Large-Sample Normal Approximation to the Posterior

Basic Result (rough statement)

For large sample size, the posterior distribution is approximately $\Theta|X_1,\dots,X_n \sim \text{Normal}(\hat{\theta}^{MLE}, \cdots)$

Binomial Model: M&M's (Lab 7b)

- As a class, we had x = 138 blue M&Ms in a sample of size n = 541.
- Data Model: $X|\Theta = \theta \sim \text{Binomial}(541, \theta)$
- Suppose we use a noninformative prior of $\Theta \sim \text{Beta}(1,1)$
- The exact posterior is $\Theta|X = 138 \sim \text{Beta}(1 + 138, 1 + 541 - 138)$
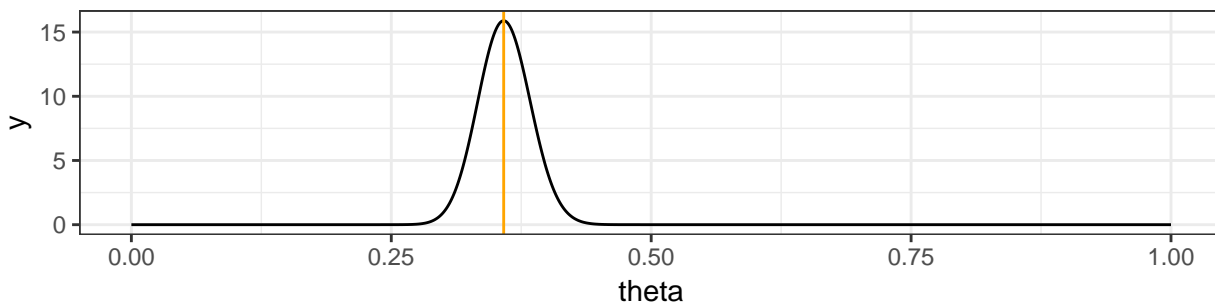- The MLE is $138/541$

```
ggplot(data = data.frame(theta = c(0, 1)), mapping = aes(x = theta)) +
  stat_function(fun = dbeta,
    args = list(shape1 = 1 + 138, shape2 = 1 + 541 - 138),
    n = 1001) +
  geom_vline(xintercept = 138/541, color = "orange", linetype = 2) +
  theme_bw()
```
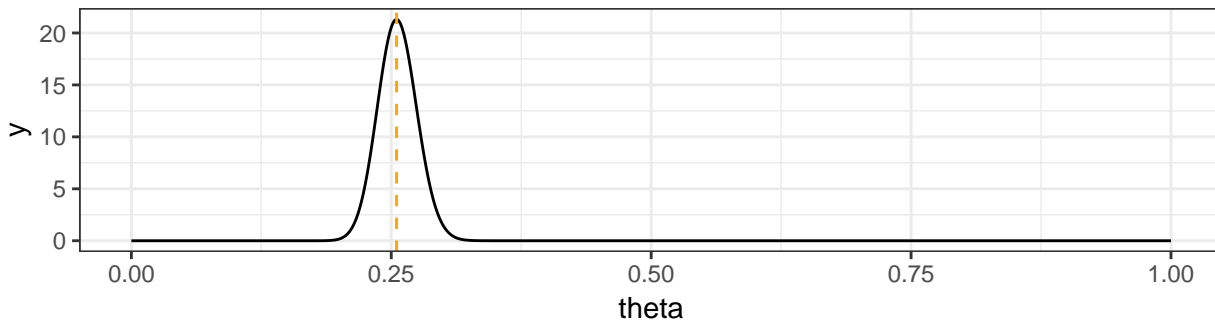

Geometric Model: Bird Hops (Lab 8)

- Observe $X_1, \dots, X_n$; X_i is the number of hops taken before bird takes off
- Data Model: $X_i|\Theta = \theta \overset{\text{i.i.d.}}{\sim} \text{Geometric}(\theta)$
- Suppose we use a noninformative prior of $\Theta \sim \text{Beta}(1,1)$
- The exact posterior is $\Theta|X_1 = x_1, \dots, X_n = x_n \sim \text{Beta}(1 + n, 1 + \sum_{i=1}^{n} x_i)$
- The MLE is $\frac{1}{1+\overline{X}}$

```
ggplot(data = data.frame(theta = c(0, 1)), mapping = aes(x = theta)) +
  stat_function(fun = dbeta,
    args = list(shape1 = 1 + nrow(bird_hops), shape2 = 1 + sum(bird_hops$num_hops)),
    n = 1001) +
  geom_vline(xintercept = 1/(1 + mean(bird_hops$num_hops)), color = "orange") +
  theme_bw()
```

