Session 1
$\mathbb{Q}$uantitative Analysis of Financial Markets
Bayesian Analysis

Christopher Ting

http://www.mysmu.edu/faculty/christophert/

Christopher Ting

: christophert@smu.edu.sg

: 6828 0364

: LKCSB 5036

October 9, 2018

Table of Contents

Introduction

- **Thomas Bayes (1702 to 176)**
- Bayesian analysis is a broad topic.
- Some very difficult problems in risk management can be tackled by conducting a Bayesian analysis.
- For two random variables, A and B, **Bayes' theorem** states that

$$\mathbb{P}\left(A \middle| B\right) = \frac{\mathbb{P}\left(B \middle| A\right) \mathbb{P}\left(A\right)}{\mathbb{P}\left(B\right)}.$$

Learning Outcomes of QA04

Chapter 6. (pp. 113-124 only)

Michael Miller, Mathematics and Statistics for Financial Risk Management, 2nd Edition (Hoboken, NJ: John Wiley & Sons, 2013).

- Describe Bayes' theorem and apply this theorem in the calculation of conditional probabilities.
- Compare the Bayesian approach to the frequentist approach.
- Apply Bayes' theorem to scenarios with more than two possible outcomes.

Overview of Applications

- **Predictive modeling is the process by which a model is created to predict an outcome.**
- Classification: for categorical outcomes
- Regression: for numerical outcomes
- Clustering: for cluster outcomes
- Association: for any interesting associations amongst observations.

An Introductory Problem

- How can you predict the value of a categorical variable (class) from an observable variable (attribute)?
- You know the prior probability $\mathbb{P}(C)$ of class C.
- Attribute is also known as the predictor. You know the predictor prior probability $\mathbb{P}(A)$ of the predictive attribute A.
- Most importantly, you know the likelihood $\mathbb{P}\left(A|C\right)$ of attribute A for a given class C.
- What is the posterior probability $\mathbb{P}\left(C|A\right)$ of class C given the attribute A?

Algorithm: Bayes Theorem

Prior

$$\mathbb{P}\left(C\middle|A\right) = \frac{\mathbb{P}\left(A\middle|C\right)\ \mathbb{P}\left(C\right)}{\mathbb{P}\left(A\right)}$$

Algorithm: Bayes Theorem

- Prior
- Likelihood

$$\mathbb{P}\left(C \middle| A\right) = \frac{\mathbb{P}\left(A \middle| C\right)\ \mathbb{P}\left(C\right)}{\mathbb{P}\left(A\right)}$$

Algorithm: Bayes Theorem

- Prior
- Likelihood

$$\mathbb{P}\left(C\middle|A\right) = \frac{\mathbb{P}\left(A\middle|C\right)\ \mathbb{P}\left(C\right)}{\mathbb{P}\left(A\right)}$$

- Predictor Prior

Algorithm: Bayes Theorem

- Prior
- Likelihood

$$\mathbb{P}\left(C\middle|A\right) = \frac{\mathbb{P}\left(A\middle|C\right)\mathbb{P}\left(C\right)}{\mathbb{P}\left(A\right)}$$

- Predictor Prior
- Posterior

Example: Bond Default

- Assume we have two bonds, Bond A and Bond B, each with a 10% probability of defaulting over the next year. Further assume that the probability that both bonds default is 6%, and that the probability that neither bond defaults is 86%.

Table: **Probability Matrix**

		Bond A		
		No Default	Default	
Bond B	No Default	86%	4%	90%
	Default	4%	6%	10%
		90%	10%	100%

- Because bond issuers are often sensitive to broad economic trends, **bond defaults are often highly correlated**.

Joint and Conditional Probabilities: Q&A

- What is the probability that Bond A defaults, given that Bond B has defaulted?
- Bond B defaults in 10% of the scenarios, but the probability that both Bond A and Bond B default is only 6%.

$$\mathbb{P}\left(A \middle| B\right) = \frac{\mathbb{P}\left(A \cap B\right)}{\mathbb{P}\left(B\right)} = \frac{6\%}{10\%} = 60\%.$$

- Answer: Bond A defaults in 60% of the scenarios in which Bond B defaults.
- Notice that the conditional probability $\mathbb{P}\left(A \middle| B\right)$ is different from the unconditional probability $\mathbb{P}\left(A\right)$ of Bond A defaulting is 10%.

