Quantitative Analysis of Financial Markets

Session 1: Probability Theory

Benjamin Ee

Week 1

Overview

The purpose of this course is to provide the participant with a **toolkit for data modelling, especially forecasting**

For a given situation / dataset / desired application:

1. **What modelling techniques should we us**e?
2. How should we **build the model and calibrate it?**
3. How should we **evaluate the model?**

Overview

We will cover **3 main categories (examinable) and 1 supplemental category which is not examinable**

1. **Probability & Statistics**
    - Discrete and continuous realizations
    - Compound events
    - Bayesian probability
    - Statistical distributions
2. **Cross sectional analysis**
    - OLS (using CAPM as example)
    - Discrete variables (logit, probit, multinomial, ordered, etc)
    - Hazard models
3. **Time series analysis**
    - Decomposing a time series
    - Modelling trends
    - Stationary ARMA models
    - ARIMA-X models (using GARCH as example)
    - VAR / ECMs
4. **Machine learning** (classes 8 and 9) – these topics are not examinable
    - SVMs for time series classification
    - LSTMs for time series forecasting

How do we apply topics covered?

- Often in financial markets, we wish to **quantify the range of future outcomes**

- i.e. we want to **‘forecast’ some future distribution**.

- For example, we want to know what is the **distribution of possible returns for a financial instrument or portfolio tomorrow**. Possible applications:
    - Asset allocation / alpha generation
    - Market risk management (“Value at Risk”)
    - Derivatives pricing and risk management
    - Insurance pricing
    - Determining credit ratings

What is the relationship between the topics covered?

- To forecast the future distribution of a financial instrument (“a variable”) or portfolio (“a collection of variables”), we can chose from one or more of the following techniques:
    - **Use all the information in the historical record of that variable (only)** or collection of variables (only)
    - **Use information in other variables** which may (based on domain specific understanding or rules) be expected to be related to, or even have a causal impact on the variable we wish to forecast. i.e. we can make use of the values of other variables we are *not* forecasting to help us predict the variable we are interested in.
    - **Incorporate our prior expectations or theories** on the distribution of the variable into our quantitative analysis

What is the relationship between the topics covered?

- To forecast the future distribution of a financial instrument (“a variable”) or portfolio (“a collection of variables”), we can chose from one or more of the following techniques:
  - **Use all the information in the historical record** of that variable (only) or collection of variables (only)
  - **Use information in other variables** which may (based on domain specific understanding or rules) be expected to be related to, or even have a causal impact on the variable we wish to forecast. i.e. we can make use of the values of other variables we are *not* forecasting to help us predict the variable we are interested in.
  - **Incorporate our prior expectations or theories** on the distribution of the variable into our quantitative analysis

Univariate / univariate time series

- Summary statistical moments
- Time Series (ARMA, ARIMA,SARIMA, SARFIMA) ECM

Multivariate

- OLS
- Probit, Logit
- Multinomial Logit
- Multinomial Probit
- Hazard

- Bayesian analysis

What is the relationship between the topics covered?

- To forecast the future distribution of a financial instrument ("a variable") or portfolio ("a collection of variables"), we can chose from one or more of the following techniques:
  - **Use all the information in the historical record** of that variable (only) or collection of variables (only)
  - **Use information in other variables** which may (based on domain specific understanding or rules) be expected to be related to, or even have a causal impact on the variable we wish to forecast. i.e. we can make use of the values of other variables we are *not* forecasting to help us predict the variable we are interested in.
  - **Incorporate our prior expectations or theories** on the distribution of the variable into our quantitative analysis

Univariate / univariate time series

- Summary statistical moments
- Time Series (ARMA, ARIMA,SARIMA, SARFIMA)
- ECM

Multivariate

- OLS
- Probit, Logit
- Multinomial Logit
- Multinomial Probit
- Hazard

- Bayesian analysis

Hybrid approaches

- VAR
- VECM
- ARMA-X
- ARIMA-X
- LSTMs
- SVMs

What is the relationship between the topics covered?

