20121865_PoissonBayes

Louise Branzuela

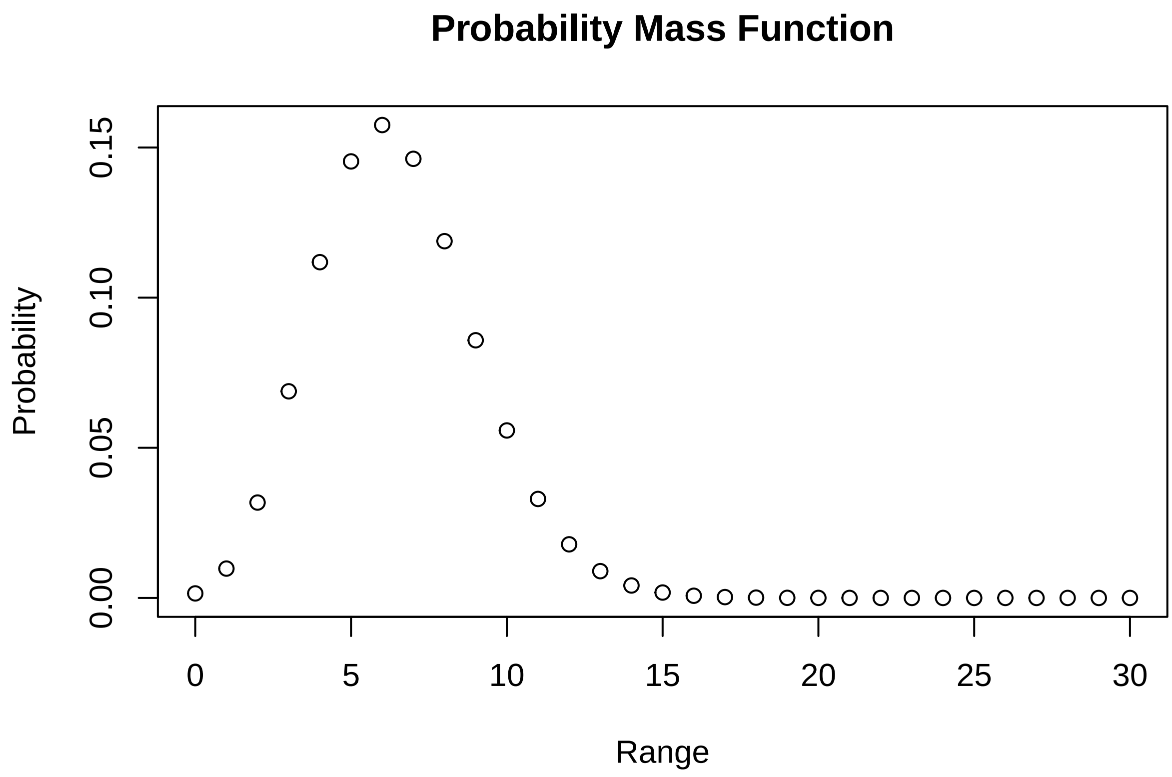
2022-05-05

Poisson Distribution Part 1

Plotting the probability mass function for the random variable

```
lambda = 6.5

x <- 0:30
plot(x,dpois(x,lambda),
     ylab = "Probability",
     xlab = "Range",
     main = "Probability Mass Function")
```



