

STAT 447 Assignment 2

Caden Hewlett

2024-01-19

Question 1 : Define a Bayesian Model

In order to perform inference on the unknown quantities, we must specify how they relate to the data; i.e., we need a probabilistic model. Assume that every Delta 7925H rocket has the same probability p of failing. For simplicity, let us assume that p is allowed to take values on an evenly space grid

$$p \in \left\{ \frac{k}{K} : k \in \{0, \dots, K\} \right\}$$

for some fixed $k \in \mathbb{N}$. Furthermore, we have access to a collection of numbers ρ_k such that:

$$\forall k \in \{0, \dots, K\} : \mathbb{P}\left(p = \frac{k}{K}\right) = \rho_k$$

Let Y_i denote a binary variable with $Y_i = 1$ encoding a success, and $Y_i = 0$ a failure. We assume that, conditionally on p , the Y_i 's are independent of each other.

We will use the following prior:

$$\rho_k \propto \frac{k}{K} \left(1 - \frac{k}{K}\right)$$

From now on, use $K = 20$

Part 1

What are the unknown quantities in this scenario? And what is the data?

Solution: In this case, we can think of two unknown quantities and one piece of known data. One unknown metric is Y_i , whether or not the rocket is going to fail. The other unknown quantity is the success probability of the rockets - this was denoted p . This is much like the “coin flip” scenario from the previous homework, where Y_i was “whether or not we flip a heads,” and the distribution of coin biases (i.e. the probability we pick a coin with bias k/K) was ρ_k .

In this case, the data is given in the setup section, and is simply that as of Jan 2024, Delta 7925H rockets have been launched 3 times, with 0 failed launches. Further, we have ρ_k , the probability that the *true* probability p is k/K for a given discrete k .

Part 2

To easily define the joint distribution, we will use the following fact discussed in class.

$$\underbrace{p_X(x)}_{\text{prior}} \underbrace{p_{Y|X}(y|x)}_{\text{likelihood}} \equiv \underbrace{p_{X,Y}(x,y)}_{\text{joint}}$$

This follows from chain rule.

Here, letting $k = 20$, we can think of the random Variable X as the probability that the rocket has failure rate k , for $k \in [0, 20]$ i.e.

$$X \sim \text{unif}(\{0, 1, 2, \dots, 20\})$$

Directly, $p_X(x) = P(X = x)$.

Similarly, the conditional distribution of Y_i given X can be written as:

$$Y | X \sim \text{bern}\left(\frac{X}{20}\left(1 - \frac{X}{20}\right)\right)$$

Directly, $p_{Y|X}(y|x) = P(Y | X = y | x)$ (note, I'm not sure if the terminology should be the random variable X or the realization x)

Then, together, these form the joint distribution $p_{X,Y}(x,y)$ by the equivalency discussed earlier.

Question 2 : Posterior and Point Estimates

To help you answer the following questions, we create the two vectors

`prior_probabilities` where entry i containing the prior probability ρ_{i-1} defined in Q1 (the minus one reflects the fact that R uses indexing starting at 1), and `realizations`, a vector of possible realizations of p in the same order, namely $(0, 1/K, 2/K \dots, 1)$.

```
K = 20
priors = sapply(0:K, function(k){(k / K)*(1 - k/K)})
prior_probabilities = priors/sum(priors)
realizations = sapply(0:K, function(k){k/K})
```

