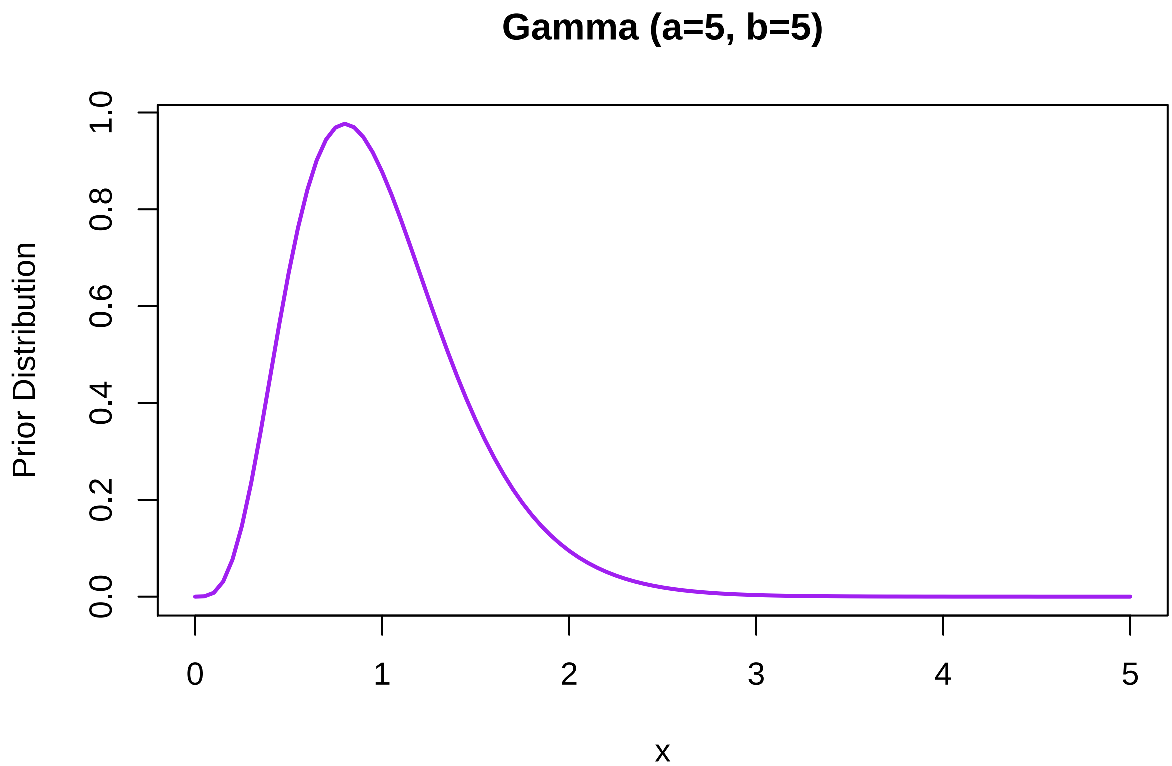
Modelling lifetimes of Electric Bulbs

Homework 2: Bayesian Data Analysis by Ankit Tewari

It is goven that experts agree on choosing as a prior distribution for our parameter λ, a $Gamma(a = 5, b = 5)$ distribution with pdf defined as follows:

$$\pi(\lambda) = \frac{b^a \lambda^{(a-1)} e^{(-b\lambda)}}{\Gamma(a)}, \lambda \geq 0$$

```r
curve(dgamma(x, shape = 5, rate = 5), col="purple", lwd=2, ylab = "Prior Distribution",
      main="Gamma (a=5, b=5)", xlim = c(0,5))
```



1. Statistical Model and Bayesian Model

- What is the statistical model and the Bayesian model when one observes a whole sample of n independent times to failure?

The statistical model for this experiment is the exponential model with the paramter λ:

$$M = \{X \sim exponential(\lambda), \lambda \geq 0\}$$

And the Bayesian model is the statistical model including the distribution for our parameter λ which is:

$$Bayesian\_Model = \{X \sim exponential(\lambda), \lambda \sim Gamma(a = 5, b = 5), x \geq 0\}$$

- What is the statistical model and the Bayesian model when only the smallest time to failure is observed?

The statistical model in this case is described as follows-

$$M = \{Y \sim n\lambda e^{-\lambda ny}, y \geq 0\}$$

and the Bayesian model is described by,

$$Bayesian\_Model = \{Y \sim n\lambda e^{-\lambda ny}, y \geq 0, \lambda \sim Gamma(a = 5, b = 5)\}$$

- What is the statistical model and the Bayesian model when only the largest time to failure is observed?

$$M = \{Z \sim n\lambda(1 - e^{-\lambda z})^{(n-1)} e^{-\lambda z}, \lambda \geq 0\}$$

and the Bayesian Model is

$$Bayesian\_Model = \{Z \sim n\lambda(1 - e^{-\lambda z})^{(n-1)} e^{-\lambda z}, \lambda \sim Gamma(a = 5, b = 5), \}$$