Bayes' Theorem

- Joint probability

$$\mathbb{P}\left(A \cap B\right) = \mathbb{P}\left(A \middle| B\right)\mathbb{P}\left(B\right)$$

can be written as

$$\mathbb{P}\left(A \cap B\right) = \mathbb{P}\left(B \middle| A\right)\mathbb{P}\left(A\right)$$

since the joint probability is symmetric:

$$\mathbb{P}\left(A \cap B\right) = \mathbb{P}\left(B \cap A\right).$$

- Combining the right-hand side of both of these equations and rearranging terms leads us to Bayes' theorem:

$$\mathbb{P}\left(A \middle| B\right) = \frac{\mathbb{P}\left(B \middle| A\right)\mathbb{P}\left(A\right)}{\mathbb{P}\left(B\right)}.$$

Sample Problem 1: A Riddle

- Suppose a skillful fund manager is just 1 in every 100.
- A test is 99% accurate in identifying that a fund manager is skillful.
- For skillful fund managers, the test is 99% correct. Similarly, for those who are not skillful, the test correctly indicates that they are not skillful with 99% of accuracy.
- If a fund manager takes the test and the result of the test is positive, what is the probability that the fund manager is actually skillful?
- It should be 99%, right? NO!

Answer to Sample Problem 1

- Using p to represent a positive test result, the probability of p can be calculated as

$$\begin{aligned}\mathbb{P}(\mathrm{p}) &= \mathbb{P}(\mathrm{p} \cap \text{ skillful}) + \mathbb{P}(\mathrm{p} \cap \text{ unskillful}) \\ &= \mathbb{P}(\mathrm{p} | \text{skillful})\, \mathbb{P}(\text{skillful}) + \mathbb{P}(\mathrm{p} | \text{unskillful})\, \mathbb{P}(\text{unskillful}) \\ &= 99\% \times 1\% + 1\% \times 99\% \\ &= 2\% \times 99\%.\end{aligned}$$

- We can then calculate the probability of having the skillful fund manager given a positive test using Bayes' theorem as follows:

$$\mathbb{P}(\text{skillful} | \mathrm{p}) = \frac{\mathbb{P}(\mathrm{p} | \text{skillful})\, \mathbb{P}(\text{skillful})}{\mathbb{P}(\mathrm{p})} = \frac{99\% \times 1\%}{2\% \times 99\%} = 50\%.$$

Sample Problem 2

- Star managers will beat the market in any given year with a probability of 75%.
- Nonstar managers are just as likely to beat the market as they are to underperform it.
- The probability of beating the market is independent from one year to the next.
- Of a given pool of managers, only 16% turn out to be stars.
- A new manager was added to your portfolio three years ago. Since then, the new manager has beaten the market every year.
  1. What was the probability that the manager was a star when the manager was first added to the portfolio?
  2. What is the probability that this manager is a star now?
  3. After observing the manager beat the market over the past three years, what is the probability that the manager will beat the market next year?

Answer to Sample Problem 2-1

- The probability that a manager beats the market given that the manager is a star is 75%:

$$\mathbb{P}\left(\mathtt{B}\middle|\mathtt{star}\right) = 75\% = \frac{3}{4}.$$

- The probability that a nonstar manager will beat the market is 50%:

$$\mathbb{P}\left(\mathtt{B}\middle|\mathtt{nonstar}\right) = 50\% = \frac{1}{2}.$$

- At the time the new manager was added to the portfolio, the probability that the manager was a star was just the probability of any manager being a star, 16%, the unconditional probability.

$$\mathbb{P}\left(\mathtt{star}\right) = 16\% = \frac{4}{25}.$$

Answer to Sample Problem 2-2

- We need to find $\mathbb{P}\left(\mathtt{star}\middle|3\mathrm{B}\right)$, the probability that the manager is a star, given that the manager has beaten the market three years in a row ($3\mathrm{B}$).

$$\mathbb{P}\left(\mathtt{star}\middle|3\mathrm{B}\right) = \frac{\mathbb{P}\left(3\mathrm{B}\middle|\mathtt{star}\right)\mathbb{P}\left(\mathtt{star}\right)}{\mathbb{P}\left(3\mathrm{B}\right)}$$

- Because outperformance is independent from one year to the next, $\mathbb{P}\left(3\mathrm{B}\middle|\mathtt{star}\right)$ is just the probability that a star beats the market in any given year to the third power.

$$\mathbb{P}\left(3\mathrm{B}\middle|\mathtt{star}\right) = \left(\frac{3}{4}\right)^3 = \frac{27}{64}.$$

Answer to Sample Problem 2-2 (cont'd)

