CGS698C - Assignment 2

Sahil Tomar (210898)

2024-06-14

Part 1 : A Simple Binomial Model

We are given:

- the data : $y = 7$
- the marginal likelihood : $\int \mathcal{L}(\theta|y) * p(\theta) d\theta = \frac{1}{11}$
- the likelihood function : $\mathcal{L}(\theta|y) = \binom{10}{y} * \theta^y * (1-\theta)^{10-y}$
- The prior assumption : $p(\theta) = \begin{cases} 1 & \text{if } 0 \le \theta \le 1 \\ 0 & \text{otherwise} \end{cases}$

1.1

Use the given data to simplify the likelihood function,

$\mathcal{L}(\theta|7) = \binom{10}{7} * \theta^7 * (1-\theta)^3 \Rightarrow \mathcal{L}(\theta|7) = 120 * \theta^7 * (1-\theta)^3$

Using $p(\theta|y) = \frac{\mathcal{L}(\theta|y) * p(\theta)}{\int \mathcal{L}(\theta|y) * p(\theta) d\theta}$,

We can see that the posterior

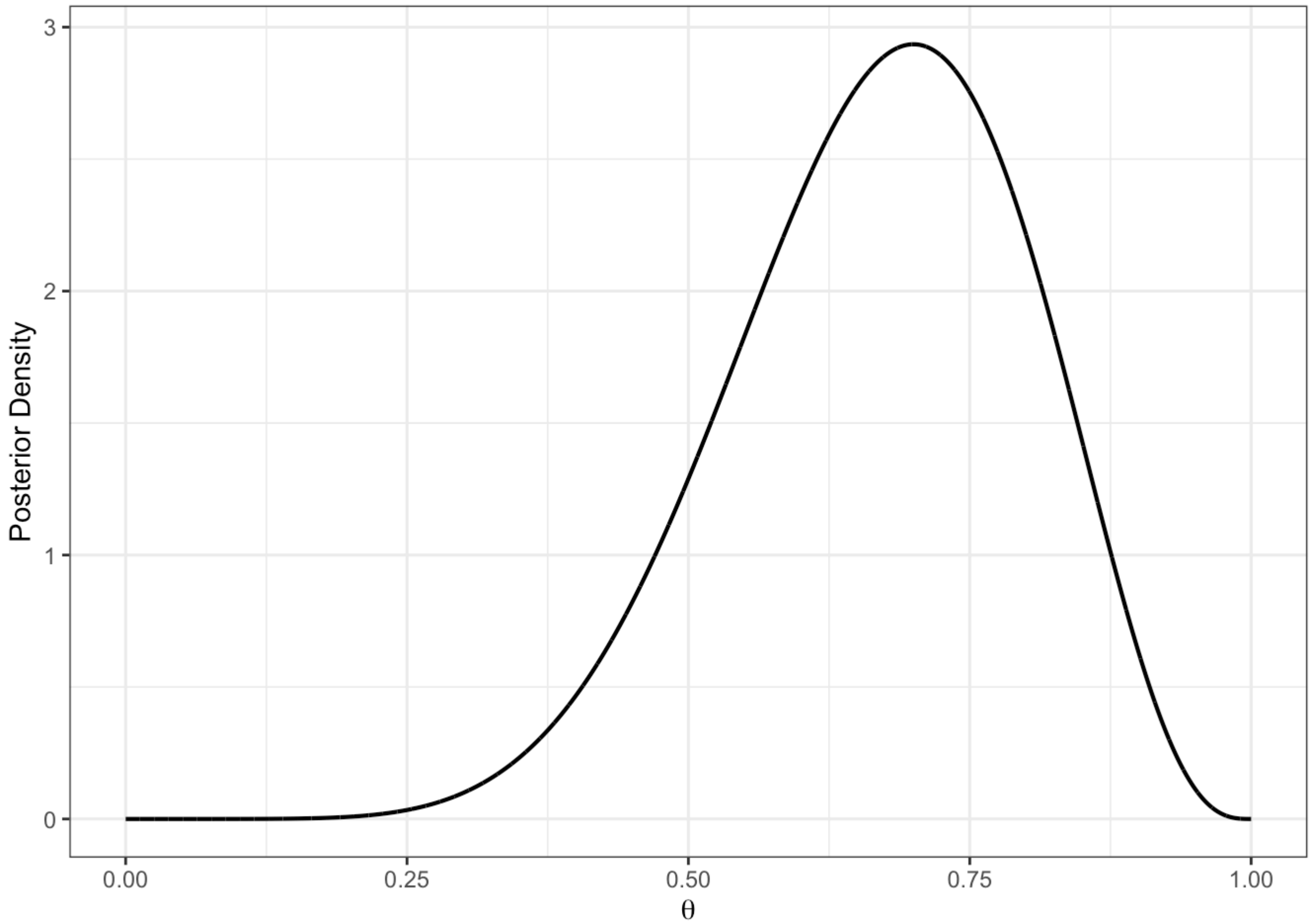
$$p(\theta|y) = \begin{cases} 11 * 120 * \theta^7 * (1-\theta)^3 & \text{if } 0 \le \theta \le 1 \\ 0 & \text{otherwise} \end{cases} \Rightarrow p(\theta|y) = \begin{cases} 1320 * \theta^7 * (1-\theta)^3 & \text{if } 0 \le \theta \le 1 \\ 0 & \text{otherwise} \end{cases}$$

