Large-Sample Normal Approximations to Posterior

Introduction

We previously considered Bayesian inference for the proportion of M&M's that are blue based on samples of size $n = 1$, $n = 10$, $n = 20$, and $n = 541$.

Our model was $X \sim \text{Binomial}(n, \theta)$

We considered a non-informative prior of $\Theta \sim \text{Beta}(1, 1)$.

In that case, the posterior is $\Theta | X = x \sim \text{Beta}(1 + x, 1 + n - x)$.

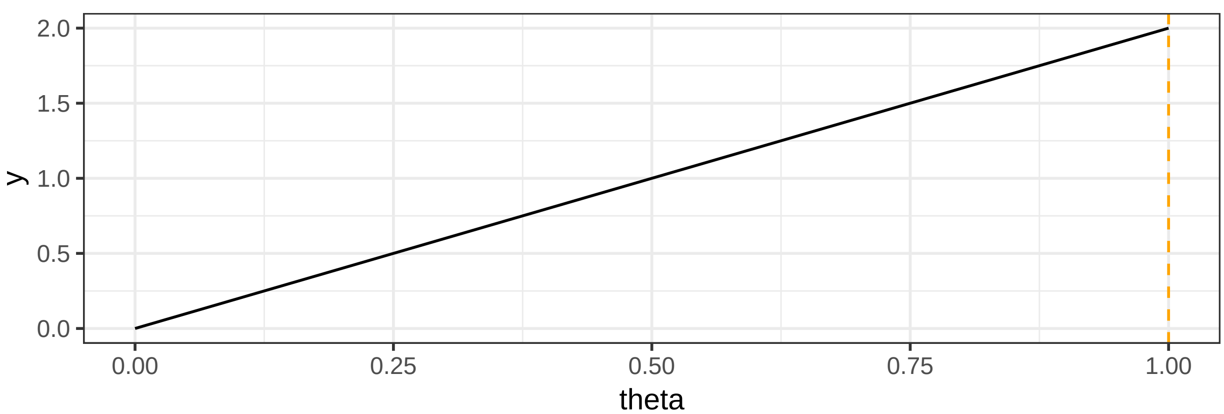
For large n, a normal approximation to the posterior is $\Theta | X = x \overset{\text{approx.}}{\sim} \text{Normal}\left(\frac{x}{n}, \left(\frac{n^2}{x} + \frac{n^2}{n-x}\right)^{-1}\right)$

n = 1 (I had x = 1 blue M&M in my sample)

Can't form the normal approximation: $n - x = 1 - 1 = 0$, so the approximation to the posterior variance is $\frac{n^2}{x} + \frac{n^2}{n-x} = 1 + \frac{1}{0} = \infty$?

```
x <- 1
n <- 1
a_posterior <- 1 + x
b_posterior <- 1 + n - x

ggplot(data = data.frame(theta = c(0, 1)), mapping = aes(x = theta)) +
  stat_function(fun = dbeta,
    args = list(shape1 = a_posterior, shape2 = b_posterior)) +
  geom_vline(xintercept = x/n, color = "orange", linetype = 2) +
  theme_bw()
```



Posterior mean and 95% posterior credible interval based on the exact Beta posterior:

```
a_posterior/(a_posterior + b_posterior)
```

```
## [1] 0.6666667
```

```
qbeta(c(0.025, 0.975), shape1 = a_posterior, shape2 = b_posterior)
```