2. Likelihood Function

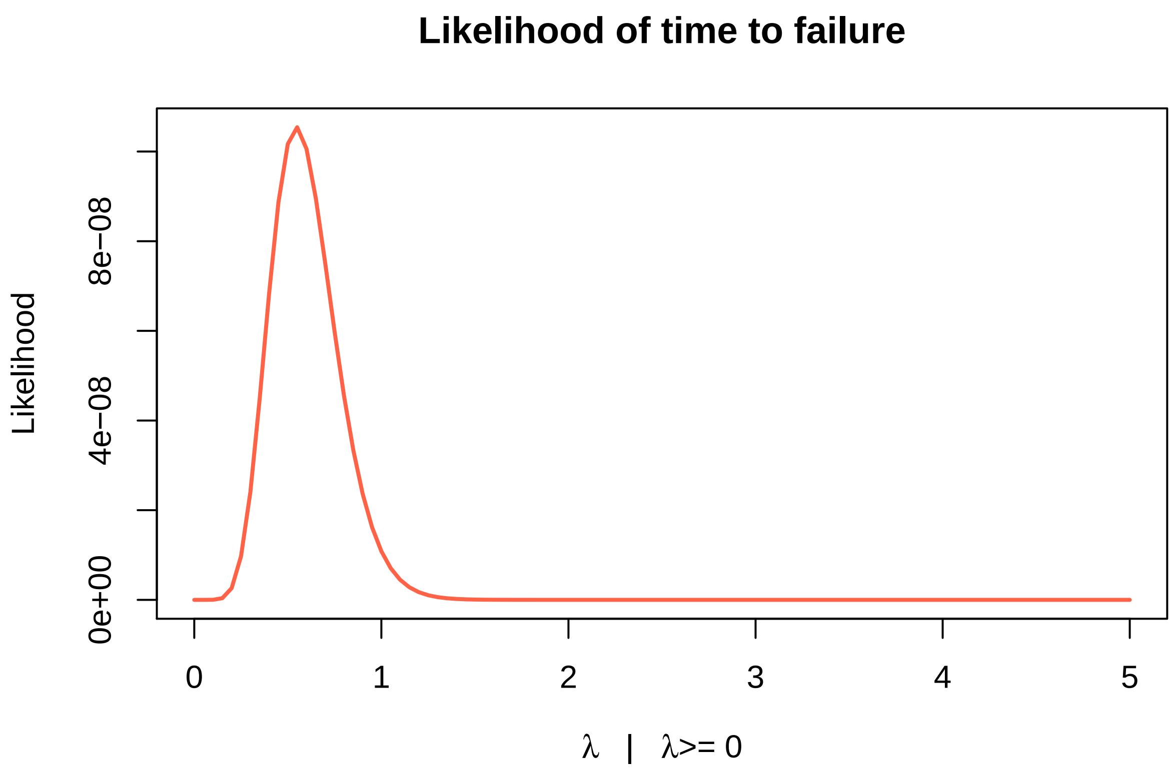
What is the likelihood function when one observes the whole sample of n = 10 times to failure, with $\sum_{i=1}^{10} xi = 18.34$?

$$p(x|\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}, x \geq 0$$

on simplifying with n=10, $\sum_{i=1}^{10} xi = 18.34$,

$$p(x|\lambda) = \lambda^{10} e^{-18.34\lambda}, x \geq 0$$

```
data <- c(2.05, 0.37, 0.47, 7.18, 0.36, 1.24, 3.95, 0.49, 1.33, 0.90)
plot( function(x) {return(x^length(data)*exp(-x*sum(data)))}, col="tomato",
      lwd=2, ylab = "Likelihood",
      xlab = expression(paste(lambda,"    " ,"|","    ",lambda,">= 0")),
      main= " Likelihood of time to failure ",
      xlim = c(0,5))
```



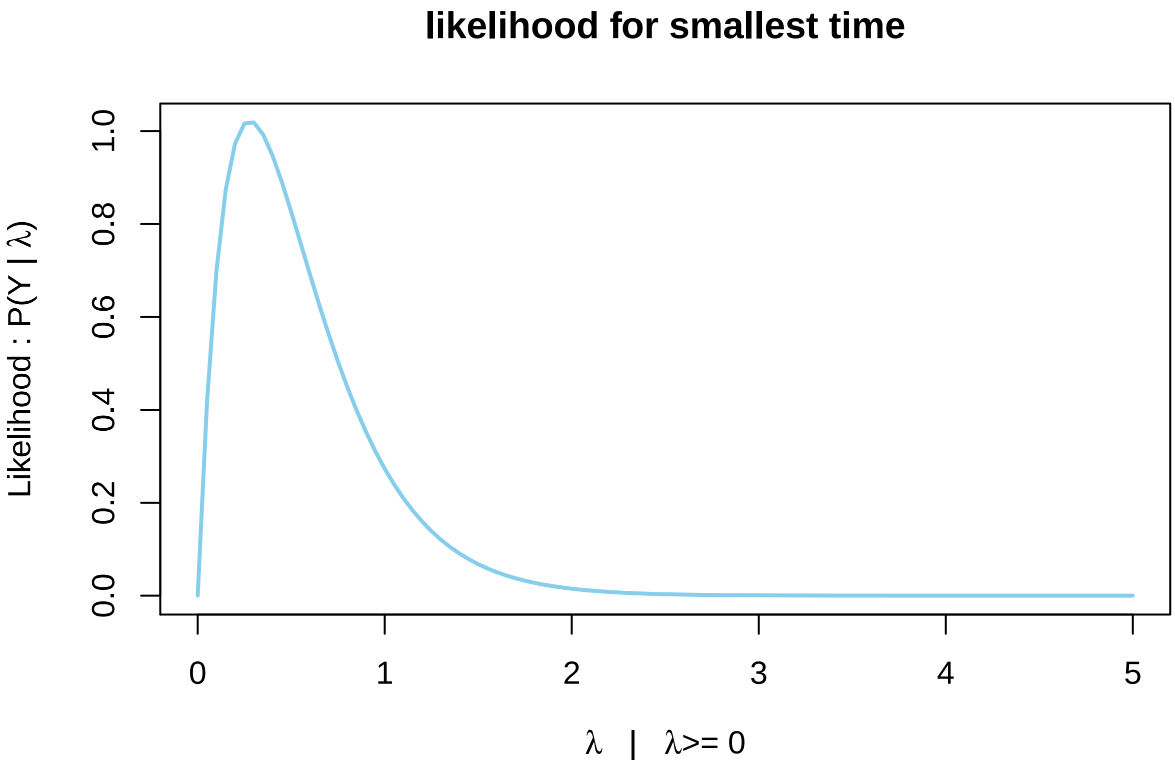
What is the likelihood function when only the smallest time to failure, $y = 0.36$, is observed?

$$p(y|\lambda) = n\lambda e^{-\lambda n y}, y \geq 0$$

on simplifying and substuting the values of n and y, we get,

$$p(y|\lambda) = 10\lambda e^{-3.6\lambda}, y \geq 0$$

```
plot(function(lambda) return(length(data)*lambda*exp(-lambda*length(data)*0.36) ),
     col = "skyblue", lwd=2, main = "likelihood for smallest time",
     ylab =expression(paste( "Likelihood : P(Y | ",lambda,")" ) ),
     xlab = expression(paste(lambda,"   " ,"|","   ",lambda,">= 0")) ,
     xlim = c(0,5))
```



