STAT 447 Assignment 1

Caden Hewlett

2024-01-11

Question 1 : Sampling from a Joint Distribution

Part 1 (INCOMPLETE)
Compute $\mathbb{E}[(1+Y_1)^X]$ mathematically (with a precise mathematical derivation).

Solution: We wish to compute $\mathbb{E}(g, X, Y_1, \dots Y_4)$. In this case, $g(x, y_1, \dots y_4) = (1+y_1)^x$. We can hence write the expectation as follows:

$$\mathbb{E}[g(X, Y_1, \dots, Y_4)] = \sum_x \sum_{y_1} \sum_{y_2} \sum_{y_3} \sum_{y_4} g(x, y_1, \dots, y_4) p(x, y_1, y_2, y_3, y_4).$$

This can be simplified slightly as follows:

$$\mathbb{E}[(1+Y_1)^X] = \sum_x \sum_{y_1} (1+y_1)^x p(x, y_1, y_2, y_3, y_4)$$

Further, we know the specific y_2, y_3, y_4 coin flip outcomes don't matter in this particular calculation, since our expectation is only in terms of X and Y_1. With this in mind, we can consider the following tree describing this situation.

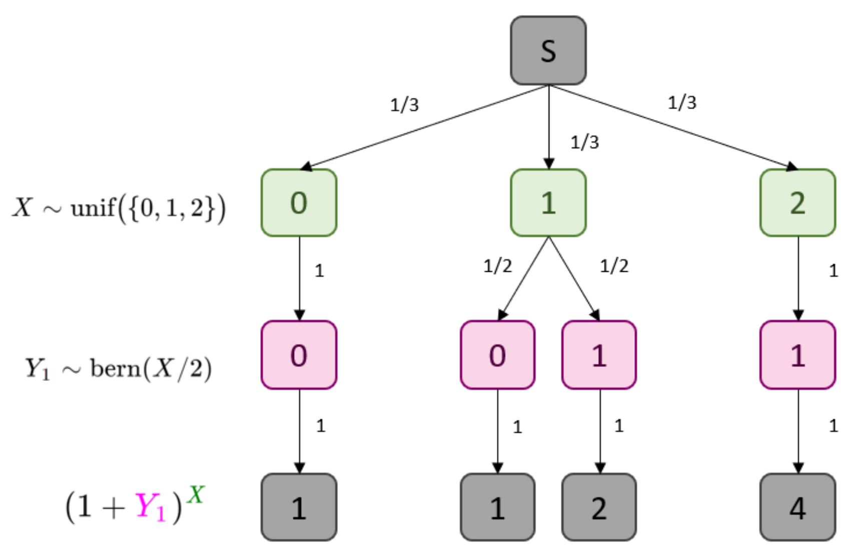


Figure 1: Decision Tree for First Coin Flip, with Function $g(x, y_1)$

Then, applying the Law of the Unconscious Statistician (as seen in the bottom level of this tree), we can find $\mathbb{E}((1+Y_1)^X)$. We use chain rule to propagate towards the probabilities aligning with the bottom leaves.

$$\mathbb{E}[(1+Y_1)^X] = \frac{1}{3}(1) + \frac{1}{6}(1) + \frac{1}{6}(2) + \frac{1}{3}(4) = \boxed{\frac{13}{6}}$$

Part 2 Write an R function called `forward_sample` that samples ("simulates") from the joint distribution of (X, Y_1, Y_2, Y_3, Y_4). As a general practice, fix the seed, and submit both the code and the output (here, a single sample).

```
library(extraDistr)
set.seed(19690720)
forward_sample <- function(){
      X = rdunif(1, 0, 2)
      return(
        c(X, sapply(1:4, function(x) rbinom(n = 1, size = 1, prob = X/2)))
      )
}
forward_sample()
```

```
## [1] 2 1 1 1 1
```

Part 3

How can your code and the law of large number be used to approximate $\mathbb{E}[(1+Y_1)^X]$