- To forecast the future distribution of a financial instrument ("a variable") or portfolio ("a collection of variables"), we can chose from one or more of the following techniques:
    - **Use all the information in the historical record** of that variable (only) or collection of variables (only)
    - **Use information in other variables** which may (based on domain specific understanding or rules) be expected to be related to, or even have a causal impact on the variable we wish to forecast. i.e. we can make use of the values of other variables we are *not* forecasting to help us predict the variable we are interested in.
    - **Incorporate our prior expectations or theories** on the distribution of the variable into our quantitative analysis

Univariate / univariate time series

- Summary statistical moments
- Time Series (ARMA, ARIMA,SARIMA, SARFIMA)
- ECM

Multivariate

- OLS **(Example: CAPM)**
- Probit, Logit
- Multinomial Logit
- Multinomial Probit
- Hazard

- Bayesian analysis

Hybrid approaches

- VAR
- VECM
- ARMA-X **(Example: GARCH)**
- **ARIMA-X**
- LSTMs
- SVMs

Course logistics

1. **Grading**
   - Project: 20%
   - Final: 40%
   - Class participation: 15%
   - HWs: 25%

2. **Please get formed in groups of around 5 by start of class 2**
   - Email your groupings to TA(s)
   - All ungrouped students by end of week 2 will be grouped by alphabetical order into sets of 5 to 7 (smaller if there is no choice)
   - Cross sectional groups (some students from G1 and some from G2) are fine

3. All class materials will be distributed under **combined "G1-G2" item on eLearn**

4. **There is a class discussion forum link in eLearn**. You can post questions there, or email to me

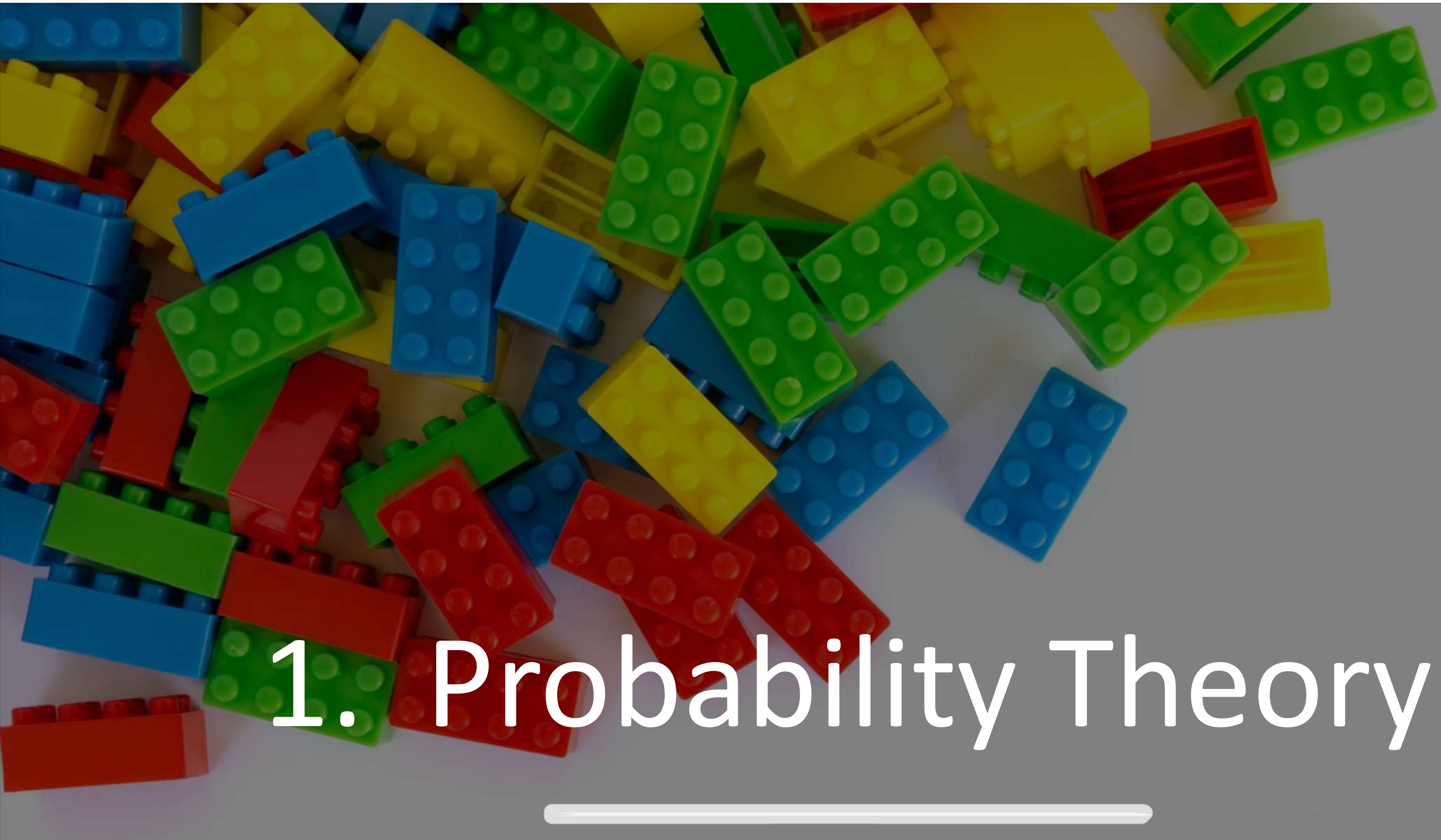
5. **Many classes are Python/R focused**. Code and data will be distributed. You will get more out of these classes if you can install Python and R beforehand

