Kuan-Ting Chen
Victoria Delk
Aaron Lee
Charles Zhang

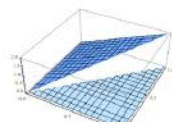
Stats 100A Homework #2

Reading Section Q1

a) We study the following bivariate density function in lecture 23:

Gasoline is to be stocked in a bulk tank once each week and then sold to customers. Let X_1 denote the proportion of the tank that is stocked in a particular week, and let X_2 denote the proportion of the tank that is sold in th same week. Due to limited supplies, X_1 is not fixed in advance but varies from week to week. Suppose that a study of many weeks shows the joint relative frequency behavior of X_1 and X_2 to be such that the following join density function provides and adequate model:

$$f(x_1,x_2)=\begin{cases}3x_1 & 0\le x_2\le x_1\le 1\\ 0 & \text{elsewhere}\end{cases}$$



like I told you, X1 is

Calculate all the density functions needed and plug them in the formula for Bayes' theorem given above. Confirm that Bayes' theorem formula holds. Show work for the ones not shown in the lecture, and use the formulas given in the lecture if they have been already calculated there.

$f(x_1,x_2)=3x_1,\quad 0\le x_2\le x_1\le 1$

Given (from lecture):

$f(x_1)=3x_1^2,\quad 0\le x_1\le 1$

$f(x_2)=\frac{3}{2}(1-x_2^2),\quad 0\le x_2\le 1$

Find the posterior probability density functions of x_1 and x_2:

$f(x_1|x_2)=\frac{f(x_1,x_2)}{f(x_2)}$ $\qquad$ $f(x_2|x_1)=\frac{f(x_1,x_2)}{f(x_1)}$

$f(x_1|x_2)=\frac{3x_1}{\frac{3}{2}(1-x_2^2)}$ $\qquad$ $f(x_2|x_1)=\frac{3x_1}{3x_1^2}$

$\boxed{f(x_1|x_2)=\frac{2x_1}{1-x_2^2}}$ $\qquad$ $\boxed{f(x_2|x_1)=\frac{1}{x_1}}$

Plug into Bayes' Theorem:

$f(x_1|x_2)=\frac{f(x_2|x_1)f(x_1)}{f(x_2)}$

$\frac{2x_1}{1-x_2^2}=\frac{\frac{1}{x_1}\cdot 3x_1^2}{\frac{3}{2}(1-x_2^2)} \Rightarrow \frac{2x_1}{1-x_2^2}=\frac{2x_1}{1-x_2^2}$ ✓

$\boxed{\text{Bayes' Theorem holds}}$

b) We study in lecture 24 the bivariate Gaussian distribution. We would like to find out what distribution the posterior density function of x given y is. Calculate all the density functions needed and plug them in the formula for Bayes' theorem given above. Confirm that Bayes' theorem formula holds. You may use formulas given already in the lecture. Make sure you do the necessary simplifications in order to identify the model for the posterior density.

Find the posterior density function $f(x|y)$

Given (from lecture):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\}$$

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}$$

$\Rightarrow$ Marginal densities of X and Y

$$\underbrace{f(y|x) = \frac{1}{\sigma_y\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left[y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right]^2\right\}}_{\text{Posterior density function of Y given X}}$$

$$\underbrace{f(x|y) = \frac{1}{\sigma_x\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{1}{2\sigma_x^2(1-\rho^2)}\left[x-\mu_x-\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right]^2\right\}}_{\text{Posterior density function of X given Y}}$$