Verifying that the probabilities sum to 1

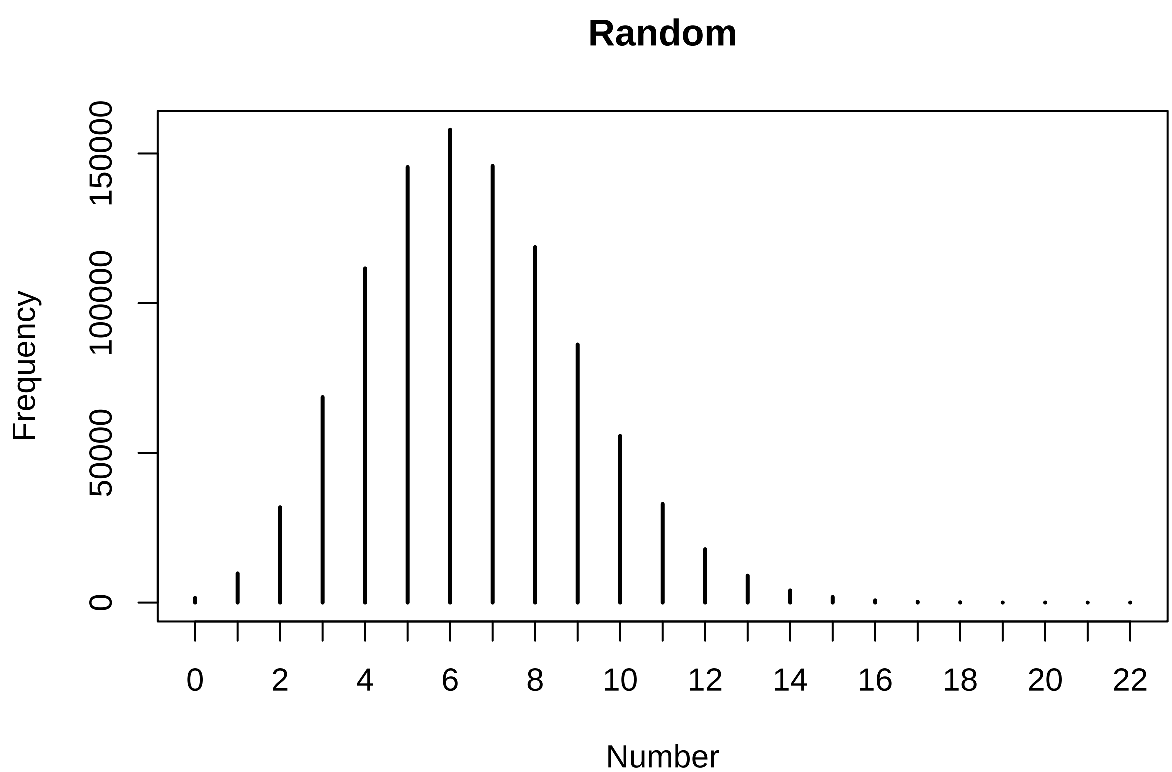
```
sum(dpois(0:30,lambda))
```

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

Mathematically, the above will be slightly less than 1 because we are not including numbers over 100; but the difference will be small.

Sampling from the distribution and plotting a table from the results

```
r = rpois(1e6, lambda)
plot(table(r),
     ylab = "Frequency",
     xlab = "Number",
     main = "Random")
```



Poisson Distribution Part 2

Precise Definition of the P-value

The precise definition of p-value is P[X >= 6]

Performing a one-sided test

```
sum(dpois(0:5, lambda=3.3))
```

```
## [1] 0.8828768
```

Since the p-value that I got was 0.8828768, I don't think the p-value is significant because it's not low enough to reject the null hypothesis.

Calculating the likelihood ratio for hypotheses

```
dpois(6, lambda=4.1)
```

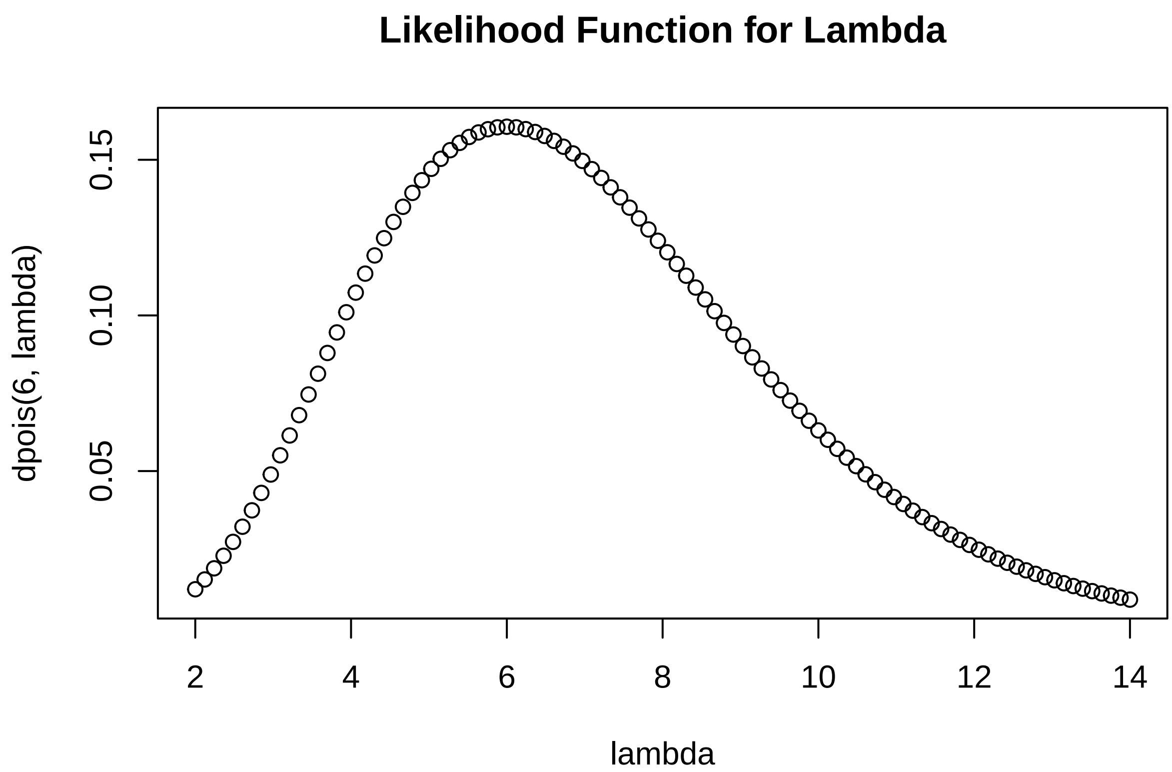
```
## [1] 0.109336
```

```
dpois(6, lambda=2.2)
```

```
## [1] 0.0174484
```