- What is the unconditional probability of beating the market for three years? It is just the weighted average probability of three market beating years over both types of managers:

$$
\begin{aligned}
\mathbb{P}\left(3\mathrm{B}\right) &= \mathbb{P}\left(3\mathrm{B}\middle|\mathtt{star}\right)\mathbb{P}\left(\mathtt{star}\right) + \mathbb{P}\left(3\mathrm{B}\middle|\mathtt{nonstar}\right)\mathbb{P}\left(\mathtt{nonstar}\right) \\
&= \left(\frac{3}{4}\right)^3 \frac{4}{25} + \left(\frac{1}{2}\right)^3 \frac{21}{25} = \frac{69}{400}.
\end{aligned}
$$

- Putting it all together, we get our final result by Bayes' theorem:

$$
\mathbb{P}\left(\mathtt{star}\middle|3\mathrm{B}\right) = \frac{\dfrac{27}{64} \times \dfrac{4}{25}}{\dfrac{69}{400}} = 39\%.
$$

Answer to Sample Problem 2-2 (cont'd)

- Our updated belief about the manager being a star, having seen the manager beat the market three times, is 39%, a significant increase from our prior belief of 16%.
- Even though it is much more likely that a star will beat the market three years in a row, we are still far from certain that this manager is a star. In fact, at 39% the odds are more likely that the manager is not a star.
- A true star needs many years to develop!

Answer to Sample Problem 2-3

- The probability that the (“new”) manager beats the market next year is just the probability that a star would beat the market plus the probability that a nonstar would beat the market, weighted by our new beliefs. Our updated belief about the manager being a star is 39% = 9/23, so the probability that the manager is not a star must be 61% = 14/23:

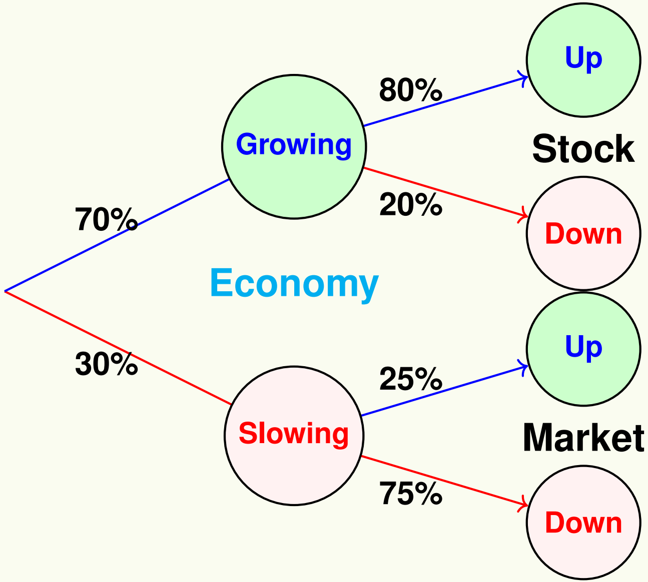
$$\begin{aligned}\mathbb{P}\left(\mathtt{B}\right) &= \mathbb{P}\left(\mathtt{B}\middle|\mathtt{star}\right)\mathbb{P}\left(\mathtt{star}\right)+\mathbb{P}\left(\mathtt{B}\middle|\mathtt{nonstar}\right)\mathbb{P}\left(\mathtt{nonstar}\right)\\ &= \frac{3}{4}\frac{9}{23}+\frac{1}{2}\frac{14}{23}=60\%.\end{aligned}$$

- The probability that the manager will beat the market next year falls somewhere between the probability for a nonstar, 50%, and for a star, 75%, but is closer to the probability for a nonstar.

Prior and Posterior Beliefs and Probabilities

- Before seeing the manager beat the market 3 times, our belief (prior probability) that the manager was a star was 16%.
- After seeing the manager beat the market 3 times, our belief (posterior probability) that the manager was a star increased to 39%.
- The probability of beating the market, assuming that the manager was a star, $\mathbb{P}\left(3\mathrm{B}\middle|\mathtt{star}\right) = 27/64$, was the **likelihood**, also known as **evidence**.
- In other words, the likelihood of the manager beating the market three times, assuming that the manager was a star, was 27/64.

