Monte Carlo Simulation

Rory Quinlan 2024-02-12

Context

A cancer lab is estimating the rate of tumorgenesis in two types of mice. Type A mice have been well stuidied , and information from other labs indicates that type A mice have tumor counts that are approximately poisson distributed with a mean of 12, and theta a has a gamma distribution as gamma(120,12). Type B mice tumor counts are unknown distribution.

Find the probability that tumorgenesis affinity, or θ, of mice A are higher than mice B given our data and using montecarlo simulation.

- There is a very high probability that mice type A has a higher affinity for tumorgenesis given the simulation

```
# Lab data for mice types
# Ya is the tumor count of 10 type A mice
# yb is the tumor count of 13 type B mice

ya <- c(12, 9, 12, 14, 13, 13, 15, 8, 15, 6)
sum_a<- sum(ya)
n_a <- length(ya)


yb <- c(11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)
sum_b <- sum(yb)
n_b <- length(yb)

# Priors from other labs [ dist ~ gamma() ]

a <- 120
a2 <- 10
b <- 12
b2 <- 1
```

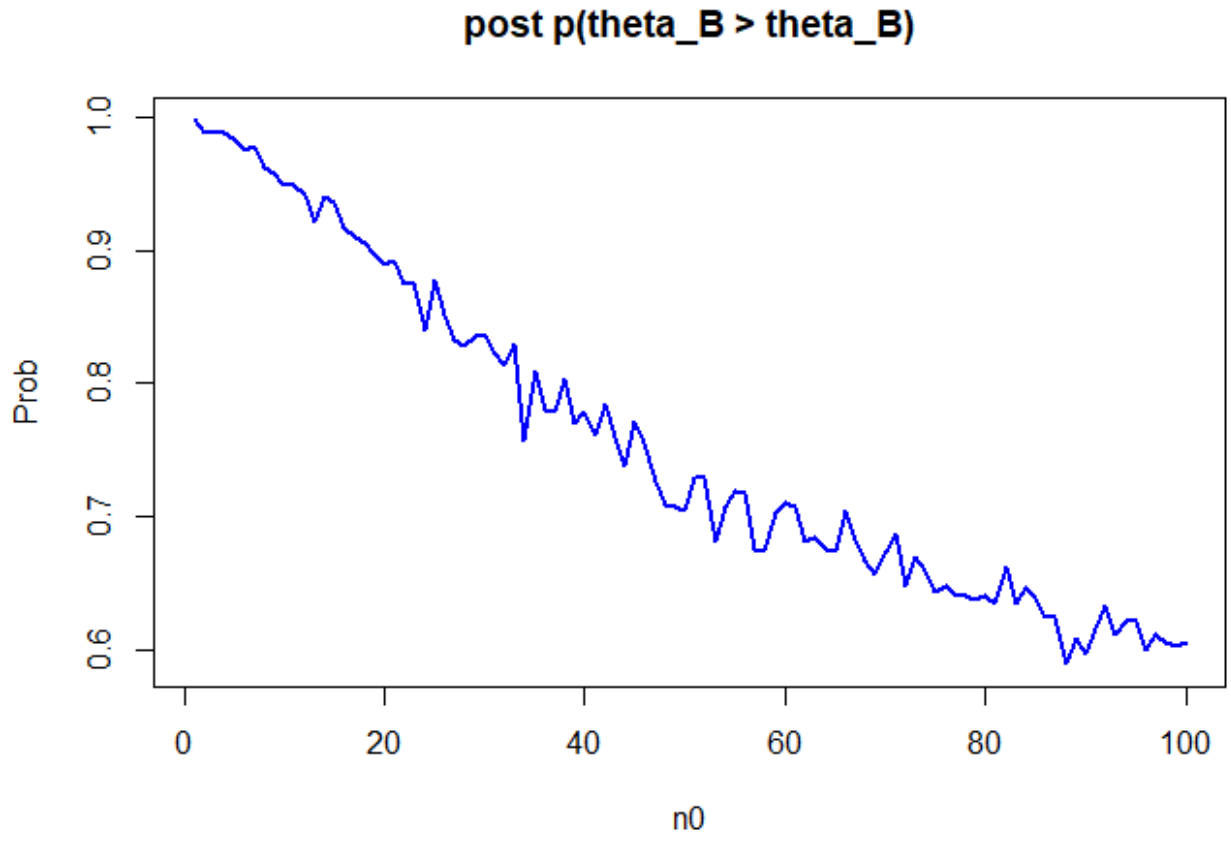
```
# Monte carlo simulation

set.seed(1000)

# set parameters for simulation with data from labs
k<- a + sum_a
k2<- b + sum_b
r<- a2 + n_a
r2<- b2 + n_b

# generate 1000 random samples from prior distribution and data
theta_a_mont <- rgamma(1000, k , r)

# generate 1000 random samples from prior distribution and data
theta_b_mont <- rgamma(1000, k2 , r2 )

# For times that theta a is larger than theta b is larger, take the mean and print as



## P(theta_b < theta_a | y_a, y_b) = 0.998
```