Poisson Model: Seedlings (Lab 8)

- Observe $X_1, \ldots, X_n$; X_i is the number of seedlings in quadrat number i.
- Data Model: $X_i | \Lambda = \lambda \overset{\text{i.i.d.}}{\sim} \text{Poisson}(\lambda)$
- Suppose we use a noninformative prior of $\Lambda \sim \text{Gamma}(1, 0.01)$
- The exact posterior is $\Lambda | X_1 = x_1, \ldots, X_n = x_n \sim \text{Gamma}(1 + \sum_{i=1}^{n} x_i, 0.01 + n)$
- The MLE is $\bar{X}$

```
ggplot(data = data.frame(theta = c(0, 2)), mapping = aes(x = theta)) +
  stat_function(fun = dgamma,
    args = list(shape = 1 + sum(seedlings$new_1993), rate = 0.01 + nrow(seedlings)), n = 1001) +
  geom_vline(xintercept = mean(seedlings$new_1993), color = "orange") +
  theme_bw()
```
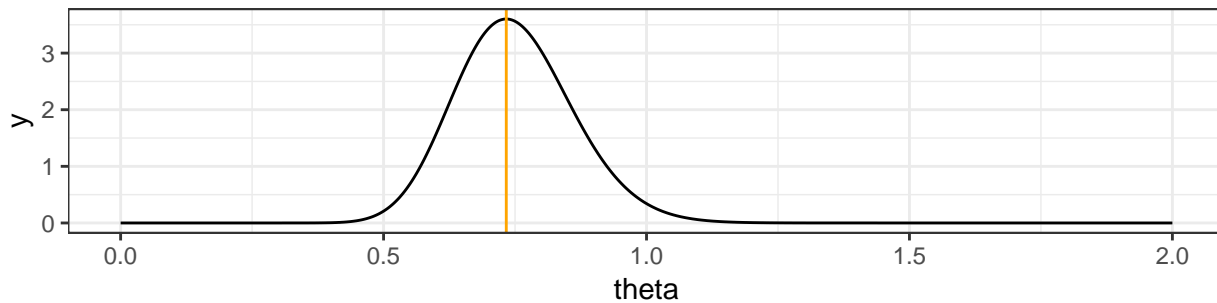

Normal Model: Cosmic Microwave Background Radiation (Lecture, March 4th)

- Observe $X_1, \ldots, X_n$, where X_i is the temperature difference in pixel i.
- Data Model: $X_i \overset{\text{i.i.d.}}{\sim} \text{Normal}(\mu, \sigma^2)$
- We approximated the posterior distributions of μ and σ by drawing samples using MCMC.
- The MLEs are $\hat{\mu}^{MLE} = \bar{X}$ and $\hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^n$

```
ggplot() +
  geom_density(data = theta_posterior_sample, mapping = aes(x = mu)) +
  geom_vline(xintercept = mean(cmb$temp_difference), color = "orange") +
  theme_bw()
```
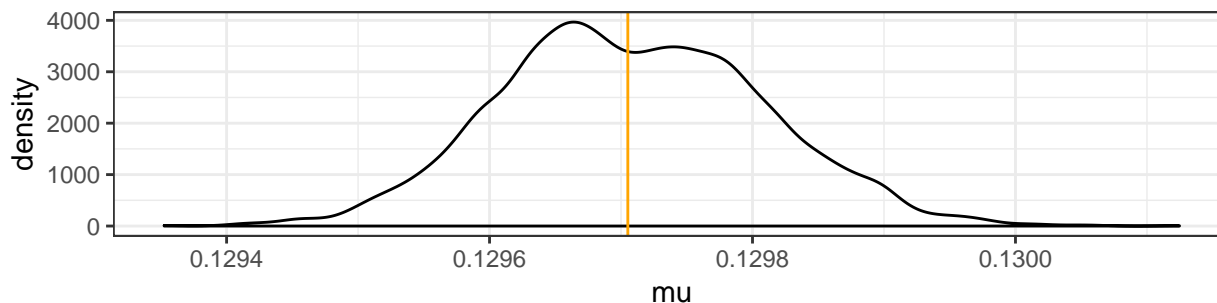


```
ggplot() +
  geom_density(data = theta_posterior_sample, mapping = aes(x = sigma_sq)) +
  geom_vline(xintercept = mean((cmb$temp_difference - mean(cmb$temp_difference))^2), color = "orange") +
  theme_bw()
```