Bayesian Analysis Problem



Tutorial Questions

1. What is the joint probability of a slowing (S) economy and the stock market going up (U)? $\mathbb{P}\left(\mathrm{S} \cap \mathrm{U}\right)$ = ____________________
2. What is the joint probability of a growing (G) economy and the stock market going down (D)? $\mathbb{P}\left(\mathrm{G} \cap \mathrm{D}\right)$ = ____________________.
3. What is the meaning of $\mathbb{P}\left(\mathrm{U} \middle| \mathrm{G}\right)$ and what is its value?
4. $\mathbb{P}\left(\mathrm{U}\right)$ = ____________________.
5. $\mathbb{P}\left(\mathrm{D}\right)$ = ____________________.
6. $\mathbb{P}\left(\mathrm{G} \middle| \mathrm{D}\right)$ = ____________________.
7. $\mathbb{P}\left(\mathrm{S} \middle| \mathrm{S}\right)$ = ____________________.
8. $\mathbb{P}\left(\mathrm{D} \middle| \mathrm{U}\right)$ = ____________________.

Many-State Problem

- In the two previous problems, each variable could exist in only one of two states: a fund manager is either skillful or unskillful; a manager was either a star or a nonstar.

- Can be generalized to three types of manager
  1. Underperformers: Beat the market only 25% of the time
  2. In-line performers: beat the market 50% of the time
  3. Outperformers: Beat the market 75% of the time

Many-State Problem (cont'd)

- Initially we believe that a given manager is most likely to be an inline performer, and is less likely to be an underperformer or an outperformer. More specifically, our prior belief is that a manager has a 60% probability of being an in-line performer, a 20% chance of being an underperformer, and a 20% chance of being an outperformer.

$$\mathbb{P}\left(p = 0.25\right) = 20\%$$
$$\mathbb{P}\left(p = 0.50\right) = 60\%$$
$$\mathbb{P}\left(p = 0.75\right) = 20\%$$

- Now suppose the manager beats the market two years in a row. What should our updated beliefs be?

Answer to Many-State Problem

- We start by calculating the likelihoods, the probability of beating the market two years in a row, for each type of manager:

$$\mathbb{P}\left(2\mathtt{B} \middle| p = 0.25\right) = \left(\frac{1}{4}\right)^2 = \frac{1}{16}$$

$$\mathbb{P}\left(2\mathtt{B} \middle| p = 0.50\right) = \left(\frac{1}{2}\right)^2 = \frac{4}{16}$$

$$\mathbb{P}\left(2\mathtt{B} \middle| p = 0.75\right) = \left(\frac{3}{4}\right)^2 = \frac{9}{16}$$

Answer to Many-State Problem (cont'd)

- The unconditional probability of observing the manager beat the market two years in a row, given our prior beliefs about p, is

$$\mathbb{P}\left(2\mathrm{B}\right) = 20\%\frac{1}{16} + 60\%\frac{4}{16} + 20\%\frac{9}{16} = \frac{44}{160} = 27.5\%.$$

- Putting this all together and using Bayes' theorem, we can calculate our posterior belief that the manager is

  1. underperformer

$$\mathbb{P}\left(p = 0.25 \middle| 2\mathrm{B}\right) = \frac{\mathbb{P}\left(2\mathrm{B} \middle| p = 0.25\right)\mathbb{P}\left(p = 0.25\right)}{\mathbb{P}\left(2\mathrm{B}\right)} = \frac{\frac{1}{16}\frac{2}{10}}{\frac{44}{160}} = \frac{1}{22}.$$

Answer to Many-State Problem (cont'd)

2. in-line performer

$$\mathbb{P}\left(p=0.50\middle|2\mathtt{B}\right)=\frac{\mathbb{P}\left(2\mathtt{B}\middle|p=0.50\right)\mathbb{P}\left(p=0.50\right)}{\mathbb{P}\left(2\mathtt{B}\right)}=\frac{\frac{4}{16}\frac{6}{10}}{\frac{44}{160}}=\frac{12}{22}.$$

3. outperformer

$$\mathbb{P}\left(p=0.75\middle|2\mathtt{B}\right)=\frac{\mathbb{P}\left(2\mathtt{B}\middle|p=0.75\right)\mathbb{P}\left(p=0.75\right)}{\mathbb{P}\left(2\mathtt{B}\right)}=\frac{\frac{9}{16}\frac{2}{10}}{\frac{44}{160}}=\frac{9}{22}.$$

A Shortcut

- For each type of manager, the posterior probability was calculated as $(x_1 = 0.25, x_2 = 0.50, x_3 = 0.75)$

$$\mathbb{P}\left(p = x_i \middle| 2\mathrm{B}\right) = \frac{\mathbb{P}\left(2\mathrm{B} \middle| p = x_i\right)\mathbb{P}\left(p = x_i\right)}{\mathbb{P}\left(2\mathrm{B}\right)}.$$