Plug into Bayes' Theorem:

$$f(x|y) = \frac{f(y|x)f(x)}{f(y)}$$

$$\frac{1}{\sigma_x\cancel{\sqrt{2\pi(1-\rho^2)}}} \exp\left\{-\frac{1}{2\sigma_x^2(1-\rho^2)}\left[x-\mu_x-\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right]^2\right\} = \frac{\frac{1}{\sigma_y\cancel{\sqrt{2\pi(1-\rho^2)}}} \exp\left\{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left[y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right]^2\right\}\frac{1}{\cancel{\sqrt{2\pi}}\sigma_x}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\}}{\frac{1}{\cancel{\sqrt{2\pi}}\sigma_y}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}}$$

$$\frac{1}{\sigma_x} \exp\left\{-\frac{1}{\cancel{2}\sigma_x^2(1-\rho^2)}\left[x-\mu_x-\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right]^2\right\} = \frac{\exp\left\{-\frac{1}{\cancel{2}\sigma_y^2(1-\rho^2)}\left[y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right]^2-\frac{1}{\cancel{2}}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\}}{\sigma_x\exp\left\{\frac{1}{\cancel{2}}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}}$$

$$\frac{1}{\sigma_x^2(1-\rho^2)}\left[x-\mu_x-\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right]^2 = \frac{1}{\sigma_y^2(1-\rho^2)}\left[y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right]^2+\left(\frac{x-\mu_x}{\sigma_x}\right)^2-\left(\frac{y-\mu_y}{\sigma_y}\right)^2$$

$$\frac{\left[x-\mu_x-\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right]^2}{\sigma_x^2(1-\rho^2)}+\left(\frac{x-\mu_x}{\sigma_x}\right)^2 = \frac{\left[y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right]^2}{\sigma_y^2(1-\rho^2)}-\left(\frac{y+\mu_y}{\sigma_y}\right)^2$$

$$\frac{\cancel{(1-\rho^2)}(x-\mu_x)^2}{\sigma_x^2\cancel{(1-\rho^2)}}+\left(\frac{x+\mu_x}{\sigma_x}\right)^2 = \frac{\cancel{(1-\rho^2)}(y-\mu_y)^2}{\sigma_y^2\cancel{(1-\rho^2)}}-\left(\frac{y+\mu_y}{\sigma_y}\right)^2$$

$$0 = 0 \checkmark$$

$$\boxed{\text{Bayes' Theorem holds}}$$

Reading Section Q2

Based solely on the prior probabilities for the models - the past history (looking at columns 1 and 2), what is the probability that player A is better than player B? Which of the players would you bet on? Explain why referring to the table.

We assume that in order for player A to be better than player B, they must win more games than player B in a given year. Since `x` is the proportion of games won in a given year by player A, this essentially means we are looking for the prior probability `P(x > 0.5)`. Using the discrete model described by the table, we can see that the domain of x is $\{ 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1 \}$ and therefore, we can say that $P(x > 0.5) = P(0.6) + P(0.7) + P(0.8) + P(0.9) + P(1)$. Plugging in the values from the table, we get $P(x > 0.5) = 0.65$. By this calculation, we can see that the probability that player A is better than player B is 65%, based solely on the prior probabilities for the models.

In order to determine who I would bet on, I would have to look at the probability that player B is better than player A. Using the same logic as the above statement, we can say that we're looking for $P(x < 0.5)$ and that $P(x > 0.5) = P(0) + P(0.1) + P(0.2) + P(0.3) + P(0.4) = 0.2$. This means that there is a 20% chance that player B is better than player A, based on the prior probabilities. As a result, if given no extra information, I would bet on player A, as there is a higher prior probability that player A is better.

Reading Section Q3

Based on the posterior probabilities, given the observed data, what is the probability that A is better than B? Which player would you bet on? What is the posterior probability that A and B are equally good? After seeing the posterior distribution, did you change the bet you suggested in question 1? Explain why using the information in the table and your answers to reading questions 1 and 2.

