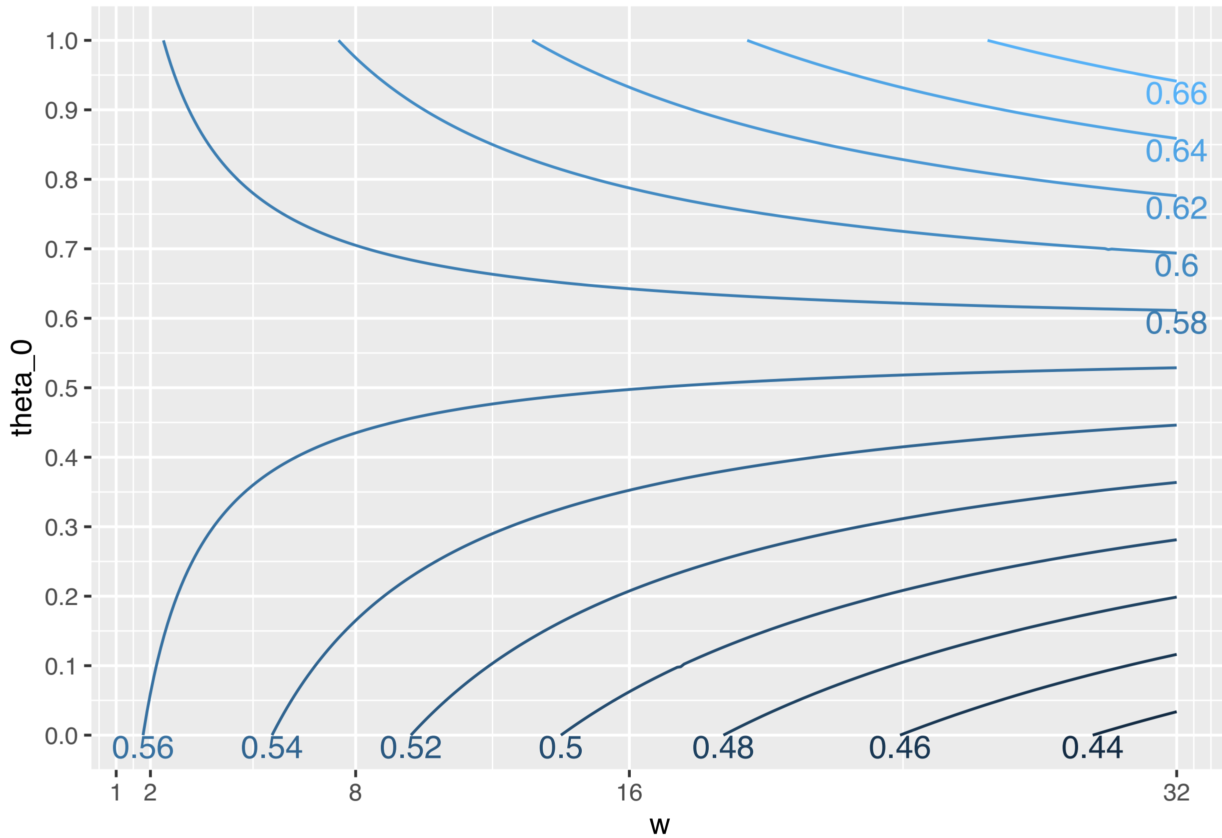
2019R1 Applied Bayesian Methods (STAT6106) Assignment 2

Yiu Chung WONG 1155017920

```r
set.seed(6106);
```

3.2.

```r
theta_0 <- seq(0, 1, by = 0.1);
w <- seq(0, 32, by = 0.1);

posterior_expectation <- function(w, n, priors, y)
{
  (w/ (w + n)) * (priors) + (n / (w+n)) * (y/n);
}

post <- outer(theta_0, w, FUN = function(t_0, w_0, n = 100, y = 57)
  {
  posterior_expectation(w_0, n, t_0, y);
  }
)

rownames(post) <- theta_0;
colnames(post) <- w;
df = reshape2::melt(post);

colnames(df) = c('theta_0', 'w', 'post_theta')

contour_plot <- ggplot(df, aes(x = w, y = theta_0, z = post_theta)) +
  stat_contour(aes(colour = ..level..)) +
  scale_color_continuous(name = '¼ Mi. Time (s)') +
  scale_x_continuous(breaks = c(1, 2, 8, 16, 32), labels = c(1, 2, 8, 16, 32)) +
  scale_y_continuous(breaks = theta_0);
directlabels::direct.label(contour_plot, 'bottom.pieces')
```



