# Assignment 2

**1a)** Calculate the probability that the turbine delivers power at any given time.

Pr(P(v) > 0) = Pr(4<v<25).

In R code:

```
prob <-
pgamma(25, shape = 0.8, scale = 6)- pgamma(4, shape = 0.8, scale = 6)
```

Gives me the output **0.4019568**.

**1b)** Compute by simulation the expected value.

I choose the sample size as 10000.

In R code (see the complete R code version on 1b)

```
# Compute approximate expected value
N <- 10000
sample <- rgamma(N, shape = 0.8, scale = 6)
P_sample <- P(sample)
approx <- 1/N * sum(P_sample)
# Compute approximate 95% CI
s <- sqrt(1 / (N-1)  * sum((P_sample - approx)^2)
ci_lower <- approx - 1.96 * s / sqrt(N)
ci_upper <- approx + 1.96 * s / sqrt(N)
print(c(ci_lower,ci_upper))
#[1] 0.2883652 0.3114523
```

Gives me the output **[0.2883652,0.3114523].**

**1c)** Compute by numerical integration

R code:_

integrate(integrand, 4, 25) (see the complete R code version on 1c)

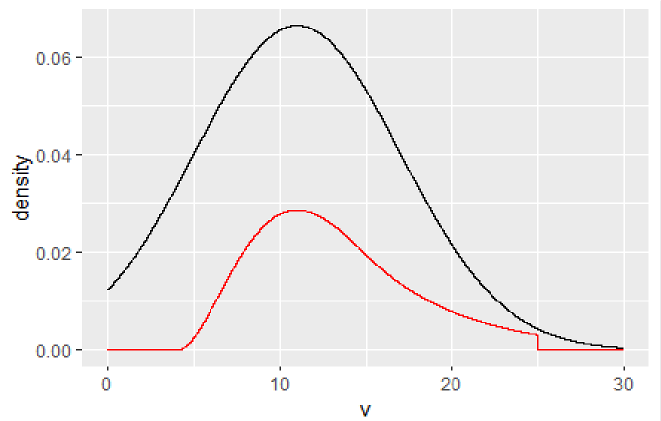
gives me the output **0.3057371 with absolute error < 3.3e-05**

R code:

```
x <- seq(4, 25, length.out=10001)
y <- P(x)*dgamma(x, 0.8, 1/6)*(x[2]-x[1])
print(sum(y))
```

gives me the output **0.3057338**

**1d)** Use importance sampling.



Product of densities of P(v) and v and choosen proposal density.

I choose a normal distribution with expectation=11 and standard deviation=6.
The black line is chosen proposal density and the red line is the product of densities of P(v) and v.

R code:

```
v <- seq(0, 30, length.out = 10000)
d1 <- P(v)*dgamma(v, 0.8, 1/6)
d2 <- dnorm(v, mean=11, sd=6)
df <- data.frame(x,d1,d2)
ggplot(df, aes(v)) +                        # basic graphical object
  geom_line(aes(y=d1), colour="red") +    # first layer
  geom_line(aes(y=d2), colour="black")+    # second layer
  labs(
       y="density",
       x="v",
       caption = "Product of densities of P(v) and v and choosen proposal density.")
```

**1e)** Repeat b) and compare with chosen proposal density.

See the complete R code from part 1e)

By R code, I get a mean **0. 3045518** with an interval **[0.3015001 0.3076036].**

**1f)** Use importance sampling to estimate the variance of the power production.

Var(P(v)) = E(P(v) ^2 ) – E(P(v))^2 .

