Bayesian Hierarchical Time Series

Leo Guelman

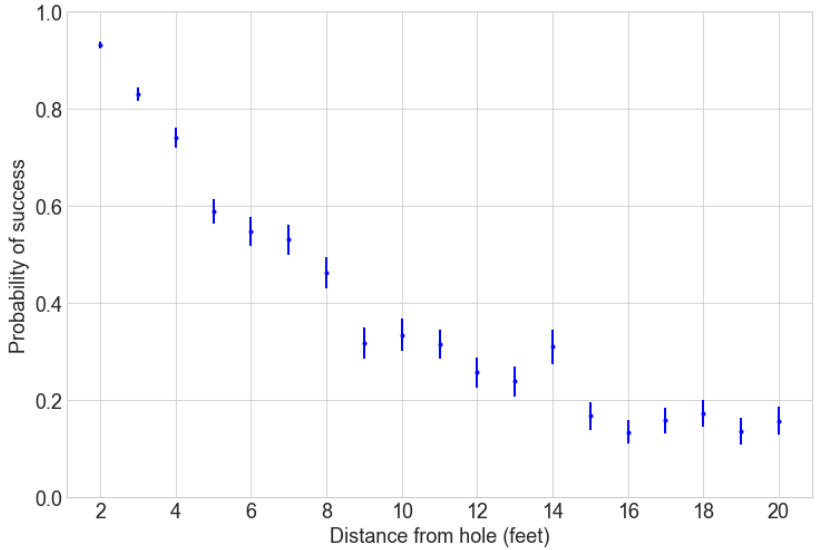
Tech Talk
DNA
RBC Royal Bank

A Metaphor...

- Read Richard McElreth
- Example decision tree, or flowchart, for selecting an appropri- ate statistical procedure. Beginning at the top, the user answers a series of questions about measurement and intent, arriving eventually at the name of a procedure. Many such decision trees are possible.
- Something similar to ML
- Probably I should put each use case a label : Prediction / Counterfactuals / Inference / Instrumental Variables (well this os not the objective but the method is to solve a problem)
- What is the goal of this presentation?
  - Often cases, thinking about the data generating process (the underlying mechanism that generated the observed data)?
  - When is this important? Think about noisy settings (with low signal to noise ratio), limited data, or when we want to predict under intervention (as opposed to under observation)
  - Contrast to using domain knowledge vs end-to-end ML o make predictions?

A Model for Golf Putting

- Aggregate data[1] from professional golfers, showing the proportion of successful putts as a function of distance.
- Error bars associated to each point j are classical standards deviations $\sqrt{\hat{p}_j(1-\hat{p}_j)/n_j}$.



[1] Source: Don Berry's 1996 textbook (Statistics, A Bayesian Perspective)