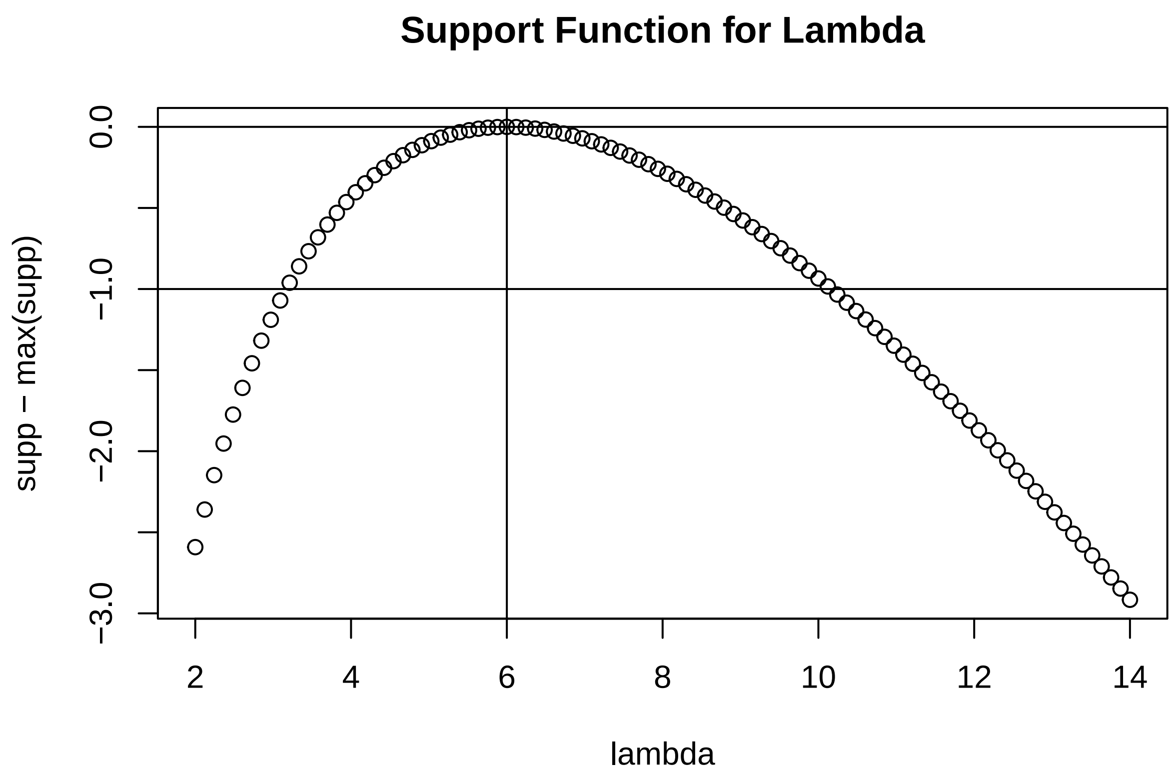
Plotting a log-likelihood for range 2 to 14

```
lambda <- seq(from=2,to=14,len=100)
plot(lambda,dpois(6,lambda),main='Likelihood Function for Lambda')
```



Calculating a credible interval using a 2 units of support criterion

```
supp <- dpois(6,lambda,log=TRUE)
plot(lambda,supp-max(supp),main='Support Function for Lambda')
abline(v=6) # best-supported value is the number of observations
abline(h=0) # maximum support = 0
abline(h=-1) # two units of support gives credible interval
```



From the graph, the credible interval is from approximately 3.1 to 10

Poisson Distribution Part 3

Personal Example

The example that I'm going to be using from my daily life would be the number of cars that go through the McDonalds drive through as I work as an order taker for McDonalds.

Since the number of people going through the drive through each day isn't fixed and can increase or decrease depending on the time of day, this example is a limiting form of a binomial distribution. It would have a large n as n is defined in this example as the number of cars that drive past Penrose McDonalds, and a small p as p is defined in this example as the small probability of any given customer ordering through the drive through in the hour in question.

Rough estimate for n and p

A rough order of magnitude estimate for the number of cars that drive past Penrose McDonalds would be 1000, and the probability that a customer will go into our drive through would be 0.03.

Reasonable value for lambda

A reasonable value for lambda would be n * p which is 1000 * 0.03, which equals to 30. This also makes sense as on average we would have 30 cars in our drive through per hour.

Using rpois() for my distribution

```
rpois(10, lambda=30)
```

```
##  [1] 28 29 36 33 32 36 31 36 36 29
```

The rpois() function is used to compute random density for poisson distribution. Therefore, the numbers that are generated by the function are random densities from the given n and lambda.

Bayesian analysis part 1

Calculating the posterior probability of the plant being a cactus

```
(0.4 * 0.05)/(0.4*0.05 + 0.6*0.24)
```

```
## [1] 0.1219512
```

From using Bayes's theorem, I've calculated that the posterior probability of the plant being a cactus would be 0.1219512.

Bayesian analysis part 2

Personal Example

A situation that I could use as an example here would be my experience with typing speeds within different age ranges in university.

A student in AUT is chosen at random. From AUT's statistics, 67% of students are under 25 years old, 23% of students are between 25 and 39 years old, and 10% of students are above 40 years old. I'm going to define a slow typer as having lower than half the average words per minute (WPM). Average WPM is around 40 WPM, therefore, a slow typer in this case would have a typing speed of less than 20 WPM. I'm going to estimate that the percentage of slow typers, according to my definition, in students under 25 years old as 5%, because of the fact that this generation grew up with computers and typing. Within the 25 and 39 years old age group of students, I estimate that there would be 8% that are slow typers, and within the 40+ years old age group, there would be 10% that re slow typers. The percentage of slow typers increase with the age group as the less they would have interacted with typing at a young age and developed the muscle memory for touch typing.

