Applied Bayesian Analysis : NCSU ST 540

Homework 4

Bruce Campbell

---

1) Poisson -Gamma posterior

Assume the likelihood $Y|\lambda \sim Poisson(\lambda)$ and prior $p(\lambda) = 1 \forall \lambda > 0$. Show that this is an improper prior, i.e., that this density function is not a valid density function. Compute the posterior, and argue that the posterior distribution is proper for any value of Y .

Let Ω_Λ be the domain of our prior, then if we try to normalize our prior by computing a constant C such that $C \int\limits_{\Omega_\Lambda} P(\lambda) d\lambda = 1$ we see that there is no solution since $\Omega_\Lambda = [0, \infty)$ causes the integral to diverge. This implies that our prior is improper - i.e. it is not a proper probability distribution function.

Now let's place our likelihood in Bayes theorem to calculate our posterior.

$$P(\lambda|Y) = \frac{P(Y|\lambda)P(\lambda)}{P(Y)} \propto P(Y|\lambda)P(\lambda) = \frac{\lambda^y}{y!} e^{-\lambda}$$

Now it's important to recognize that the last expression is a function of λ. The right hand side is the kernel of a $Gamma(y + 1, 1)$ distribution so our posterior must be $Gamma(y + 1, 1)$. Since the domain of the parameters for a $Gamma(a, b)$ distribution is $[0, \infty) \times [0, \infty)$ any $Y \in \Omega_Y = [0, 1, 2, ...]$ will yield a proper posterior distribution. Recall the domain of a Poisson random variable is the non negative integers.

2) Exponential - Gamma posterior

Say that $Y \sim Exp(\lambda)$ so that the likelihood is $p(y|\lambda) = \lambda exp(-\lambda y)$ Assuming the prior is $\lambda \sim Gamma(a, b)$,find the posterior of λ.

$$P(\lambda|Y) = \frac{P(Y|\lambda)P(\lambda)}{P(Y)} \propto P(Y|\lambda)P(\lambda) = \lambda e^{-\lambda y} \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$$

Gathering terms and dropping constants not related to λ we have that

$$P(\lambda|Y) = \lambda^a e^{-(y+b)\lambda}$$

which we recognize to be a $Gamma(a + 1, y + b)$ distributed random variable.

3) Hurricane Bayesian analysis.

The file hurricanes.csv on the course webpage has the year and Saffr-Simpson intensity category of all Atlantic hurricanes that made landfall between 1990 and 2016. The counts are downloaded from http://www.aoml.noaa.gov/hrd/hurdat/. Break the years into two intervals: 1990-2002 and 2003-2016. Compute the posterior probability that the average number of category k storms per year has increased from 1990-2002 to 2003-2016. Describe and justify your method and carry it out separately for each $k = 1, ..., 5$, and for all storms combined.

For this analysis we'll be using a Poisson likelihood and a conjugate Gamma prior. We were going to consider two priors, an uninformative one a and one informed be the overall mean. There are two standard options for the uninformative prior - the uniform prior and the Jefferies prior. The uniform prior gives the likelihood as the posterior. The Jeffries prior for the Gamma is actually complicated. The ploygamma function is involved and the information is 2 dimensional. We'll revisit the use of the Jeffries prior if there's time.

We divide the data sets into H_0 and H_a and calculate $P(\lambda_0|Y_0)$ and $P(\lambda_a|Y_a)$ - the posteriors for our two sets. Then we can use these to do a simulation to calculate $P(\lambda_a > \lambda_0)$.

As before we have

$$P(\lambda|Y) = \frac{P(Y|\lambda)P(\lambda)}{P(Y)} \propto P(Y|\lambda)P(\lambda) = \frac{\lambda^y}{y!}e^{-\lambda}\frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}$$

but now we have an iid sample constituting the category k storms in each set H_0 H_a so

$$P(\lambda_0|Y) \propto \lambda_0^{(\sum y_i)}e^{-n\ \lambda_0}\lambda_0^{a-1}e^{-b\lambda_0}$$