6. **University policy requires full in-person attendance.** I will still try to live-stream / record all classes, as previous batches have found recordings to be useful. This may be discontinued at any time if it impacts experience for those attending in person

For today

PROBABILITY THEORY

BAYESIAN ANALYSIS



1. Probability Theory

Probability theory is essential to forecasting because we need a ‘confidence bound’

1. If we invest $100 in the stock market today, we expect probability that it is worth more in 10 years is >50%, because we expect the **distribution of equity returns** is positive

2. Saying there is a >50% chance it will be “worth more in 10 years” is **not that useful**. Try raising funds from an investor using this statement and you will see what I mean!

3. More useful is saying “**there is a 95% chance it will be worth between $300 and $600 in 10 years time**”.

4. To make the above statement, we need to also have a view on the volatility of underlying returns, which is (once again) a **feature of the underlying distribution**

5. Another example: we want to use one variable to predict another. E.g. price/earnings ratios to predict stock returns. Our estimate is that for every unit increase in P/E ratio, stock returns fall by 1%. What is the probability that this estimate is “**better than white noise**”?

6. Overall:
   - **a) All forecasting is only useful if we also have the probability of the forecast being correct**
   - **b) Determining a “confidence interval” for the forecast is based on a foundation of the underlying statistical distribution**

We will study both individual and compound probabilities

Probability of individual event. i.e. a single flip of a coin. What is probability of heads?

Probability of compound events. i.e. two flips of a coin, or probability of "raining AND morning"

One reason to study compound events is to consider applications that we encounter in finance.

1. We may wish to estimate the probability that "all of the estimated parameters in a model" are jointly not different from noise [multiple variables]
2. Estimate probability that a bond will default *given* that there is a recession [2 variables]
3. Confidence interval for future returns *given* that recent returns are positive [2 variables]
4. etc

What do we measure probability over? Random variables

1. We measure probability over the same items we wish to forecast. i.e. random variables
2. A random variable is a mapping from the space of random events to an outcome which is a real number
   a) E.g. from the space of all possible heights of a person (say 0 to 3 meters) to their actual height
   b) E.g. from the space of all possible 1 day returns of a stock (say -100% to INF) to its actual return
   c) Each of these is an *outcome* of a random event
3. Mutually exclusive outcomes are outcomes that cannot happen at the same time for an underlying random variable.
4. Exhaustive outcomes are a set of outcomes that include all possible eventualities
5. A probability distribution describes the probabilities of all possible outcomes for a random variable.