Part 1

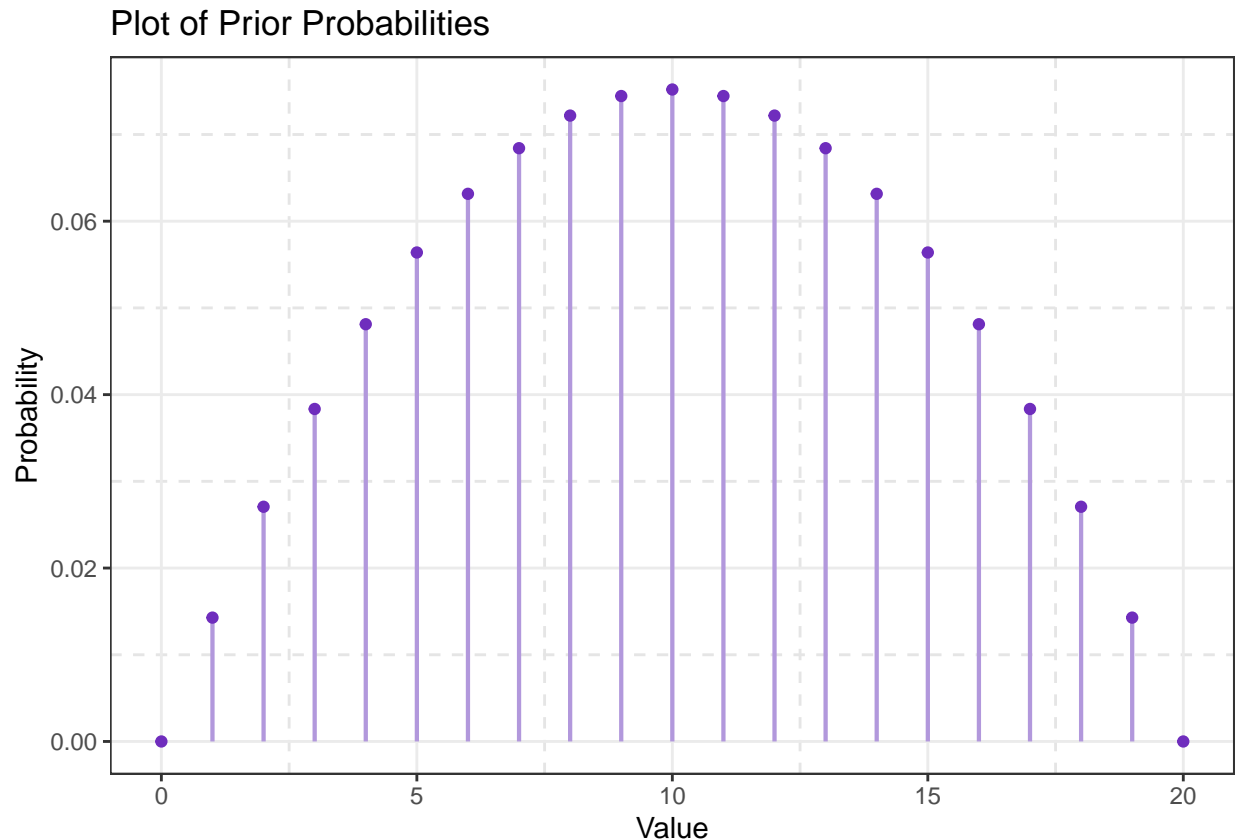
Plot the prior PMF. Do you think this is a reasonable prior?

```
plot_data <- data.frame(
  value = 0:K,
  probability = prior_probabilities
)
# plot prior
p <- ggplot(plot_data, aes(x = value, y = 0, yend = probability)) +
  geom_segment(aes(xend = value), color = "#b298dc", linewidth = 0.75) +
  geom_point(aes(x = value, y = probability), color = "#6f2dbd") +
  labs(x = "Value", y = "Probability") +
  ggtitle("Plot of Prior Probabilities") + theme_bw() +
```

```

theme(panel.grid.minor = element_line(
  color = "grey90",
  linetype = "dashed",
  linewidth = 0.5
))
print(p)

```



I'm not sure that this prior is exactly reasonable for this specific situation, but it's not bad. My main issue with it is the lower probability placed on higher outcomes. For example, p_{20} , i.e. $P(p = 20/K = 1)$ is assigned the same (zero) probability as p_0 , the probability of the rocket being guaranteed to fail. This symmetry in the fail probability likelihoods doesn't exactly square with our existing data (which is fine because we will calculate the posterior). However, we know some information about the history of *all* rocket launches in the past, and we could in theory use this to see if there is, in reality, a "heavier tail" for greater values of k . Also, as was discussed in lecture, it's best not to assume any of the values in the prior are exactly zero.