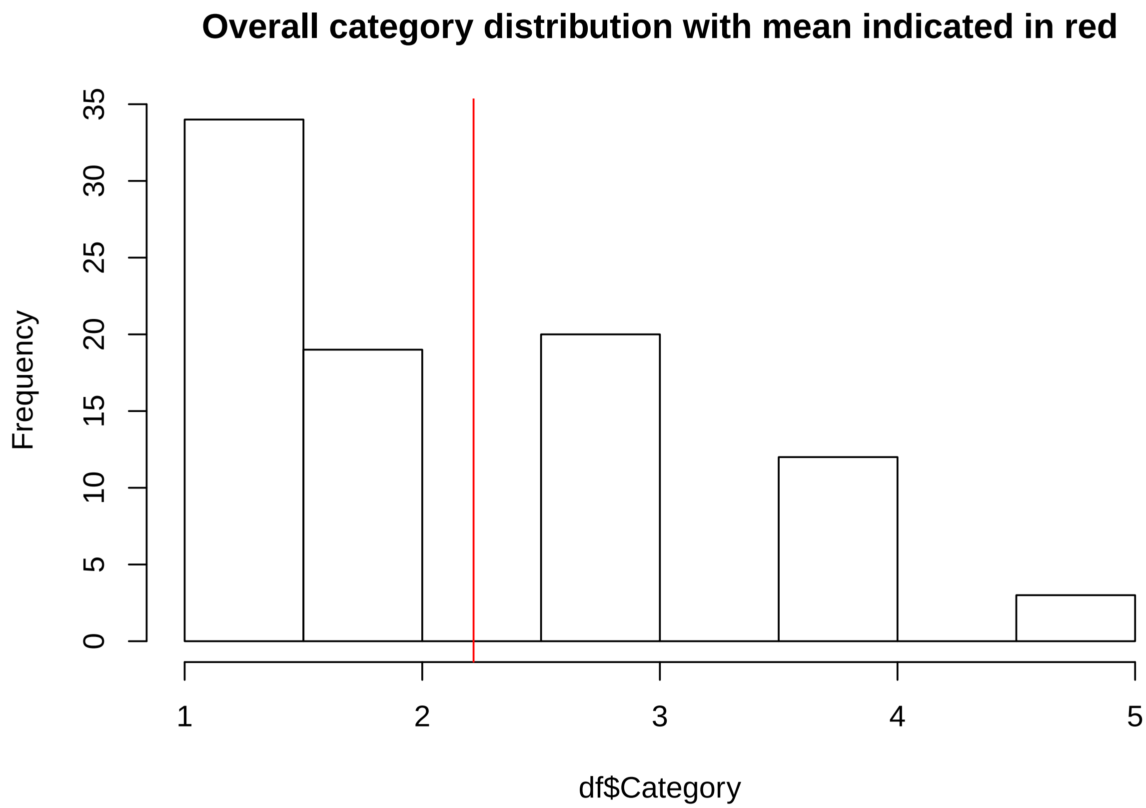
$$P(\lambda_a|Y) \propto \lambda_a^{(\sum y_i)}e^{-n\ \lambda_a}\lambda_a^{a-1}e^{-b\lambda_a}$$

The posteriors are then $Gamma(\sum y_i + a, b + n)(\lambda)$ where the sums and count n come the category k elements in H_0 and H_a

In the code below we set a,b so the mean of the prior is at the overall mean of the categories, and we'll set the variance to the overall sample variance.