Since the Law of Large numbers basically states that average of many *iid* random samples will approach the truth (if it exists), we can use our newly-created `forward_sample()` function to approximate via Monte Carlo simulation. We can conduct m simulations, and let $\hat{\hat{G}}_M = (1+y_1)^x$. Then, the average given by $\hat{G}_M = \frac{1}{M}\sum_{m=1}^{M} G^{(m)}$ is approximately equal to $\mathbb{E}((1+Y_1)^X)$. The actual code to do this is below:

```
set.seed(19690720)
# declare number of iterations
num_iterations = 20000
# generate all simulations
simulations = sapply(1:20000, function(x){
    # generate a sample using the forward_sample() function
    sample_i = forward_sample()
    # calculate g(x, y) using the generated sample
    g_x = (1 + sample_i[2])^sample_i[1]
    # return this iteration
    return(g_x)
})
```

Part 4

Compare the approximation from your code with you answer in part 1.

Here, we actually conduct the simulations and compare the results to those found by arithmetic operations.

```r
# generate the estimate of E(g(X, Y))
simulated_mean = mean(simulations)
# value from simulations
print(paste("Simulated Value:", simulated_mean))
```

```
## [1] "Simulated Value: 2.1529"
```

```r
# value from part 1
print(paste("Calculated Value:", round(13 / 6, 4)))
```

```
## [1] "Calculated Value: 2.1667"
```

```r
# (observed - expected)/(expected) as a percent
print(paste("Percent Difference: ",
            round(abs(
              simulated_mean - 13 / 6
            ) / (13 / 6) * 100, 4), "%", sep = ''))
```

```
## [1] "Percent Difference: 0.6354%"
```

So we can see that the result found from simulation is very close to that found in **Part 1**.

Question 2 : Computing a Conditional

Suppose now that you observe the outcome of the 4 coin flips, but not the type of coin that was picked. Say you observe: "heads", "heads", "heads", "heads" = `[0, 0, 0, 0]`.

Note: We will use precise and careful notation to solve, not skipping any steps.

Part 1

Write mathematically: "Given you observe 4 heads, what is the probability that you picked the standard coin?"

We know that the coin flips are Y_1, Y_2, Y_3, Y_4 respectively. Based on the question description, for arbitrary $i \in [1, 4]$ the event that a coin is "heads" is $Y_i = 0$. Therefore, we can all four flips as the random vector (Y_1, Y_2, Y_3, Y_4). Understanding that selecting the standard coin is the event that $X = 1$, the sentence can be succinctly described as follows:

$$P(X = 1|(Y_1, Y_2, Y_3, Y_4) = (0, 0, 0, 0))$$

Part 2 Now, solve the expression.

For simplicity, we will let event E be the event that (Y_1, Y_2, Y_3, Y_4) is equal to $(0, 0, 0, 0)$, where again a zero indicates a heads.

We wish to find $P(X = 1|E)$. We will implement Bayes' Theorem.

$$P(X = 1|E) = \frac{P(E|X = 1)P(X = 1)}{P(E)}$$

We can compute the probability of picking Coin #1 directly.

$$P(X=1) = P(\text{unif}(\{0,1,2\})=1) = \frac{1}{3}$$

Assuming each Y_i is iid for $i \in [1,4]$, we can compute $P(E|X=1)$:

$$P(E|X=1) = P((Y_1,Y_2,Y_3,Y_4)=(0,0,0,0)|X=1) = \prod_{i=1}^{4} P(Y_i=0|X=1) = \left(\frac{1}{2^4}\right) = \frac{1}{16}$$

Now, we can compute $P(E)$ via the Law of Total Probability. Trivially, we know $P(E|X=2)=1$ and $P(E|X=0)=0$.

$$P(E) = \sum_{i=0}^{2} P(E|X=i)P(X=i) = (0)\left(\frac{1}{3}\right) + \left(\frac{1}{16}\right)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{3}\right) = \frac{1}{48} + \frac{1}{3} = \frac{17}{48}$$

Putting this all together, we have the following:

$$P(X=1|E) = \frac{P(E|X=1)P(X=1)}{P(E)} = \frac{(1/3)(1/16)}{(17/48)} = \boxed{\frac{1}{17}}$$

Question 3 : Non-Uniform Prior on Coin Types

We now modify the problem as follows: I stuffed the bag with 100 coins: 98 standard (fair) coins, 1 coin with only heads, and 1 coin with only tails. The rest is the same: pick one of the coins, flip it 4 times.