a. posterior density for $\theta = 0.75$ is - $p(0.075|7) = 1320 * 0.75^7 * (1-0.75)^3 = 1320 * 0.075^7 * (0.25)^3 = 2.753105164$
b. posterior density for $\theta = 0.25$ is - $p(0.025|7) = 1320 * 0.25^7 * (1-0.25)^3 = 1320 * 0.025^7 * (0.75)^3 = 0.03398895264$
c. posterior density for $\theta = 1$ is - $p(1|7) = 1320 * 1^7 * (1-1)^3 = 0$

1.2

posterior distribution of θ, that is $p(\theta|y)$ -

```
y <- 7
N <- 10
marginal_likelihood <- 1/11

likelihoods <- data.frame(theta = seq(from=0, to=1, by=0.0001))

likelihoods$lkl <- dbinom(y, N, likelihoods$theta)

likelihoods$prior <- rep(1, nrow(likelihoods))

likelihoods$posterior_unnorm <- likelihoods$lkl * likelihoods$prior

likelihoods$posterior <- likelihoods$posterior_unnorm / marginal_likelihood

ggplot(likelihoods, aes(x=theta, y=posterior)) + geom_line(linewidth=0.75) + theme_bw
() + xlab(expression(theta)) + ylab("Posterior Density")
```