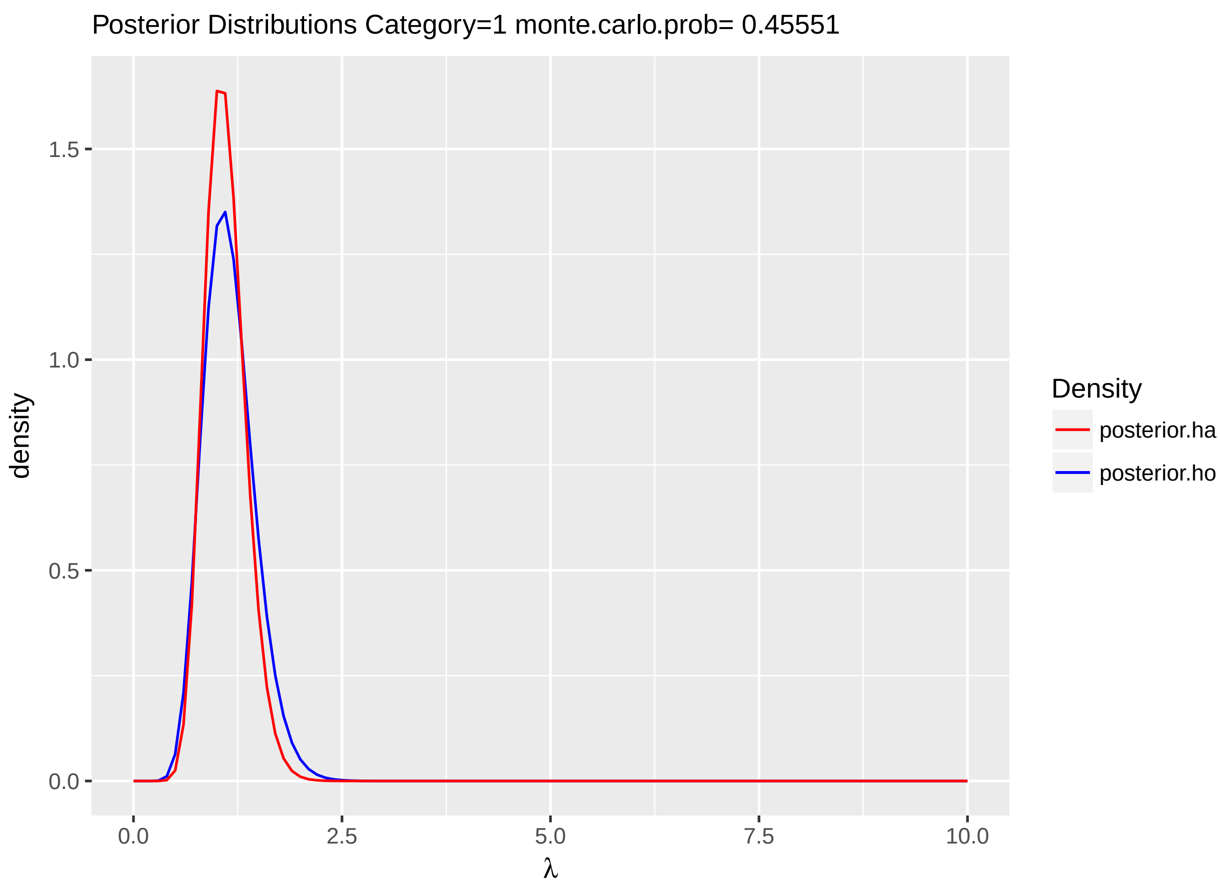
$\mu = \frac{a}{b}$ $\sigma^2 = \frac{a}{b^2} = \frac{\mu}{b}$