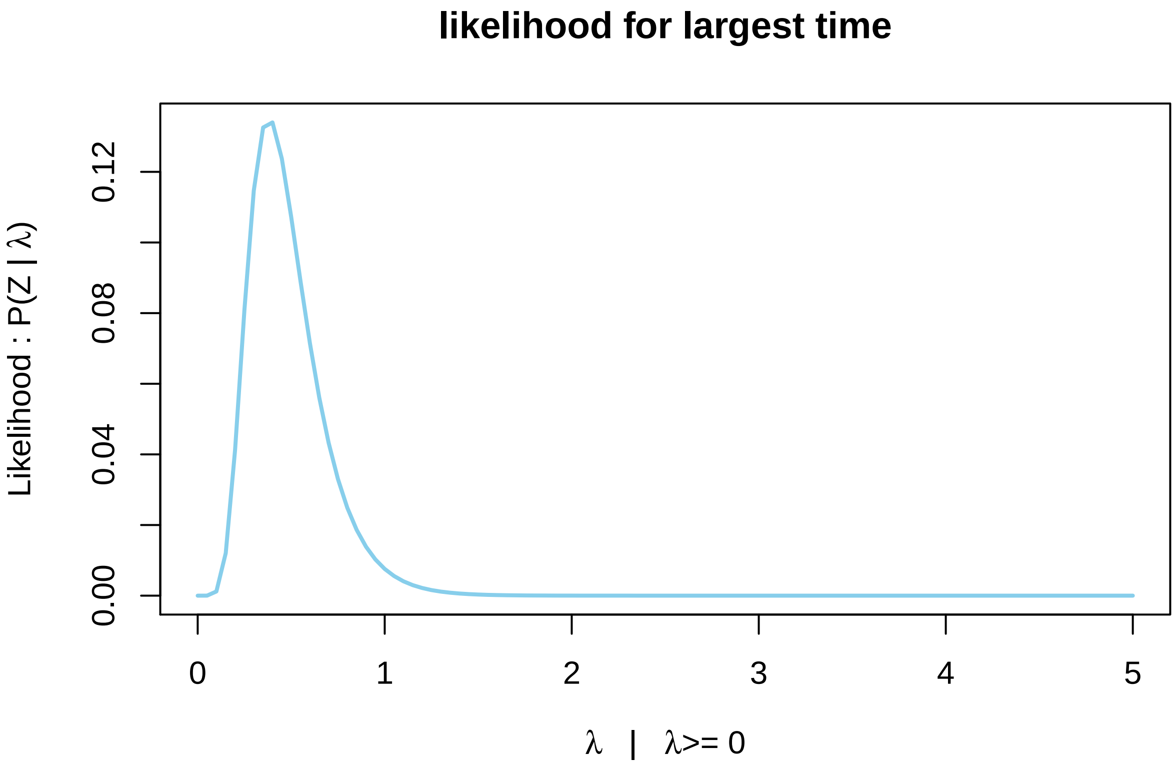
What is the likelihood function when only the largest time to failure, z = 7.18, is observed?

$$p(z|\lambda) = n\lambda(1 - e^{-\lambda z})^{(n-1)}e^{-\lambda z}, z \geq 0$$

simplifying with z=7.18 and n=10, we get,

$$p(z|\lambda) = 10\lambda(1 - e^{-7.18\lambda})^{(9)}e^{-7.18\lambda}$$

```
z = 7.18
plot(function(lambda)
  return( length(data)*lambda* (1 - exp(-lambda*z))^(length(data)-1)*exp(-lambda*z) ),
     col = "skyblue", lwd=2, main = "likelihood for largest time",
     ylab =expression(paste( "Likelihood : P(Z | ",lambda,")" ) ),
     xlab = expression(paste(lambda,"   " ,"|","   ",lambda,">= 0")),
  xlim = c(0,5))
```



3. Posterior Distributions

Now, if we consider the posterior distributions, we will observe that in the case of `whole sample`,

$$p(\lambda|x) = \frac{p(x|\lambda).\pi(\lambda)}{p(x)} = \frac{\lambda^n.e^{-\lambda\sum_{i=1}^n x_i}\frac{b^a\lambda^{(a-1)}e^{(-b\lambda)}}{\Gamma(a)}}{\int_0^\infty \lambda^n.e^{-\lambda\sum_{i=1}^n x_i}\frac{b^a\lambda^{(a-1)}e^{(-b\lambda)}}{\Gamma(a)}d\lambda}$$

on simplifying and using the values of a, b, n and $\sum_{i=1}^n x_i = 18.34$, we get,

$$p(\lambda|x) = \lambda^{14}e^{-23.34\lambda}\frac{(23.34)^{15}}{\Gamma(15)}$$

Also, note that,

$$p(\lambda|x) \propto p(x|\lambda).\pi(\lambda) = \lambda^{a+n-1}e^{-\lambda(b+\sum_{i=1}^n x_i)}\frac{b^a}{\Gamma(a)}$$

and,

$$p(x|\lambda).\pi(\lambda) \propto \lambda^{a+n-1}e^{-\lambda(b+\sum_{i=1}^n x_i)}$$

and hence, we can arrive at the coclusion that our posterior in this case also follows the gamma distribution with parameters described as follows-

$$p(x|\lambda).\pi(\lambda) \propto Gamma(\alpha = a+n, \beta = b + \sum_{i=1}^n x_i) = p(\lambda|x)$$

Similarly, we consider the posterior distribution for `smallest time to failure`,

$$p(\lambda|y) = p(y|\lambda).\pi(\lambda)$$

or,

$$p(\lambda|y) = \frac{n\lambda e^{-\lambda ny}.\frac{b^a\lambda^{(a-1)}e^{(-b\lambda)}}{\Gamma(a)}}{\int_0^\infty n\lambda e^{-\lambda ny}.\frac{b^a\lambda^{(a-1)}e^{(-b\lambda)}}{\Gamma(a)}d\lambda} = \frac{\lambda^5 e^{-8.6\lambda}(8.6)^6}{\Gamma(5)}$$

Finally, we consider the `largest time to failure` , the posterior in this case can be described as follows-

$$p(\lambda|z) = p(z|\lambda).\pi(\lambda)$$

or, we can express the psterior in terms of expressions of $p(z|\lambda)$ and $\pi(\lambda)$ to obtain