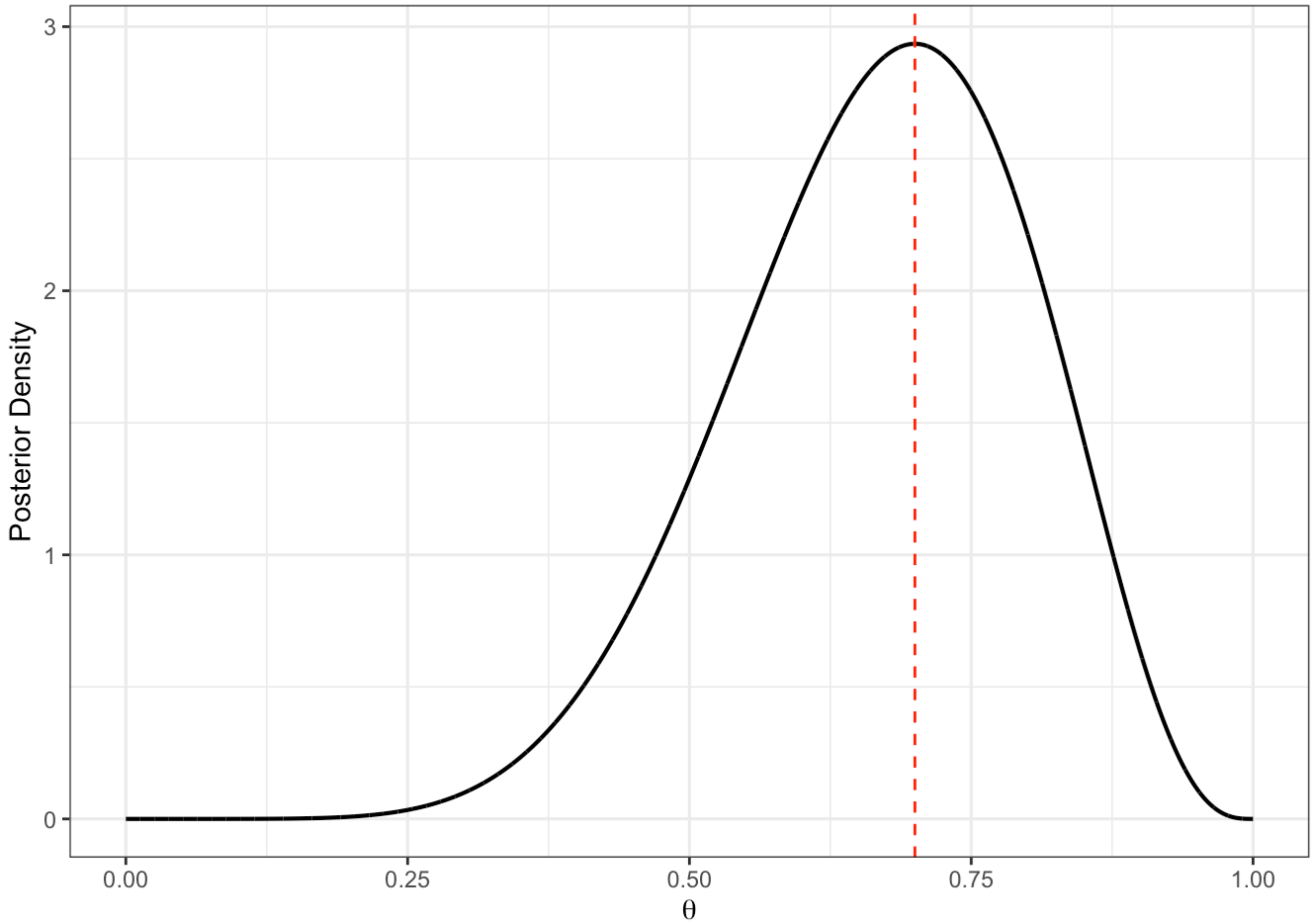
1.3

The value of θ that has the maximum posterior density, i.e.,

arg max $p(\theta|y)$ = 0.7 (Calculated using `likelihoods$theta[which(likelihoods$posterior == max(likelihoods$posterior))]`)

Differentiating the function $p(\theta|y)$ also gives the same maxima.

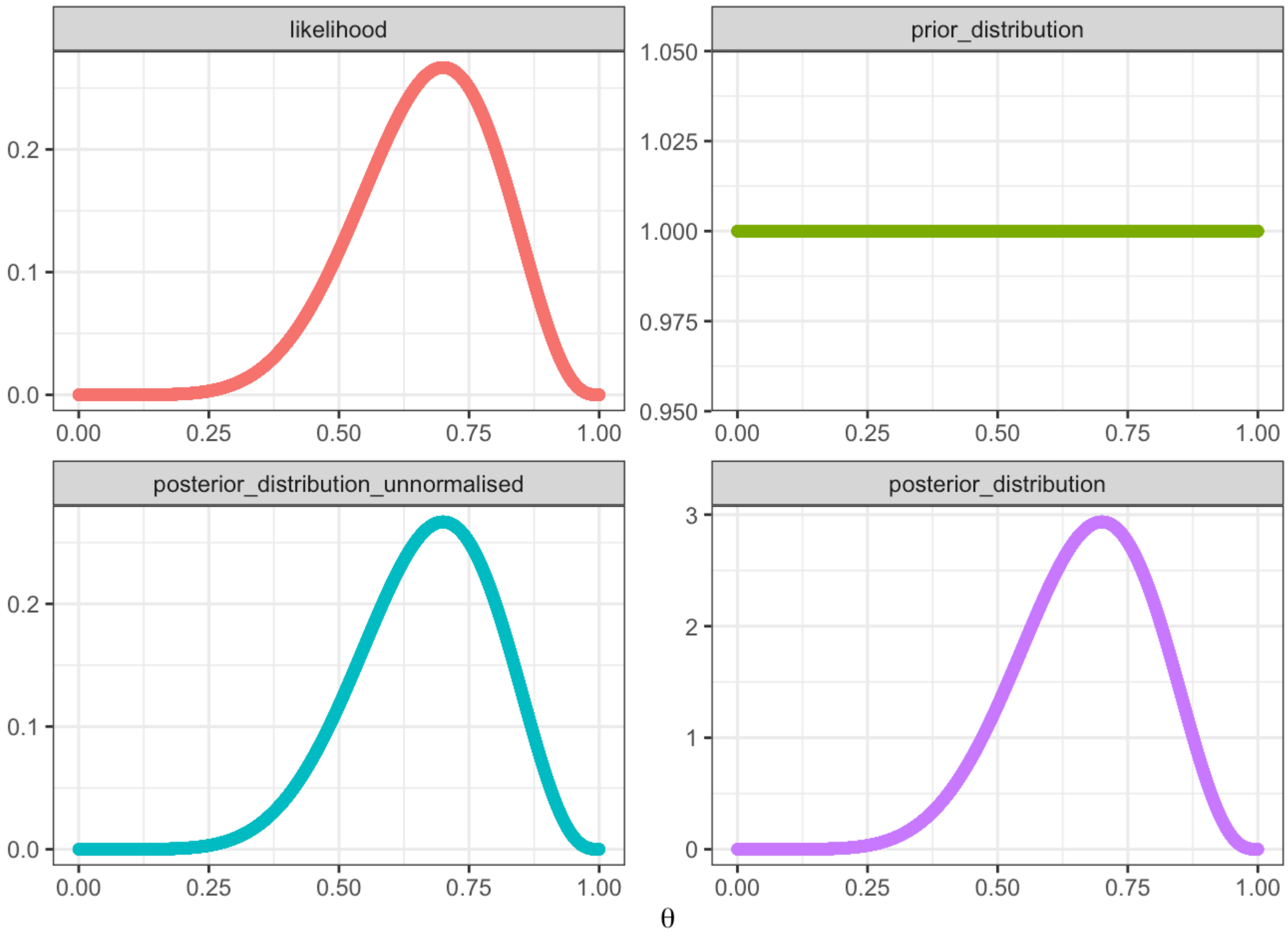
```
ggplot(likelihoods, aes(x=theta, y=posterior)) + geom_line(linewidth=0.75) + theme_bw
() + xlab(expression(theta)) + ylab("Posterior Density") + geom_vline(xintercept = 0.
7, linetype = "dashed", color = "red")
```