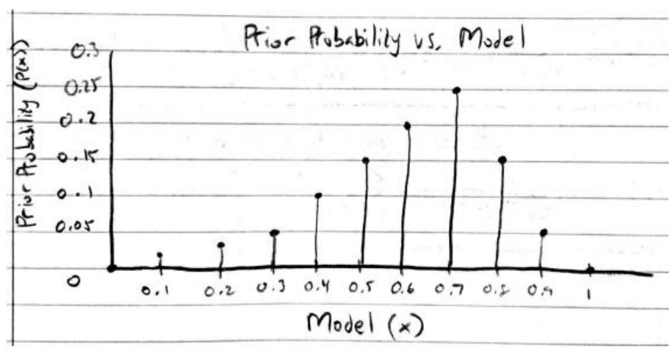
To calculate the probability that A is better than B, we repeat the process used in Q2, but this time, we use the posterior probabilities in the table rather than the prior ones. As a result, we see that $P(x > 0.5|y) \approx 0.47$. Therefore, the approximate probability that A is better than B, given the observed data, is 47%. Once again, to determine which player I would bet on, I would look at the probability that player B is better than player A. We would do this using the same process as in Q2, but this time, we use the posterior probabilities rather than the prior probabilities. This calculation tells us $P(x < 0.5|y) \approx 0.32$, or that player B has approximately a 32% chance of being better than player A. As a result, we would once again bet on player A, as they have a higher posterior probability of being the better player, given the observed data.

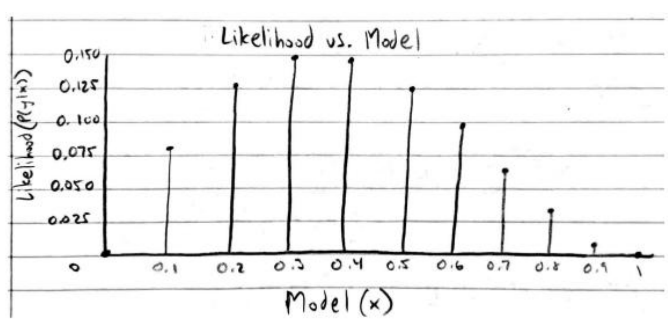
In order to determine the posterior probability that A and B are equally good, we must find the posterior probability that they win the same number of games. This means we must locate $P(x = 0.5|y)$ in the posterior probability column in the table. Doing so tells us that $P(x = 0.5|y) \approx 0.22$, or that there is approximately a 22% chance that A and B are equally good, given the observed data.

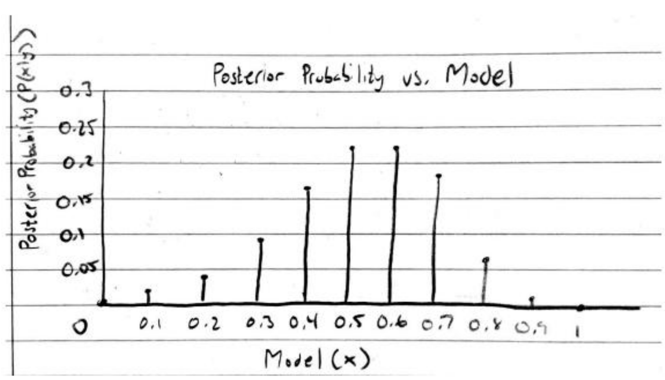
We didn't change the player we bet on after seeing the posterior distribution. Based on the posterior probabilities calculated using Bayes' Theorem for the given data, we see that there is still a higher probability of player A being the better player, even after observing the new data.

Reading Section Q4

a) Plot the prior probability against the model in one graph, using only lines to represent the probabilities (see the graph in the title slide of lecture 16, week 5; you must use that type of graph, not bar graphs). Plot the Likelihood against model in a separate graph right below the first graph, and plot the posterior probability right below the other two, also against the model. In all the plots, the horizontal axis is the model column. Label the axes. Use the same scale on the vertical axis of the prior and the posterior. You may not use a bar graph. Use instead a line graph like the one on the front of the pre-recorded lecture on the negative binomial.







b) Compare the posterior with the prior distribution plots. Has the data had any effect on the prior distribution? Explain why you answer the way you answer, referring to your graphs.

No, the data has had no effect on the prior distribution. By definition, the prior distribution is created using the probabilities given *prior* to any data collection. Therefore, any new data cannot possibly be reflected in the prior distribution, which will always be represented by the first graph for this given dataset. Instead, the new data influences the creation of the posterior distribution. As seen by comparing the first and third graphs, the given data that player A wins game 1 and loses games 2 and 3 pushes the distribution to the left. Intuitively, this makes sense, since player A is guaranteed to have at least 2 losses, resulting in a smaller number of potential wins to gain. Therefore, it becomes increasingly likely for player A to win a fewer number of games compared to the prior distribution.