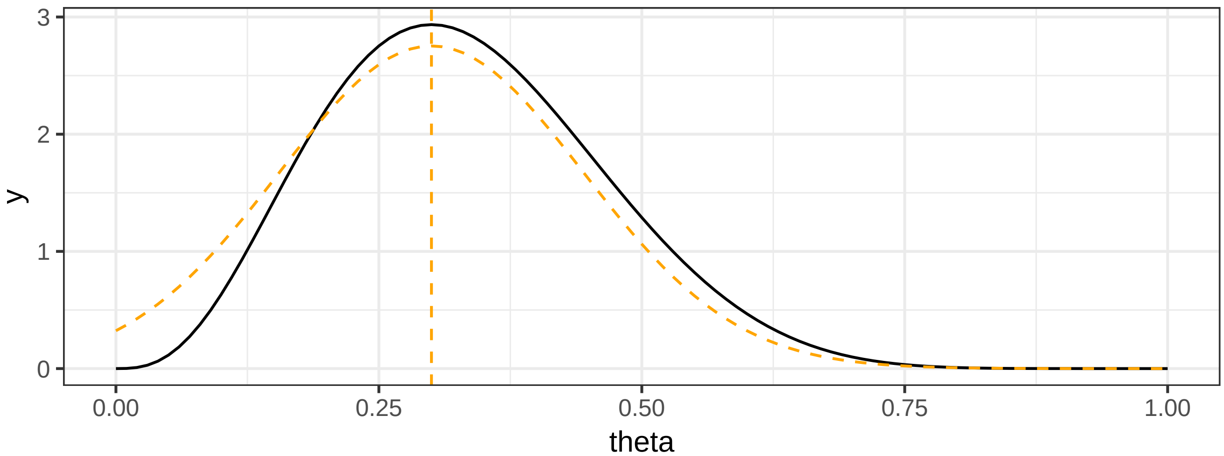
```
## [1] 0.1581139 0.9874209
```

Can't get anything out of the normal approximation to the posterior

n = 10 (I had x = 3 blue M&Ms in my sample)

```
x <- 3
n <- 10
a_posterior <- 1 + x
b_posterior <- 1 + n - x

ggplot(data = data.frame(theta = c(0, 1)), mapping = aes(x = theta)) +
  stat_function(fun = dbeta,
    args = list(shape1 = a_posterior, shape2 = b_posterior)) +
  stat_function(fun = dnorm,
    args = list(mean = x/n, sd = sqrt(1/(n^2 / x + n^2 / (n - x)))),
    color = "orange",
    linetype = 2) +
  geom_vline(xintercept = x/n, color = "orange", linetype = 2) +
  theme_bw()
```



Posterior mean and 95% posterior credible interval based on the exact Beta posterior:

```
a_posterior/(a_posterior + b_posterior)
```

```
## [1] 0.3333333
```

```
qbeta(c(0.025, 0.975), shape1 = a_posterior, shape2 = b_posterior)
```

```
## [1] 0.1092634 0.6097426
```

Approximate posterior mean and 95% posterior credible interval based on the approximate normal posterior:

```
x/n
```

```
## [1] 0.3
```

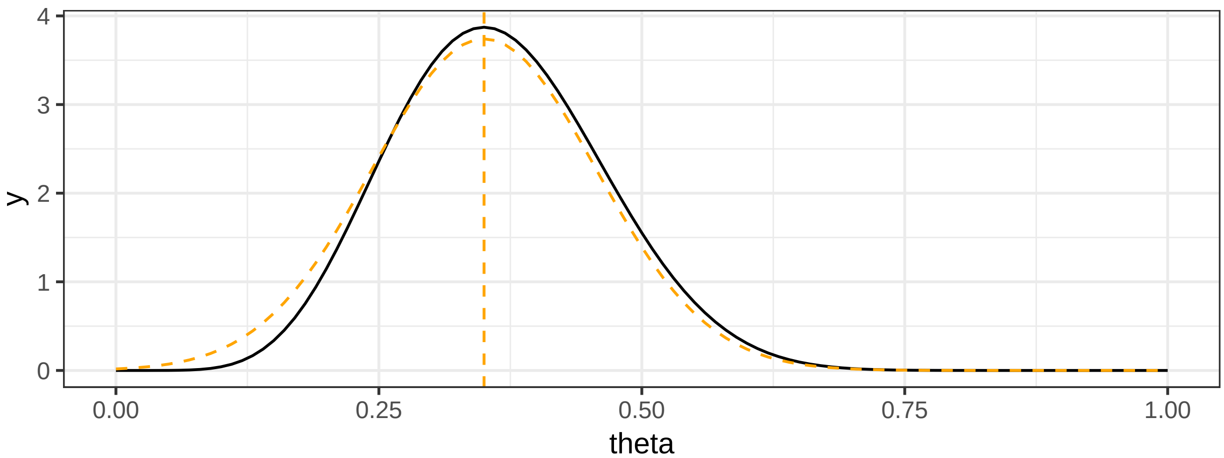
```
qnorm(c(0.025, 0.975), mean = x/n, sd = sqrt(1/(n^2 / x + n^2 / (n - x))))
```

```
## [1] 0.01597423 0.58402577
```

n = 20 (I had x = 7 blue M&Ms in my sample)

```
x <- 7
n <- 20
a_posterior <- 1 + x
b_posterior <- 1 + n - x

ggplot(data = data.frame(theta = c(0, 1)), mapping = aes(x = theta)) +
  stat_function(fun = dbeta,
    args = list(shape1 = a_posterior, shape2 = b_posterior)) +
  stat_function(fun = dnorm,
    args = list(mean = x/n, sd = sqrt(1/(n^2 / x + n^2 / (n - x)))),
    color = "orange",
    linetype = 2) +
  geom_vline(xintercept = x/n, color = "orange", linetype = 2) +
  theme_bw()
```