See the complete R code from part 1f)

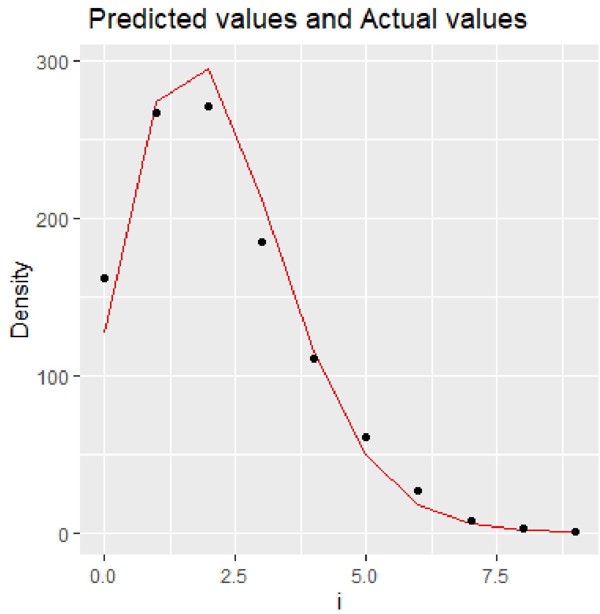
I select rnorm(N, 15, 7) as my new integral. Then I get the output 0.3591714. Compare the one from part b 0.3527757, the two values are very similar.

No matter how we choose the integral to compute, we will always get similar output.

**2a)** Compute the maximum likelihood estimate for λ.

From the R code, lambda is **[2.156934].**

Below is the plot of the predicted counts using this λ together with the actual counts.




**2b)** Write down a function proportional to) the likelihood for $Y_0, \ldots\ldots, Y_9$; $Z_0; \ldots\ldots, Z_9$ given the parameters $\lambda 1$, $\lambda 2$, p.

2b) The probability of day $i$ with $z_i$

$$P_1 = p\,\frac{\lambda_1^{i} e^{-\lambda_1}}{i!} \quad \& \quad \text{its notation } N_1 = \sum_{i=0}^{9} z_i$$

The probability of day $i$ with $Y_i - z_i$

$$P_2 = (1-p)\,\frac{\lambda_2^{i} e^{-\lambda_2}}{i!} \quad \& \quad \text{its notation } N_2 = \sum_{i=0}^{9} (Y_i - z_i)$$

The likelihood is

$$\pi(Y, z \mid p, \lambda_1, \lambda_2) = \frac{N!}{N_1!\, N_2!} \prod_{i=0}^{9} P_1^{z_i} P_2^{Y_i - z_i}$$

**2c)** Write down the posterior[1] for p given all the other observations and variables.

2c) posterior probability can be writen as

posterior probability $\propto$ Likelihood $\times$ ~~poster~~ prior probability

( get from 2b)

$\Rightarrow$ posterior pobability:

$$\pi(Y, z \mid p, \lambda_1, \lambda_2)\, \pi(p)\, \pi(\lambda_1)\, \pi(\lambda_2)$$

$$\propto_p p^{N_1} (1-p)^{N-N_1}$$

$$\Rightarrow p \mid Y, z, \lambda_1, \lambda_2 \sim Beta(N_1+1, N-N_1+1)$$

2d)

2d) also get the Likelihood from 2b)

posterior probablity:

$$\pi(Y, z \mid p, \lambda_1, \lambda_2)\, \pi(p)\, \pi(\lambda_1)\, \pi(\lambda_2) \propto_{\lambda_1} \exp(-\lambda_1 (N_1+1) + \log(\lambda_1) M_1)$$

$$\pi(Y, z \mid p, \lambda_1, \lambda_2)\, \pi(p)\, \pi(\lambda_1)\, \pi(\lambda_2) \propto_{\lambda_2} \exp(-\lambda_2 (\overset{N-N_1}{\cancel{N-M_1}}) + \log(\lambda_2) M-M_1)$$

$$\Rightarrow \lambda_1 \mid Y, z, p, \lambda_2 \sim Gamma(M_1+1, N_1+1)$$

$$\lambda_2 \mid Y, z, p, \lambda_1 \sim Gamma(M-M_1+1, N-N_1+1)$$

2e)