Reading Section Q5

Suppose that `X` can be modeled by a Beta distribution and suppose that `Y` follows a Binomial model `P(Y=y|X=x,n=3)`. What would be the formula for the posterior distribution of `X` given `Y`? To answer, substitute the formulas for these models in the formula for Bayes theorem given above, in the numerator. The Beta is a density function, and the Binomial is a pmf. So notice that we can mix distributions in the Bayes formula too. By looking at the formula for the posterior, and comparing with the formulas used, can you identify the density model for the posterior? Use the Wikipedia parameterization of the Beta density function (at the top of Wikipedia's page after you google Beta distribution). Observe in Wikipedia's plots what the density looks like.

Beta distribution: $\frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$

Binomial model: $\binom{n}{k} p^k (1-p)^{n-k}$

Likelihood:

$$P(Y=y \mid X=x, n=3) = \binom{3}{y} x^y (1-x)^{3-y}$$

Prior:

$$P(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

Postelor:

$$P(x \mid Y, n=3) = \frac{\left[\binom{3}{y} x^y (1-x)^{3-y}\right]\left[\frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\right]}{P(Y)}$$

$$P(x \mid Y, n=3) = \underbrace{\left[\frac{\binom{3}{y}}{B(\alpha,\beta)P(Y)}\right]}_{\text{Constant}}\left[x^y(1-x)^{3-y} x^{\alpha-1}(1-x)^{\beta-1}\right]$$

$$P(x \mid Y, n=3) \propto x^{y+\alpha-1}(1-x)^{3-y+\beta-1}$$ Beta distribution

Let $\gamma = y+\alpha$ and $\delta = 3-y+\beta$

$$\boxed{P(x \mid Y, n=3) = \frac{x^{\gamma-1}(1-x)^{\delta-1}}{B(\gamma,\delta)}}$$

The density model for the posterior is clearly a beta distribution. However, unlike the beta distribution parameterized with α and β that was used for the prior of X, the posterior's beta distribution is parameterized with γ and δ, as defined in the work shown above. The new parameters account for the likelihood through Bayes' Theorem to produce the posterior distribution.

Reading Section Q6

Suppose that instead of having the data "A wins the first game" we had the data "A wins 1 game of the three games already played this current year." That is very different information than the one we used to construct the table provided. Modify the calculations in the table, and write the whole table again, with the new information for all columns that will need change. Compare the new prior and posterior probabilities and determine which player is better now, A or B, based on this information. How do these new calculations change your answers to questions 3 and 4?

Based on the new information, we cannot simply use the product rule for independent events to calculate the likelihood as we did in Q2. Instead, we must now use a binomial model. Doing so generates the following table:

Model	Prior	Likelihood	Prior × Likelihood	Posterior	Model × Posterior
0	0	$({}_3C_1)(0)(1)^2 = 0$	0	0	0
0.1	0.02	$({}_3C_1)(0.1)(0.9)^2 = 0.243$	0.00486	0.01880	486/258480
0.2	0.03	$({}_3C_1)(0.2)(0.8)^2 = 0.384$	0.01152	0.04457	2304/258480
0.3	0.05	$({}_3C_1)(0.3)(0.7)^2 = 0.441$	0.02205	0.08531	6615/258480
0.4	0.10	$({}_3C_1)(0.4)(0.6)^2 = 0.432$	0.04320	0.16713	17280/258480
0.5	0.15	$({}_3C_1)(0.5)(0.5)^2 = 0.375$	0.05625	0.21762	28125/258480
0.6	0.20	$({}_3C_1)(0.6)(0.4)^2 = 0.288$	0.05760	0.22284	34560/258480
0.7	0.25	$({}_3C_1)(0.7)(0.3)^2 = 0.189$	0.04725	0.18280	33075/258480
0.8	0.15	$({}_3C_1)(0.8)(0.2)^2 = 0.096$	0.01440	0.05571	11520/258480
0.9	0.05	$({}_3C_1)(0.9)(0.1)^2 = 0.027$	0.00135	0.00522	1215/258480
1	0	$({}_3C_1)(1)(0)^2 = 0$	0	0	0
Sum	1.00		0.25848	1	135180/258480