1.4

Plot the graphs of the likelihood function, the prior distribution, and the posterior distribution -

```
colnames(likelihoods) <- c("theta", "likelihood", "prior_distribution", "posterior_di
stribution_unnormalised", "posterior_distribution")
likelihoods.m <- melt(likelihoods, id = c("theta"))
ggplot(likelihoods.m,aes(x=theta,y=value,group=variable,color=variable))+ geom_point
()+facet_wrap(~variable,scales="free",nrow=3) + theme_bw() + xlab(expression(theta))
+ ylab("") + theme(legend.position = "none")
```



Part 2 : A Gaussian model of reading

We are given:

- the data : $y = \{y_1, y_2, \ldots, y_8\} = \{300, 270, 390, 450, 500, 290, 680, 450\}$
- the likelihood assumptions : y_i ~ Normal(μ, σ).
- the joint likelihood : $\mathcal{L}(\mu, \sigma | y) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2}$
- the prior assumptions : $\sigma = 50$ and μ ~ $Normal(250, 25)$

2.1

To calculate unnormalized posterior density, we use $p'(\mu|\sigma, y) = L(\mu, \sigma|y)p(\mu)$

```
y <- c(300, 270, 390, 450, 500, 290, 680, 450)
unnorm_post <- function(mu) {
  log_lkl<- sum(dnorm(y, mean = mu, sd = 50, log = TRUE))
  joint_lkl <- exp(log_lkl)
  prior <- dnorm(mu, mean=250, sd=25)
  return (joint_lkl * prior)
}
```