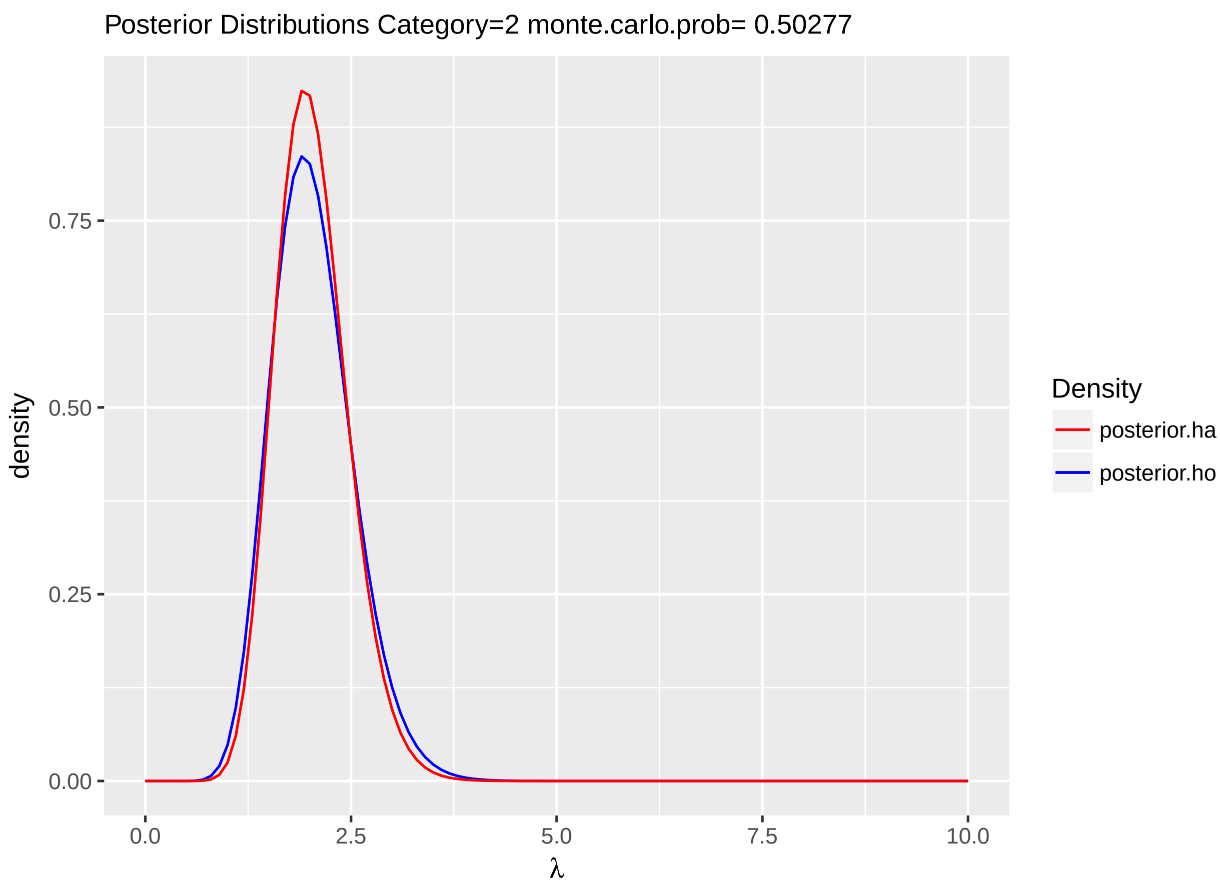
```r
rm(list = ls())
setwd("C:/E/brucebcampbell-git/bayesian-learning-with-R")
df <- read.csv("hurricanes.csv", header = TRUE)

sample.mean <- mean(df$Category)
sample.var <- var(df$Category)

prior.b <- sample.mean/sample.var
prior.a <- sample.mean * prior.b

hist(df$Category, main = "Overall category distribution with mean indicated in red")
abline(v = sample.mean, col = "red")
```



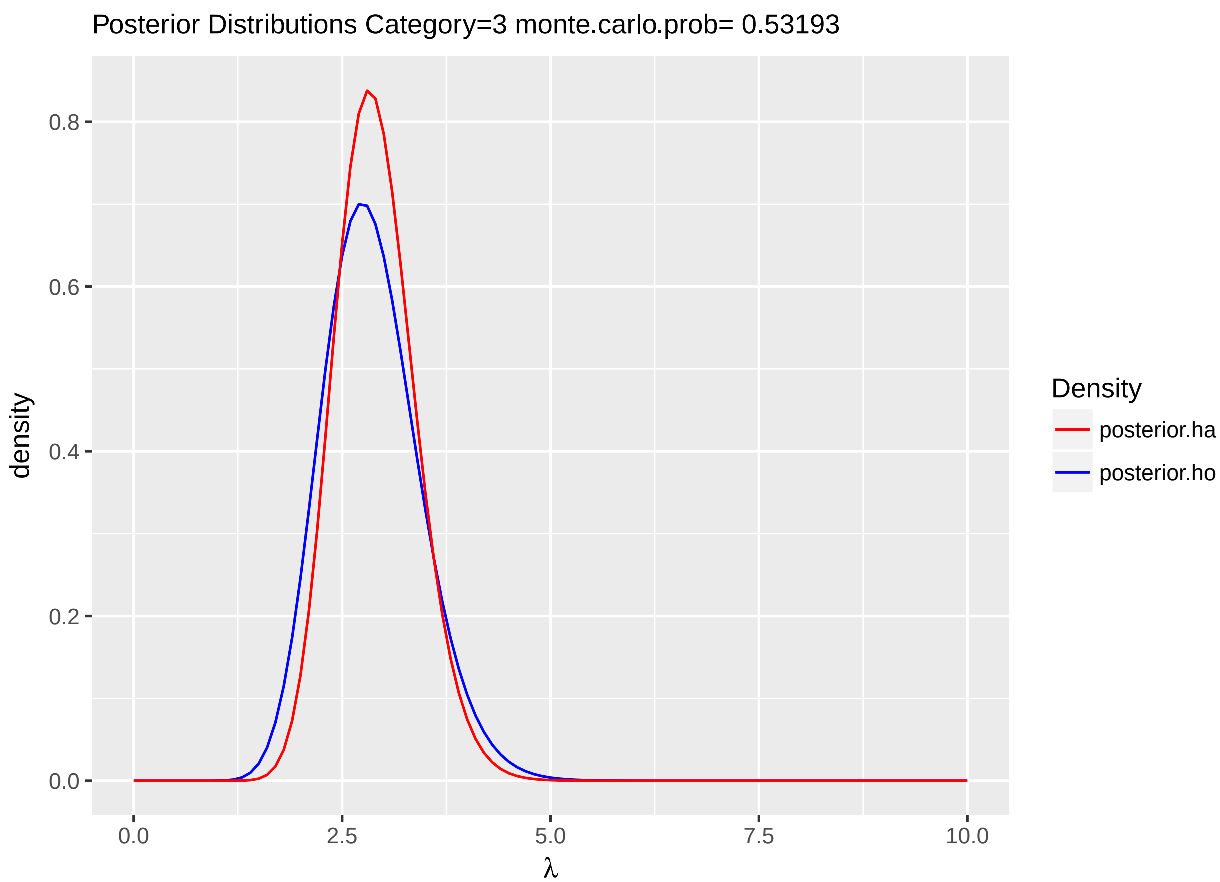

```r
# Split the data
Ho <- df[df$Year > 1990 & df$Year < 2002,
    ]
Ha <- df[df$Year > 2003 & df$Year < 2016,
    ]

monte.carlo.probs <- list()

for (k in 1:5) {
    # Create data frames for category k
    Hok <- Ho[Ho$Category == k, ]
    Hak <- Ha[Ha$Category == k, ]

    # Calculate components of posterior
    # parameters
    sum.ho <- sum(Hok$Category)
    sum.ha <- sum(Hak$Category)

    n.ho <- nrow(Hok)
    n.ha <- nrow(Hak)

    if (n.ho < 10 || n.ha < 10)
```