Given n = 100 and success = 57, a sigificant portion of teh contour plot consist of posterior probabilities between 0.5 and 0.66; only a small portion of the plot assigns probabilities lower than 0.5 when both w and priod are at extreme ends. Therefore, one can be confident to say that the theta is bigger or equal than 0.5.

3.3a.

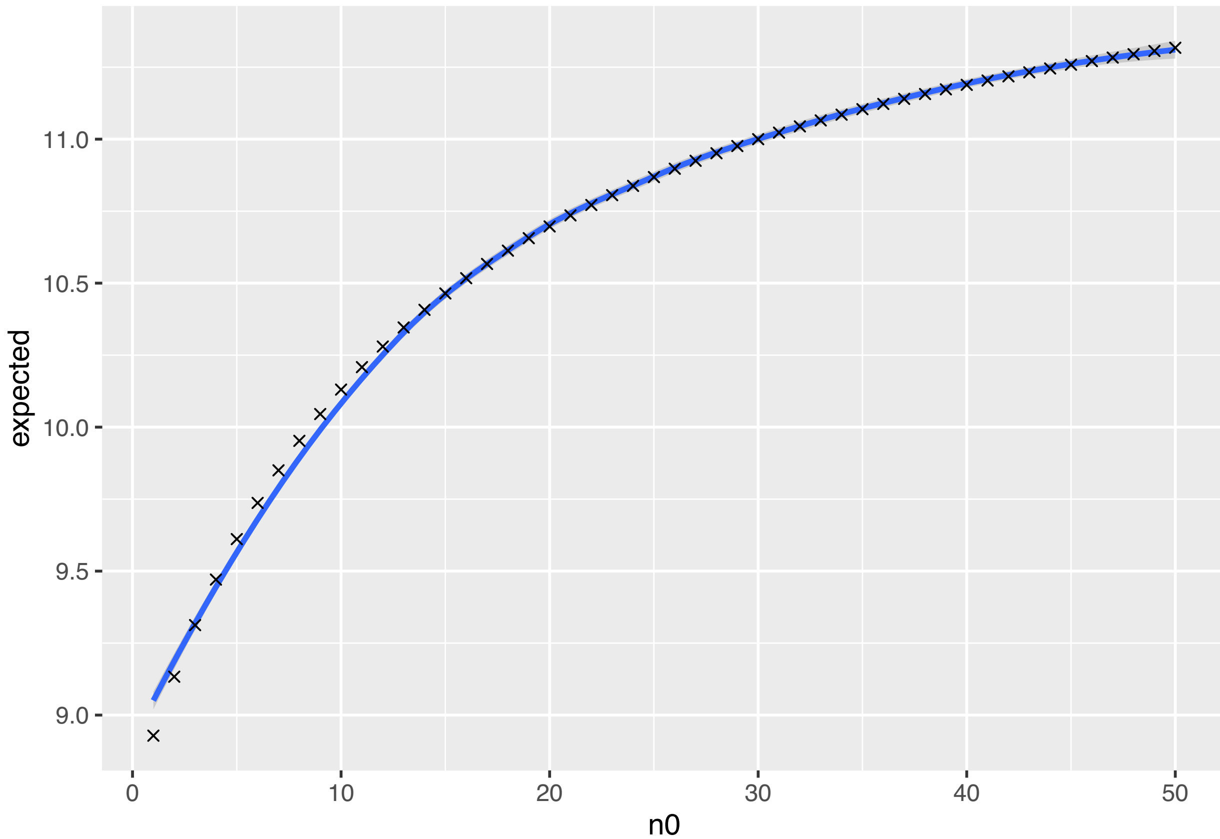
```
ya <- c(12, 9, 12, 14, 13, 13, 15, 8, 15, 6);
yb <- c(11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7);

ya_mean <- (120 + sum(ya)) / (10 + length(ya));
ya_var <- (120 + sum(ya)) / (10 + length(ya))^2;
yb_mean <- (12 + sum(yb)) / (1 + length(yb));
yb_var <- (12 + sum(yb)) / (1 + length(yb))^2;

ya_CI <- qgamma(c(.025, .975), 120 + sum(ya), 10 + length(ya));
yb_CI <- qgamma(c(.025, .975), 12 + sum(yb), 1 + length(yb));
```