2e)

$$\pi(z_i \mid Y, \overset{\text{all } z_i}{\cancel{\boxed{z}}}, p, \lambda_1, \lambda_2) \propto_{z_i} \frac{p^{z_i}(1-p)^{Y_i - z_i}}{z_i!\,(Y_i - z_i)!} \lambda_1^{i z_i} \lambda_2^{i(Y_i - z_i)} \exp(-z_i \lambda_1 - (Y_i - z_i)\lambda_2)$$

This proves that the distribution of $z_i$ given all ~~pa~~ parameters.

which is similar to <u>Binomial distribution</u>

$$\frac{n!}{(n-x)!\,x!} \cdot p^x \cdot q^{n-x}$$

I can rewrite the $\pi(z_i \mid Y, \text{all } z_i, p, \lambda_1, \lambda_2) \propto z_i$

as

$$\lambda_1^{i z_i} \lambda_2^{i(Y_i - z_i)} e^{-z_i\lambda_1 - (Y_i - z_i)\lambda_2}$$

$$= (\lambda_1^i e^{\lambda_1})^{z_i} \cdot (\lambda_2^i \cdot e^{-\lambda_2})^{Y_i - z_i}$$

$$\frac{p^{z_i}(1-p)^{Y_i - z_i}}{(z_i)!\,(Y_i - z_i)!} \cdot (\lambda_1^i e^{\lambda_1})^{z_i} \cdot (\lambda_2^i \cdot e^{-\lambda_2})^{Y_i - z_i}$$

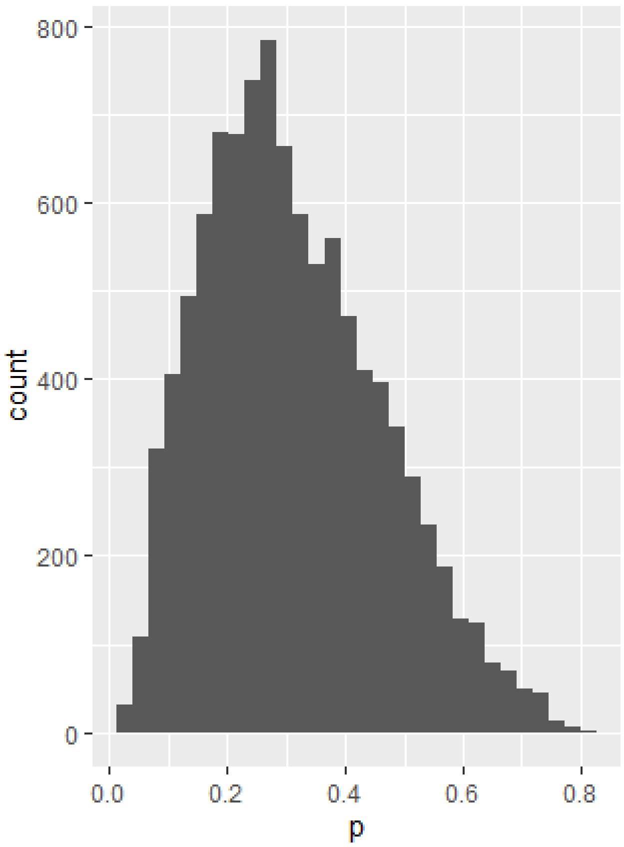
$$= \frac{1}{(z_i)!\,(Y_i - z_i)!} \cdot (p\lambda_1^i e^{\lambda_1})^{z_i} \cdot ((1-p)(\lambda_2^i \cdot e^{-\lambda_2}))^{Y_i - z_i}$$

