

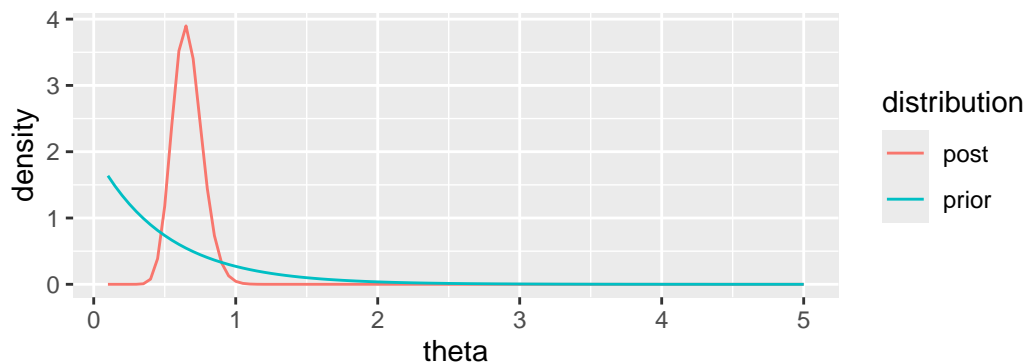
STAT 412: Problem Set 2 (R)

```
library(tidyverse)
library(patchwork)
prussian1 <- read_csv("../handouts/prussianHorses1.csv")
prussian2 <- read_csv("../handouts/prussianHorses2.csv")
```

```
a <- 1
b <- 2
qgamma(0.95, a, b)
```

```
[1] 1.497866
```

```
y <- prussian1$deaths
n <- length(prussian1$deaths)
theta_seq <- seq(0.1, 5, 0.05)
prior <- dgamma(theta_seq, a, b)
a_post <- a + sum(y)
b_post <- b + n
post <- dgamma(theta_seq, a_post, b_post)
bayes_df <- data.frame(theta = theta_seq, prior = prior, post = post) |>
  pivot_longer(cols= 2:3, names_to = "distribution", values_to = "density")
bayes_df |>
  ggplot(aes(x = theta, y = density, col = distribution)) +
  geom_line()
```



```
post_prob <-pgamma(0.5, a_post, b_post)
```

The posterior probability $\Pr(\theta < 0.5 = y) = 0.0487284$. Since this is relatively small, I would say that there is not strong evidence to suggest that the average rate of deaths by horsekick in a given year and cavalry is less than 0.5.

```
set.seed(2)
y_new <- rnbino(220, size = a_post, prob = (b_post)/(b_post + 1))
p1 <- data.frame(y_new) |>
  ggplot(aes(x = y_new)) +
  geom_bar() +
  labs(x = "Deaths", title = "Draws from posterior predictive distribution")

p2 <- prussian2 |>
  ggplot(aes(x = deaths)) +
  geom_bar() +
  labs(x = "Deaths", title = "Held-out observations")

p1 + p2
```