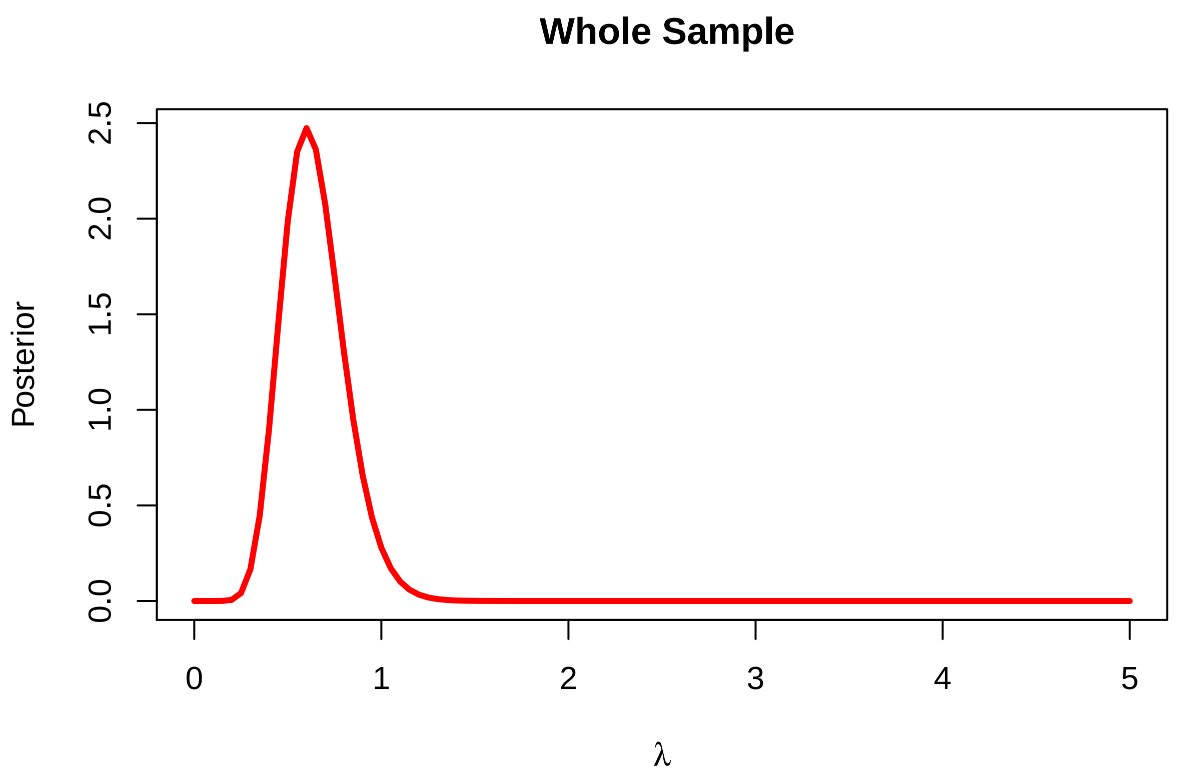
$$p(\lambda|z) = \frac{n\lambda(1-e^{-\lambda z})^{(n-1)}e^{-\lambda z}.\frac{b^a\lambda^{(a-1)}e^{(-b\lambda)}}{\Gamma(a)}}{\int_0^\infty n\lambda(1-e^{-\lambda z})^{(n-1)}e^{-\lambda z}.\frac{b^a\lambda^{(a-1)}e^{(-b\lambda)}}{\Gamma(a)}d\lambda} = \frac{(1-e^{-7.18\lambda})^9\lambda^5 e^{-12.18\lambda}}{\int_0^\infty(1-e^{-7.18\lambda})^9\lambda^5 e^{-12.18\lambda}d\lambda}$$

Upon performing the integral numerically, we get the value as,

```
a=5
b=5
K1 <- integrate( function(lambda) lambda^(a+length(data)-1)*exp(-lambda*(b + sum(data)))*(b^a)/(gamma(a
print(K1)
```