Using the above code to get the unnormalised posterior -

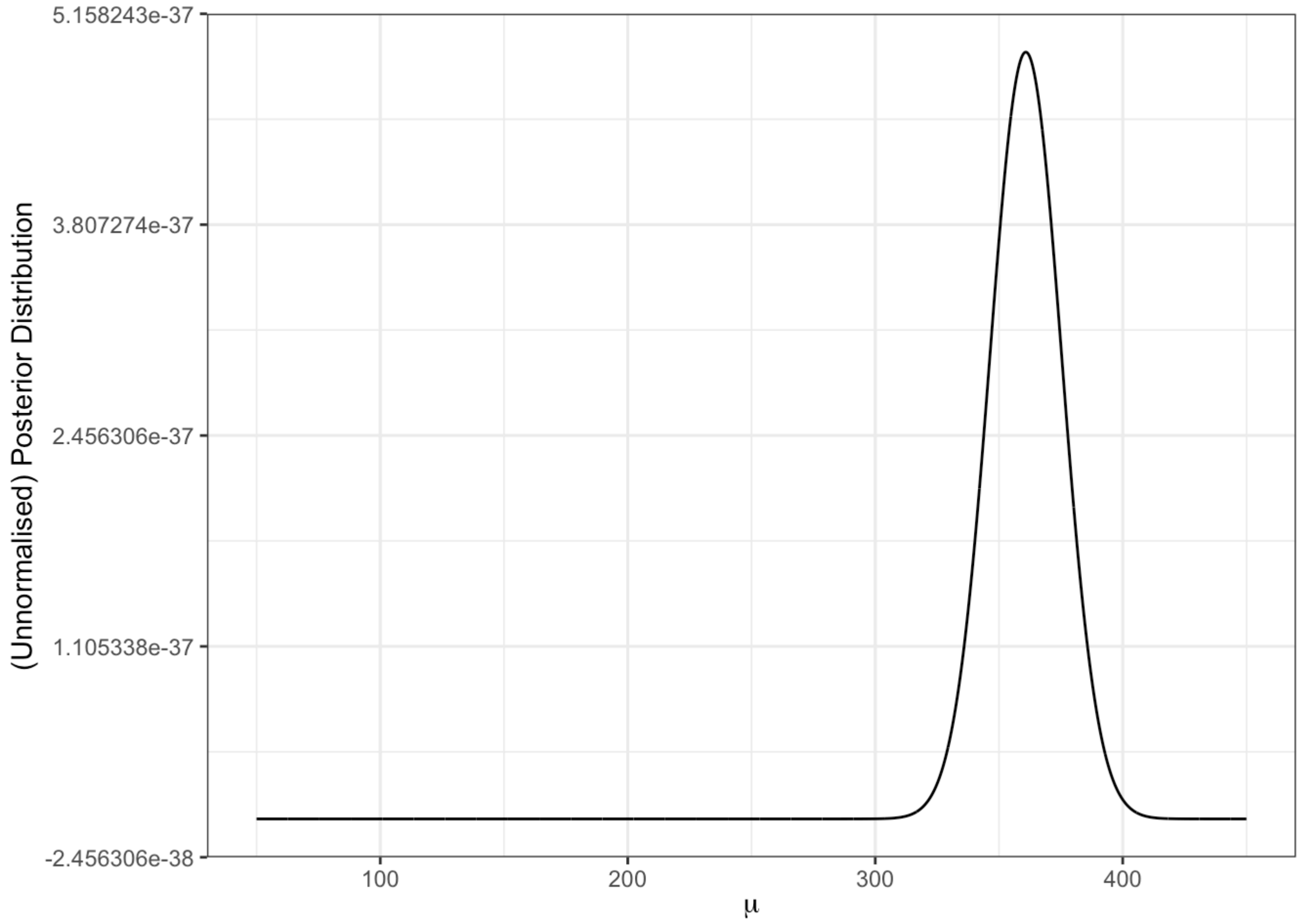
a. (unnormalised) posterior density for $\mu = 300$ is 6.82e-41
b. (unnormalised) posterior density for $\mu = 900$ is 0.00e+00