However, there is nuance to this problem (as in all real-world scenarios.) This prior distribution is quite conservative - i.e., it isn't placing too much potentially risky weight on the historical successes, which is *also* a sensible choice given the cost of these rockets.

Part 2

Let $\pi_k = \mathbb{P}(p = k/K | Y_{1:3} = \vec{1})$ denote the posterior probabilities, for $k \in \{0, 1, 2, \dots, K\}$. Create a vector `posterior_probabilities` where entry i is π_{i-1} . Plot the posterior PMF.

We can start by computing $P(Y_{1:3} = \vec{1} \mid p = \frac{k}{20})$ using the `realizations` vector from earlier. This

probability decomposes as follows (letting $k/20 = p_{\text{obs}}$ for simplicity):

$$P(Y_{1:3} = \vec{1} \mid p = p_{\text{obs}}) = P(Y_1 = 1 \mid p = p_{\text{obs}})P(Y_2 = 1 \mid Y_1 = 1, p = p_{\text{obs}})P(Y_3 = 1 \mid Y_2 = 1, Y_1 = 1, p = p_{\text{obs}})$$

Which can simplify nicely into:

$$P(Y_{1:3} = \vec{1} \mid p = p_{\text{obs}}) = P(Y_1 = 1 \mid p = p_{\text{obs}})P(Y_2 = 1 \mid p_{\text{obs}})P(Y_3 = 1 \mid p = p_{\text{obs}})$$

Then, assuming the rocket launches are *iid* $\text{bern}(p_{\text{obs}})$ as discussed previously, with success probability p_{obs} , this can be simplified as:

$$P(Y_{1:3} = \vec{1} \mid p = p_{\text{obs}}) = (p_{\text{obs}})(p_{\text{obs}})(p_{\text{obs}}) = (p_{\text{obs}})^3$$

This is really just a long-winded way of saying we simply take the cube of our **realizations** vector to get the likelihood.

```
likelihood = (realizations)^3
```

Then, $\gamma(k/20)$ for $k \in \{0, 1, \dots, 20\}$ is simply the product of the prior probabilities with the likelihood.