A Model for Golf Putting – Logistic Regression

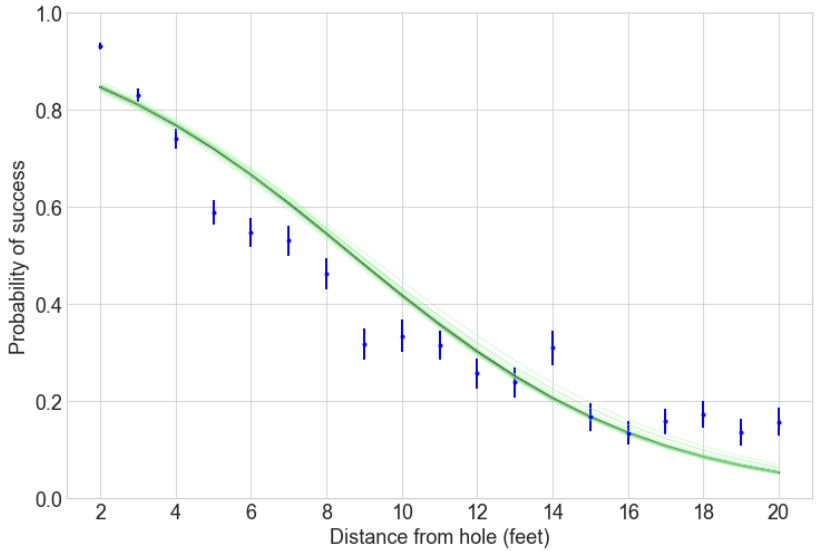
The Model

$$
\begin{aligned}
y_j &\sim \text{Binomial}(n_j, p_j) \\
\text{logit}(p_j) &= a + bx_j; \text{ for } j = \{1, \ldots, J\} \\
a &\sim \text{Normal}(0, 1) \\
b &\sim \text{Normal}(0, 1)
\end{aligned}
$$

A Model for Golf Putting – Logistic Regression

The Model

$$
\begin{aligned}
y_j &\sim \text{Binomial}(n_j, p_j) \\
\text{logit}(p_j) &= a + bx_j; \text{ for } j = \{1, \ldots, J\} \\
a &\sim \text{Normal}(0, 1) \\
b &\sim \text{Normal}(0, 1)
\end{aligned}
$$

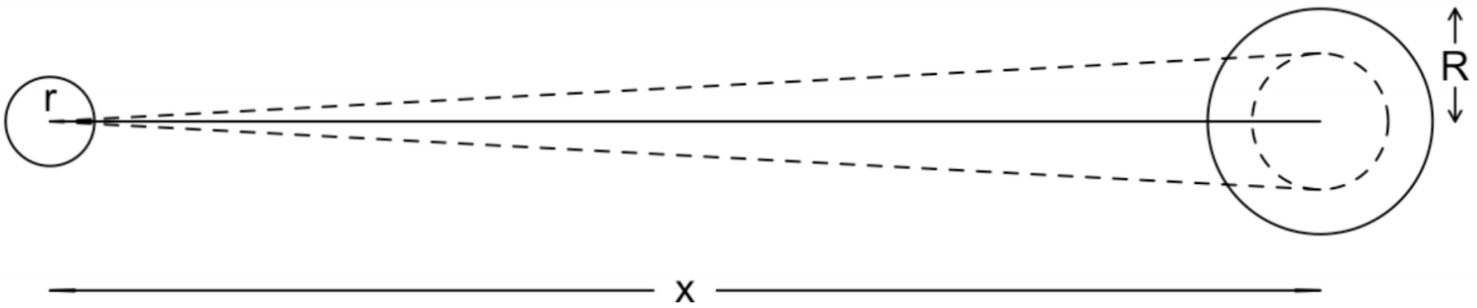


	mean	std	median	5.0%	95.0%	n_eff	r_hat
a	2.23	0.06	2.22	2.13	2.32	2274.20	1.00
b	-0.26	0.01	-0.26	-0.27	-0.24	2294.94	1.00

Number of divergences: 0

A Model for Golf Putting – Principled Approach

- The dotted line represents the angle within which the ball of radius r must be hit so that it falls within the hole of radius R.



- The threshold angle is given by $\sin^{-1}((R-r)/x)$.
- Next, we need to model the human error:

$$\text{Angle} \sim \text{Normal}(0, \sigma)$$

- The probability the ball goes in the hole $\Rightarrow$ Probability that the angle is less than the threshold

$$\Pr\left(|\text{angle}| \leq \sin^{-1}\left((R-r)/x\right)\right) = 2\Phi\left(\frac{\sin^{-1}\left((R-r)/x\right)}{\sigma}\right) - 1$$

A Model for Golf Putting – Principled Model (cont'd)

The Model

$$y_j \sim \text{Binomial}(n_j, p_j)$$

$$p_j = 2\Phi\left(\frac{\sin^{-1}\left((R-r)/x_j\right)}{\sigma}\right) - 1$$

where r=1.68; R=4.25 inches.

