SCI 2025: Homework 1

Setup

Imagine that you have performed an experiment where $N = 25$ chimpanzees watch a video of an interaction between an "agent" and a "patient". The agent leads the action of the scene, and the patient recieves or is affected by the action. Afterwards, the video is paused and the chimpanzees can touch either the agent or patient side of the screen. At the most basic level, you want to know: "Are chimpanzees more likely to touch the agent or patient side of the screen?".

Load the data into your R session:

```r
N_agent <- 14 # number of chimpanzees that touched the agent side
N_patient <- 11 # number of chimpanzees that touched the patient side
N_trials <- N_agent + N_patient
```

Question 1

Using grid-approximation, compute the posterior distribution of the probability p that the chimpanzees touch the agent side of the screen. The posterior should be represented with samples. Use a uniform prior for p.

```r
p_grid <- seq(0, 1, length.out = 1000)
prior <- rep(1, length(p_grid))
likelihood <- dbinom(N_agent, size = N_trials, prob = p_grid)
posterior <- likelihood * prior
posterior <- posterior / sum(posterior)
posterior_samples <- sample(p_grid, size = 1e4, replace = TRUE, prob = posterior)
```

Question 2

With your posterior samples from Question 1:

(a) Compute and report the 90% credible interval (equal-tailed) for p.

```r
quantile(posterior_samples, probs = c(0.05, 0.95))
```

```
       5%       95% 
0.4014014 0.7087087 
```

(b) Compute and report the 90% highest posterior density interval for p.

```r
rethinking::HPDI(posterior_samples, prob = 0.9)
```

```
     |0.9      0.9| 
0.4024024 0.7087087 
```

(c) Compute and report the posterior mean for p.

```r
mean(posterior_samples)
```

```
[1] 0.5557302
```

(d) Compute and report the posterior probability that p is greater than 0.5.

```r
mean(posterior_samples > 0.5)
```

```
[1] 0.7223
```

Question 3

Repeat all the computations from Questions 1-2, but this time use a non-uniform prior. You are free to choose the prior, but you must justify your choice.

```r
# I'm sure the chimps have some intermediate preference for the agent
prior2 <- ifelse(p_grid < 0.2 | p_grid > 0.8, 0, 1)
likelihood2 <- dbinom(N_agent, size = N_trials, prob = p_grid)
posterior2 <- likelihood2 * prior2
posterior2 <- posterior2 / sum(posterior2)
posterior_samples2 <- sample(p_grid, size = 1e4, replace = TRUE, prob = posterior2)

quantile(posterior_samples2, probs = c(0.05, 0.95))
```

```
       5%       95%
0.4014014 0.7067067
```

```r
rethinking::HPDI(posterior_samples2, prob = 0.9)
```

```
     |0.9      0.9|
0.4134134 0.7167167
```

```r
mean(posterior_samples2)
```

```
[1] 0.5563226
```

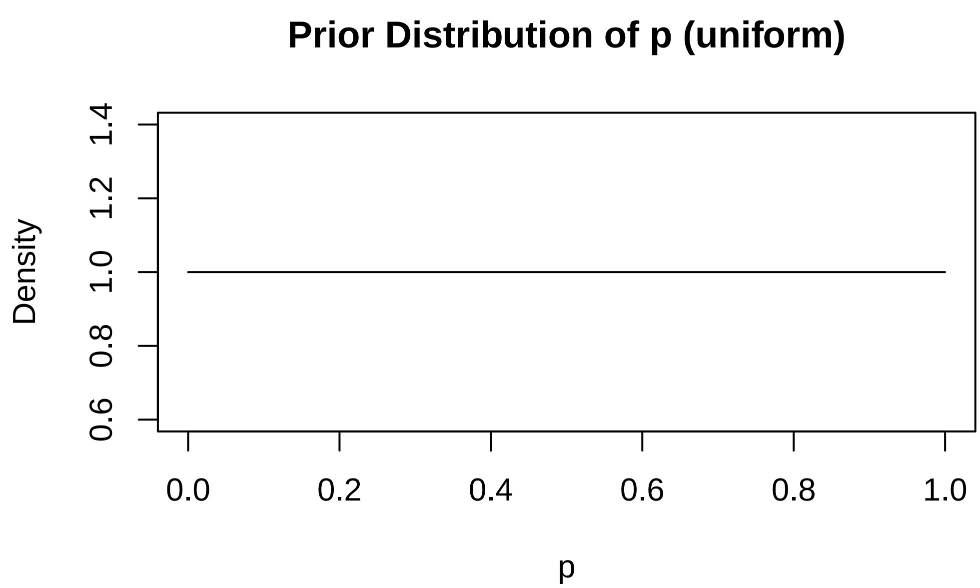
```r
mean(posterior_samples2 > 0.5)
```

```
[1] 0.7264
```