Hypotheses

My hypotheses are as follows: The probability that a student under 25 is chosen is 67%. The probability that a student between 25 and 39 is chosen is 23%. The probability that a student above the age of 40 is chosen is 10%.

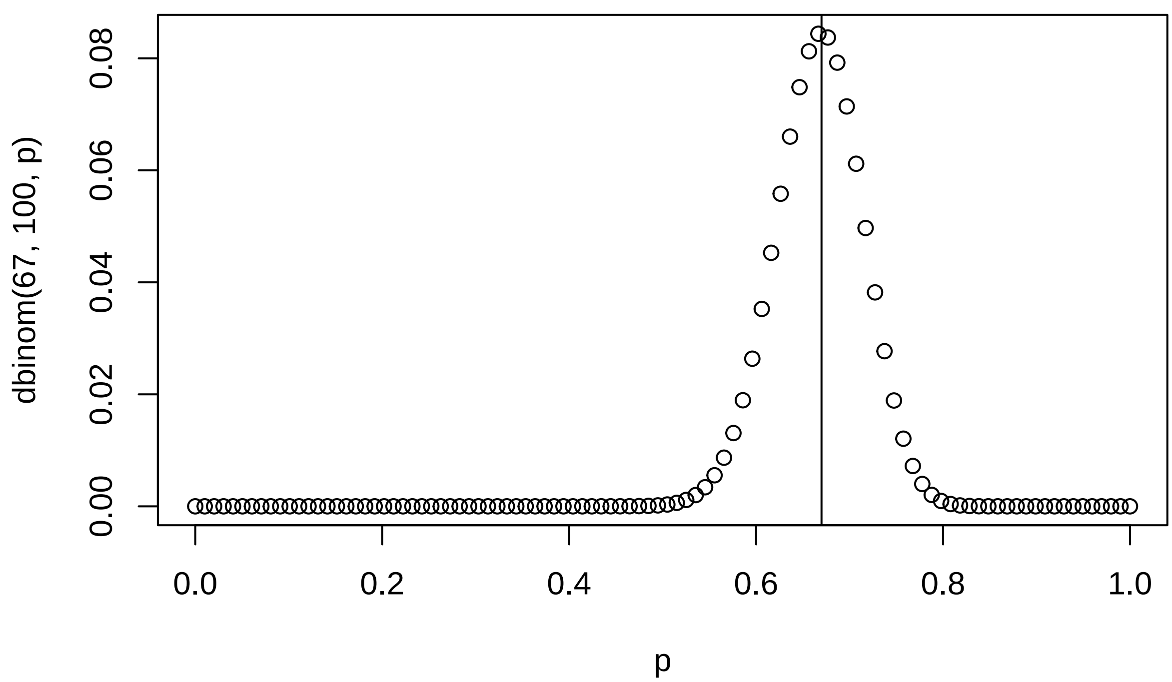
Data

My data is as follows: 5% of students under 25 years old are slow typers. 8% of students between 25 and 39 years old are slow typers. 10% of students above 40 years old are slow typers.

Likelihood Functions

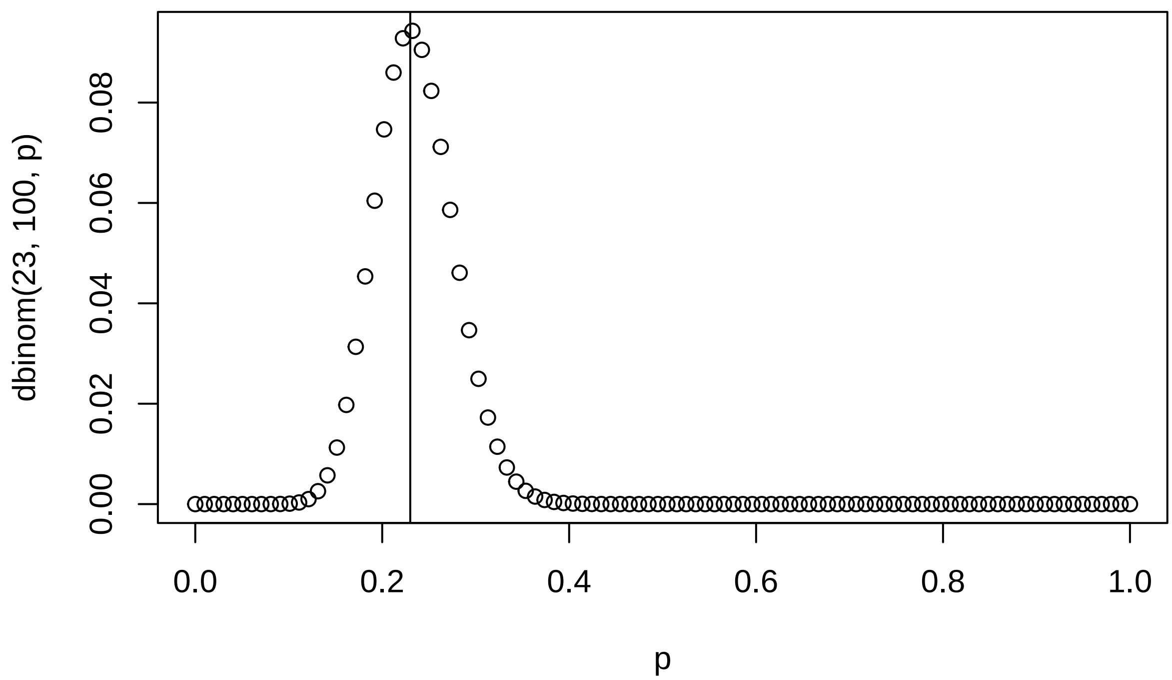
Likelihood Function of a student under 25 years old is chosen:

```
p <- seq(from=0, to=1, len=100)
plot(p,dbinom(67,100,p))
abline(v=0.67)
```