Discrete and Continuous Random Variables

- There are two types of random variables, discrete and continuous

- A discrete random variable X can take on only a countable number of values.

$$P(X = x_i) = p_i, \qquad i = 1, 2, \ldots, n.$$

- A continuous random variable X can take on any value within a given range.

$$P(r_1 < X < r_2) = p.$$

The probability of X being between r_1 and r_2 is equal to p.

Examples of discrete and continuous random variables

Examples

- Getting infected by Covid-19 or not (1/0)
- Moody's long-term credit rating
  - 21 outcomes: Aaa, Aa1, Aa2, Aa3, A1, A2, A3, Baa1, Baa2, Baa3, Ba1, Ba2, Ba3, B1, B2, B3, Caa1, Caa2, Caa3, Ca, C.
- Corporate Earnings
  - Three outcomes: better than, equal to, or worst than consensus estimate
- Daily return r given daily volatility σ
  - Outcomes: $-1 < r \leq -3\sigma$, $-3\sigma < r \leq -2\sigma$, $-2\sigma < r \leq -\sigma$,
  - $-\sigma < r \leq 0$, $0 < r \leq \sigma$, $\sigma < r \leq 2\sigma$, $2\sigma < r \leq 3\sigma$, $r > 3\sigma$

Discrete or Continuous?

Probability Density Functions

- For a **continuous random variable** we can define a probability density function (PDF), which tells us the likelihood of outcomes occurring between any two points.

- For the probability p of X lying between r_1 and r_2, we define the probability density function $f(x)$ as follows:

$$\int_{r_1}^{r_2} f(x)\, dx = p.$$

Sample Problem

- Suppose probability density function for price of a zero coupon bond in percentage point x of par (or face value) is,

$$f(x) = \frac{8}{9}x,$$

for $0 < x \leq 3/2$.

1. Do probabilities sum to 1?

   Yes, int_0^{3/2} f(x) dx = (8/9)[x^2/2]_0^{3/2} = 4/9 * (3/2)^2 = 1

2. What is probability that price of the bond is between 90% and 100%?

   Yes, int_{0.9}^1 f(x) dx

Cumulative Distribution Functions (CDF)

- A cumulative distribution function tells us probability of a random variable being less than a certain value.
- By integrating probability density function from its lower bound to upper bound a, CDF is obtained.

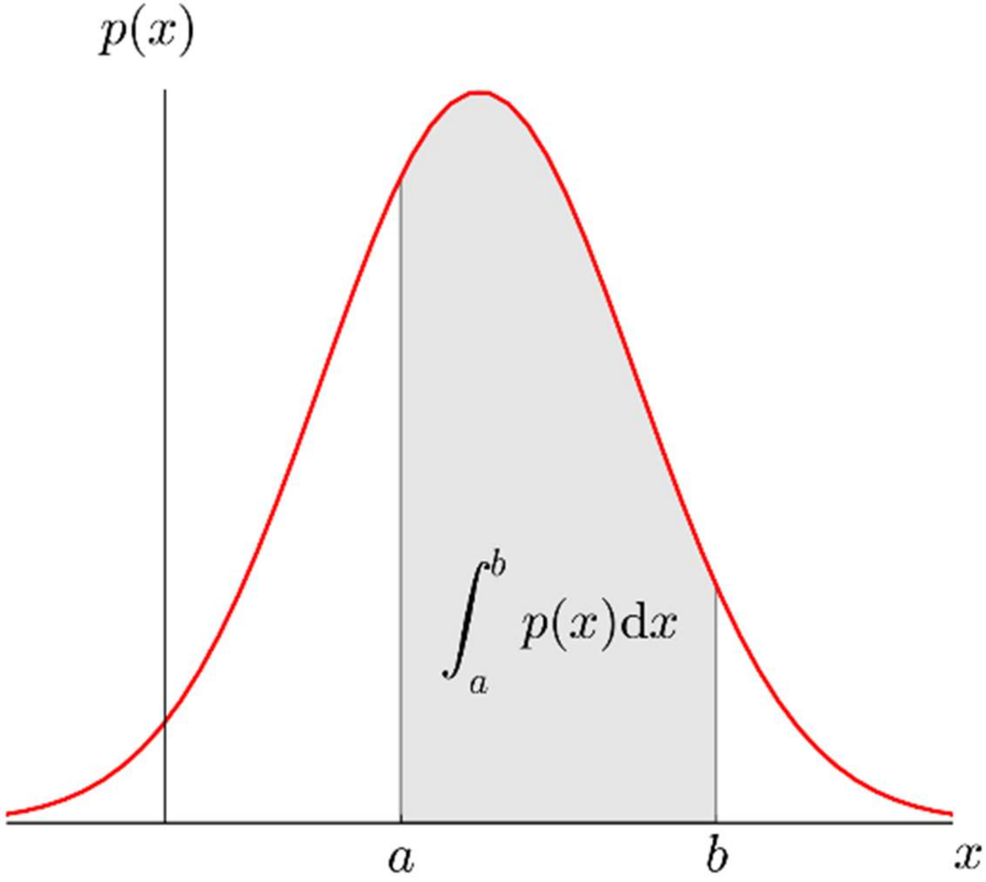
$$F(a) = \int_{-\infty}^{a} f(x)\,dx = \mathrm{P}(X \le a).$$