posterior mean and variance of y_a are 11.85and 0.5925 posterior mean and variance of y_b are 8.9285714 and 0.6377551

95% quantile- based confidence intervals for θ_a is 10.3892382, 13.4054483 95% quantile- based confidence intervals for θ_b is 7.4320642, 10.5603081

3.3b.

```
n0 <- 1:50;
expected_theta <- (12 * n0 + sum(yb)) / (n0 + length(yb));
ggplot(data = data.frame(expected = expected_theta, n0 = n0), aes(x = n0, y = expected)) + geom_smooth(
```



MLE of θ_a is 11.85. For posterior expectation of θ_b to be close to this number, a prior with large n_0 is required.

3.3c.

- The fact that the two types of mouse are related should be incorpoated into the priod of θ_b.

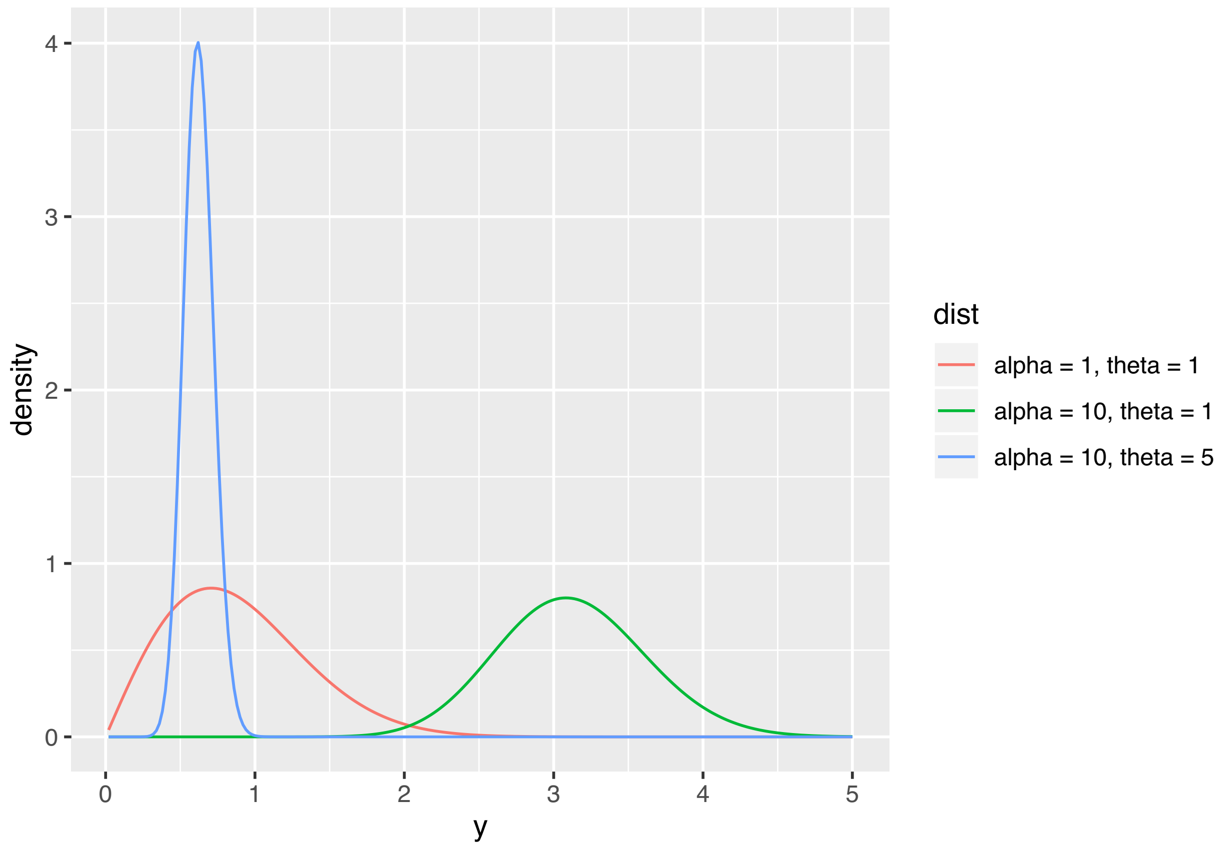
3.9