**For a range of n_0 values find the probability that $\theta_B < \theta_A$ from lab data priors and assuming gamma θ_B also follows gamma like type A mice **

- as n_0 increases the probability that θ_a is greater than θ_b decreases. conclusions are not sensitive to prior, because at a $n_0 = 100$, a large prior still has a posterior probability of 0.6; above .5

```
set.seed(1000)

# Range of values for n0
n0 <- c(1:100)

# Empty list to fill with loop
prob <- c()

# Calculate the probability that theta a is greater than theta b for each value in th

for(i in 1:length(n0)){
  new_b <- b * n0[i]
  new_b2 <- b2 * n0[i]
  theta_a_mont2 <- rgamma(1000, k, r)
  theta_b_mont2 <- rgamma(1000, new_b + sum_b, new_b2 + n_b)
```

```
  mean <- mean(theta_a_mont2>theta_b_mont2)
  prob <- c(prob, mean)
}


# Create data for graph


# Plot graph
plot(new_p$n0, new_p$prob, lwd=2, col="blue", pch=19, type="l", main="post p(theta_B
```



**Use montecarlo simulation to find the probability that $\tilde{Y}_B < \tilde{Y}_A$ samples from posterior distribution. Where $\tilde{Y}_A$ and $\tilde{Y}_B$ are samples from the posterior distribution **

```
# Theta_a_mont and theta_b_mont from above
# select 1000 random samples from a poisson distribution with the thetas calculated f
```