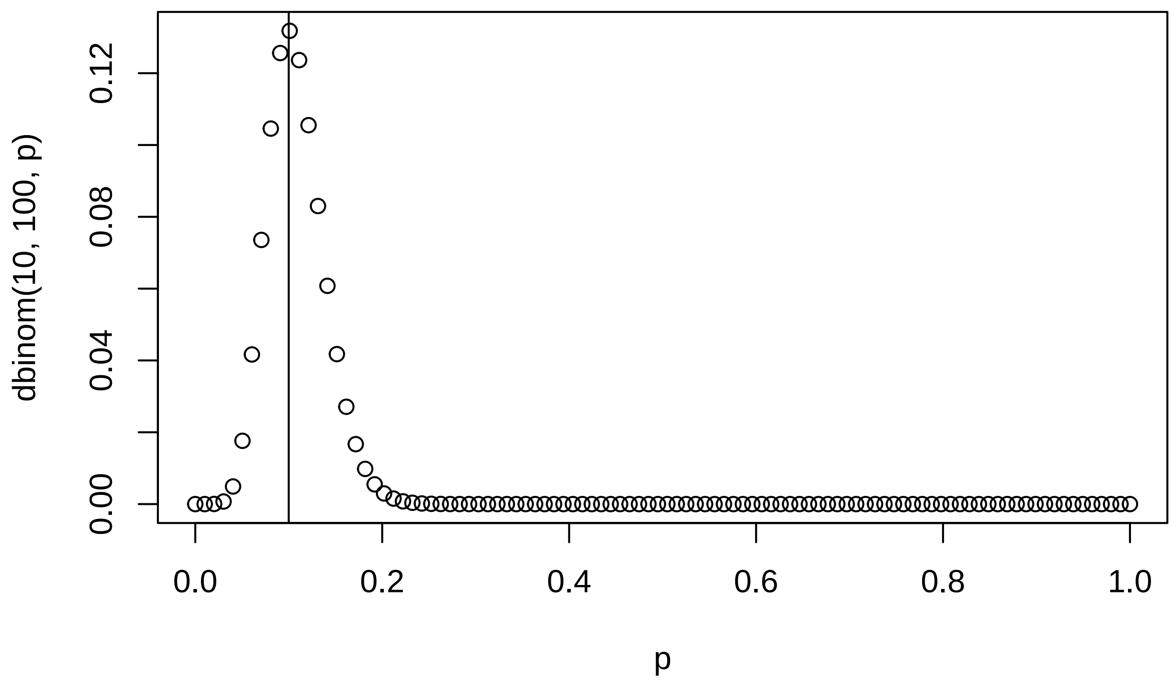
Likelihood Function of a student between 25 and 39 years old is chosen:

```
p <- seq(from=0, to=1, len=100)
plot(p,dbinom(23,100,p))
abline(v=0.23)
```



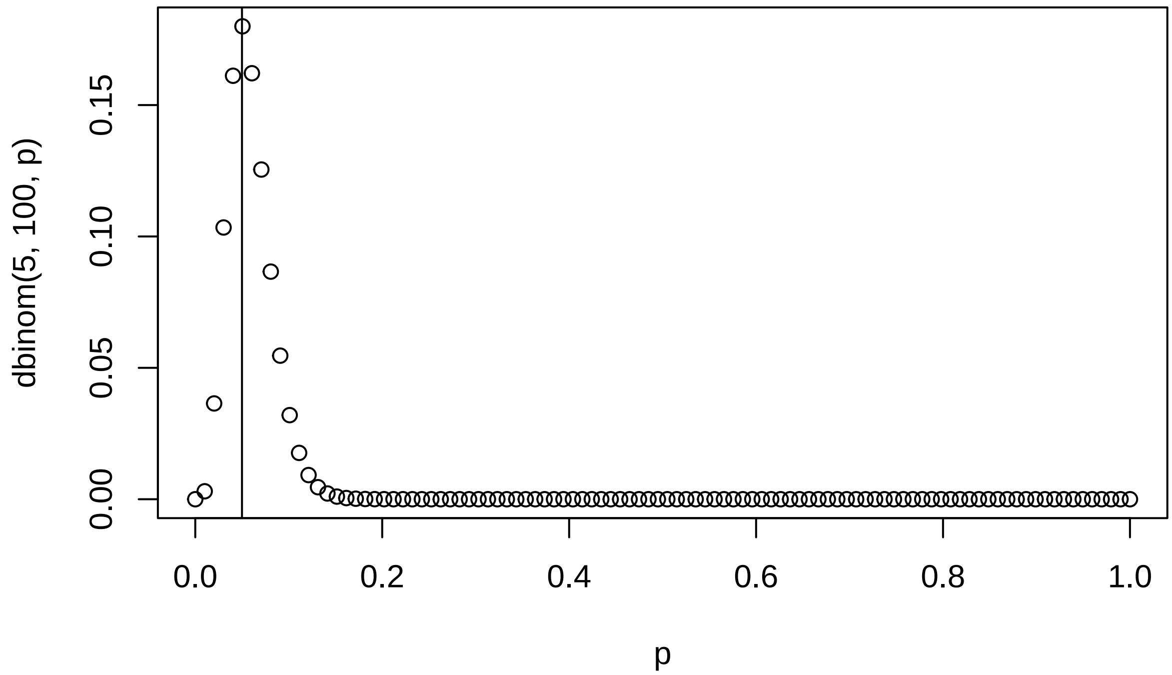
Likelihood Function of a student above 40 years old is chosen:

```
p <- seq(from=0, to=1, len=100)
plot(p,dbinom(10,100,p))
abline(v=0.1)
```