- We emphasize that $0 \le F(x) \le 1$, and $F(x)$ is non-decreasing.
- By the fundamental theorem of calculus, $f(x) = \dfrac{dF(x)}{dx}$.
- Moreover,

$$\mathrm{P}(a < X \le b) = \int_{a}^{b} f(x)\,dx = F(b) - F(a).$$

- Probability that a random variable is greater than a certain value a is

$$\mathrm{P}(X > a) = 1 - F(a).$$


$p(x)$
$\int_a^b p(x)\mathrm{d}x$
a
b
x

Inverse Cumulative Distribution Functions

- Let $F(a)$ be the cumulative distribution function. We define the inverse function $F^{-1}(p)$, the inverse cumulative distribution, as follows:

$$F(a) = p \Longleftrightarrow F^{-1}(p) = a.$$

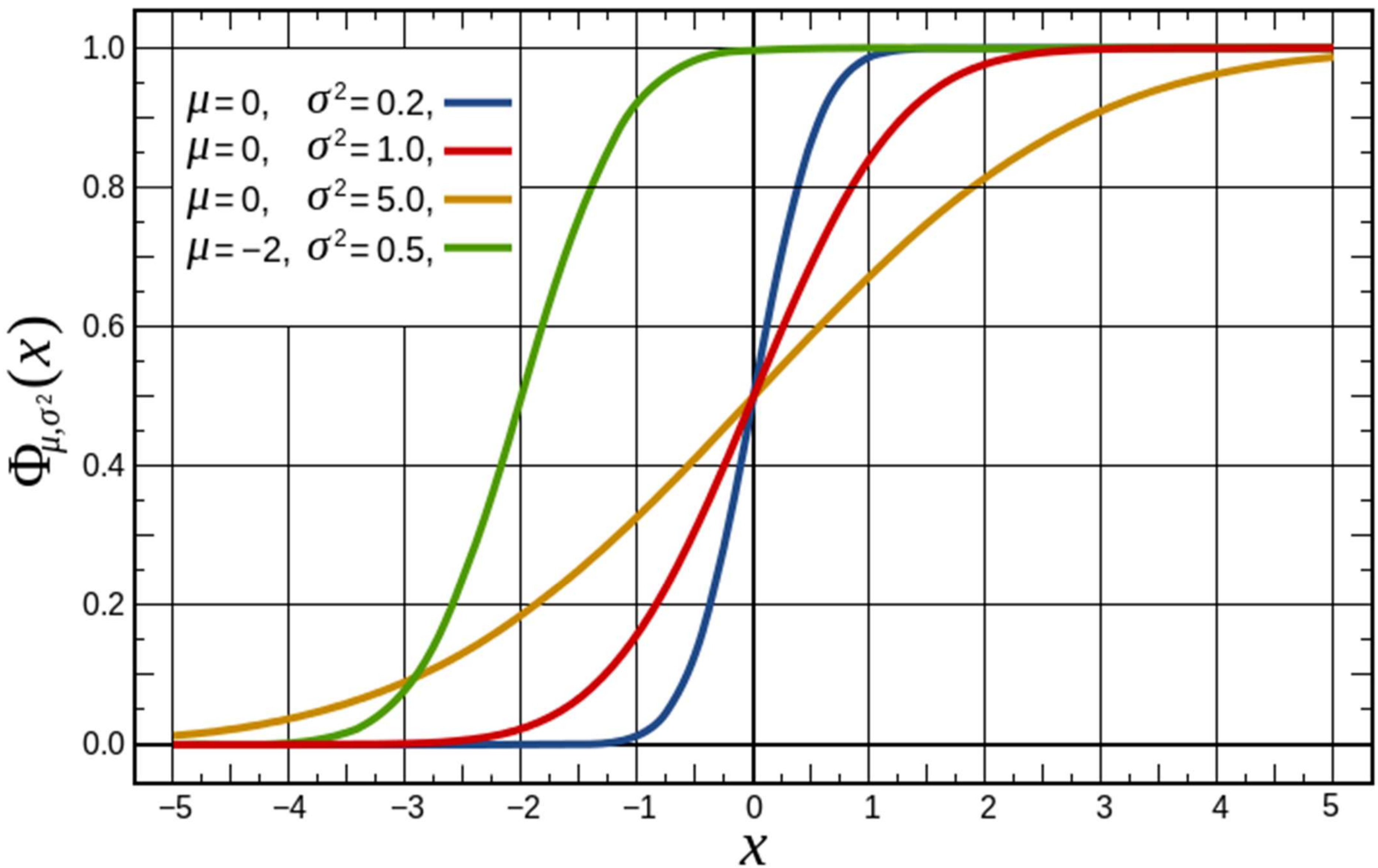
- The inverse distribution function is also called the quantile function.
- The 95-th percentile is $F^{-1}(0.95)$.
- Properties of $F^{-1}(p)$

  $F^{-1}(p)$ is non-decreasing

  $F^{-1}(y) \leq x$ if and only if $y \leq F(x)$.

  If Y has a uniform distribution in the interval $[0, 1]$, then $F^{-1}(Y)$ is a random variable with distribution F.

Normal Distribution CDF



Source: `https://en.wikipedia.org`

Sample Problem

- Consider the cumulative distribution function for $0 \leq a \leq 10$,

$$F(a) = \frac{a^2}{100}. = p$$

1. Derive the inverse cumulative distribution function.
2. Find the value of a such that 25% of the distribution is less than or equal to a.

Sample Problem 2

- The cumulative distribution function of Exponential(λ) (i.e. intensity λ and expected value (mean) $1/\lambda$) is