Part 1 Write the joint distribution of this modified model. Use the notation as in Equation 1. Hint: use a `Categorical` distribution.

Now, we know that X is no longer Uniform. Using the notation for the categorical distribution used in class, we write that

$$X \sim \text{categorical}\big\{(0,1,2), (\frac{1}{100}, \frac{98}{100}, \frac{1}{100})\big\}$$

Since the rules for flipping the coins remains the same, we still have that:

$$Y_i \mid X \sim \text{bern}(X/2)$$

Where X is now the categorical distribution.

Part 2 Compute the probability that you picked one of the fair coins, given you see four heads.

<u>**NOTE**</u>: We will let the event of a heads be represented with a **1** rather than a **0**, to comply with the `dunif` function used in the succeeding questions.

We wish to find $P(X=1 \mid Y_{1:4}=\vec{\mathbf{1}})$.We will implement Bayes' Theorem.

$$P(X=1 \mid Y_{1:4}=\vec{\mathbf{1}}) = \frac{P(Y_{1:4}=\vec{\mathbf{1}}|X=1)P(X=1)}{P(Y_{1:4}=\vec{\mathbf{1}})}$$

We note from the Categorical Distribution defined above, that $P(X=1)=\frac{98}{100}$. Similarly, by the same logic applied in **Question 2**, we know that:

$$P(Y_{1:4}=\vec{\mathbf{1}}|X=1) = \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

Now, we can find $P(Y_{1:4} = \vec{\mathbf{1}})$ via the Law of Total Probability.

$$P(Y_{1:4} = \vec{\mathbf{1}}) = \sum_{i=0}^{2} P(X = i)P(Y_{1:4} = \tilde{\mathbf{1}} \mid X = i)$$
$$P(Y_{1:4} = \vec{\mathbf{1}}) = \frac{1}{100}(0) + \frac{98}{100}\left(\frac{1}{16}\right) + \frac{1}{100}(1)$$
$$P(Y_{1:4} = \vec{\mathbf{1}}) = \frac{1}{100} + \frac{98}{1600} = \frac{114}{1600}$$

Now, we can bring this all together to solve for $P(X = 1 \mid Y_{1:4} = \vec{\mathbf{1}})$

$$P(X = 1 \mid Y_{1:4} = \vec{\mathbf{1}}) = \frac{P(Y_{1:4} = \vec{\mathbf{1}}|X = 1)P(X = 1)}{P(Y_{1:4} = \vec{\mathbf{1}})} = \frac{\left(\frac{1}{16}\right)\left(\frac{98}{100}\right)}{\left(\frac{114}{1600}\right)} = \boxed{\frac{49}{57}}$$

Where $\frac{49}{57}$ is approximately $0.859649\ldots$ as a decimal.

Question 4

We now generalize to having $K + 1$ types of coins such that:

- coin type $k \in \{0, 1, \ldots, K\}$ has bias k/K
- the fraction of coins in the bag of type k is ρ_k.

We consider the same observation as before: "you observe 4 heads". We want to find the conditional probability π_k, for all k that we picked coin type $k \in \{0, 1, \ldots, K\}$ from the bag given the observation.

Part 1 Write an R function called `posterior_given_four_heads` taking as input a vector $\rho = (\rho_0, \rho_1, \ldots, \rho_K)$ and returning $\pi = (\pi_0, \pi_1, \ldots, \pi_K)$.

```r
posterior_given_four_heads <- function(rho){
  kvals = 0:(length(rho)-1)
  K = length(kvals)
  # kvals = ( 0:(K-1) )
  biases = kvals / (K-1)
  # in this case, the denominator is from LOTP again
  p_E =  sum( biases^4 * rho )
  # then numerator is P(X)P(E | X)
  prior_likeli =  sapply(1:K, function(k){ biases[k]^4 * rho[k]} )
  # then return the ratio
  return(prior_likeli / p_E)
}

# this returns q2
posterior_given_four_heads(rho = c(1/3, 1/3, 1/3))
```

```
## [1] 0.00000000 0.05882353 0.94117647
```