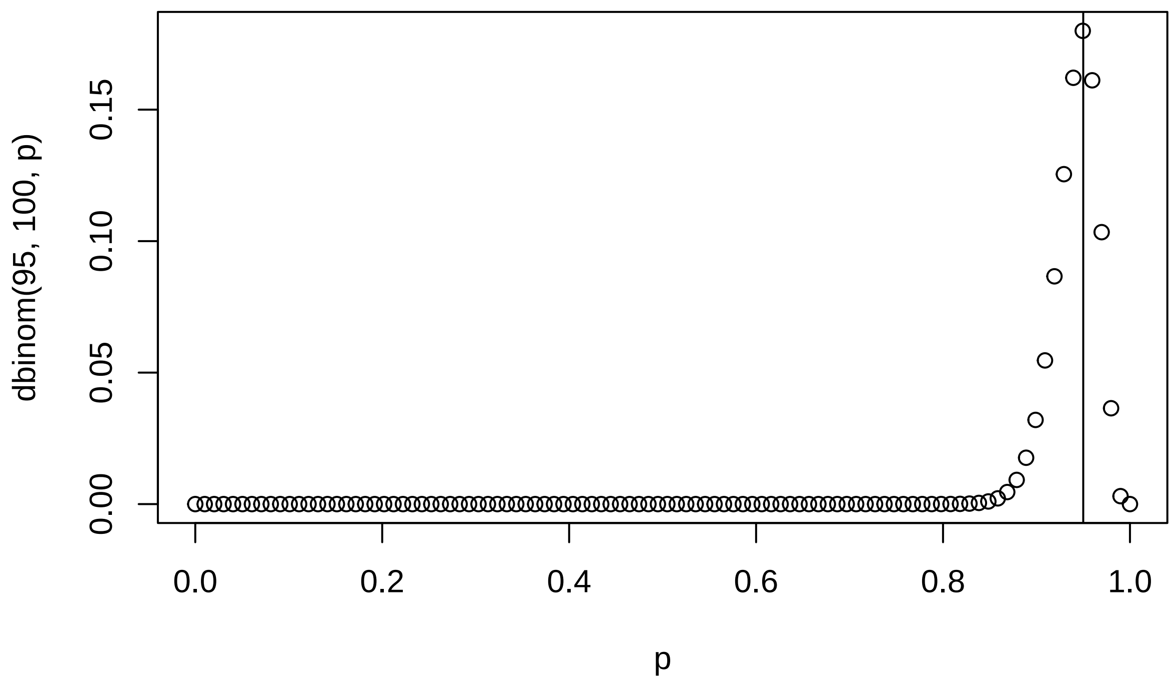
Likelihood Function of a slow typer given that they are under 25 years old is chosen:

```
p <- seq(from=0, to=1, len=100)
plot(p,dbinom(5,100,p))
abline(v=0.05)
```



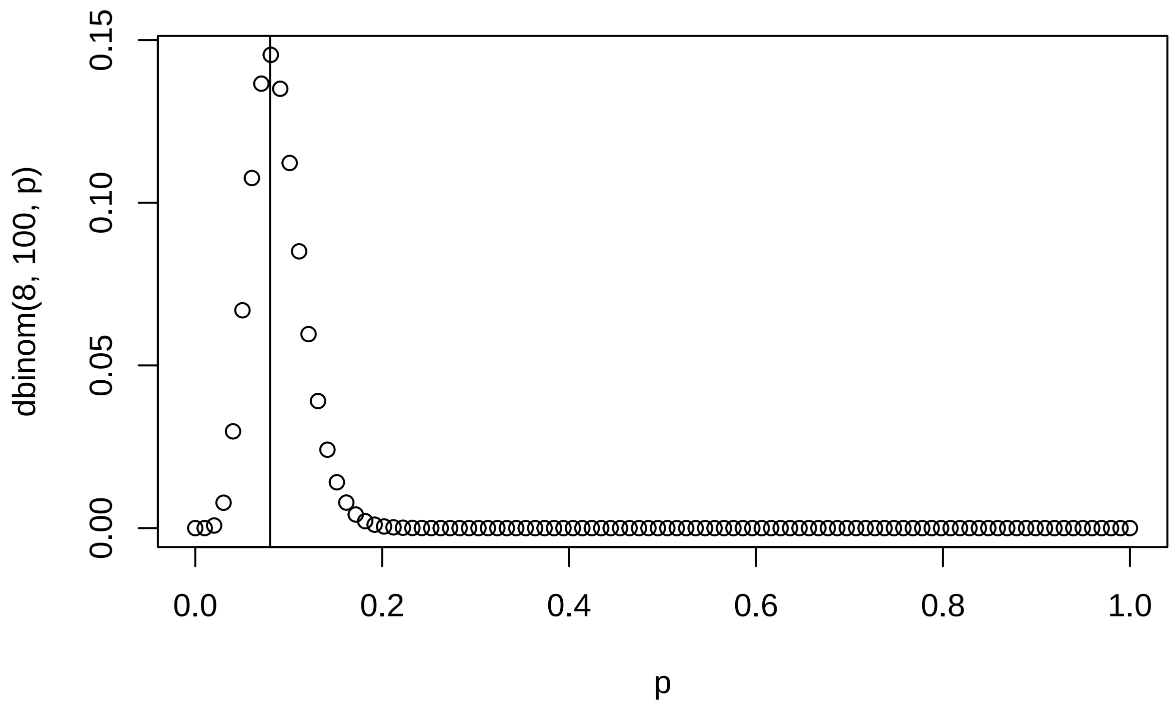
Likelihood Function of a non-slow typer given that they are under 25 years old is chosen:

```
p <- seq(from=0, to=1, len=100)
plot(p,dbinom(95,100,p))
abline(v=0.95)
```