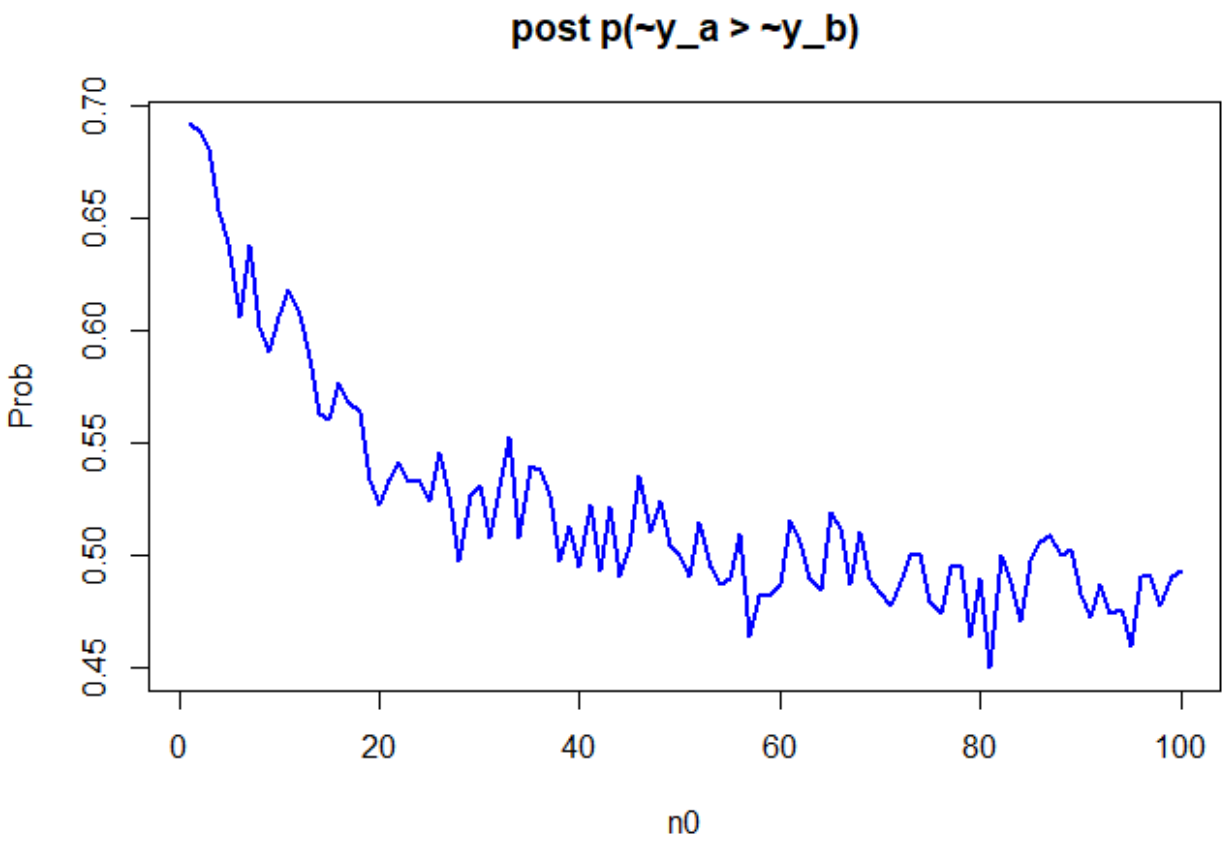
```
y_a_mont <- rpois(1000, theta_a_mont)
y_b_mont <- rpois(1000, theta_b_mont)


# Print probability
cat("P(theta_b < theta_a | y_a, y_b) =",mean(y_a_mont > y_b_mont))
```

```
## P(theta_b < theta_a | y_a, y_b) = 0.692
```

**For a range of n_0 values find the probability that $\tilde{Y}_B < \tilde{Y}_A$ from lab data priors and assuming gamma θ_B also follows gamma like type A mice **

- as $n_0 $increases the probability that $\tilde{Y}_A$ is greater than $\tilde{Y}_B$ decreases. conclusions are much more sensitive to prior, because at a $n_0 = 100$, a large prior only has a posterior probability of 0.6; ~ 0.49 below .5

```
# Sample 1000 posterior theta_a and b for each value of n0 (1 to 100)

prob <- c()
for(i in 1:length(n0)){
  new_b <- b * n0[i]
  new_b2 <- b2 * n0[i]
  theta_a_mont2 <- rgamma(1000, k, r)
  theta_b_mont2 <- rgamma(1000, new_b + sum_b, new_b2 + n_b)
  y_a_mont <- rpois(1000, theta_a_mont2)
  y_b_mont <- rpois(1000, theta_b_mont2)

  mean <- mean(y_a_mont > y_b_mont)
  prob <- c(prob, mean)
}

# Create data to plot
new_p2 <- data.frame(n0 = n0, probability = prob)


plot(new_p2$n0, new_p2$prob, lwd=2, col="blue", pch=19, type="l", main="post p(~y_a >
```