Let the new data be represented by z. Repeating the process used in Q3 to find the probability of A being the better player based on the given data, we see that $P(x > 0.5|z) \approx 0.47$ and $P(x < 0.5|z) \approx 0.32$. From this, we can see that the probabilities for A being better and for B being better are both the same as in Q3. This can be confirmed by the fact that the posterior probabilities for all models are the same in the new table as in the original table. This can be attributed to how, in the context of the problem, it doesn't matter which of the first 3 games player A won, it only matters that player A won 1 out of 3 games. As a result the probability that A is better than B would remain the same given both sets of data. For the reasons listed above, our answer to Q3 wouldn't change, since all posterior probabilities remain the same. For Q4, the prior and posterior probability charts would remain the same, but the likelihood chart would increase its scale by 3x, as it is ${}_3C_1 = 3$ times as likely for z to occur than y. This can be shown mathematically by the addition of the ${}_3C_1$ factor to the likelihood calculation, or intuitively by how z may occur if A wins the first, second, or third game and loses the rest, but y may only occur if A wins the first game and loses the rest.

Lecture Video Q1

In Module 4 lesson 11B, we fitted a lognormal density model to the radon data. We said there that we used a method called maximum likelihood, which is a method used by statisticians to estimate parameters of probability models from data. The function calculated is the likelihood function, which is the joint distribution of the data observed given the parameters, the $f(y|x)$-equivalent seen in the Bayes formula given at the beginning of the reading section in this homework. That is obtained by doing the calculation we did in week 6, the video supplementing lecture 21, but looking at the joint distribution as a function of the parameters.

a) After you watch the video "Introduction to Bayesian Data Analysis Part 1," posted right above, compare the two ways of estimating the parameters of a distribution, and indicate what else would we need in the estimation of the lognormal distribution for the radon data in order to do the estimation of the parameters in a Bayesian one. Refer specifically to the examples seen in lecture 11B, the "Introduction to Bayesian Data Analysis Part 1" video, and the example of the two players in the reading.

The video makes a distinction between the underlying rate at which people sign up, or the prior probability, and the conditional probability. The maximum likelihood model is used for the estimation of the μ and σ parameters for the radon lognormal distribution example seen in lecture. The "Introduction to Bayesian Data Analysis" mentions that data, a generative model, and priors are necessary in finding unknowns. The process of "fitting" the generative model and fine-tuning the parameters is performed by comparing the prior and posterior probabilities and distributions. Specifically, in machine learning, this is referred to as backpropagation. The example of the two players tells us that a prior probability and model are necessary in order to calculate the posterior probabilities.

b) The video before this "Introduction to Bayesian Data Analysis" posted above mentions several applications of Bayesian data analysis that are crucial to today's progress in discovery in the world. Give those examples mentioned in the video. As you watch that video you probably are also thinking about the slide in the video posted in our "Getting Ready" module, the video titled "Why is probability so important in machine learning?" There, the authors mention conditional probability. When the author says in this last video that machine learning uses conditional probability, which conditional probabilities is it referring to in the Bayes formula? Justify your answer by providing examples of applications of that conditional probability mentioned in the video.

Bayesian data analysis uses probability to present uncertainty in all parts of a statistical model. In order to perform this analysis, data, a generative model, and prior probabilities are necessary. The generative model example used in the video is the Swedish Fish signup which attempts to predict how many people sign up to buy the Swedish Fish product. The video explains that Bayesian probability in generative models applies to the sign-up problem because there is a prior and posterior distribution, which are probabilities calculated given an existing distribution. Bayes' Formula, states `P(X | Y) = P(X, Y) / P(Y)`; in the Swedish Fish example, the prior underlying sign up rate, `P(Y)`, is `0.55`. Additionally, in the video about conditional probability in machine learning, event `X` refers to the event that the human actually has a disease, while event `Y` is the event that the test returns a positive result. The video shows that the conditional probability that the individual actually has the disease considering the inaccuracies of the test are relatively low. Bayes' Theorem is crucial to understanding heuristics in AI and how generative vs. discriminatory models are constructed.

c) Give a couple of generative model examples given in the video posted above ("Introduction to Bayesian Data Analysis"). What is the difference between a generative model and the models we are studying in this class (probability mass functions and density models).