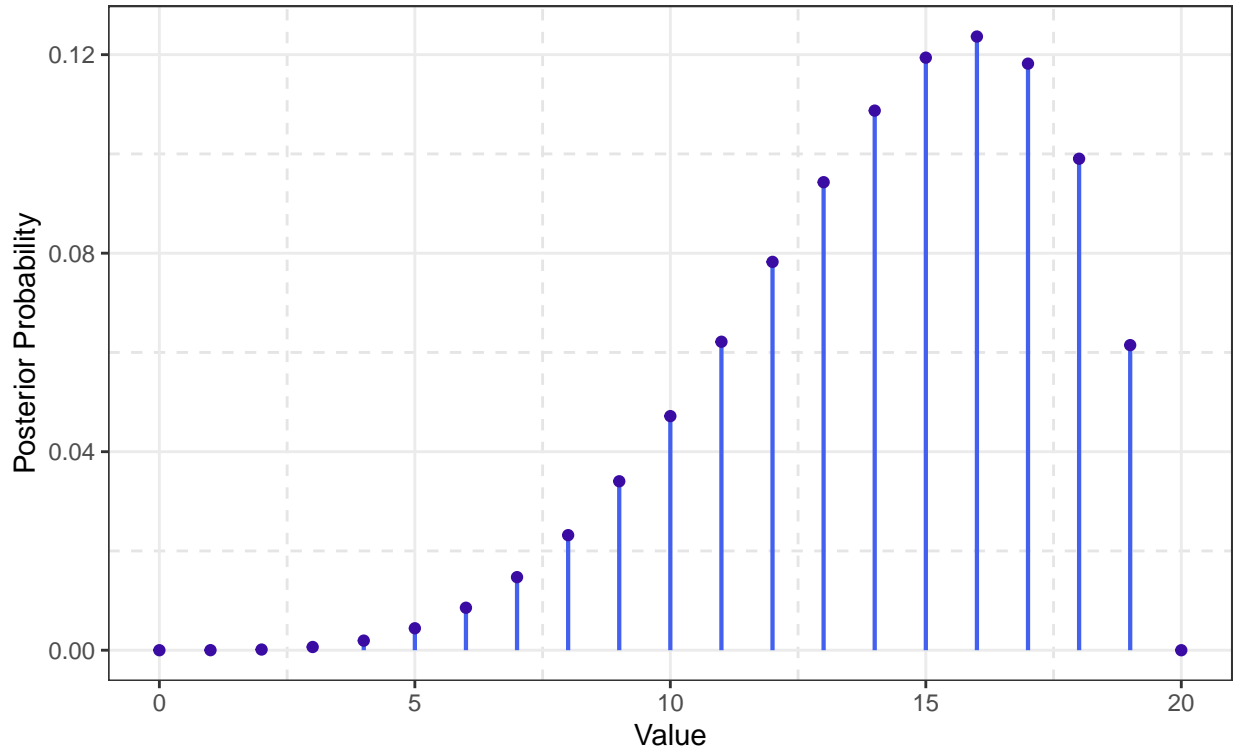
```
gamma = prior_probabilities*likelihood
```

Finally, we divide by the normalizing constant Z so the probabilities sum to 1. This gives π_k , which we calculate and plot below.

```
pi = gamma/sum(gamma)

plot_data_2 <- data.frame(
  value = 0:K,
  probability = pi
)
# plot posterior
p2 <- ggplot(plot_data_2, aes(x = value, y = 0, yend = probability)) +
  geom_segment(aes(xend = value), color = "#4361ee", linewidth = 0.75) +
  geom_point(aes(x = value, y = probability), color = "#3a0ca3") +
  labs(x = "Value", y = "Posterior Probability",
       title = "Plot of Posterior Probabilities",
       subtitle = "Given the Three Successful Rocket Launches") + theme_bw() +
  theme(panel.grid.minor = element_line(
    color = "grey90",
    linetype = "dashed",
    linewidth = 0.5
  ))
print(p2)
```

Plot of Posterior Probabilities
Given the Three Successful Rocket Launches



Part 3

What is the posterior mode?

By definition, we know that:

$$\text{mode}(\pi(x)) = \underset{x \in X}{\operatorname{argmax}} [\pi(x)]$$

For our prior distribution, this is found to be:

```
(0:K)[which.max(pi)]
```

```
## [1] 16
```

Corresponding to $k = 16$, or $p = 16/20$.

Part 4

Write a function that compute the posterior mean of p . (Hint: you should obtain $\mathbb{E}[p|Y_{1:3} = \vec{1}] \approx 0.7$.)

We know that the posterior mean $\mathbb{E}[p|Y_{1:3} = \vec{1}]$ can be found by calculating

$$\sum_{\{x \text{ s.t. } \pi(x) > 0\}} x \pi(x)$$

Thankfully, with our `realizations` and `posterior` distributions, we already have most of the tools at our disposal to make this calculation.

```
get_posterior_mean <- function(){
  return(sum(realizations*pi))
}
get_posterior_mean()
```

```
## [1] 0.7124978
```

Question 3 : Bayes Action

Let $a \in \{0, 1\}$ be a binary variable denoting the decision of buying the insurance ($a = 1$) or not ($a = 0$.)