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases}$$

Find the probability density function.

Find the quantile function for Exponential(λ).

Suppose $\lambda = \ln(2)$. Find the median.

Mutually Exclusive Outcomes

- For a given random variable, the probability of any of two mutually exclusive outcomes A and B occurring is just the sum of their individual probabilities.

$$P(A \cup B) = P(A) + P(B).$$

- It is the probability of either A or B occurring, which is true only for mutually exclusive events.

- Question: Calculate the probability that a stock return is either below -10% or above 10%, given that

$$P(R < -10\%) = 14\%, \qquad P(R > 10\%) = 17\%,$$

Independent Events and Joint Probability

- What happens when we have more than one random variable?
- Event: It rains tomorrow and the return on MSFT is greater than 0.5%?
- If the outcome of one random variable is not influenced by the outcome of the other random variable, then we say those variables are independent. The joint probability of W and R is such that

$$\begin{aligned} \mathrm{P}\,(W = \text{rain and } R > 0.5\%) &= \mathrm{P}\,(\text{rain} \cap R > 0.5\%) \\ &= \mathrm{P}\,(\text{rain}) \times \mathrm{P}\,(R > 0.5\%). \end{aligned}$$

- Question: According to the most recent weather forecast, there is a 20% chance of rain tomorrow. The probability MSFT returns more than 0.5% on any given day is 40%. The two events are independent. What is the probability that it rains and MSFT returns more than 0.5% tomorrow?