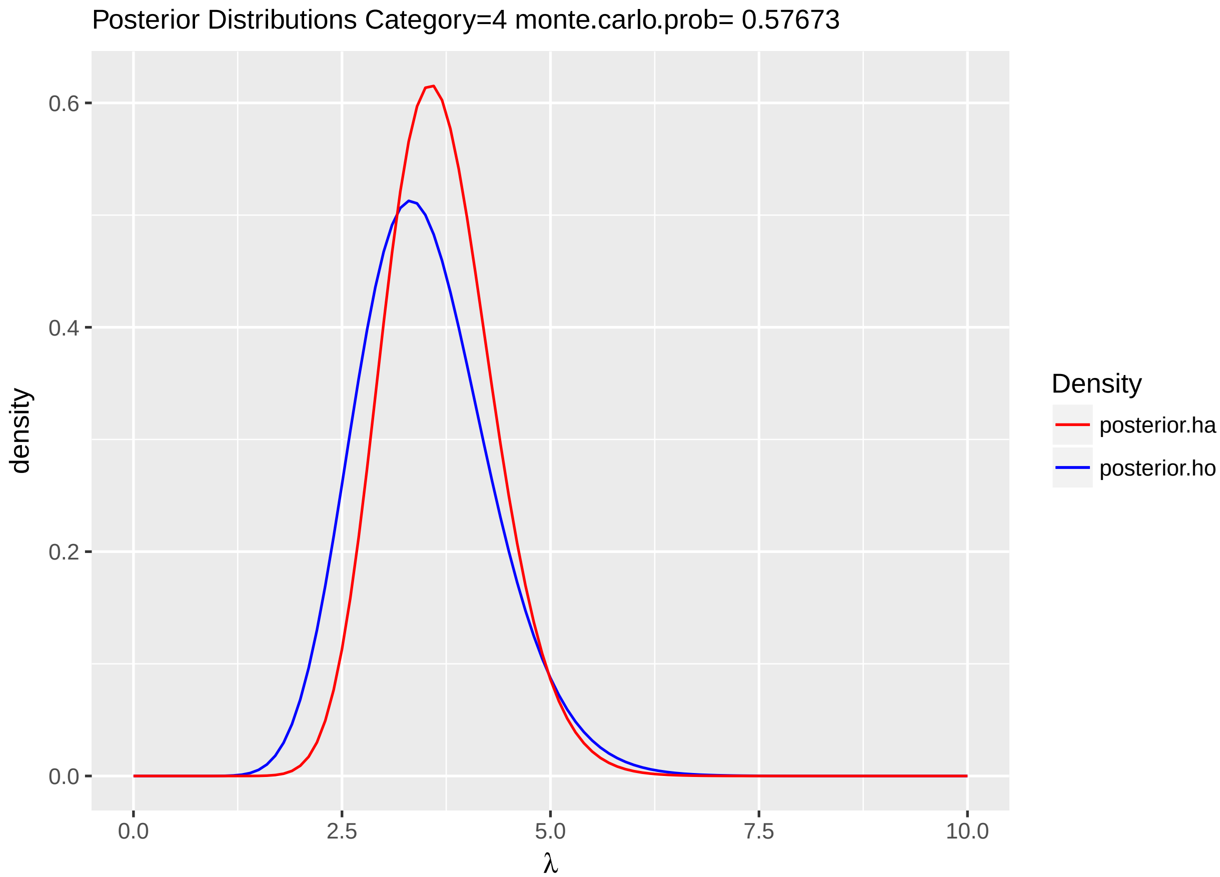
```r
        warning("Data Warning : n<10")

    a.posterior.ho <- sum.ho + prior.a
    b.posterior.ho <- n.ho + prior.b

    a.posterior.ha <- sum.ha + prior.a
    b.posterior.ha <- n.ha + prior.b

    # Now we use Monte Carlo sampling to
    # determine the proportion of time that
    # lambda_0 < lambda_a and if it's
    # larger than 1-p we'll claim that the
    # rate has increased.
    ho.sample <- rgamma(1e+05, shape = a.posterior.ho,
        rate = b.posterior.ho)
    ha.sample <- rgamma(1e+05, shape = a.posterior.ha,
        rate = b.posterior.ha)

    proportion.k <- mean(ho.sample < ha.sample)

    monte.carlo.probs[[k]] <- proportion.k

    p1 <- ggplot(data.frame(x = c(0, 10)),
        aes(x)) + stat_function(fun = function(x) dgamma(x,
        shape = a.posterior.ho, rate = b.posterior.ho),
        aes(colour = "posterior.ho")) +
        ylab("density") + xlab(TeX("$\\lambda$"))
    p2 <- p1 + stat_function(fun = function(x) dgamma(x,
        shape = a.posterior.ha, rate = b.posterior.ha),
        aes(colour = "posterior.ha")) +
        ggtitle(paste("Posterior Distributions Category=",
            k, " monte.carlo.prob= ", proportion.k,
            sep = "")) + scale_colour_manual("Density",
        values = c("red", "blue")) + theme(plot.title = element_text(size = 11))
    print(p2)

}
```