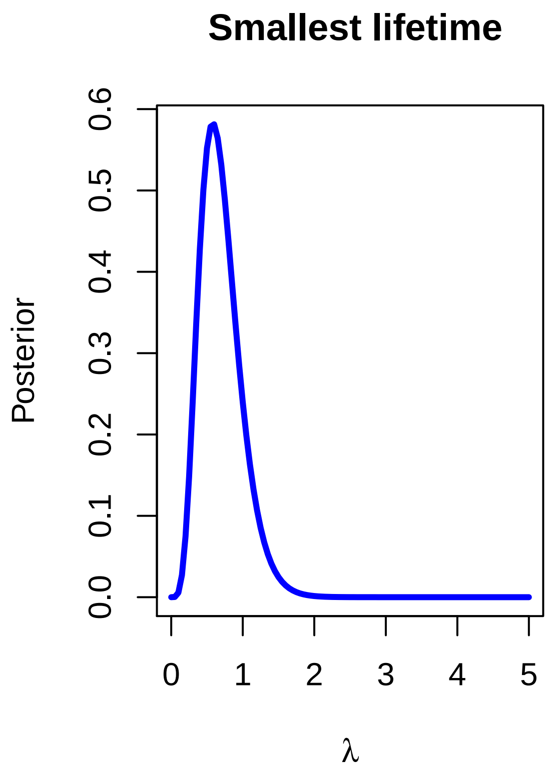
```
## [1] 3.41613e-08
```

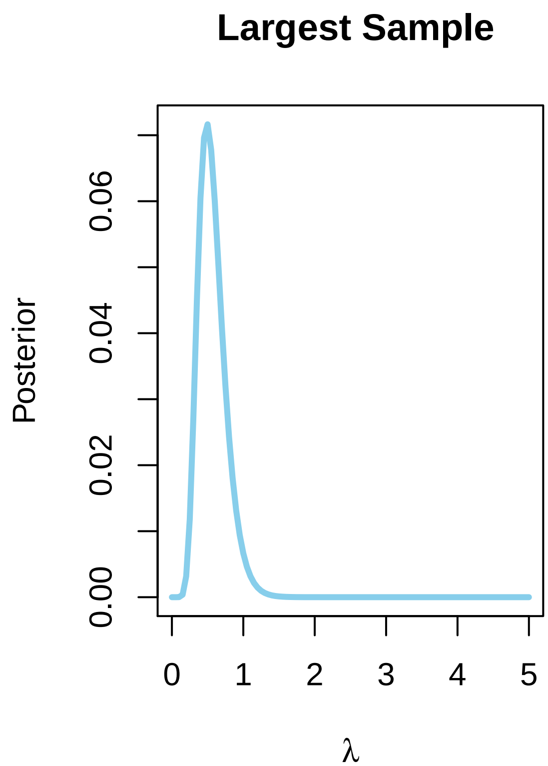
```
a=5
b=5
plot(function(lambda) (lambda^(a+length(data)-1)*exp(-lambda*(b + sum(data))*(b^a)/(gamma(a)))/K1, xlin
     xlab=expression(lambda), ylab = "Posterior",
     main=" Whole Sample", col="red", lwd=3)
```




```
par(mfrow=c(1,2))
y = 0.36
```

```r
plot( function(lambda)
  (lambda^a*exp(-lambda*(b+length(data)*y))*length(data)*(b^a)/gamma(a)),
  xlab=expression(lambda), ylab = "Posterior",
     main=" Smallest lifetime", col="blue", lwd=3, xlim = c(0,5))
z= 7.18
plot( function(lambda)
  (length(data)*(lambda^a)*((1-exp(-lambda*z))^(length(data)-1))*
     (exp(-lambda*(z+b)))*(b^a)/(gamma(a)) ), xlim = c(0,5),
  xlab=expression(lambda), ylab = "Posterior",
     main=" Largest Sample", col="skyblue", lwd=3)
```





(4)

- What are the prior predictive and the posterior predictive distributions, when one observes the whole sample of n = 10 times to failure, with $\sum_{i=1}^{10} x_i = 18.34$?
- What are the prior predictive and the posterior predictive distributions, when only the smallest time to failure, y = 0.36, is observed?
- What are the prior predictive and the posterior predictive distributions, when only the largest time to failure, z = 7.18, is observed?

4 (1) Prior Predictive Distributions

We can describe the prior predictive as follows-

$$p(\hat{x}) = \int_0^\infty p(\hat{x}|\lambda)\pi(\lambda)d\lambda$$

for the case of `whole sample` of 10 data: on substituting $a = 5, b = 5, \sum_{i=1}^{10} x_i = 18.34$, we get,

$$p(x) = \int_0^\infty \frac{b^a \lambda^{(a+n-1)} e^{-\lambda(b+\sum_{x=1}^{10} x_i)}}{\Gamma(a)} = \int_0^\infty \frac{5^5 \lambda^{(n+5-1)} e^{-\lambda(\sum_{i=1}^{10} \hat{x_i}+5)}}{4!} d\lambda$$

Similarly, for the case of `smallest time to failure` :

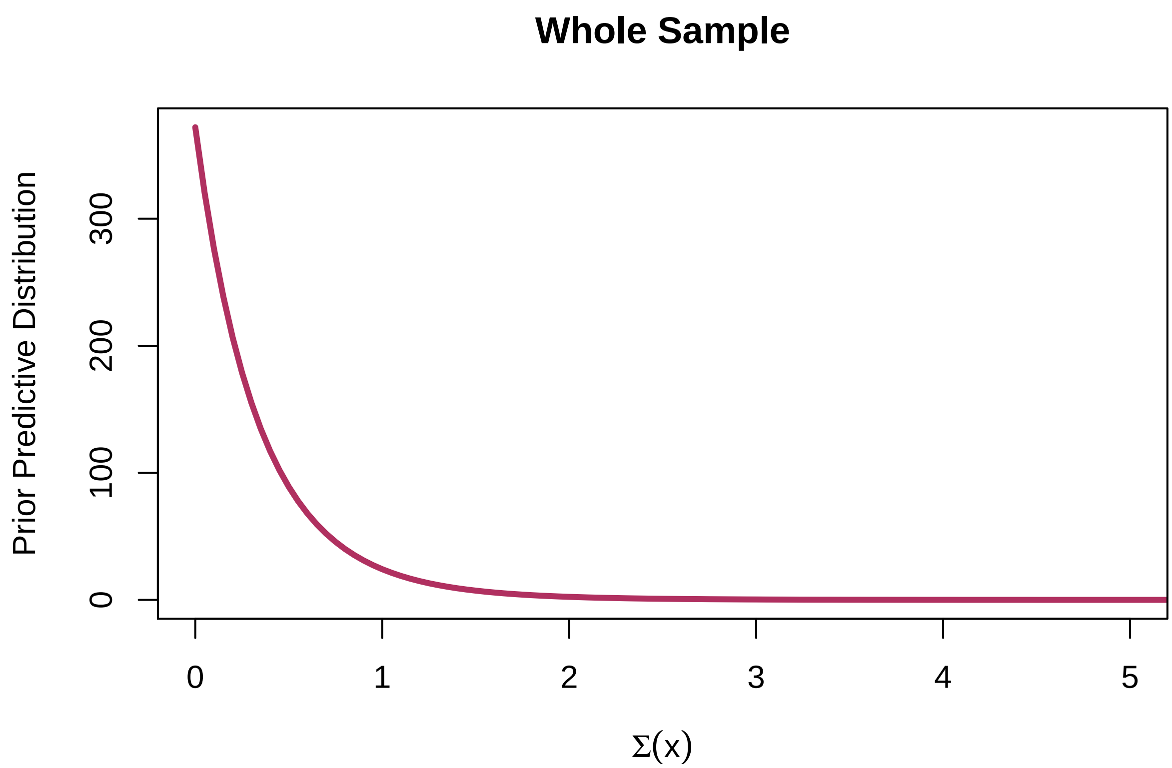
$$p(\hat{y}) = \int_0^\infty n\lambda e^{-\lambda \hat{y} n} \frac{5^5 \lambda^{(4)} e^{(-5\lambda)}}{\Gamma(a)} d\lambda$$

finally, for the case of `largest time to failure`, z=7.18:

$$p(\hat{z}) = \int_0^\infty n\lambda (1 - e^{-\lambda \hat{z}})^{(n-1)} e^{-\lambda \hat{z}} \frac{5^5 \lambda^{(4)} e^{(-5\lambda)}}{\Gamma(a)} d\lambda$$