- In each case, the denominator on the right-hand side is the same, $\mathbb{P}\left(2\mathrm{B}\right) = 44/160$. We can then rewrite this equation in terms of a constant, c

$$\mathbb{P}\left(p = x_i \middle| 2\mathrm{B}\right) = c \times \mathbb{P}\left(2\mathrm{B} \middle| p = x_i\right)\mathbb{P}\left(p = x_i\right).$$

- Now, the sum of all the posterior probabilities must equal one:

$$c\sum_{i=1}^{3}\mathbb{P}\left(2\mathrm{B} \middle| p = x_i\right)\mathbb{P}\left(p = x_i\right) = 1.$$

A Shortcut (cont'd)

- For our current example, we have

$$c\left(\frac{1}{16}\frac{2}{10}+\frac{4}{16}\frac{6}{10}+\frac{9}{16}\frac{2}{10}\right)=c\frac{44}{160}=1.$$

  Solving for c, we obtain $c = 160/44$.

- We then use this result to calculate each of the posterior probabilities.

- For example, the posterior probability that the manager is an underperformer is

$$\mathbb{P}\left(p=0.25\middle|2\mathrm{B}\right)=c\,\mathbb{P}\left(2\mathrm{B}\middle|p=0.25\right)\mathbb{P}\left(p=0.25\right)=\frac{160}{44}\frac{1}{16}\frac{2}{10}=\frac{1}{22}.$$

Sample Problem 3

- Using the same prior distributions as in the preceding example, what would the posterior probabilities be for an underperformer, an in-line performer, or an outperformer, if instead of beating the market two years in a row, the manager beat the market in 6 of the next 10 years?

- Answer:
  For each possible type of manager, the likelihood of beating the market 6 times out of 10 can be determined using a **binomial distribution**

$$\mathbb{P}\left(6\mathrm{B}\middle|p\right) = \binom{10}{6} p^6 (1-p)^4.$$

Answer to Sample Problem 3

- Using the shortcut method, we first calculate the posterior probabilities in terms of an arbitrary constant, c.
- If the manager is an underperformer,

$$\begin{aligned}\mathbb{P}\left(p = 0.25 \middle| 6\mathtt{B}\right) &= c\,\mathbb{P}\left(6\mathtt{B} \middle| p = 0.25\right)\mathbb{P}\left(p = 0.25\right) \\ &= c\binom{10}{6}\left(\frac{1}{4}\right)^6\left(\frac{3}{4}\right)^4 \times \frac{2}{10} \\ &= c\binom{10}{6}\frac{2 \times 3^4}{10 \times 4^{10}}.\end{aligned}$$

Answer to Sample Problem 3 (cont'd)

- Similarly, if the manager is an in-line performer or outperformer, we have

$$\mathbb{P}\left(p = 0.50 \middle| 6\mathrm{B}\right) = c\binom{10}{6}\frac{6 \times 2^{10}}{10 \times 4^{10}}$$

$$\mathbb{P}\left(p = 0.75 \middle| 6\mathrm{B}\right) = c\binom{10}{6}\frac{2 \times 3^{6}}{10 \times 4^{10}}$$

- Because all of the posterior probabilities sum to one, we have

$$\mathbb{P}\left(p = 0.25 \middle| 6\mathrm{B}\right) + \mathbb{P}\left(p = 0.50 \middle| 6\mathrm{B}\right) + \mathbb{P}\left(p = 0.75 \middle| 6\mathrm{B}\right) = 1.$$

Consequently,

$$c = \frac{1}{\binom{10}{6}}\frac{10 \times 4^{10}}{2 \times 3}\frac{1}{1294}.$$

Answer to Sample Problem 3 (cont'd)

- Substituting back into the equations for the posterior probabilities, we have

$$\mathbb{P}\left(p = 0.25 \middle| 6\mathtt{B}\right) = c\binom{10}{6}\frac{2\times 3^4}{10\times 4^{10}} = \frac{3^3}{1294} = 2.09\%$$

$$\mathbb{P}\left(p = 0.50 \middle| 6\mathtt{B}\right) = c\binom{10}{6}\frac{6\times 2^{10}}{10\times 4^{10}} = \frac{2^{10}}{1294} = 79.13\%$$

$$\mathbb{P}\left(p = 0.75 \middle| 6\mathtt{B}\right) = c\binom{10}{6}\frac{2\times 3^6}{10\times 4^{10}} = \frac{3^3}{1294} = 18.78\%$$

Answer to Sample Problem 3 (cont'd)

- In this case, the probability that the manager is an in-line performer has increased from 60% to 79.13%.
- The probability that the manager is an outperformer decreased slightly from 20% to 18.78%.
- It now seems very unlikely that the manager is an underperformer (2.09% probability compared to our prior belief of 20%).

