STAT 447 Assignment 4

Caden Hewlett

2024-02-02

Question 1 : Logistic Rocket Improvement

Recall the Rocket data from last week:

```
launches = c(1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1)
```

Recall that we discussed a model where the reliability of the rocket changes in time. This will allow us to incorporate, for example, the fact that engineering teams implement fixes based on past launches and therefore the probability of success should increase.

Part 1

Write a function called `logistic_regression` containing a `simPPLe` probabilistic programming description of the model described in class.

We recall that the model was of the following form, where ℓ is the launch and $\ell \in \{1, 2, \ldots 11\}$.

For algebraic conciseness, we let slope $= \beta_1$ and intercept $= \beta_0$.

Therefore, by the reasoning dicsussed in class,

$$\beta_1 \sim N(0, 1)$$
$$\beta_0 \sim N(0, 1)$$
$$\theta(\ell) = \text{logistic}(\beta_1 \cdot \ell + \beta_0) = \left(1 + \exp(-(\beta_1 \cdot \ell + \beta_0))\right)^{-1}$$

Let's begin to design our function.

```
logistic_regression = function() {
  beta_1 = simulate(distr::Norm(0, 1))
  beta_0 = simulate(distr::Norm(0, 1))
  sapply(1:length(launches),
        function(L) {
          observe(launches[L],
                  Bern(plogis(beta_0 + beta_1 * L)))
        })
  next_p = plogis(beta_0 + beta_1 * 12)
  return(c(beta_0, beta_1, simulate(Bern(next_p))))
}
```

Your function should return a vector containing 3 elements in the following order:

The intercept, $\beta_0 \in \mathbb{R}$, the slope ($\beta_1 \in \mathbb{R}$), a prediction if one more launch would have been successful (1) or a failure (0) ($s \in \{0, 1\}$).

```
logistic_regression()
```

```
## [1] -0.7863147  1.3384654  1.0000000
```

Part 2

Follow the instructions in the appendix below to get some helper functions. Use these functions to reproduce the lecture's bivariate posterior plot over the intercept and slope parameters.

The code to produce the bivariate posterior plot is below:

```
set.seed(11292023)
post = posterior_particles(logistic_regression, 1000)
weighted_scatter_plot(post,
                      plot_options =
                        list(
    xlab = TeX(r'($\hat{\beta}_0, Intercept \; Terms$)'),
    ylab = TeX(r'($\hat{\beta}_1, Slope \; Terms$)'),
    main = TeX(r'(Weighted Forward Simulated Values for m\in\{1, 2, ..., M\})')
                        ))
```


Weighted Forward Simulated Values for m ∈ {1, 2, ..., M}

2
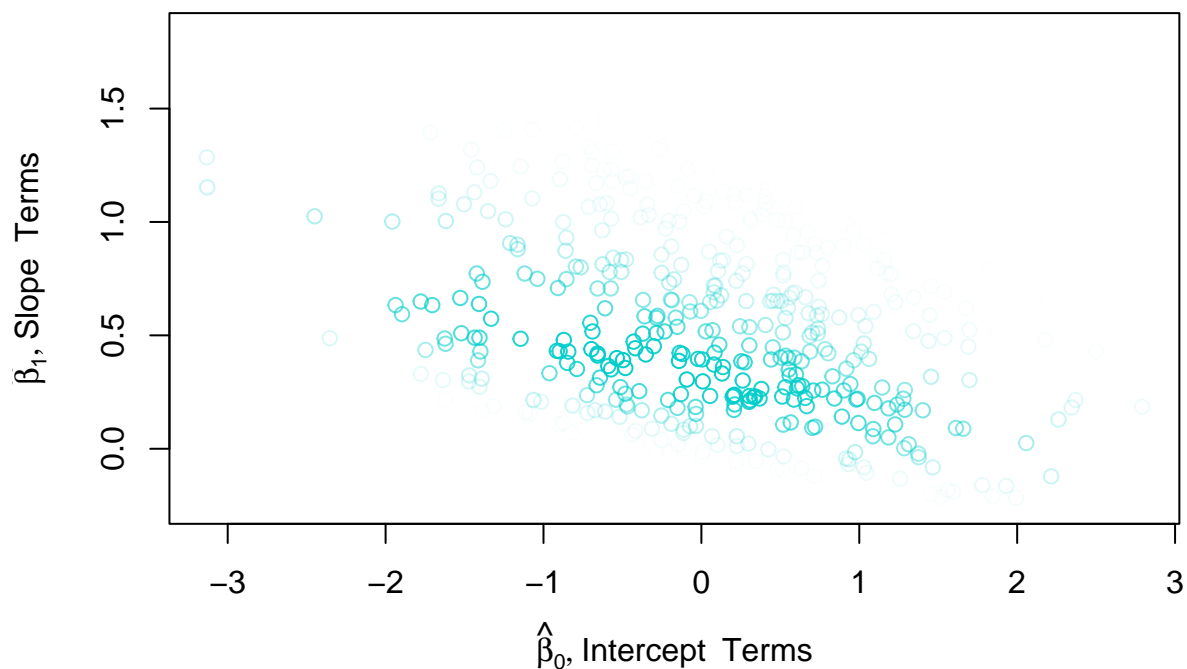
Part 3

Estimate the probability that the next launch is a success given the data under the **logistic model**.

We know that the launches so far can be thought of as realizations of the random vector $\vec{Y} = Y_{1:11} = \{Y_1, Y_2, Y_3, \ldots, Y_{11}\}$. From these, we have the specific realizations (i.e. "our data") $\vec{y} = \{1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1\}$.

Note that there are $n = 11$ launches so far.

We wish to find:

$$P(Y_{12} = 1 \mid Y_1 = 1, Y_2 = 0, Y_3 = 1, \ldots Y_{11} = 1)$$

Or, in a simpler sense, using the variables defined earlier,

$$P(Y_{12} = 1 \mid Y_{1:12} = \vec{y})$$

In order to find this probability, all we need to do is add this outcome of Y_{12} and use our forward simulator.

We do this as follows:

```
launches = c(launches, 1)
next_launch_post = posterior(logistic_regression, 1000)
rbind(c("beta_0", "beta_1", "success"), next_launch_post)
```

```
##                      [,1]                  [,2]                 [,3]
##                      "beta_0"              "beta_1"             "success"
## next_launch_post "-0.163923670961287"  "0.436866706377377"  "0.980854206698202"
```

Part 4

Create a variant of the same model but where the **slope is set to zero**. Estimate the probability that the next launch is a success given the data under this simplified model.

Question 2 : Choosing a Model

You debate with your friend whether the logistic model or the simplified model (with slope equals to zero) should be preferred. To stop that debate, write a **unified** model which gives probability $1/2$ to the simplified model, and $1/2$ to the logistic model.

Estimate the posterior probability that the logistic model is preferred under the unified model given the same data as in Q.1.

Where $\psi \in \{0, 1\}$ is whether we choose no slope ($\psi = 0$) or to include slope ($\psi = 1$.) We begin by assuming each outcome is equally likely.

$$\beta_1 \sim N(0, 1)$$
$$\beta_0 \sim N(0, 1)$$
$$\psi \sim \mathrm{bern}(1/2)$$
$$\theta(\ell) = \left(1 + \exp(-\left(\beta_1 \cdot \psi \cdot \ell\right) + \beta_0)\right)^{-1}$$

trivariate posterior I think