c. (un-normalised) posterior density for $\mu = 50$ is 9.69e-138

2.2

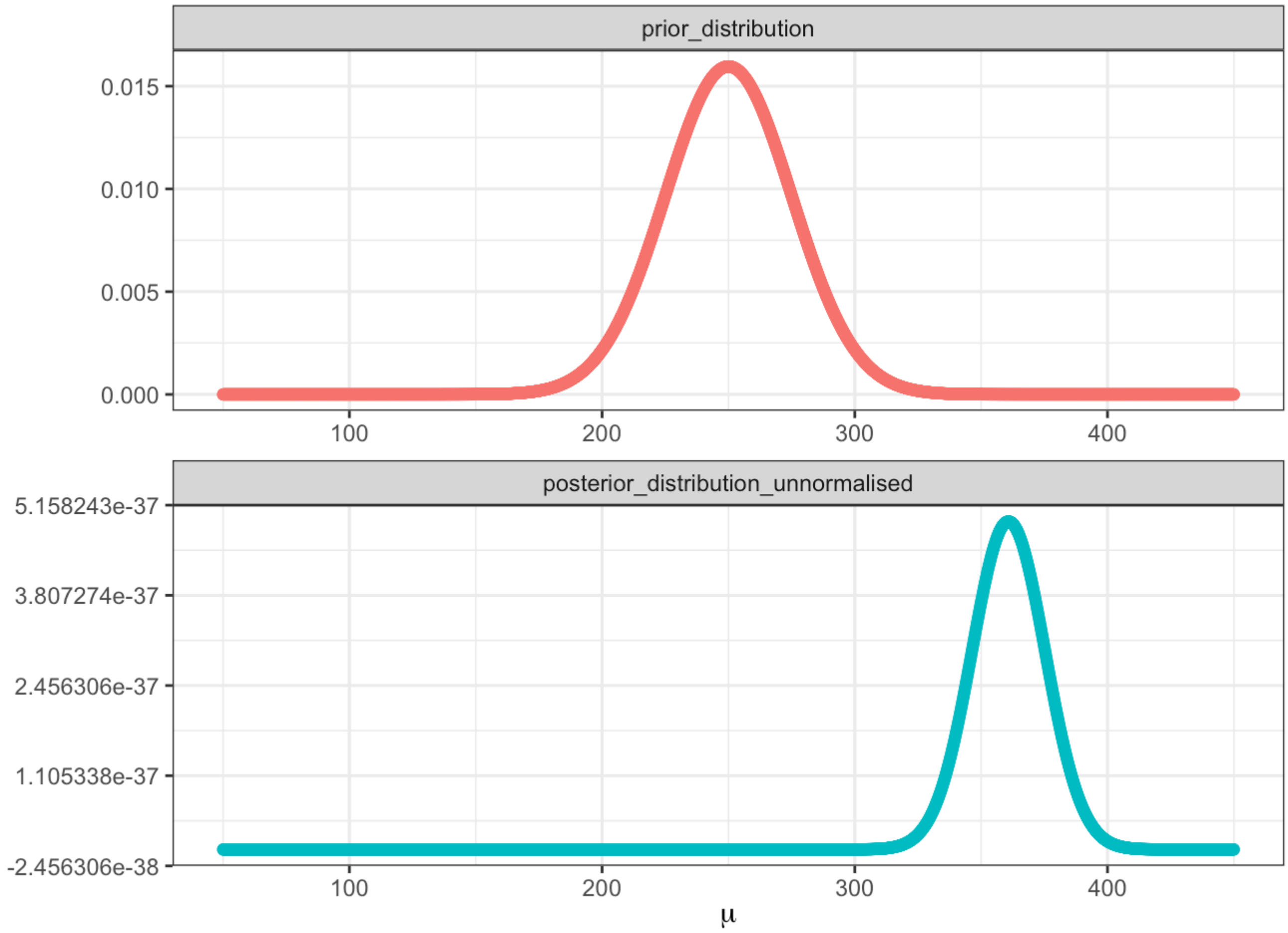
Plotting the Unnormalised Posterior Density of μ, i.e. $p'(\mu|\sigma, y)$ -

```
likelihoods <- data.frame(mu = seq(from=50, to=450, by=0.1))
likelihoods$prior_distribution <- dnorm(likelihoods$mu, mean=250, sd=25)
likelihoods$posterior_distribution_unnormalised <- NA
for (i in 1:nrow(likelihoods))
{
  likelihoods$posterior_distribution_unnormalised[i] <- unnorm_post(likelihoods$mu
[i])
}

ggplot(likelihoods, aes(x=mu, y=posterior_distribution_unnormalised)) +
  geom_line() + theme_bw() + xlab(expression(mu)) + ylab("(Unnormalised) Posterior Di
stribution")
```



2.3

```
likelihoods.m <- melt(likelihoods, id = c("mu"))
ggplot(likelihoods.m,aes(x=mu,y=value,group=variable,color=variable))+ geom_point()+f
acet_wrap(~variable,scales="free",nrow=3) + theme_bw() + xlab(expression(mu)) + ylab
("") + theme(legend.position = "none")
```



Part 3 - The Bayesian learning

We are given:

- the data (number of accidents) : $k = \{k_1, k_2, k_3, k_4\} = \{25, 20, 23, 27\}$
- the prior for day 1 : λ_1 ~ $Gamma(40, 2)$.
- The posterior for day i λ_i ~ $Gamma(\alpha_i, \beta_i)$ is posterior distribution of day i+1 : λ ~ $Gamma(\alpha_i + k_i, \beta_i + 1)$ which is also the prior for the next day
- The likelihood function (Poisson Distribution) : $\mathcal{L}(\lambda|k) = \frac{\lambda^k e^{-\lambda}}{k!}$

Using the recurrence relation between the data we can find out the posterior for day 5

```r
k <- c(25, 20, 23, 27)

accident.df <- data.frame(k=k)
accident.df$alpha_prior <- NA
accident.df$alpha_posterior <- NA
accident.df$beta_prior <- NA
accident.df$beta_posterior <- NA

accident.df$alpha_prior[1] <- 40
accident.df$beta_prior[1] <- 2

for (i in 1:length(k)) {
  if (i == 1) {
    accident.df$alpha_posterior[i] <- accident.df$alpha_prior[i] + accident.df$k[i]
    accident.df$beta_posterior[i] <- accident.df$beta_prior[i] + 1
  } else {
    accident.df$alpha_prior[i] <- accident.df$alpha_posterior[i - 1]
    accident.df$beta_prior[i] <- accident.df$beta_posterior[i - 1]
    accident.df$alpha_posterior[i] <- accident.df$alpha_prior[i] + accident.df$k[i]
    accident.df$beta_posterior[i] <- accident.df$beta_prior[i] + 1
  }
}

alpha5 <- accident.df$alpha_posterior[length(k)]
beta5 <- accident.df$beta_posterior[length(k)]
```

So, the prior distribution on the 5th day is λ ~ $Gamma$(135, 6)

Prediction of the model would be the expected value of the random variable k, which is λ itself as it follows the poisson distribution.