Part 1

Based on the problem description from the Setup Section, define a loss function $L(a, y)$ that summarizes the cost of having taken decision $a \in \{0, 1\}$ depending on whether the next launch is successful ($y = 1$) or not ($y = 0$). Hint: use indicator functions (i.e. binary functions taking either the value zero or one)

Let's first approach this problem by making a table designating potential losses based on outcomes. Here we let the numbers be in millions of dollars. (i.e. 100 = \$100M)

	$y = 1$	$y = 0$
$a = 1$	Loss of 2 . You don't get your insurance deposit back.	Loss of -98 . The rocket crashes and you get your money back. Still paid to be insured.
$a = 0$	Loss of 0 . You spend nothing on insurance, and the rocket is fine.	Loss of 100 . The rocket crashes and you were not insured.

With this in mind, we can write our loss function as follows. **NOTE** the command `\mathbb{1}` doesn't work in my version of latex, so I used `\mathbb{I}` for the indicator function instead.

$$L(a, y) = 2\mathbb{I}(a = 1, y = 1) + 0\mathbb{I}(a = 0, y = 1) - 98\mathbb{I}(a = 1, y = 0) + 100\mathbb{I}(a = 0, y = 0)$$

Which simplifies to the following, if we eliminate the term with zero:

$$L(a, y) = 2(\mathbb{I}(a = 1)\mathbb{I}(y = 1)) - 98(\mathbb{I}(a = 1)\mathbb{I}(y = 0)) + 100(\mathbb{I}(a = 0)\mathbb{I}(y = 0))$$

Part 2

We now consider the expected loss under the posterior predictive distribution:

$$\mathcal{L}(a) := \mathbb{E}[L(a, Y_4) | Y_{1:3} = \vec{1}]$$

We wish to write $\mathcal{L}(a)$ in terms of $P(Y_4 = 1 | Y_{1:3} = \vec{1})$, using the fact that $P(Y_4 = 1 | Y_{1:3} = \vec{1}) = \mathbb{E}[p | Y_{1:3} = \vec{1}]$, a special property of a Bernoulli likelihood model.

We know in general that"

$$\mathbb{E}[L(a, X) | Y = y] = \sum_x L(a, x) p_{X|Y}(x|y)$$

Hence,

$$\mathbb{E}[L(a, Y_4)|Y_{1:3} = \vec{1}] = \sum_{y_4 \in \{0,1\}} L(a, y_4) p_{Y_4|Y_{1:3}}(y_4|Y_{1:3} = \vec{1})$$

Which can be written as:

$$\mathbb{E}[L(a, Y_4)|Y_{1:3} = \vec{1}] = \sum_{y_4 \in \{0,1\}} L(a, y_4) P(Y_4 = y_4 | Y_{1:3} = \vec{1})$$

Expanding the sum, we have:

$$\mathbb{E}[L(a, Y_4)|Y_{1:3} = \vec{1}] = L(a, 0)P(Y_4 = 0 | Y_{1:3} = \vec{1}) + L(a, 1)P(Y_4 = 1 | Y_{1:3} = \vec{1})$$

Since events $y = 0$ and $y = 1$ are exhaustive, we can write:

$$\mathbb{E}[L(a, Y_4)|Y_{1:3} = \vec{1}] = L(a, 0)(1 - P(Y_4 = 1 | Y_{1:3} = \vec{1})) + L(a, 1)P(Y_4 = 1 | Y_{1:3} = \vec{1})$$

Then, using $P(Y_4 = 1|Y_{1:3} = \vec{1}) = \mathbb{E}[p|Y_{1:3} = \vec{1}]$, we have:

$$\mathbb{E}[L(a, Y_4)|Y_{1:3} = \vec{1}] = L(a, 0)(1 - \mathbb{E}[p|Y_{1:3} = \vec{1}]) + L(a, 1)\mathbb{E}[p|Y_{1:3} = \vec{1}]$$