Bayes versus Frequentists

- In Sample Problem 2, we were presented with a portfolio manager who beat the market three years in a row. Shouldn't we have concluded that the probability that the portfolio manager would beat the market the following year was 100% (3/3 = 100%), and not 60%?

- Taking three out of three positive results and concluding that the probability of a positive result next year is 100%, is known as the frequentist approach. The conclusion is based only on the observed frequency of positive results.

- The conclusion is different because the Bayesian approach starts with a prior belief about the probability.

Bayes versus Frequentists (cont'd)

- Which is correct?

- In the example with the portfolio manager, we had only three data points. Using the Bayesian approach for this problem made sense. If there were 1,000 data points, most practitioners would probably utilize frequentist analysis.

- In risk management, performance analysis and stress testing are examples of areas where we often have very little data, and the data we do have is noisy. These areas are likely to lend themselves to Bayesian analysis.

Coin-Flipping Problem

- Binary mutually exclusive event: head is 1 and tail is 0.
- How biased is a coin?
- Toss it a number of times and record how many heads and tails.
- How biased is a coin?
- Bias parameter θ: bias of 1 (0) will always land heads (tails)
- Problem formulation. Let y be the number of heads in N tosses.

$$\mathbb{P}\left(\theta \middle| y\right) \propto \mathbb{P}\left(y \middle| \theta\right) \mathbb{P}(\theta).$$

Choosing the Likelihood

- Assumption: Each coin toss is independent of all other tosses.
- Candidate of likelihood

$$\mathbb{P}\left(y \middle| \theta\right) = \frac{N!}{y!(N-y)!}\theta^{y}(1-\theta)^{N-y}$$

- The binomial distribution is a discrete distribution returning the probability of getting y heads out of N coin tosses given a fixed value of θ
- θ indicates how likely it is that we will obtain a head.

Choosing the Prior

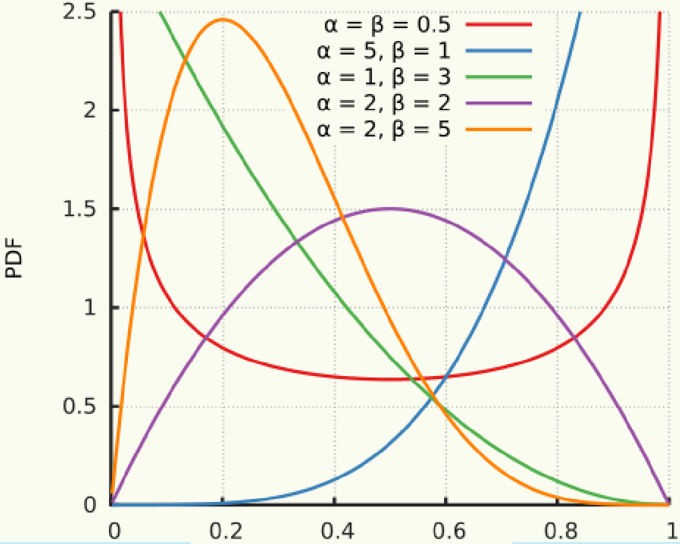
- As a prior we will use a **beta probability density function**, $f(\theta)$, which is commonly used in Bayesian statistics.

$$f\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

- Why use beta probability density function?
  - θ is restricted to be between 0 and 1 .
  - The shape of $f(\theta)$ is versatile.
  - β pdf is the **conjugate prior** of the binomial distribution (which we are using as the likelihood).

- A **conjugate prior of a likelihood** is a prior that, when used in combination with the given likelihood, returns a posterior with the same functional form as the prior.