$$p(\theta \mid n_0, t_0) \propto \text{dgalenshore}\left(\theta, an_0 + 1, \sqrt{-n_0 t_0}\right)$$

a.

```
dgalenshore = function(y, a, theta) {
  (2 / gamma(a)) * theta^(2 * a) * y^(2 * a - 1) * exp(-1 * (theta^2) * y^2)
}
y = seq(0.02, 5, by = 0.02)
df = rbind(
  data.frame(y = y, density = dgalenshore(y, 1, 1), dist = 'alpha = 1, theta = 1'),
  data.frame(y = y, density = dgalenshore(y, 10, 1), dist = 'alpha = 10, theta = 1'),
```

```r
  data.frame(y = y, density = dgalenshore(y, 10, 5), dist = 'alpha = 10, theta = 5')
)

ggplot(df, aes(x = y, y = density, group = dist, color = dist)) +
  geom_line()
```



b.

$$\text{Galenshore}\left(a(n_0+n)+1, \sqrt{-(n_0+n)(n_0t_0+n\bar{t}(\mathbf{y}))}\right)$$

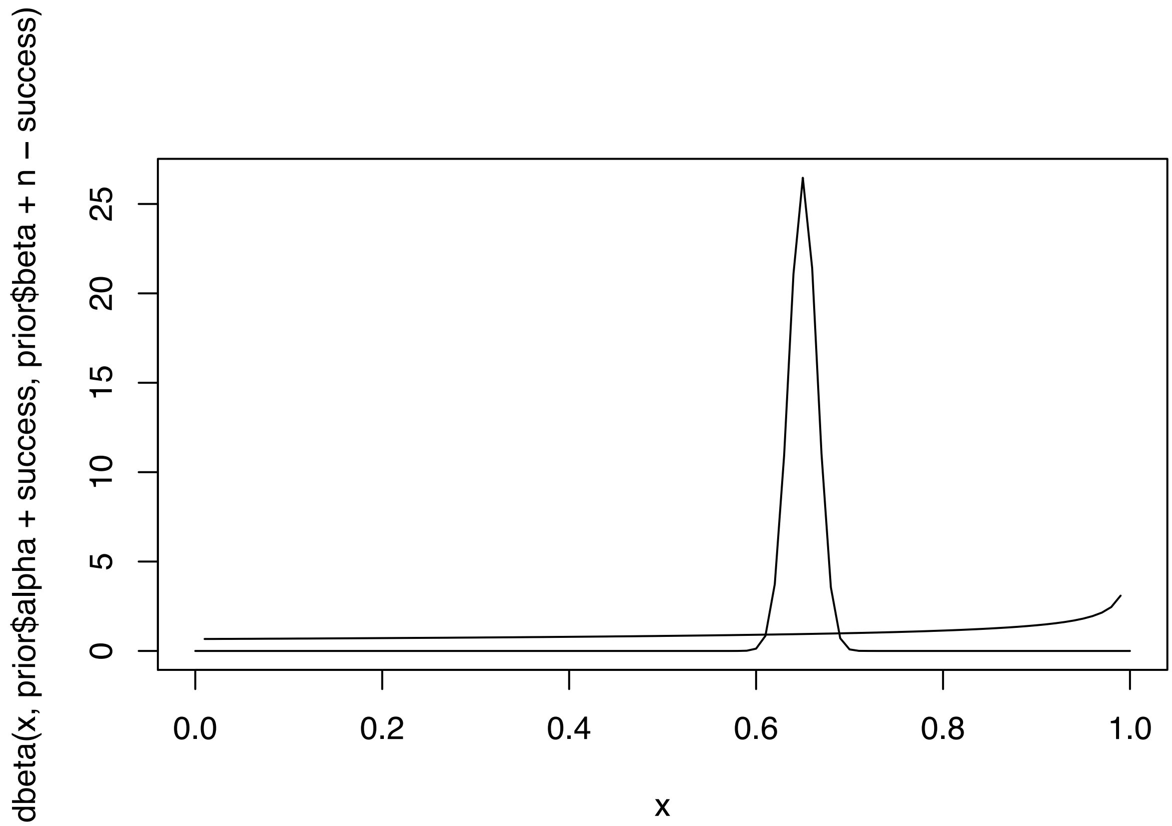
d.