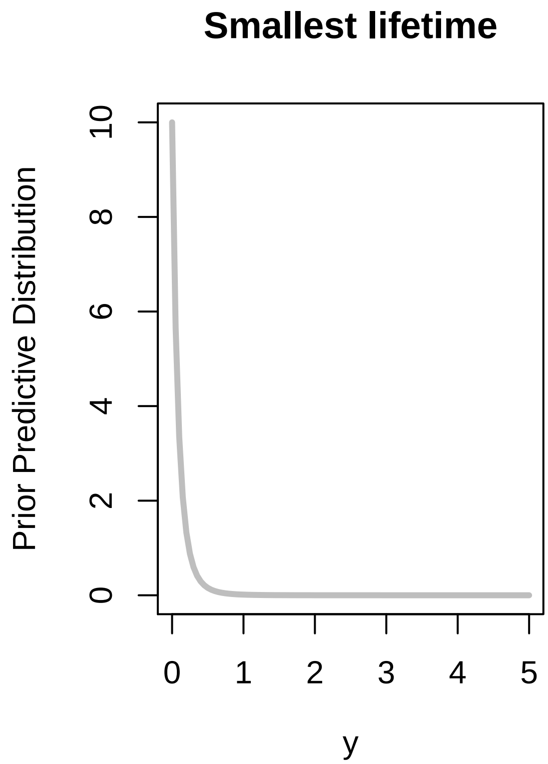
```
prior<-function(lambda)(5^5*lambda^(5-1)*exp(-5*lambda))/factorial(5-1)
likelihood1<-function(lambda,sumx)lambda^10*exp(-lambda*sumx)
likelihood2<-function(lambda,y)10*lambda*exp(-lambda*10*y)
likelihood3<-function(lambda,z)10*lambda*(1-exp(-lambda*z))^(10-1)*exp(-lambda*z)
#initiating sum x, y, z and the prior predictive vectors.
sumx<-seq(0,6.5,by = 0.05)
predictive_prior_sumx<-numeric(length(sumx))
y<-seq(0,5,by = 0.05)
predictive_prior_y<-numeric(length(y))
z<-seq(0,5,by = 0.05)
predictive_prior_z<-numeric(length(z))
#Making the prior predictive vectors
for (i in 1:length(sumx)){
temp<-function(lambda)likelihood1(lambda,sumx[i])*prior(lambda)
predictive_prior_sumx[i]<-integrate(temp,0,Inf)$value
}
for (i in 1:length(y)){
temp<-function(lambda)likelihood2(lambda,y[i])*prior(lambda)
predictive_prior_y[i]<-integrate(temp,0,Inf)$value
}
for (i in 1:length(z)){
temp<-function(lambda)likelihood3(lambda,z[i])*prior(lambda)
predictive_prior_z[i]<-integrate(temp,0,Inf)$value
}
```

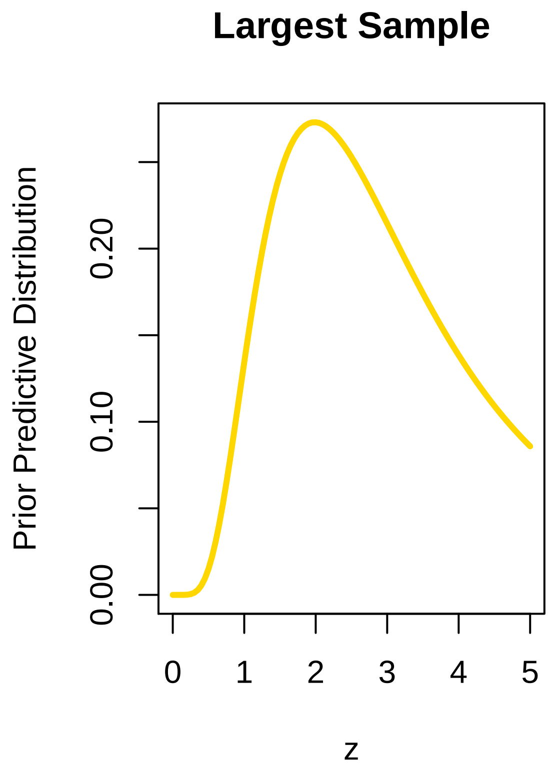
```
plot(sumx,predictive_prior_sumx,type="l", xlab = expression(Sigma (x)),
     ylab = "Prior Predictive Distribution",
     main = "Whole Sample",
     col="maroon", lwd=3, xlim = c(0,5))
```




```
par(mfrow=c(1,2))
plot(y,predictive_prior_y,type="l", xlab = expression(y),
     ylab = "Prior Predictive Distribution",
     main = "Smallest lifetime",
     col="gray", lwd=3, xlim = c(0,5))

plot(z,predictive_prior_z,type="l", xlab = expression(z),
     ylab = "Prior Predictive Distribution",
     main = "Largest Sample",
     col="gold", lwd=3, xlim = c(0,5))
```





4 (2) Posterior Predictive Distributions

The posterior predictive distribution is defined in general as follows-

$$p(\hat{x}|x) = \int_0^{\infty} p(\hat{x}|\lambda)p(\lambda|x)d\lambda$$

So, in case of our `whole data sample`, posterior predictive are defined by-

$$p(\hat{x}|x) = \int_0^{\infty} \lambda^n e^{-\lambda(\sum_{i=1}^{n} \hat{x_i})} \lambda^{14} e^{-23.34\lambda} \frac{(23.34)^{15}}{\Gamma(15)} d\lambda = \frac{(23.34)^{15}\Gamma(14+n+1)}{(23.34 + \sum_{i=1}^{n} \hat{x_i})^{14+n+1}\Gamma(15)}$$

similarly, in case of `smallest lifetime observed`,

$$p(\hat{y}|y) = \int_0^{\infty} n\lambda e^{-\lambda\hat{y}n} \frac{\lambda^5 e^{-8.6\lambda}(8.6)^6}{\Gamma(6)} d\lambda$$