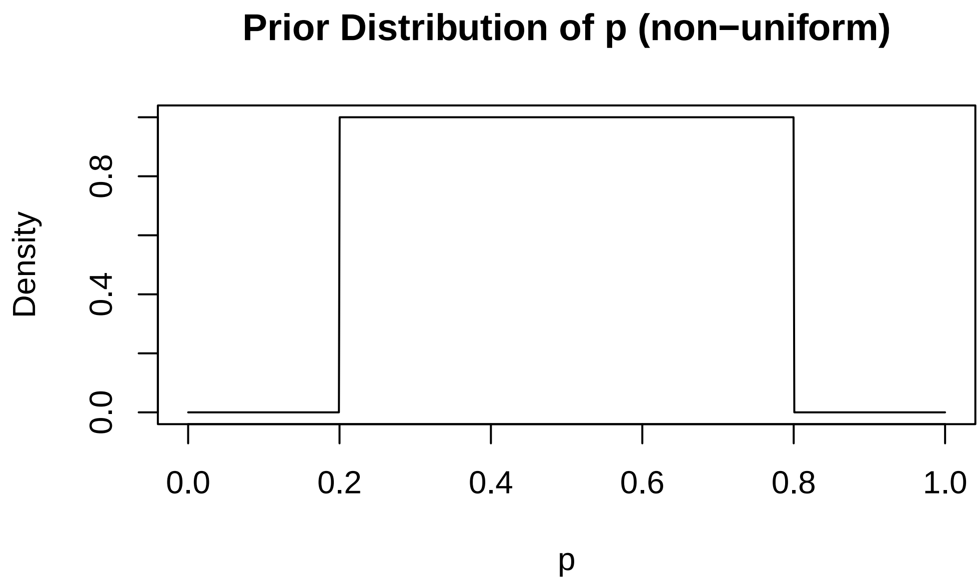
Question 4

Visualize the prior and posterior distributions of p from Questions 1-3.

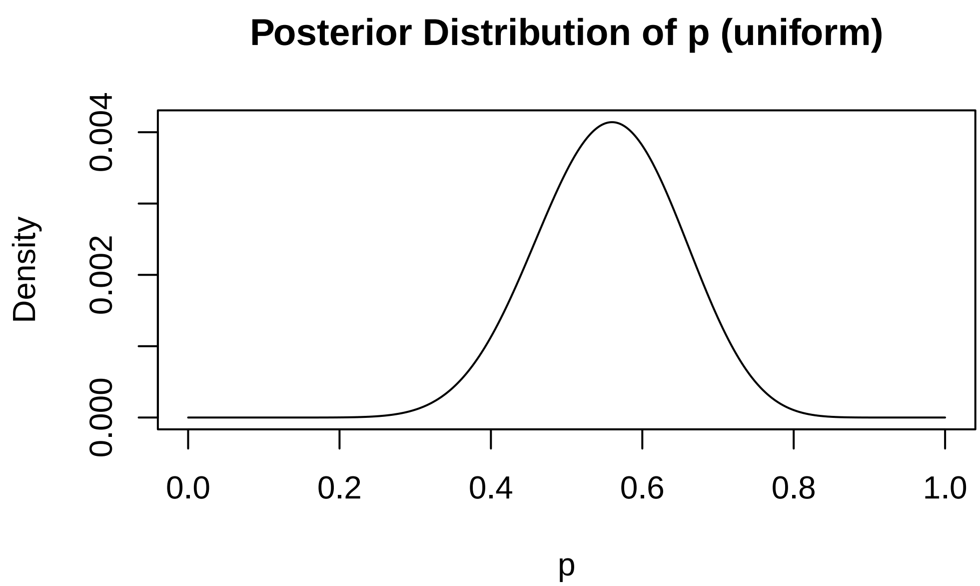
```r
plot(prior ~ p_grid, type = "l", xlab = "p", ylab = "Density",
 main = "Prior Distribution of p (uniform)")
```



