STAT 447C: Bayesian Statistics **Fall 2024**

Exercise 2

Junsong Tang *junsong.tang@stat.ubc.ca*

1 define a Bayesian model

1. The unknown quantity is the value of p, and the data is the observed value of Y_i's.
2.

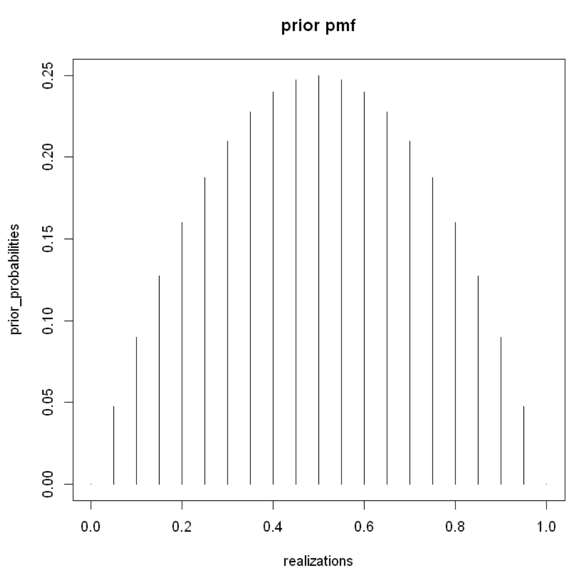
$$X \sim \rho, \rho = (\rho_k)_{k=1}^{K}$$
$$Y|X \sim \text{Bernoulli}(X)$$

2 posterior and point estimates

1.

```
prior_probabilities = NULL
realizations = NULL
for (k in (0 : K)) {
    prior_probabilities = c(prior_probabilities, (k/K)*(1-k/K))
    realizations = c(realizations, k/K)
}
prior_probabilities
realizations
plot(realizations, prior_probabilities, type="h", main = "prior pmf")

```

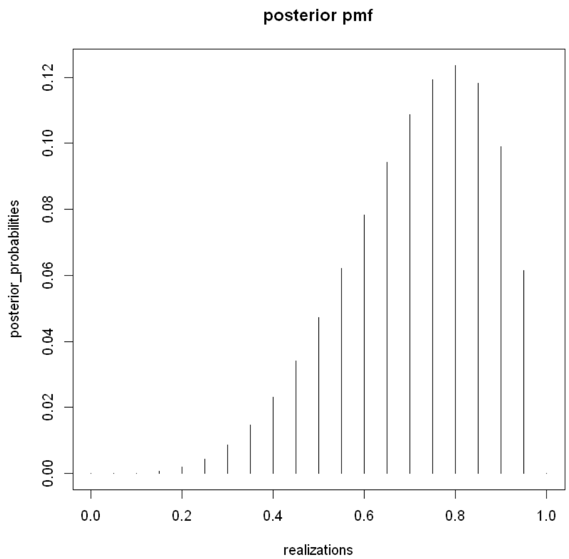


2.

```
posterior_probabilities = NULL
Z = 0
for (k in (0 : K)) {
  # likelihood = P(X = k/K) * P(Y = (1,1,1) | X = k/K)
  likelihood = realizations[k+1]**3 * prior_probabilities[k+1]
  Z = Z + likelihood
  posterior_probabilities = c(posterior_probabilities, likelihood)
}
```

```r
posterior_probabilities = posterior_probabilities / Z
posterior_probabilities
plot(realizations, posterior_probabilities, type = "h", main = "posterior pmf")

```



3. From the plot, we can see the posterior mode is at $X = 0.8$
4.
```r
# posterior mean
m = 0
for (k in (0 : K)) {
  x = realizations[k+1]
  pi_k = posterior_probabilities[k+1]
  m = m + x * pi_k
}
m #0.712497759455099

```

3 Bayes action

1. Set the money unit into $\$M$

$$\begin{aligned} L(a, Y_4) &= \mathbf{1}_{a=0} \cdot \$0 + \mathbf{1}_{a=1;Y_4=1} \cdot \$2 + \mathbf{1}_{a=1;Y_4=0} \cdot (\$2 - \$100) \\ &= 2\mathbf{1}_{a=1;Y_4=1} - 98\mathbf{1}_{a=1;Y_4=0} \end{aligned}$$

2. By LOTUS,

$$\begin{aligned} \mathcal{L}(a) &= \mathbb{E}[L(a, Y_4)|Y_{1:3} = (1,1,1)] \\ &= \sum_{y_4 \in \{0,1\}} L(a, Y_4 = y_4) \cdot \mathbb{P}(Y_4 = y_4|Y_{1:3} = (1,1,1)) \\ &= 2\mathbf{1}_{a=1} \cdot \mathbb{P}(Y_4 = 1|Y_{1:3} = (1,1,1)) - 98\mathbf{1}_{a=1} \cdot \mathbb{P}(Y_4 = 0|Y_{1:3} = (1,1,1)) \\ &= 2\mathbf{1}_{a=1} \cdot p - 98\mathbf{1}_{a=1} \cdot (1-p);\ \text{where } p := \mathbb{P}(Y_4 = 1|Y_{1:3} = (1,1,1)) \end{aligned}$$

3. Plug in the value of p that has been computed in q2(4), then we have $\mathcal{L}(a) = -26.75\mathbf{1}_{a=1}$, so $\mathcal{L}(1) < \mathcal{L}(0)$, meaning we need to buy insurance.

However, if $\mathcal{L}(0) < \mathcal{L}(1) \Rightarrow (2p - 98(1-p)) > 0 \Rightarrow p > 0.98$, then we don't need to buy insurance for the posterior at least 98%.

4 Challenges

1.

```
# challenge question
# signature: posteriors, threshold -> index set
# overall complexity:
# time: O(nlogn + 2n)
# space: O(n)

highest_prob_set = function(posterior_probabilities, alpha) {
    n = length(posterior_probabilities)
    # create hashmap that maps prob -> index
    # time complexity O(n) and space complexity O(n)
    dict = c()
    for (i in (1:n)) {
        # set k-v pair as (prob, index)
        p = posterior_probabilities[i]
        dict[as.character(p)] = i
    }
    # sort complexity: O(nlogn)
    posterior_probabilities = sort(posterior_probabilities, decreasing = TRUE)

    # linear search complexity: O(n)
    j = 1
    sum = 0
    hps = NULL
    while (j <= n && sum < alpha) {
        p = posterior_probabilities[j]
        hps = c(hps, dict[as.character(p)])
        j = j + 1
        sum = sum + p
    }

    return (hps)
}
```

2.

```
print(highest_prob_set(posterior_probabilities = posterior_probabilities, alpha=0.75))
# 17, 16, 18, 15, 19, 14, 13, 12
```