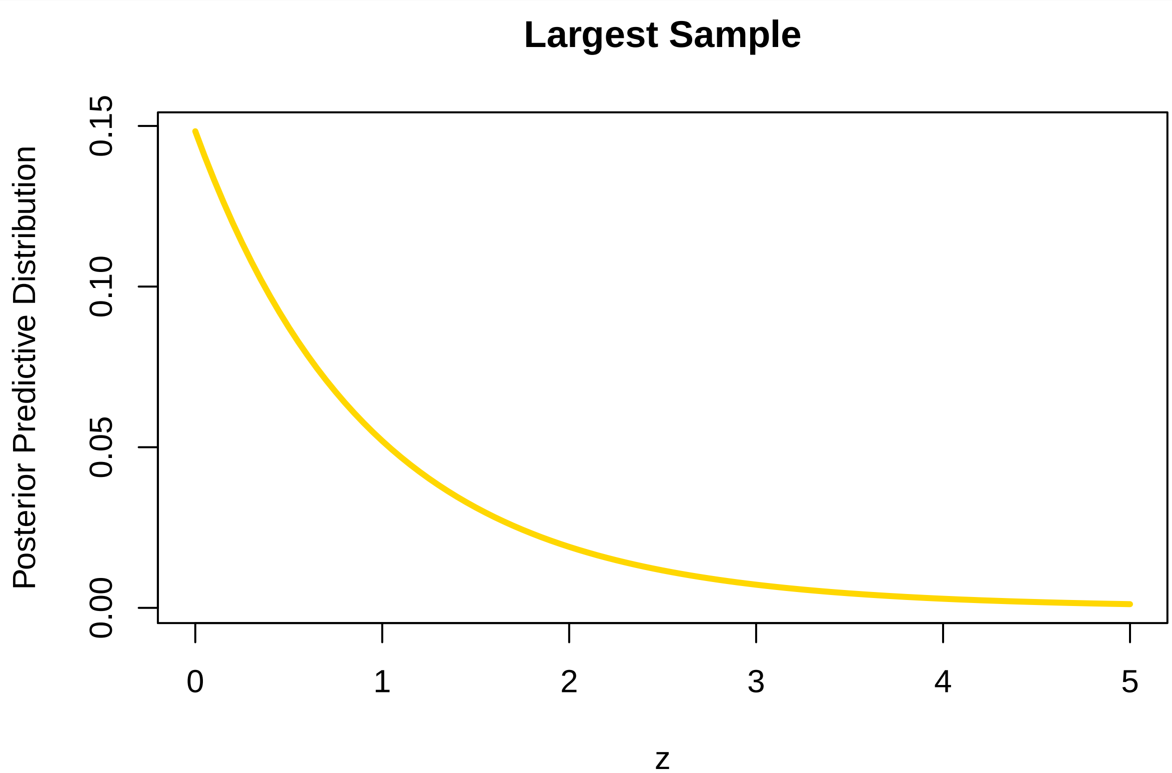
finally, in case of `largest time to failure`, we have,

$$p(\hat{z}|z) = \int_{\lambda\in\Omega} n\lambda(1 - e^{-\lambda\hat{z}})^{(n-1)} e^{-\lambda\hat{z}} \frac{(1 - e^{-7.18\lambda})^9 \lambda^5 e^{-12.18\lambda}}{\int_0^{\infty}(1 - e^{-7.18\lambda})^9 \lambda^5 e^{-12.18\lambda} d\lambda} d\lambda$$

```
sumx<-seq(0,5,by=0.05)
y<-seq(0,5,by=0.05)
z<-seq(0,5,by=0.05)
posterior_predictive_sumx<-numeric(length(sumx))
posterior_predictive_y<-numeric(length(y))
posterior_predictive_z<-numeric(length(z))
for (i in 1:length(sumx)){
posterior_predictive_sumx[i]<-
```

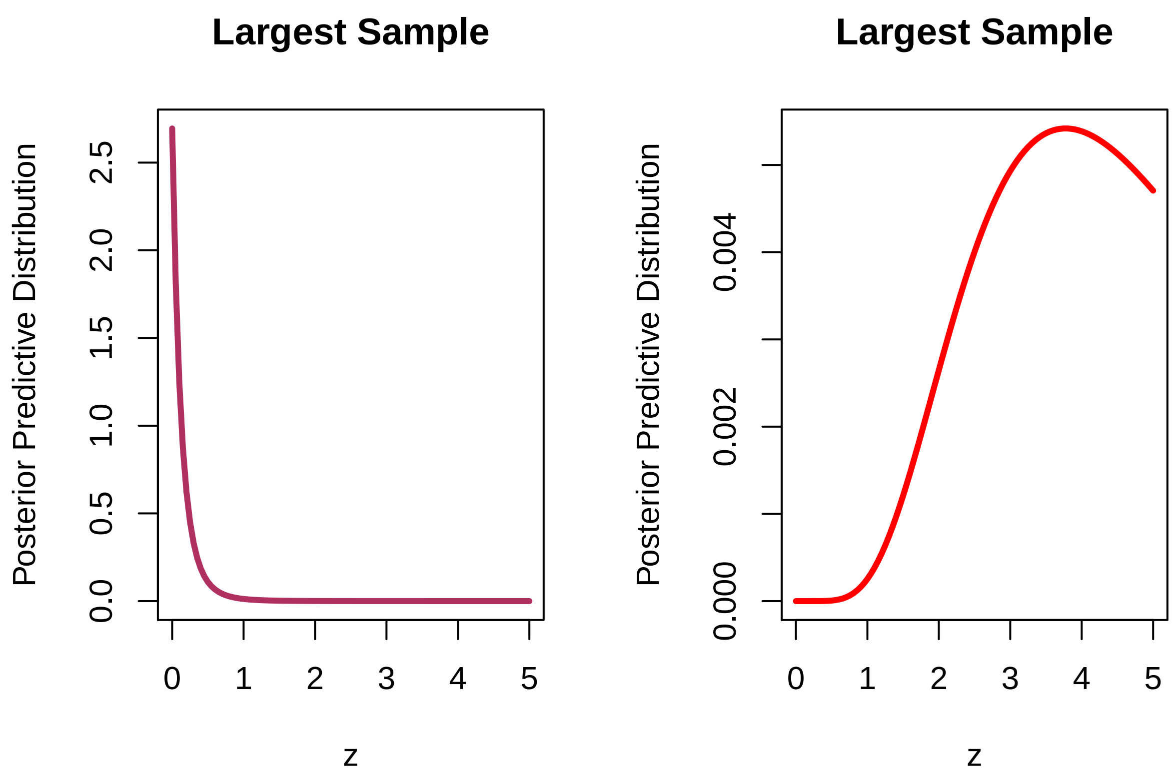
```r
  integrate(function(lambda)lambda^10*exp(-lambda*sumx[i])*
              (5^5*lambda^(10+5-1)*exp(-lambda*(18.34+5)))
            /factorial(4)/K1,0,Inf)$value
}

# division by K2 pending
for (i in 1:length(y)){
posterior_predictive_y[i]<-
  integrate(function(lambda)10*lambda*
              exp(-lambda*10*y[i])*(10*5^5*lambda^5*
                                        exp(-lambda*(10*0.36+5)))/factorial(4),0,Inf)$value
}
# division by K3 pending
for (i in 1:length(z)){
posterior_predictive_z[i]<-
  integrate(function(lambda)10*lambda*(1-exp(-lambda*z[i]))^(10-1)*
              exp(-lambda*z[i])*(10*5^5*lambda^5*exp(-lambda*(5+7.18))*
                                     (1-exp(-lambda*7.18))^(10-1))/factorial(4),0,Inf)$value
}
plot(sumx,posterior_predictive_sumx, type ="l",
      xlab = expression(z), ylab = "Posterior Predictive Distribution",
     main = "Largest Sample",
     col="gold", lwd=3, xlim = c(0,5))
```




```r
par(mfrow=c(1,2))
plot(y, posterior_predictive_y, type = 'l',
      xlab = expression(z),
     ylab = "Posterior Predictive Distribution",
     main = "Largest Sample",
```

```r
      col="maroon", lwd=3, xlim = c(0,5))
plot(z, posterior_predictive_z, type = 'l',
       xlab = expression(z),
      ylab = "Posterior Predictive Distribution",
      main = "Largest Sample",
      col="red", lwd=3, xlim = c(0,5))
```