Beta PDF



Three Priors by Beta PDF

(1) **Uniform**

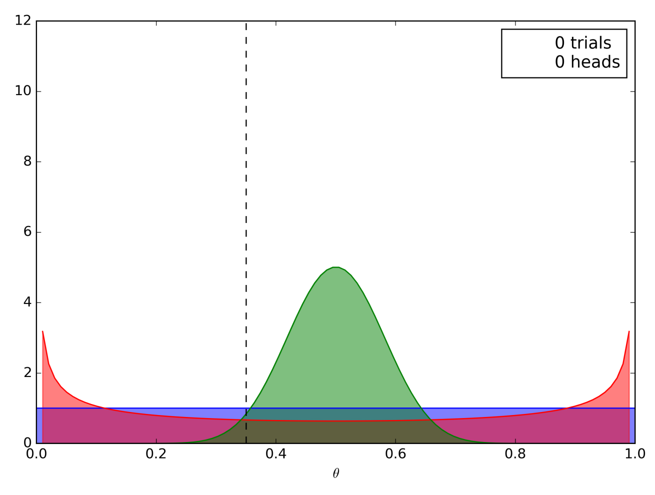
$(\alpha, \beta) = (1, 1)$

(2) **more on 0 and 1**

$(\alpha, \beta) = (0.5, 0.5)$

(3) **about normal**

$(\alpha, \beta) = (20, 20)$



Getting the Posterior

- Bayes' theorem says that the posterior is proportional to the likelihood times the prior:

$$\rho(\theta|y) \propto \frac{N!}{y!(N-y)!}\theta^y(1-\theta)^{N-y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

- We can drop terms that do not involve θ:

$$\rho(\theta|y) \propto \theta^y(1-\theta)^{N-y}\theta^{\alpha-1}\theta^{\beta-1} = \theta^{y+\alpha-1}(1-\theta)^{N-y+\beta-1}$$

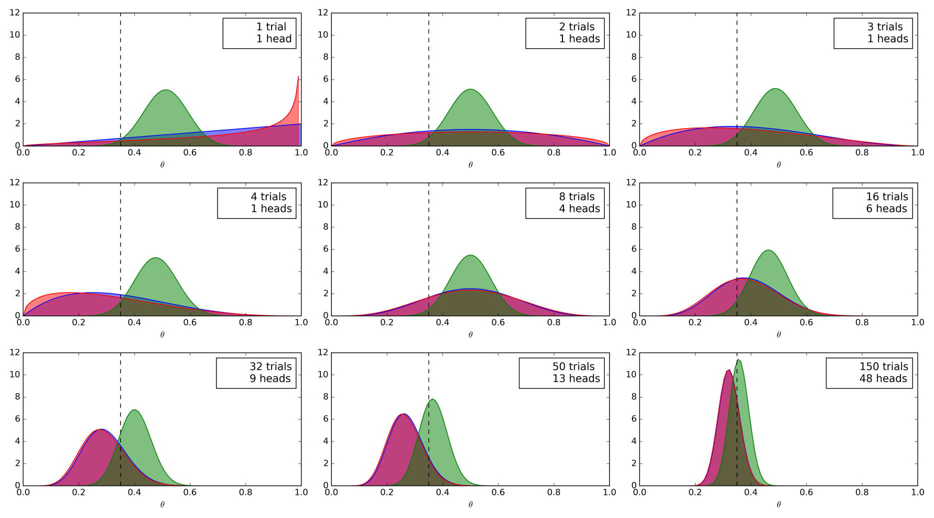
- This expression has the same functional form of a beta distribution (except for the normalization) with

$$\alpha_{\text{posterior}} = \alpha_{\text{prior}} + y, \qquad \beta_{\text{posterior}} = \beta_{\text{prior}} + N - y.$$

- The posterior for our problem is the beta distribution:

$$p(\theta|y) = \text{Beta}\big(\alpha_{\text{prior}} + y, \beta_{\text{prior}} + N - y\big).$$

Updated Priors $\Longrightarrow$ Posteriors



Remarks on Bayesian Analysis

- The result of a Bayesian analysis is the posterior distribution, not a single value but a distribution of plausible values given the data and our model.
- The most probable value is given by the mode of the posterior (the peak of pdf).
- The spread of the posterior is proportional to the uncertainty about the value of a parameter.
- The spread of 1 head out of 2 trials is larger than that of 4 out of 8, because there are more data that support our inference.
- Given a sufficiently large amount of data, two or more Bayesian models with different priors will tend to converge to the same result.

Prior Dependent?

- We see from the toy model that priors influence the analysis result.
- Some researchers are nervous about how to choose priors, because they do not want the prior to act as a censor that does not let the data speak for itself.
- **Data don't really speak; at best, they murmur.**
- Same data sets lead people to think differently about the same topics! (**Additional Reading**)
- Data only make sense in the light of models, both mathematical and mental.