Probability Matrices

- When dealing with the joint probabilities of two variables, it is often convenient to summarize the various probabilities in a probability matrix or probability table.

- Example: Stock Grading by Equity Analyst and Credit Rating Agency

		Stock		
		Outperform	Underperform	Total %
Bonds	Upgrade	15%	5%	20%
	No Change	30%	25%	55%
	Downgrade	5%	20%	25%
	Total %	50%	50%	100%

Sample Problem

- Bonds versus Stock Matrix

		Stock		
		Outperform	Underperform	Total %
Bonds	Upgrade	5%	0%	5%
	No Change	40%	20Y%	60Z%
	Downgrade	5 X%	30%	35%
	Total %	50%	50%	100%

- What are the values of X, Y, and Z?

- What is the probability that the stock will outperform given that there is a bond upgrade?

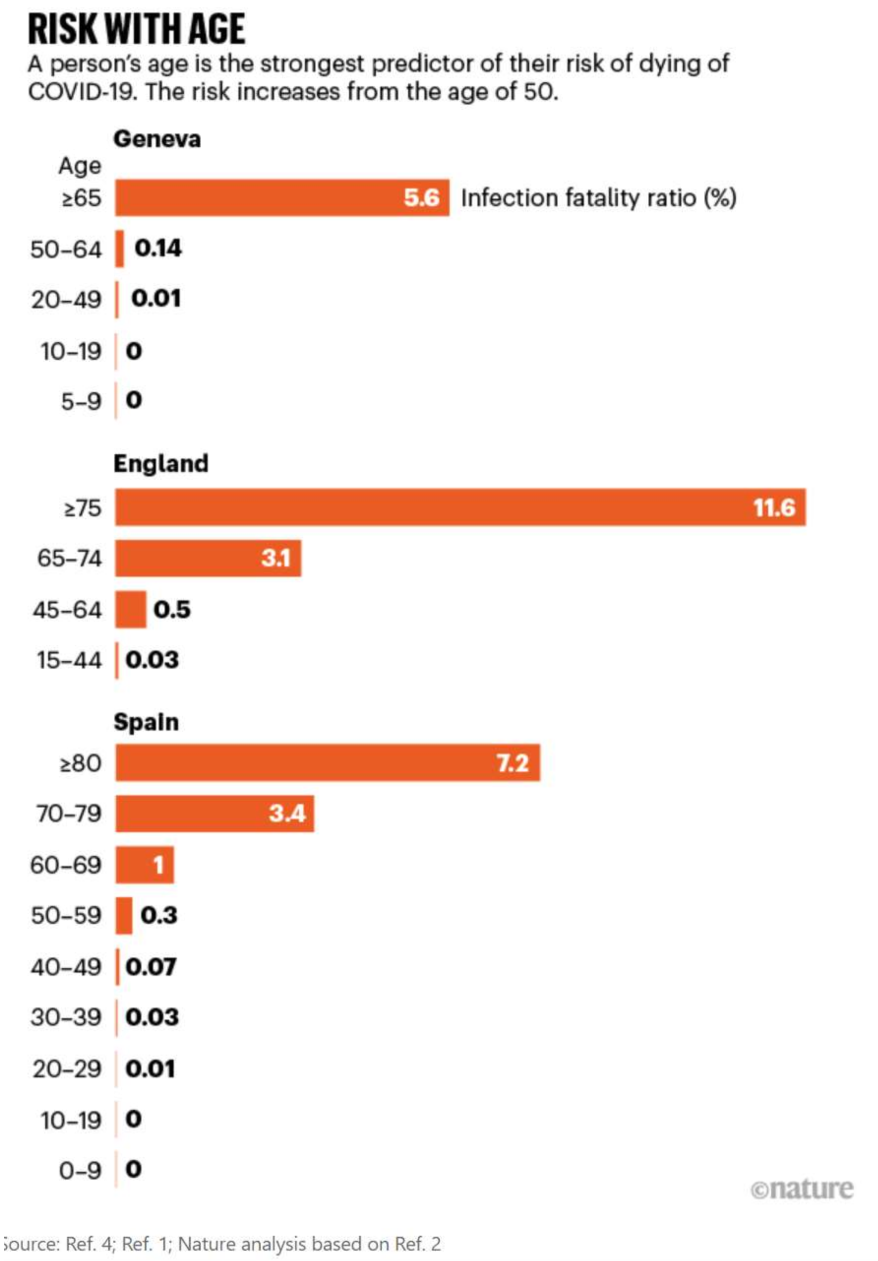
Conditional Probability

- What is the probability that the stock is up **given that** it is downgraded?

$$P\ (M = \text{up} \mid W = \text{downgraded})$$

- Using the conditional probability, we can calculate the probability that it will be downgraded and that the stock will be up.

$$P\ (M = \text{up and } W = \text{downgraded}) = P\ (M = \text{up} \mid W = \text{downgraded})\ P\ (W = \text{downgraded})$$



Source: Nature (https://www.nature.com/articles/d41586-020-02483-2)

Independence

- One way to define the concept of independence: If $P(M = \text{up} \mid W = \text{dg}) = P(M = \text{up})$, the two random variables, M and W, are independent.

- Show that if M and W are independent, then

$$P(M = \text{up and } W = \text{dg}) = P(M = \text{up})\, P(W = \text{dg}).$$

2. Bayesian Analysis

Applications of Bayesian Analysis

Overview - Bayesian analysis can enable us to conduct forecasting in ‘low data’ situations

We can make use of the Bayesian prior (an assumption based on external theory) to compensate

Bayes Theorem

Bayesian analysis is based on Bayes Theorem, which can be derived by conditional probability

Example on next few slides illustrates …

Illustration: Bond Default

- Assume we have two bonds, Bond A and Bond B, each with a 10% probability of defaulting next year. Further assume probability both bonds default is 6%, and probability neither defaults is 86%.