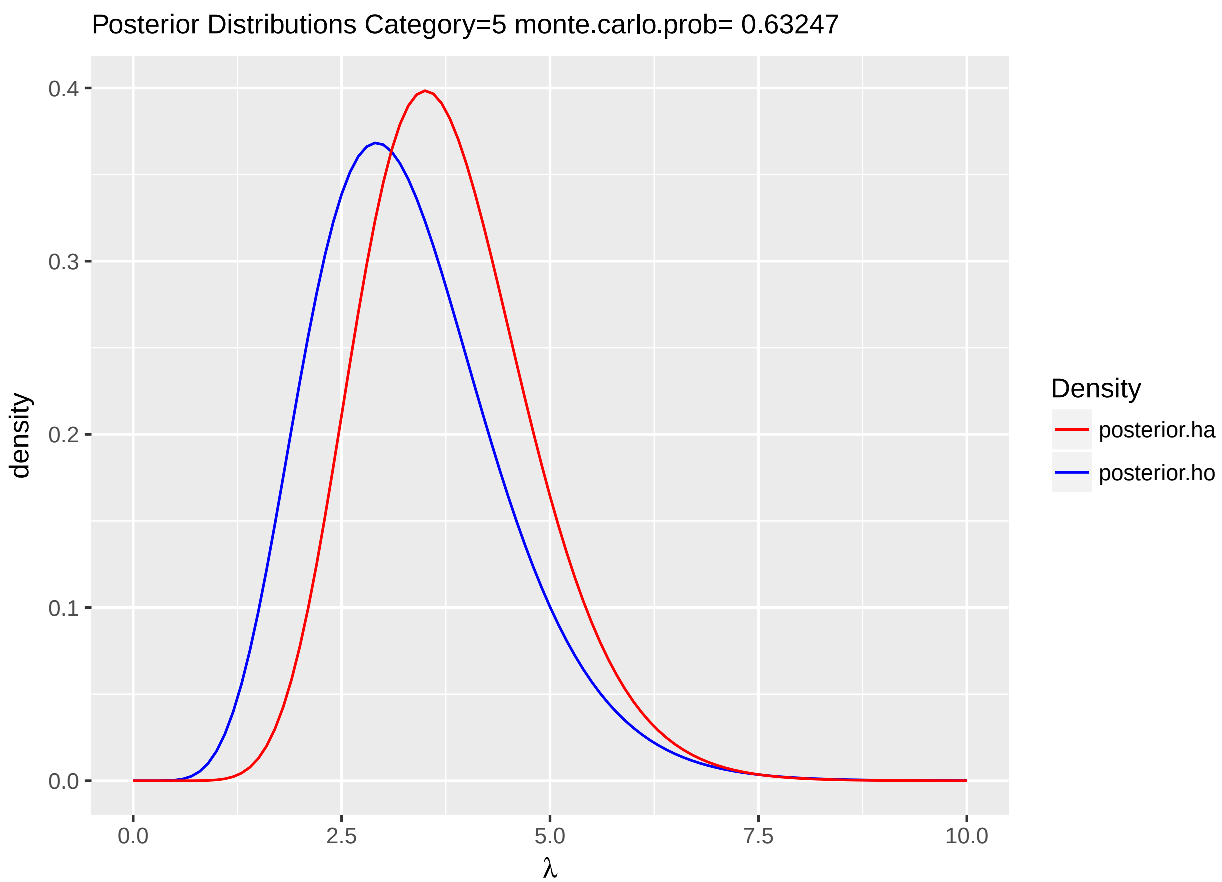

Posterior Distributions Category=1 monte.carlo.prob= 0.45551
density
1.5
1.0
0.5
0.0
0.0
2.5
5.0
7.5
10.0
λ
Density
posterior.ha
posterior.ho



Posterior Distributions Category=2 monte.carlo.prob= 0.50277
density
0.75
0.50
0.25
0.00
0.0
2.5
5.0
7.5
10.0
λ
Density
posterior.ha
posterior.ho


Posterior Distributions Category=3 monte.carlo.prob= 0.53193
density
0.0
0.2
0.4
0.6
0.8
0.0
2.5
5.0
7.5
10.0
λ
Density
posterior.ha
posterior.ho


Posterior Distributions Category=4 monte.carlo.prob= 0.57673
density
0.0
0.2
0.4
0.6
0.0
2.5
5.0
7.5
10.0
λ
Density
posterior.ha
posterior.ho