Generative models (in artificial intelligence and other applications) predict outputs based on prior probabilities and data distributions, while considering joint probabilities. Discriminative models use Bayesian statistics and conditional probabilities, `P(Y | X = x)`. In the field of machine learning, generative models would create a set of pixels resembling an object, for example, a cat, given a prior training dataset that helps the model learn and adjust values of parameters to successfully generate an image of a cat with the characteristics in common (values of trained parameters) of the original dataset.

Conventional classification models operate on the discriminative principle using probability mass functions and density models, which differ from generative models. In simple terms, Bayes Theorem can be applied for classification or discriminative models with these probabilities: `P(class | data) = P(data | class) × P(class) / P(data).` Here, the joint probability, `P(data, class)` is equal to `P(data | class) × P(class)`, which is the numerator of the Bayes' Theorem. `P(data)` represents the prior probability of data the model has seen or trained on, which is then used to find the conditional probability, `P(class | data)`, of the object falling into the "class" category. Essentially, discriminative models calculate the probability that an object falls into a certain category, given the prior training datasets of similar objects that it sees. This concept is extensively used in image classification in artificial intelligence.

Additional Activities Q1

a) Your team will also need to think about a sport that involves two key players (horses, humans, etc.) in real life and add what type of data and information you would need to collect in order to perform a Bayesian data analysis. In your description, be specific as to where you would look for that data and what that data might look like. Specific examples preferred.

The sport that we will be taking a look at is Formula 1 racing. We would like to look at a driver's chance of winning a race, given his starting position. In order to perform a Bayesian data analysis, we will need three items, as per the Lecture Video: data, a generative model, and a prior. To gather our data, we can simply take a look at the results of all the past races, which is available online. From this, we can find the actual percentage that a driver finished first, given any starting position. The second item we will need is a generative model. For this, we can use a binomial model to determine the driver's finishing position. Since each race is a Bernoulli trial (the driver either wins or he doesn't) and that each race is independent of each other, we can use this model. The last thing we need is a prior, or probabilities to find the result before we have any data. For this, we will use a uniform density, allowing us to assume that each starting position has an equal likelihood of winning the race. The data for starting positions and race placements can be found online on many websites, but officially on the Formula 1 site. There are tables that give the starting positions of each driver, the place in which they finished.

b) Gathertown has a game going on now, Escape the Island, that can be played with your team. I think you have to login with a guest account in order to access it. I am writing here the link to access it. Link (links to an external site). Preferably 4 people should play, so coordinate with your team to play it. After playing it, do you think Bayesian statistics could be used to predict the outcome of this game? Why or why not? You may want to introduce how the game goes first, because the reader needs to know.

Gathertown's Escape the Island, is an escape room type game where players have to solve puzzles to, as the title suggests, "escape the island". The game begins with you and your team obtaining tickets to an island. Once there, the team needs to solve several puzzles and complete challenges to escape. The game requires between 2-4 players, so a Bayesian analysis that could be done is what is the chance of escape given the number of players. One would assume that more people playing would lead to more insights on the puzzles, and therefore increase the chances of escape. To perform a Bayesian data analysis, we would need three things. First is our data, or the proportion of teams who have beat the game by the number of players. This is something that can only be obtained from the Gathertown company. Next is our generative model. In an escape room, you either escape or you don't, making it a Bernoulli trial. And assuming that every team operates independently of each other, we can use the binomial model to determine the chance that a team with 2, 3, or 4 players beats the game. Finally, we need a prior, which in this case will be the uniform model, which assumes that the chance of escaping the island is the same for any number of players. If we have all three of these components, we can perform a Bayesian data analysis to find the probability that a team wins given their number of players.