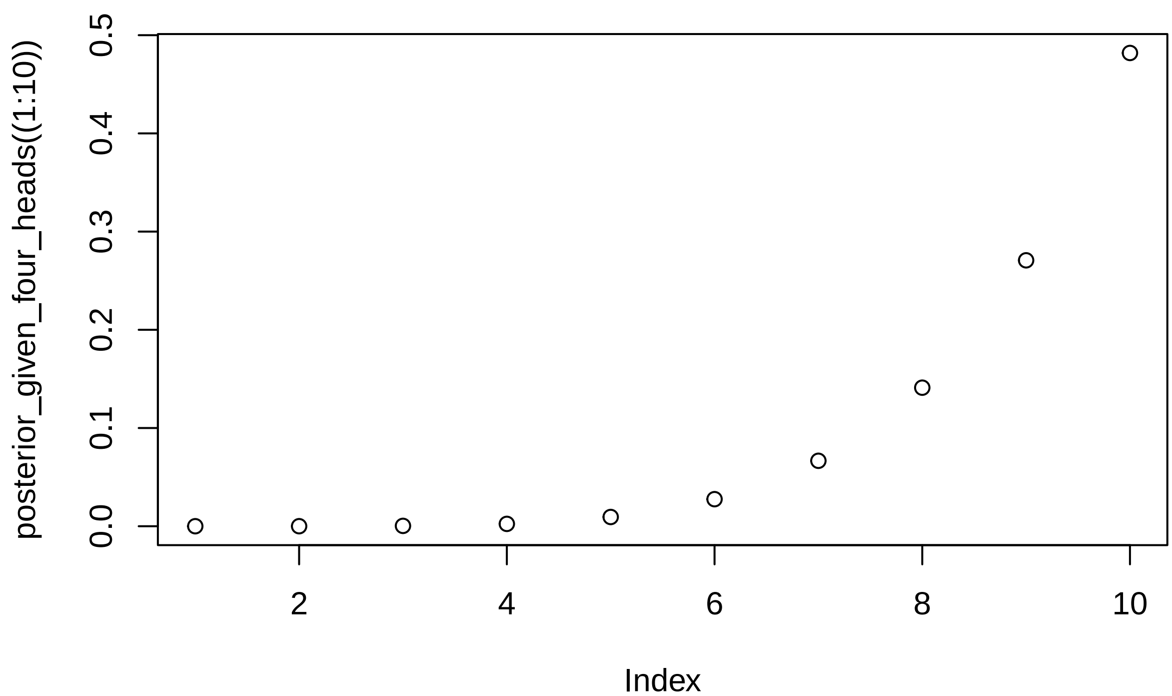
Part 2 Test your code by making sure you can recover the answer in Q. 3 as a special case. Report what values of K and ρ you used.

Part 3 (INCOMPLETE) Show the output for $\rho \propto (1, 2, 3, \ldots, 10)$. Here $\propto$ means "proportional to"; try to infer what it means in this context.

```r
posterior_given_four_heads((1:10))
```

```
##  [1] 0.000000e+00 1.468882e-05 3.525316e-04 2.379588e-03 9.400843e-03
##  [6] 2.754153e-02 6.662848e-02 1.410714e-01 2.707443e-01 4.818667e-01
```

```r
plot(posterior_given_four_heads((1:10)))
```



What does "proportional to" mean in this context?

```r
# note that the whole 1 to 10 thing is non-normalized
# so we know the proportionality to 1:10 is the same as
# actually passing the following probabilities
1:10/(sum(1:10))
```

```
##  [1] 0.01818182 0.03636364 0.05454545 0.07272727 0.09090909 0.10909091
##  [7] 0.12727273 0.14545455 0.16363636 0.18181818
```

```r
# also code output is the same
posterior_given_four_heads( 1:10/(sum(1:10)))
```

```
##  [1] 0.000000e+00 1.468882e-05 3.525316e-04 2.379588e-03 9.400843e-03
##  [6] 2.754153e-02 6.662848e-02 1.410714e-01 2.707443e-01 4.818667e-01
```

Q.5: Generalizing Observations

We now generalize Q. 4 as follows: instead of observing 4 "heads" out of 4 observations, say we observe `n_heads` out of `n_observations`, where `n_heads` and `n_observations` will be additional arguments passed into a new R function.

Part 1 (INCOMPLETE) Write the joint distribution of this modified model. Use the $\sim$ notation as in Equation 1. Hint: use a Binomial distribution.