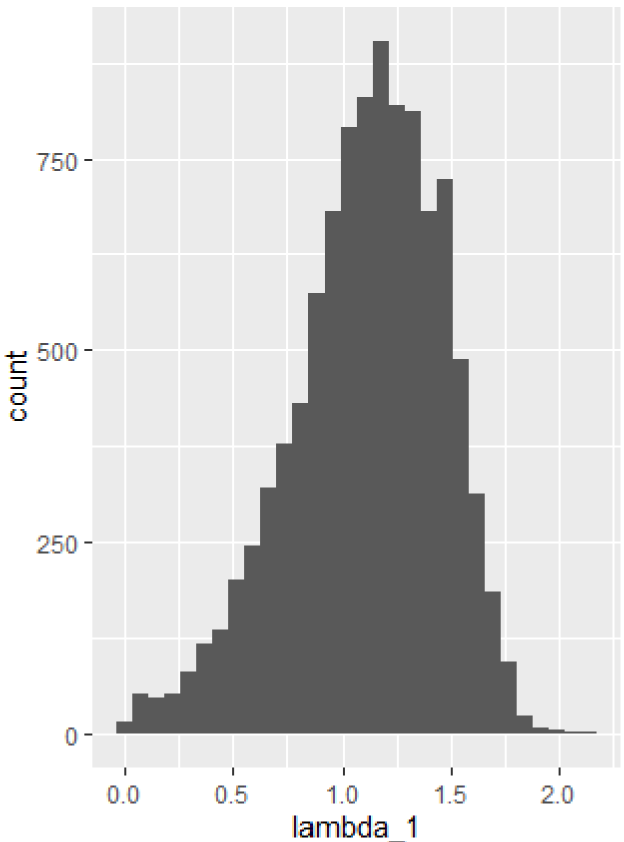
I can add some constant to make them equal to each other.

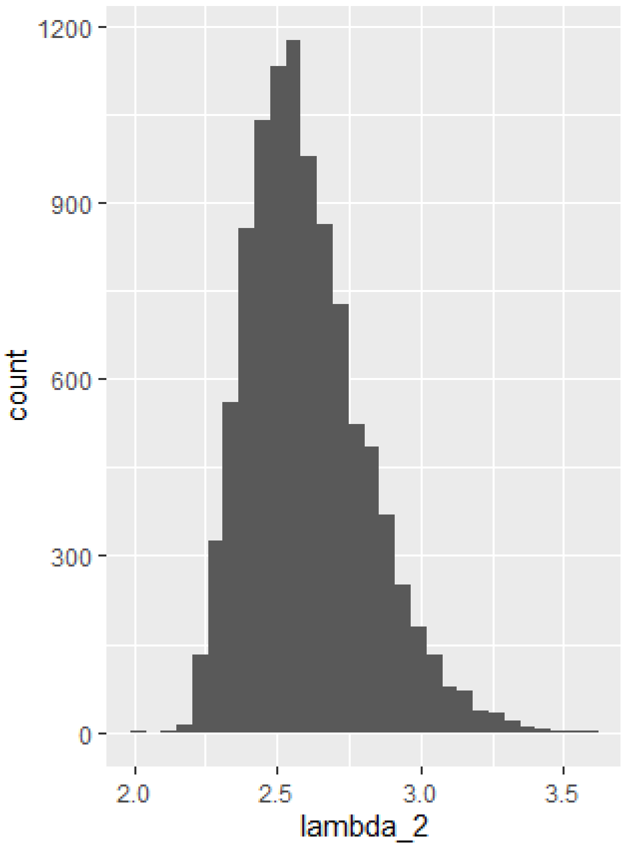
$$\Rightarrow \pi(z_i \mid Y, \text{all } z_i, p, \lambda_1, \lambda_2) \propto z_i \; C_1 \frac{Y_i}{(Y_i - z_i)!\,z_i!} (p\lambda_1^i e^{\lambda_1})^{z_i} ((1-p)(\lambda_2^i \cdot e^{-\lambda_2}))^{Y_i - z_i}$$

## 2f)

Below are the plots of each parameters.


800
600
400
200
0
count
0.0
0.2
0.4
0.6
0.8
p


750
500
250
0
count
0.0
0.5
1.0
1.5
2.0
lambda_1


1200
900
600
300
0
count
2.0
2.5
3.0
3.5
lambda_2