Then, expanding the $L(a, 0)$ and $L(a, 1)$ function calls one at a time:

$$\begin{aligned} L(a, 0) &= \left(2(\mathbb{I}(a = 1)\mathbb{I}(0 = 1)) - 98(\mathbb{I}(a = 1)\mathbb{I}(0 = 0)) + 100(\mathbb{I}(a = 0)\mathbb{I}(0 = 0)) \right) \\ L(a, 0) &= -98 \mathbb{I}(a = 1) + 100\mathbb{I}(a = 0) \end{aligned}$$

Then, similarly expanding $L(a, 1)$:

$$\begin{aligned} L(a, 1) &= \left(2(\mathbb{I}(a = 1)\mathbb{I}(1 = 1)) - 98(\mathbb{I}(a = 1)\mathbb{I}(1 = 0)) + 100(\mathbb{I}(a = 0)\mathbb{I}(1 = 0)) \right) \\ L(a, 1) &= 2 \mathbb{I}(a = 1) \end{aligned}$$

Then, we can substitute these back into our equation for $\mathcal{L}(a)$.

$$\mathcal{L}(a) = (-98 \mathbb{I}(a = 1) + 100\mathbb{I}(a = 0))(1 - \mathbb{E}[p|Y_{1:3} = \vec{1}]) + 2 \mathbb{I}(a = 1)\mathbb{E}[p|Y_{1:3} = \vec{1}]$$

Then, we can let $\mathbb{E}[p|Y_{1:3} = \vec{1}] = 0.7$ as an approximation for simplicity.

$$\mathcal{L}(a) = (-98 \mathbb{I}(a = 1) + 100\mathbb{I}(a = 0))(1 - 0.7) + 2 \mathbb{I}(a = 1)(0.7)$$

Then we can expand the terms.

$$\mathcal{L}(a) = -98(0.3) \mathbb{I}(a = 1) + 100(0.3)\mathbb{I}(a = 0) + 2 \mathbb{I}(a = 1)(0.7)$$

Finally, we begin to simplify.

$$\mathcal{L}(a) = 100(0.3)\mathbb{I}(a = 0) + (2(0.7) - 98(0.3)) \mathbb{I}(a = 1)$$

This simplifies nicely!

$$\mathcal{L}(a) = 30\mathbb{I}(a = 0) - 28 \mathbb{I}(a = 1)$$

If we wanted to, we could also use exact values and write it as an R function!

```

L <- function(a){
  pmean = get_posterior_mean()
  if(a == 0){
    return(100*(1 - pmean))
  }
  else if(a == 1){
    return(
      -98*(1 - pmean) + 2*pmean
    )
  }
}
# show results
sapply(0:1, function(a){ L(a) })

```

```
## [1] 28.75022 -26.75022
```

Part 4

Note: we continue to use our approximation that $P(Y_4 = 1 | Y_{1:3} = \vec{1}) \approx 0.7$.

We will now formulate a recommendation to the owner of the satellite using the Bayes' Estimator:

$$\operatorname{argmin}_{a \in \{0,1\}} (\mathbb{E}[L(a, X) | Y = y]) = \operatorname{argmin}_{a \in \{0,1\}} (\mathcal{L}(a))$$

We can note the formal definition of argmin with respect to $\mathcal{L}(a)$:

$$\operatorname{argmax}_{a \in \{0,1\}} (\mathcal{L}(a)) = \{a \text{ s.t. } \mathcal{L}(a) \leq \mathcal{L}(y), \forall a, y \in \{0,1\}\}$$

Thus, we see that the only $a \in \{0,1\}$ that satisfies this condition is $a = 1$, as $\mathcal{L}(1) = -28$ is less than $\mathcal{L}(0) = 30$.

We can verify this claim by using the R function we defined earlier!

```
(0:1)[which.min(sapply(0:1, function(a){ L(a) }))]

```

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

Hence, we would recommend action $a = 1$ to the satellite owner - get the insurance! From our calculations, we learned that the loss associated with an uninsured failure outweighs the potential money saved from an insured flight success, despite the fact the rocket is more likely to succeed than fail.