Part 2 Write an R function called `posterior` taking three input arguments in the following order: a vector ρ as in Q. 4, as well as two integers, `n_heads` and `n_observations`.

```
posterior <- function(rho, n_heads, n_observations){
  # this part is the same
  kvals = 0:(length(rho)-1)
  K = length(kvals)
  biases = kvals / (K-1)
  # we need the overall probability of nheads
  p_E =  sum(dbinom(x = n_heads, size = n_observations, p = biases)*rho)
  # then, across the k coins that can be picked
  # then numerator is still P(X)P(E | X)
  prior_likeli =  sapply(1:K, function(k){
    dbinom(x = n_heads, size = n_observations, p = biases[k]) * rho[k]} )
  # then return the ratio
  return(prior_likeli / p_E)
}
```

```
posterior(c(1/3, 1/3, 1/3), 4, 4)
```

Part 3 (MAYBE INCOMPLETE)

```
## [1] 0.00000000 0.05882353 0.94117647
```

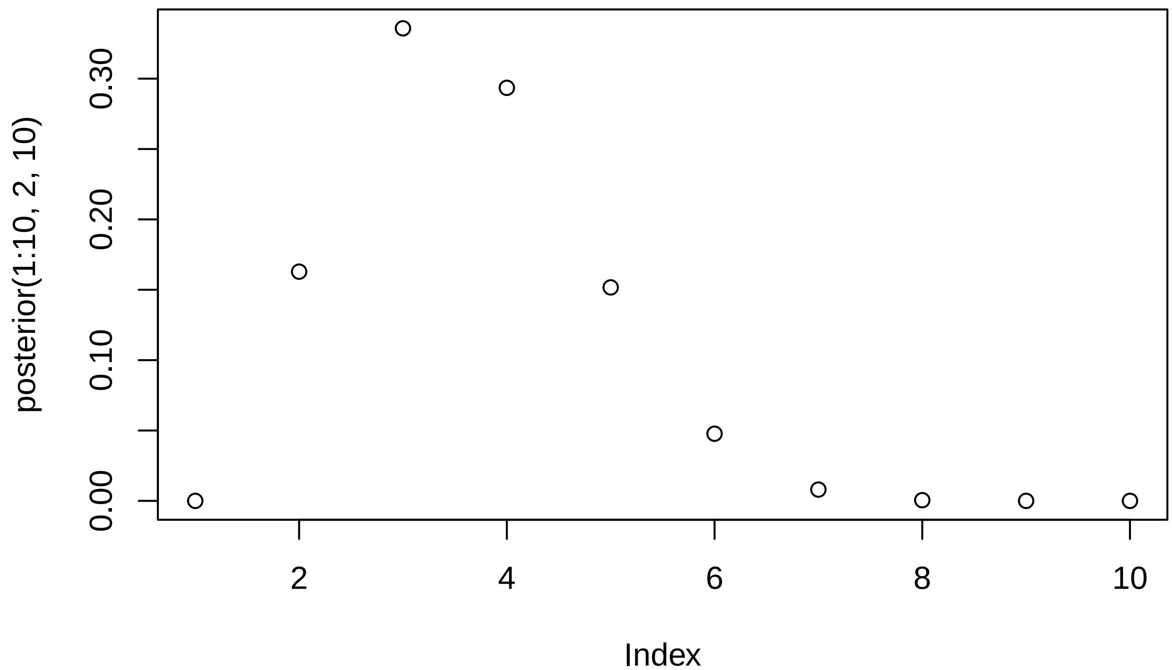
Test your code by making sure you can recover the answer in Q. 3 as a special case.

Part 4 (MAYBE INCOMPLETE) Show the output for $\rho \propto (1, 2, 3, \ldots, 10)$ and `n_heads = 2` and `n_observations = 10`.

```
posterior(1:10, 2, 10)
```

```
##  [1] 0.000000e+00 1.628596e-01 3.357600e-01 2.934782e-01 1.516748e-01
##  [6] 4.771276e-02 8.024795e-03 4.870678e-04 2.795670e-06 0.000000e+00
```

```
plot(posterior(1:10, 2, 10))
```


posterior(1:10, 2, 10)
0.00
0.10
0.20
0.30
2
4
6
8
10
Index