5. Point Estimate

How would you compute a point estimate for λ based on the whole sample of n = 10 times to failure, with $\sum_{i=1}^{10} xi = 18.34$?

We know that our posterior distribution follows a $Gamma(a + n = 15, b + \sum_{i=1}^{10} xi = 23.34)$. So, we can use the expected value of our posterior distribution as our point estimate. Now, for the `gamma` distribution, we can define it's expected value as

$$E(\lambda|x) = \frac{a+n}{b+\sum_{i=1}^{10} xi} = \frac{15}{23.34} = 0.6426735$$

6. Credibility Interval

How would you compute a 90% posterior credibility for λ based on the whole sample of n = 10 times to failure, with $\sum_{i=1}^{10} = 18.34$?

We know that the 90% credibility intervals can be computed using the idea of integrating the posterior between the limits of 0 and fifth percentile and between 95th percentile and ∞. This can be described as follows-

$$p(\lambda|x) = \int_0^a \lambda^{14} e^{-23.34\lambda} \frac{(23.34)^{15}}{\Gamma(15)} d\lambda$$

$$p(\lambda|x) = \int_b^\infty \lambda^{14} e^{-23.34\lambda} \frac{(23.34)^{15}}{\Gamma(15)} d\lambda$$

These two integrals can be resolved by setting the values of first integral equal to 0.05 and that of second equal to 0.95 in order to obtain the values of a and b so that the 90% credible interval will be [a,b]

7. Hypothesis Testing

How would you choose between $H_1 : \lambda < 1, H_2 : 1 \leq \lambda < 2$ and $H_3 : \lambda \geq 2$, based on the whole sample of n = 10 times to failure, with $\sum_{i=1}^{10} = 18.34$? (How would you answer this question if you were not willing to be Bayesian?)

Consider the three following hypothesis-

- H_1: $\lambda < 1$
- H_2: $1 \leq \lambda < 2$
- H_3: $\lambda \geq 2$

We are going to perform the `Maximum Aposteriori Test (MAP)`. Here, we will be choosing the hypothesis with the highest probability, it is relatively easy to show that the error probability is minimized.

Now, corresponding to the first hypothesis, we have to compute the probability $P(x|H_1)$ which can be described as-

$$p(H_1|x) = \int_0^1 \lambda^{14} e^{-23.34\lambda} \frac{(23.34)^{15}}{\Gamma(15)} d\lambda$$

Similarly, we can describe and compute the second hypothesis as-

$$p(H_2|x) = \int_1^2 \lambda^{14} e^{-23.34\lambda} \frac{(23.34)^{15}}{\Gamma(15)} d\lambda$$

and finally, we can compute the third hypothesis as-

$$p(H_3|x) = \int_2^\infty \lambda^{14} e^{-23.34\lambda} \frac{(23.34)^{15}}{\Gamma(15)} d\lambda$$

```
H1 <- integrate( function(lambda)
  lambda^(14)*exp(-lambda*(23.34))*(23.34^15)/(gamma(15)), lower = 0, upper = 0.99999999)$value

H2 <- integrate( function(lambda)
  lambda^(14)*exp(-lambda*(23.34))*(23.34^15)/(gamma(15)), lower = 1, upper = 1.99999999)$value

H3 <- integrate( function(lambda)
  lambda^(14)*exp(-lambda*(23.34))*(23.34^15)/(gamma(15)), lower = 2, upper = Inf)$value


print(paste("H1:", H1))
```

```
## [1] "H1: 0.973267846181667"
```

```
print(paste("H2:", H2))
```

```
## [1] "H2: 0.0267321308882538"
```

```
print(paste("H3:", H3))
```

```
## [1] "H3: 2.01454807906362e-08"
```

Therefore, based on the MAP Test, we can assert that the hypothesis H_1 is most probable and we would like to choose the H_1

If I was a non-Bayesian, I would do differents hypothesis tests taking the hypothesis H1, H2 and H3 taken two at a time.

8. Invariance Principal

How would you compute a point estimate and a 90% posterior credibility for the expected time to failure of such a light bulb, $\theta = 1/\lambda$? (How would you compute a 90% confidence interval for $\theta = 1/\lambda$?)

Since we know from the invariance property that credible intervals are invariant to reparameterization of the equal tailed interval, so, we can conclude that the credible intervals will remain the same when we do reparamterization.

9.

Firstly, This question can not be answered if we adopt the non-bayesian paradigm becuase it involves computation of probability which is straight forward using the bayesian approach.

Secondly, we can use the posterior of whole sample as the prior for computing the probability that the smallest time to failure of a future sample of n = 10 times to failure was smaller than 0.20 years. In this case, the probability of time to failure of a future sample is the posterior probability.

10.

I anticipate that the distribution of the whole sample serves our purpose better. It is so because it has more information and therefore reduces the variance in our estimates and predictions.