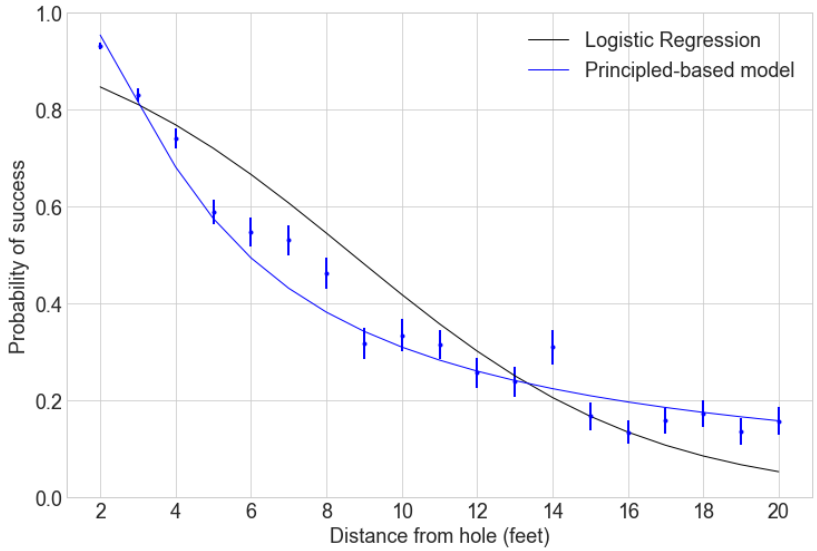
A Model for Golf Putting – Principled Model (cont'd)

The Model

$$y_j \sim \text{Binomial}(n_j, p_j)$$

$$p_j = 2\Phi\left(\frac{\sin^{-1}\left((R-r)/x_j\right)}{\sigma}\right) - 1$$

where r=1.68; R=4.25 inches.

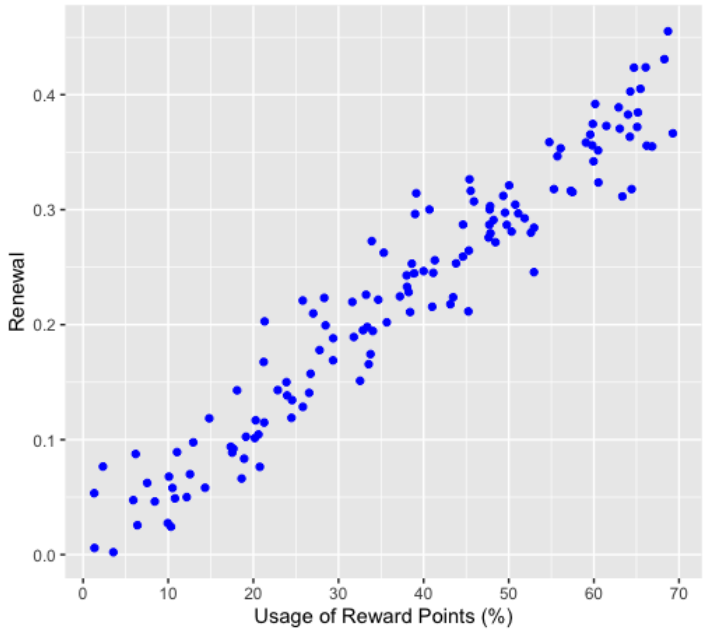


	mean	std	median	5.0%	95.0%	n_eff	r_hat
sigma	0.03	0.00	0.03	0.03	0.03	2562.56	1.00

Number of divergences: 0

A Client Attrition Model

- **Consider the following problem**: we are tasked to build a client attrition model for RBC Visa Avion Cards, and we find the following "insight":