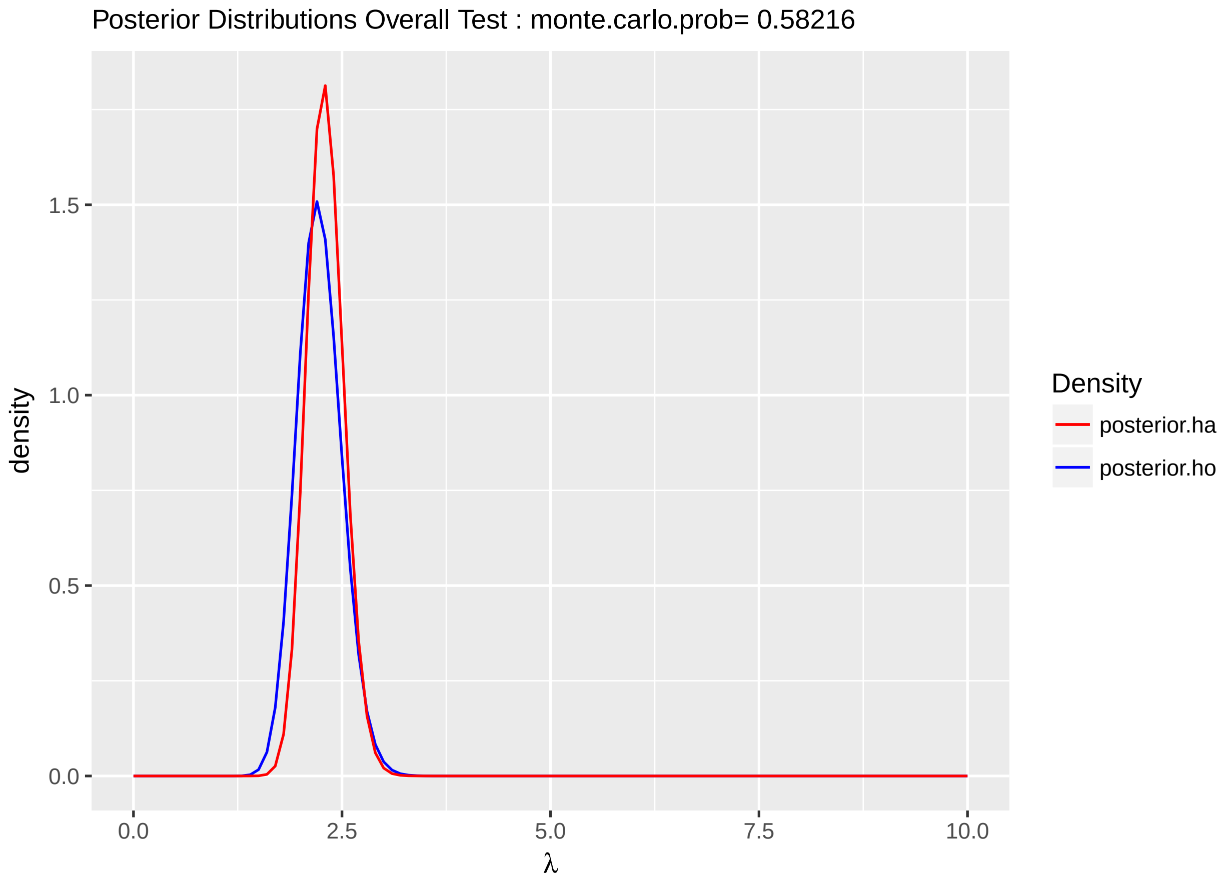

```
monte.carlo.probs <- as.data.frame(do.call("rbind",
    monte.carlo.probs))
names(monte.carlo.probs) <- "probability"
monte.carlo.probs$category <- 1:5
pander(monte.carlo.probs)
```

probability	category
0.4555	1
0.5028	2
0.5319	3
0.5767	4
0.6325	5

We see evidence that the probability of higher category hurricanes are increasing. We will now examine the case that the overall intensity increased.

```
Hok <- Ho
Hak <- Ha

# Calculate components of posterior
# parameters
sum.ho <- sum(Hok$Category)
```

```
sum.ha <- sum(Hak$Category)

n.ho <- nrow(Hok)
n.ha <- nrow(Hak)

if (n.ho < 10 || n.ha < 10) warning("Data Warning : n<10")

a.posterior.ho <- sum.ho + prior.a
b.posterior.ho <- n.ho + prior.b

a.posterior.ha <- sum.ha + prior.a
b.posterior.ha <- n.ha + prior.b

# Now we use Monte Carlo sampling to
# determine the proportion of time that
# lambda_0 < lambda_a and if it's
# larger than 1-p we'll claim that the
# rate has increased.
ho.sample <- rgamma(1e+05, shape = a.posterior.ho,
    rate = b.posterior.ho)
ha.sample <- rgamma(1e+05, shape = a.posterior.ha,
    rate = b.posterior.ha)

proportion.k <- mean(ho.sample < ha.sample)

monte.carlo.probs[[k]] <- proportion.k

p1 <- ggplot(data.frame(x = c(0, 10)),
    aes(x)) + stat_function(fun = function(x) dgamma(x,
    shape = a.posterior.ho, rate = b.posterior.ho),
    aes(colour = "posterior.ho")) + ylab("density") +
    xlab(TeX("$\\lambda$"))
p2 <- p1 + stat_function(fun = function(x) dgamma(x,
    shape = a.posterior.ha, rate = b.posterior.ha),
    aes(colour = "posterior.ha")) + ggtitle(paste("Posterior Distributions Overall Test : monte
    proportion.k, sep = "")) + scale_colour_manual("Density",
    values = c("red", "blue")) + theme(plot.title = element_text(size = 11))
print(p2)
```



We see some evidence that the overall rate has increased.