$$\mathbb{E}(\theta \mid y_1, \ldots, y_n) = \frac{\Gamma\left(\frac{1}{2}a(n_0+n)+2\right)}{\sqrt{-(n_0+n)(n_0t_0+n\bar{t}(\mathbf{y}))}\Gamma\left(a(n_0+n)+1\right)}$$

1.

```r
estBetaParams <- function(mu, var) {
  alpha <- ((1 - mu) / var - 1 / mu) * mu ^ 2;
  beta <- alpha * (1 / mu - 1);
  return(params = list(alpha = alpha, beta = beta));
```

```
}
prior <- estBetaParams(.6, .3^2);

n <- 1000;
success <- .65 * n;
posterior <- list(alpha = prior$alpha + success, beta = prior$beta + n);

curve(dbeta(x, prior$alpha + success, prior$beta + n - success));
curve(dbeta(x, prior$alpha, prior$beta), add = TRUE );
```



* Prior α is 1 and prior beta is 0.6666667.

2.

```
Pr_C1_Prior <- 0.5;
Pr_C2_Prior <- 0.5;

Pr_C1_success <- 0.6;
Pr_C2_success <- 0.4;

Pr_C1_Post <- (Pr_C1_Prior * (1 - Pr_C1_success)^2) /
  (Pr_C1_Prior * (1 - Pr_C1_success)^2 + Pr_C2_Prior * (1 - Pr_C2_success)^2 );

Pr_C2_Post <- (Pr_C2_Prior * (1 - Pr_C2_success)^2) /
  (Pr_C1_Prior * (1 - Pr_C1_success)^2 + Pr_C2_Prior * (1 - Pr_C2_success)^2 );
```

```
C1_expected_count <- 1/(Pr_C1_success);
C2_expected_count <- 1/(Pr_C2_success);

total_expected_count <- Pr_C1_Post * C1_expected_count + Pr_C2_Post * C2_expected_count
```

- Expected count until first head is 2.2435897 spins.

3.1

$$\mathcal{N}\left(\frac{\frac{180}{40^2}+\frac{150n}{20^2}}{\frac{1}{40^2}+\frac{n}{20^2}}, \frac{1}{\frac{1}{40^2}+\frac{n}{20^2}}\right)$$