To get the value of λ we can use its expected value to predict for that day.

So predicted number of accidents on day 5 would be $E[E[k]]$ = $E[\lambda]$ = $\frac{\alpha}{\beta}$ = 135/ 6 = **22.5**

So predicted value would be 22 or 23 as it cannot be fractional.

Part 4 - Model building in the Bayesian framework

4.1

The research problem - Is the mean recognition time for the non-words larger than the mean recognition time for the words?

4.2

Null hypothesis - The mean recognition time for the words is equal to the mean recognition time for the non-words.

Lexical-access hypothesis - The mean recognition time for the words is longer than the mean recognition time for the non-words.

4.3 -

Null Hypothesis Model

- Likelihood Assumptions: T_w ~ $Normal(\mu, \sigma)$ and T_{nw} ~ $Normal(\mu + \delta, \sigma)$
- Prior Assumptions : μ ~ $Normal(300, 50)$, $\sigma = 60$ and $\delta = 0$

Lexical-access Model

- Likelihood Assumptions: T_w ~ $Normal(\mu, \sigma)$ and T_{nw} ~ $Normal(\mu + \delta, \sigma)$
- Prior Assumptions : μ ~ $Normal(300, 50)$, $\sigma = 60$ and δ ~ $Normal_+(0, 50)$

4.4

The data -

```
# recognition.csv from GitHub
dat <- read.table(
  "https://raw.githubusercontent.com/yadavhimanshu059/CGS698C/main/notes/Module-2/rec
ognition.csv",
  sep=",",header = T)[,-1]

dat$Tw = as.numeric(dat$Tw)
dat$Tnw = as.numeric(dat$Tnw)

flextable(head(dat))
```

Tw	Tnw
285.0780	296.8060
267.5184	280.1157
289.9203	310.4417
399.0674	324.8276
359.9884	373.8152
403.3993	269.8220

4.5

4.5.1

The unnormalised posterior distribution of μ will be calculated using the following formula -
$p'(\mu, \delta | T_w, T_{nw}) = \mathcal{L}(\mu, \delta, \sigma | T_w)\mathcal{L}(\mu, \delta, \sigma | T_{nw})p(\mu)p(\delta)$

Note that in the case of Null Hypothesis, δ is also fixed at $\delta = 0$. Therefore, $p(\delta)$ is also fixed as
$p(\delta) = \begin{cases} 1 & \text{if } \delta = 0 \\ 0 & \text{otherwise} \end{cases}$