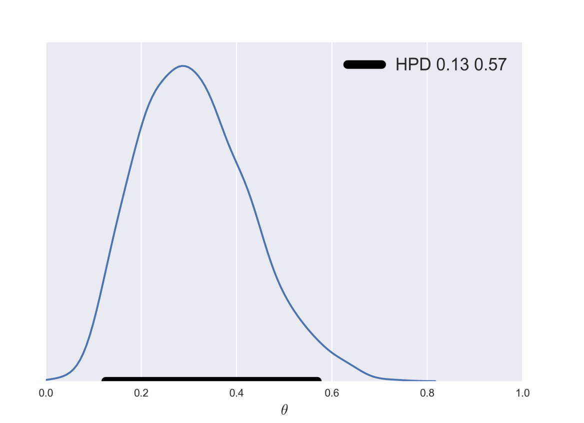
Types of Priors

- **Non-informative** priors (aka **flat**, **vague**, or **diffuse** priors)
  These priors have the least possible amount of impact on the analysis.

- **Weakly informative** priors (also know as **regularizing** priors)
  If you know something about the values a parameter can take, e.g., restricted to being positive, or the approximate range , e.g., close to zero or above/below some value.

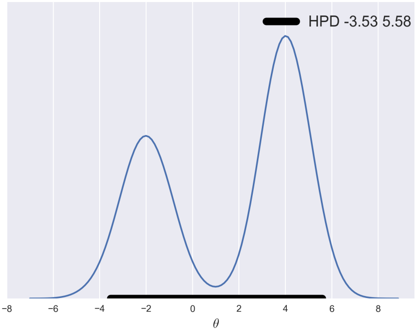
- **Informative priors**
  If you are experienced working in the same field and thus you have reliable prior information.

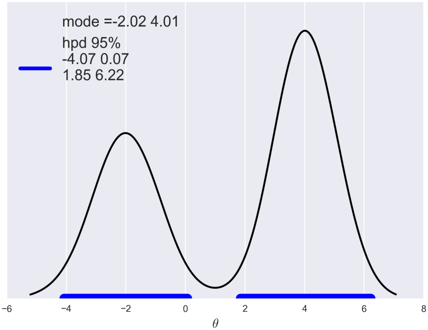
Highest Posterior Density Interval



- The result of a Bayesian analysis is the posterior distribution, which contains all the information about our parameters.
- **Highest Posterior Density (HPD)** interval is the shortest interval containing a given portion (e.g. 95%) of the probability density.

Bimodal Posteriors



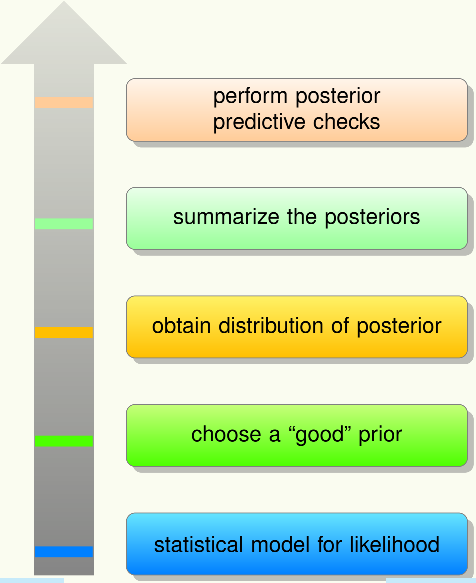


- When posteriors' pdf is bimodal, you need to have two HPD intervals.

Posterior Predictive Checks

- Once we have a posterior, it becomes possible to use the posterior to generate future data y, i.e., predictions. For example, you can select a median $\widehat{\theta}$ and compute the predicted y by feeding it to the model.
- Reserve some observed data for checking (**out-of-sample test**).
- The generated data and the observed data should look more or less similar, otherwise something is wrong.
- Knowing which part of our problem/data the model is capturing well and which it is not is valuable information even if we do not know how to improve the model.

Bayesian Data Analysis Algorithm



Takeaways

- The **prior distribution** is what we know about the value of some parameter before seeing the data.
- The **likelihood** is an expression of the plausibility of the data given the parameters.
- The **posterior distribution** is the result of the Bayesian analysis and reflects all that we know about a problem (given our data and model). It is an updated prior in the light of empirical (new) data.
- For data analysis and machine learning,

$$\mathbb{P}\left(C\middle|A\right) \propto \mathbb{P}\left(A\middle|C\right) \times \mathbb{P}\left(C\right)$$

Additional Reading

1. We Gave Four Good Pollsters the Same Raw Data. They Had Four Different Results.