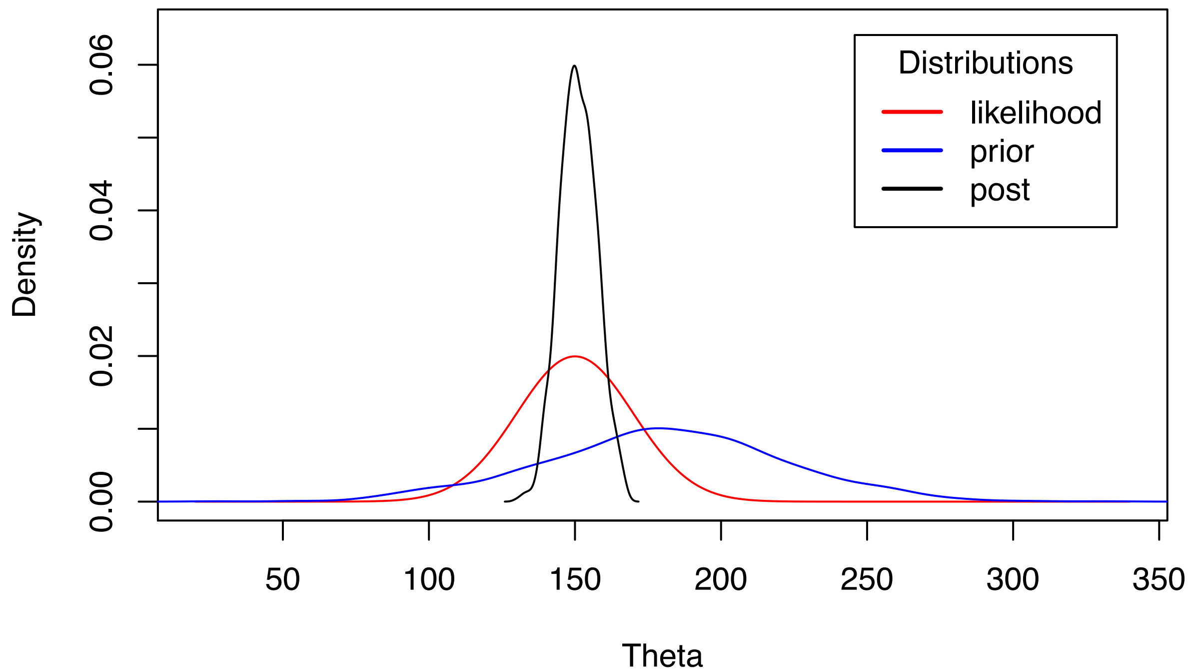
3.2

```
n <- 10;
theta_post <- ((180/(40^2)) + ((150*n)/(20^2))) / ((1/40^2)+(n/20^2));
var <- 1 / ((1/40^2)+(n/20^2));

CI <- qnorm(c(.025, .975), theta_post, sqrt(var));
```

95% posterior interval is between 138.4879094 and 162.9755053

3.3

```
thetas <- seq(180 - 4*40, 180 + 4*40, length.out = 1000);

prior <- rnorm(1000, 180, 40);
likelihood <- dnorm(150, mean = thetas, sd = 20);
post <- rnorm(1000, theta_post, sqrt(var));

colors <- c("red", "blue", "black");
labels <- c("likelihood", "prior", "post")

#plot(density(post), xlim = c(min(thetas), max(thetas)), ylim = c(0, 0.08), type = 'l', col = colors[1]
plot(likelihood, x = thetas, type = 'l', col = colors[1], xlab = "Theta", ylab = "Density", xlim = c(mi
lines(density(prior), col = colors[2]);
lines(density(post), col = colors[3]);

legend("topright", inset=.05, title="Distributions",
  labels, lwd=2, lty=c(1, 1, 1, 1, 2), col=colors)
```



4a.

$$\mathcal{L}(p_i \mid 100, y_i) = \binom{100}{y_i} \prod_{i=1}^{6} p_i^{y_i}$$

4b.

- Dirichlet distribution

$$f\left(x_1, \ldots, x_K; \alpha_1, \ldots, \alpha_K\right) = \frac{1}{\mathrm{B}(\boldsymbol{\alpha})} \prod_{i=1}^{K} x_i^{\alpha_i - 1}$$

where

$$\mathrm{B}(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}, \qquad \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K).$$

4c.

- In binomial case people tend to use Beta(1,1) as prior. Hence we can use Dirichlet(1,1,1,1,1,1) as prior.

4d.

$$\text{Dirichlet}(1+10, 1+10, 1+10, 1+20, 1+10, 1+40)$$