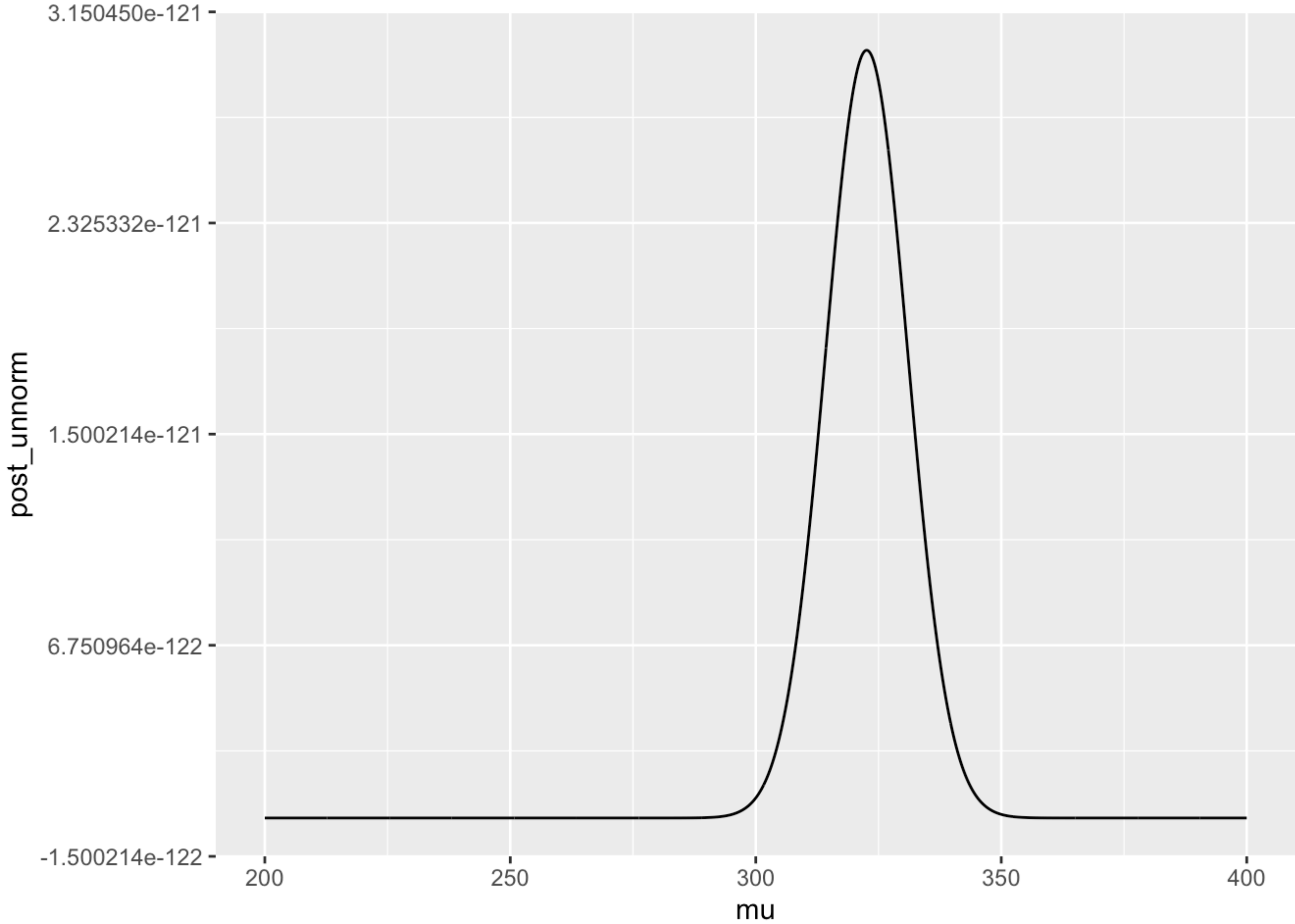
So the formula boils down to $p'(\mu | \delta, T_w, T_{nw}) = \mathcal{L}(\mu | \delta, \sigma, T_w)\mathcal{L}(\mu | \delta, \sigma, T_{nw})p(\mu)$

We already know that $\sigma = 60$, μ ~ $Normal(300, 50)$, T_w ~ $Normal(\mu, \sigma)$ and T_{nw} ~ $Normal(\mu + \delta, \sigma)$

```r
# Using the following formulae to get the unnormalised posterior
log_lkl_tw <- sum(dnorm(as.numeric(dat$Tw), mean=mu, sd=60, log=TRUE))
log_lkl_tnw <- sum(dnorm(as.numeric(dat$Tnw), mean=mu, sd=60, log=TRUE))
log_prior_mu <- dnorm(mu, mean=300, sd=50, log=TRUE)
unnorm_post <- (exp(log_lkl_tnw+log_lkl_tw+log_prior_mu))
```

Plotting the Unnormalised Posterior Distribution

```r
null.lkls <- data.frame(mu=seq(from=200, to=400, by=0.1))

null.lkls$post_unnorm <- NA
null.lkls$log_lkl_tw <- NA
null.lkls$log_lkl_tnw <- NA
null.lkls$log_prior_mu <- NA

for (i in 1:nrow(null.lkls))
{

  null.lkls$log_lkl_tw <- exp(sum(dnorm(dat$Tw, mean=null.lkls$mu[i], sd=60, log=TRU
E)))
  null.lkls$log_lkl_tnw <- sum(dnorm(dat$Tnw, mean=null.lkls$mu[i], sd=60, log=TRUE))
  null.lkls$log_prior_mu <- dnorm(null.lkls$mu[i], mean=300, sd=50, log=TRUE)

  null.lkls$post_unnorm[i] <- (exp(
    null.lkls$log_prior_mu[i] +
    null.lkls$log_lkl_tnw[i] +
    null.lkls$log_lkl_tw[i]
    ))
}

ggplot(null.lkls, aes(x=mu, y=post_unnorm)) + geom_line()
```


3.150450e-121
2.325332e-121
1.500214e-121
6.750964e-122
-1.500214e-122
post_unnorm
200
250
300
350
400
mu