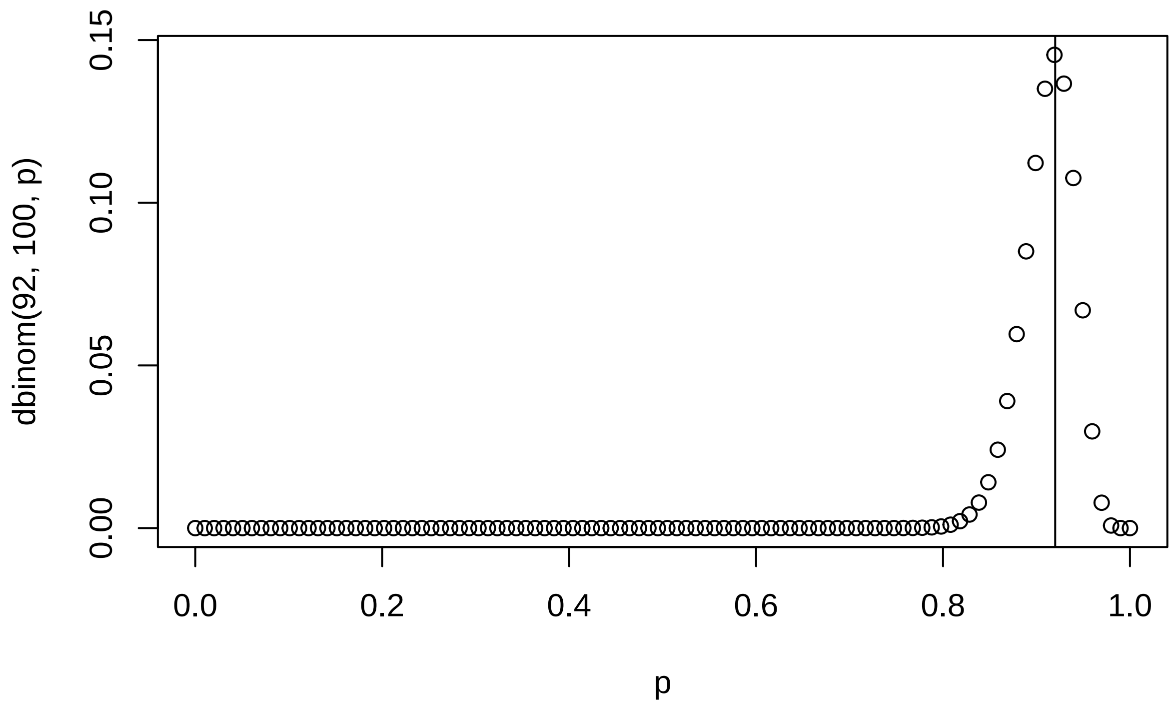
Likelihood Function of a slow typer given that they are between 25 and 39 years old is chosen:

```
p <- seq(from=0, to=1, len=100)
plot(p,dbinom(8,100,p))
abline(v=0.08)
```