Posterior mean and 95% posterior credible interval based on the exact Beta posterior:

```
a_posterior/(a_posterior + b_posterior)
```

```
## [1] 0.3636364
```

```
qbeta(c(0.025, 0.975), shape1 = a_posterior, shape2 = b_posterior)
```

```
## [1] 0.1810716 0.5696755
```

Approximate posterior mean and 95% posterior credible interval based on the approximate normal posterior:

```
x/n
```

```
## [1] 0.35
```

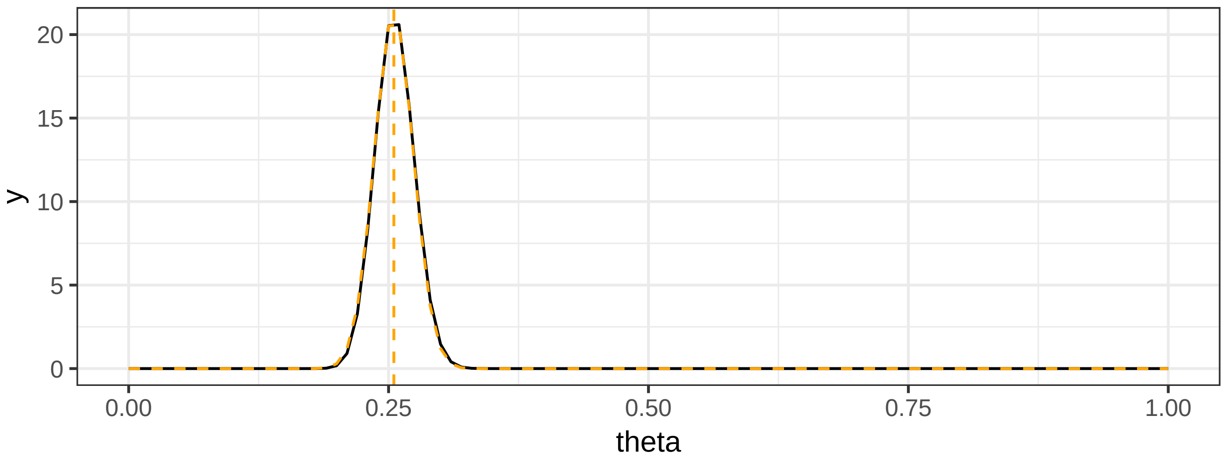
```
qnorm(c(0.025, 0.975), mean = x/n, sd = sqrt(1/(n^2 / x + n^2 / (n - x))))
```

```
## [1] 0.1409627 0.5590373
```

Sample of large size (As a class, we had x = 138 blue M&Ms in a sample of size n = 541)

```
x <- 138
n <-541
a_posterior <- 1 + x
b_posterior <- 1 + n - x

ggplot(data = data.frame(theta = c(0, 1)), mapping = aes(x = theta)) +
  stat_function(fun = dbeta,
    args = list(shape1 = a_posterior, shape2 = b_posterior)) +
  stat_function(fun = dnorm,
    args = list(mean = x/n, sd = sqrt(1/(n^2 / x + n^2 / (n - x)))),
    color = "orange",
    linetype = 2) +
  geom_vline(xintercept = x/n, color = "orange", linetype = 2) +
  theme_bw()
```



Posterior mean and 95% posterior credible interval based on the exact Beta posterior:

```
a_posterior/(a_posterior + b_posterior)
```

```
## [1] 0.2559853
```

```
qbeta(c(0.025, 0.975), shape1 = a_posterior, shape2 = b_posterior)
```

```
## [1] 0.2201851 0.2934879
```

Approximate posterior mean and 95% posterior credible interval based on the approximate normal posterior:

```
x/n
```

```
## [1] 0.2550832
```

```
qnorm(c(0.025, 0.975), mean = x/n, sd = sqrt(1/(n^2 / x + n^2 / (n - x))))
```

```
## [1] 0.2183512 0.2918152
```