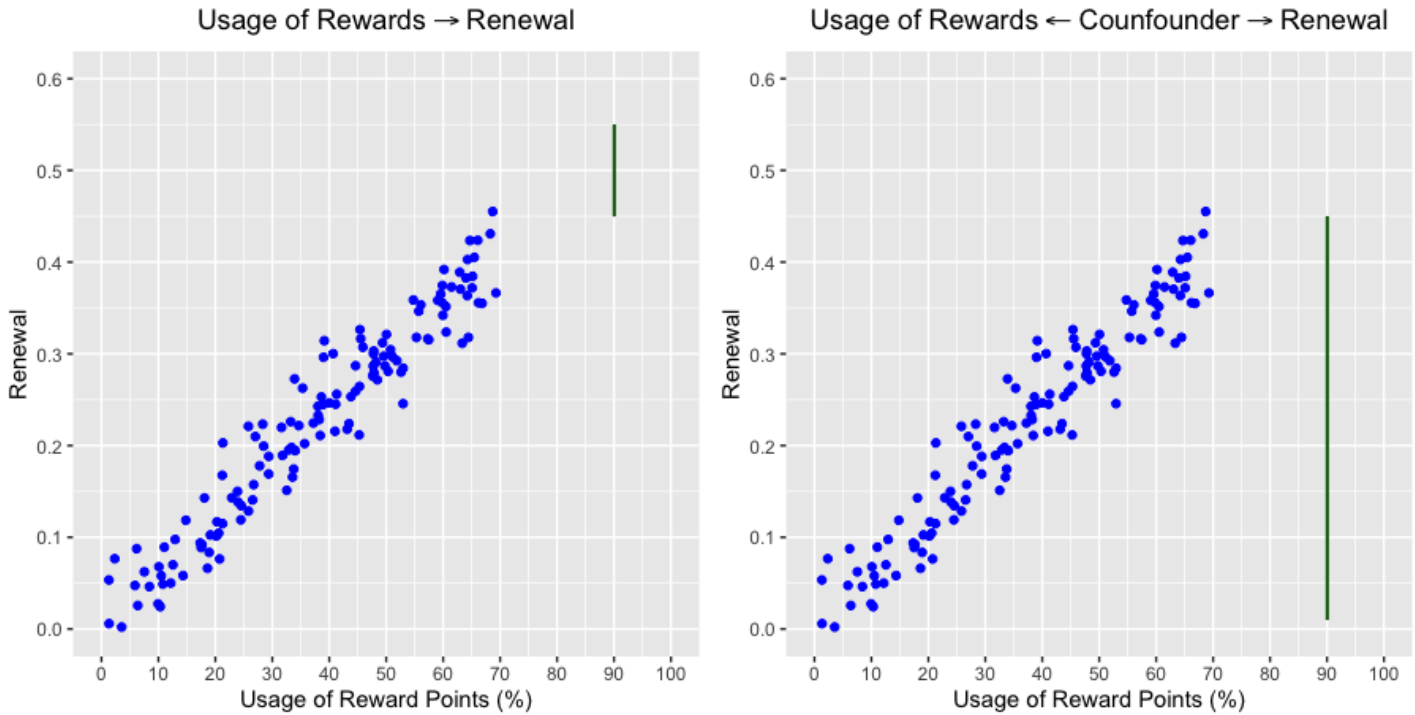
- **What is our best prediction of renewal** if:
  1. we were to observe a client using 90% of her reward points?
  2. create an incentive program to promote usage, and get this client to use 90% of her reward points?

A Client Attrition Model (cont'd)

- **Case 1** is a well-studied problem in ML: we just need to fit a model and use it to make a prediction under the new X (renewal points).
- **Case 2** requires us to think about the data generating process, and in particular, the optimal prediction should depend on the underlying causal structure.

A Client Attrition Model (cont'd)

- **Case 1** is a well-studied problem in ML: we just need to fit a model and use it to make a prediction under the new X (renewal points).
- **Case 2** requires us to think about the data generating process, and in particular, the optimal prediction should depend on the underlying causal structure.



- If we do not accept any form of causal notion, we cannot distinguish between these two cases and our best prediction must be: "I do not know."!