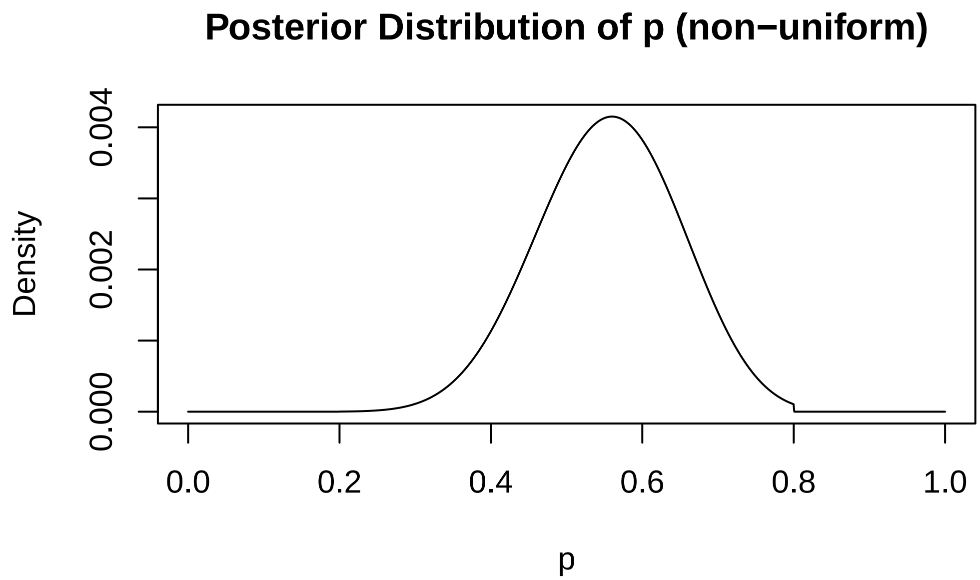

```r
plot(prior2 ~ p_grid, type = "l", xlab = "p", ylab = "Density",
 main = "Prior Distribution of p (non-uniform)")
```




```r
plot(posterior ~ p_grid, type = "l", xlab = "p", ylab = "Density",
 main = "Posterior Distribution of p (uniform)")
```



```r
plot(posterior2 ~ p_grid, type = "l", xlab = "p", ylab = "Density",
 main = "Posterior Distribution of p (non-uniform)")
```

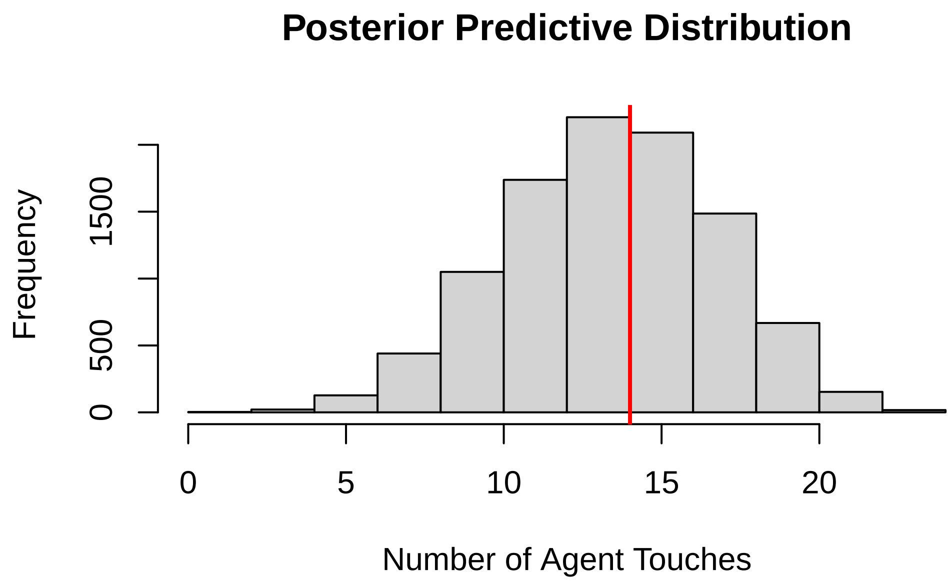


Question 5

Simulate 25 new observations of agent/patient touches, using the posterior distribution of p from Question 3. How many of these simulations are agent vs patient? How closely does this compare to the actual data?

```r
simulated_data <- rbinom(n = length(posterior_samples2),
 size = 25, prob = posterior_samples2)

hist(simulated_data,
 main = "Posterior Predictive Distribution",
  xlab = "Number of Agent Touches")
abline(v = N_agent, col = "red", lwd = 2)
```


Posterior Predictive Distribution
Frequency
0
500
1500
0
5
10
15
20
Number of Agent Touches