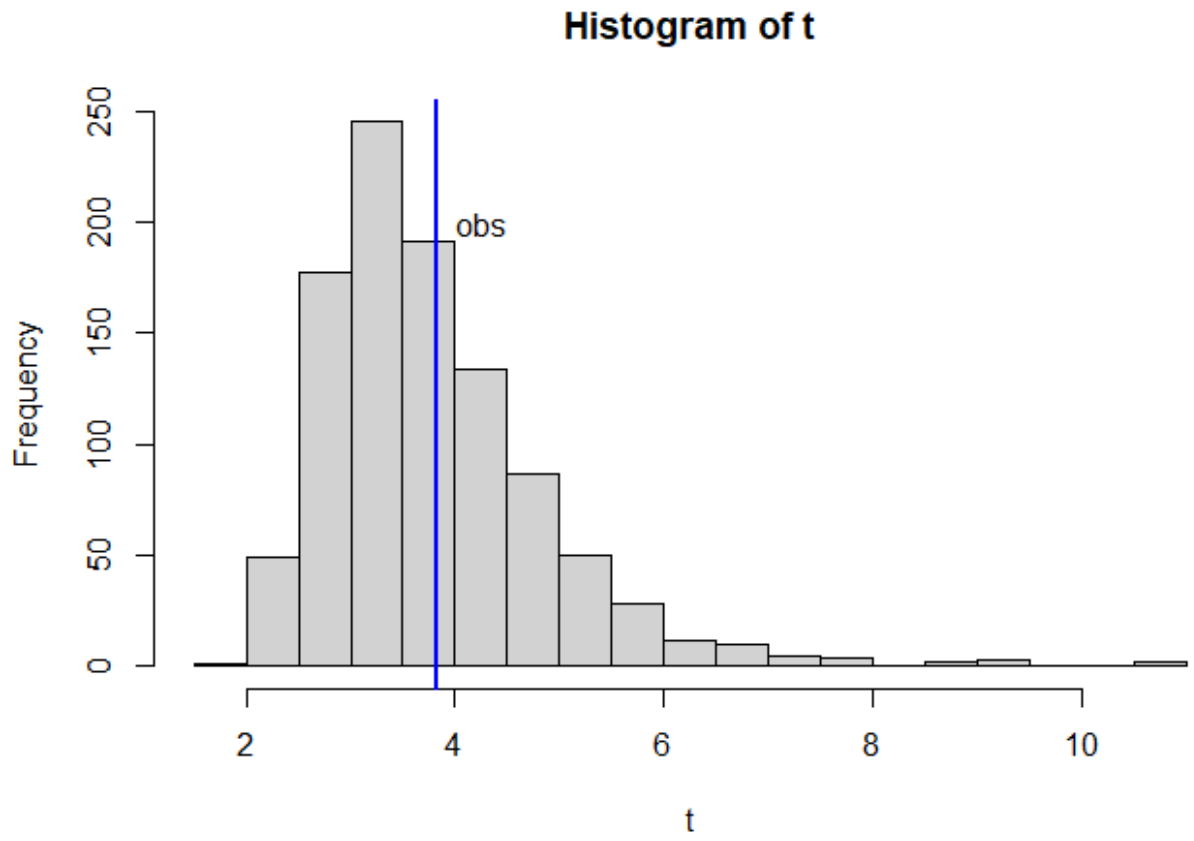


** Evaluate accuracy of our poisson model)**

- The model is a good fit because the observed value (blue line) is close to the mode of the histogram.

```
t_mc <- c()

# for random samples from our poisson distribution 1000 samples of 10 with our theta
# Calculate statistic t from each of 1000 sample
for(s in 1:1000){
  theta1 <- rgamma(1, a + sum_a, a2 + n_a)
  y1_mc <- rpois(10, theta1)
  t_mc <- c(t_mc, mean(y1_mc)/sd(y1_mc))
}

# T obs if the observed value of statistics from lab data
t_obs <- mean(ya)/sd(ya)

# Create histogram of simulated statistics (1000 samples)
# Add actual observed value
hist(t_mc,  main ="Histogram of t", xlab = "t", breaks= 13)
```

```
abline(v = t_obs, col = "blue",lwd = 2)
text(x=4.25,y=200, "obs")
```



** Evaluate accuracy for data in type B mice**

- This model is not a good fit because the observed value (blue line) is not close to the histograms mode (observed stat sits on the right tail).

```
# Type B data
yb <- c(11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)
sum_b <- sum(yb)
n_b <- length(yb)

b <- 12
b2 <- 1

t_mc <- c()

# same as above, except with data for B now
for(s in 1:1000){
  theta1 <- rgamma(1, b + sum_b, b2 + n_b)
```

```
  y2_mc <- rpois(10, theta1)
  t_mc <- c(t_mc, mean(y2_mc)/sd(y2_mc))
}

t_obs <- mean(yb)/sd(yb)

hist(t_mc, xlab = "t")
abline(v = t_obs, col = "blue",lwd = 2)
text(x=6,y=200, "obs")
```