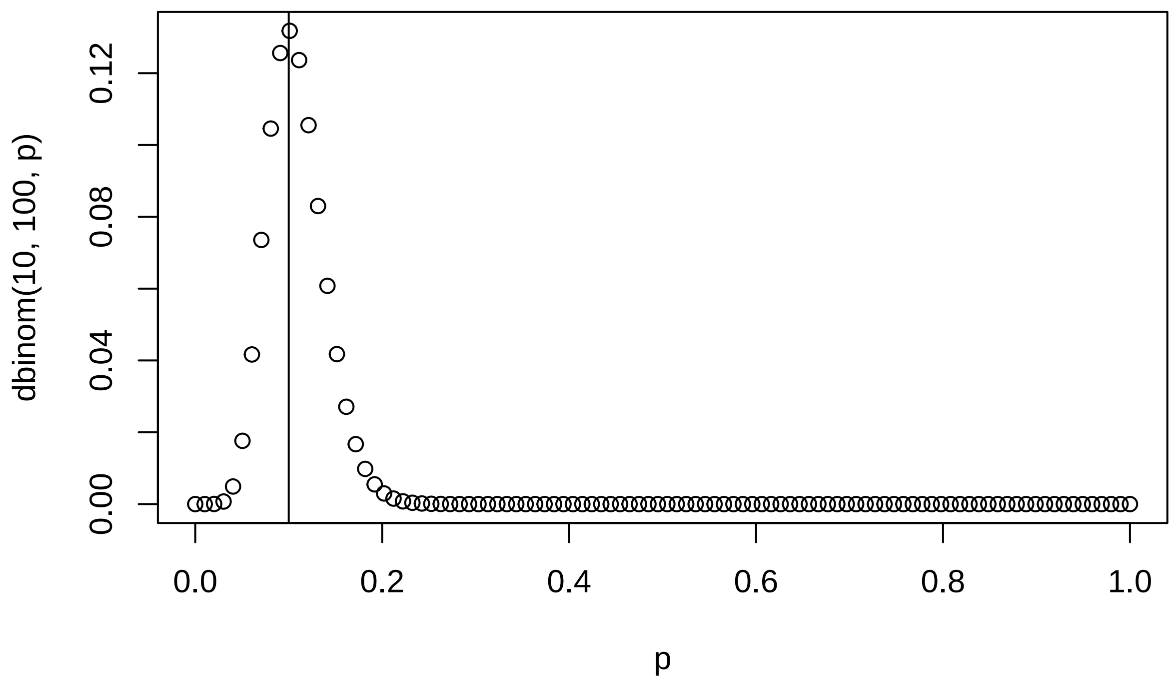
Likelihood Function of a non-slow typer given that they are between 25 and 39 years old is chosen:

```
p <- seq(from=0, to=1, len=100)
plot(p,dbinom(92,100,p))
abline(v=0.92)
```



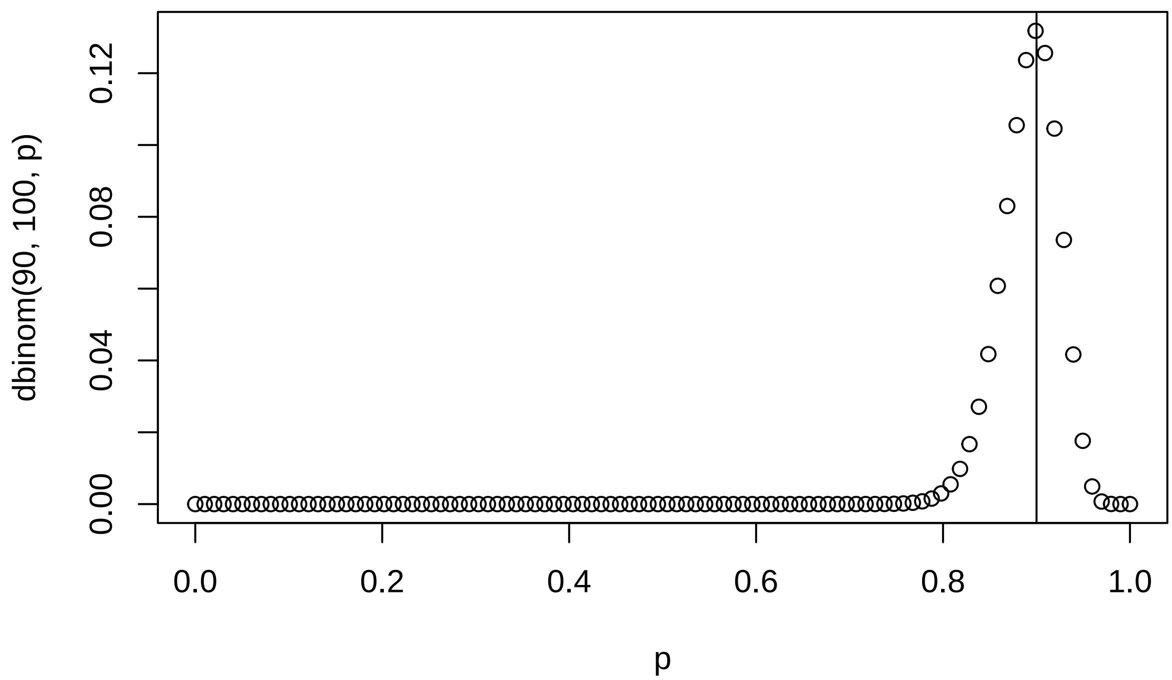
Likelihood Function of a slow typer given that they are above 40 years old is chosen:

```
p <- seq(from=0, to=1, len=100)
plot(p,dbinom(10,100,p))
abline(v=0.1)
```



Likelihood Function of a non-slow typer given that they are above 40 years old is chosen:

```
p <- seq(from=0, to=1, len=100)
plot(p,dbinom(90,100,p))
abline(v=0.9)
```



Posterior Probability

Posterior Probability of a student under 25 years old is chosen given they are a slow typer:

```r
(0.67*0.05)/(0.67*0.05 + 0.23*0.08 + 0.1*0.1)
```

```
## [1] 0.5411955
```

Posterior Probability of a student between 25 and 39 years old is chosen given they are a slow typer:

```r
(0.23*0.08)/(0.67*0.05 + 0.23*0.08 + 0.1*0.1)
```

```
## [1] 0.2972536
```

Posterior Probability of a student above 40 years old is chosen given they are a slow typer:

```r
(0.1*0.1)/(0.67*0.05 + 0.23*0.08 + 0.1*0.1)
```

```
## [1] 0.1615509
```