- Because bond issuers are affected by similar economic trends, **defaults are positively correlated**.

Probability Matrix

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

Joint and Conditional Probabilities: Q&A

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

$$P(A \mid B) = \frac{P(A \text{ n } B)}{P(B)} = \frac{6\%}{10\%} = 60\%$$

- Answer: Bond A defaults in 60% of scenarios in which B defaults.
- Notice conditional probability P(A | B) is different from unconditional probability *P(A)* of A defaulting, which is 10%.

Deriving Bayes Theorem from Conditional Probability

- Joint probability $$P(A \cap B) = P(A \mid B)\, P(B)$$

  can be written as $$P(A \cap B) = P(B \mid A)\, P(A)$$

  since the joint probability is symmetric:

$$P(A \cap B) = P(B \cap A)$$

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

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Introduction

For two random variables, *A* and *B*, **Bayes' theorem** states that

$$P(A \mid B) = \frac{P(B \mid A)\,P(A)}{P(B)}$$

An Introductory Problem

- How can you predict the value of an **unobserved** variable (class) from an **observable** variable (attribute)?
- You know the **prior probability** P(*C*) of class *C*. This is based on theoretical assumptions [we will discuss later]
- **Observed** attribute is also known as the predictor. You know the predictor prior probability P(*A*) of the predictive attribute *A*.
- Most importantly, you know the likelihood P(A|C) of attribute A for a given class C.
- What is the posterior probability P (*C*|*A)* of class *C* given the attribute A?

Algorithm: Bayes Theorem

Prior

$$P(C \mid A) = \frac{P(A \mid C)\, P(C)}{P(A)}$$

Prior P(C) is our unconditional expectation of the class incidence. In practical models, choice of this variable is clearly subject to debate

Algorithm: Bayes Theorem

Prior

Likelihood

$$P(C \mid A) = \frac{P(A \mid C)\,P(C)}{P(A)}$$

Likelihood P(A|C) is the probability of observing the attribute A given the class C.

Algorithm: Bayes Theorem

Prior

Likelihood

$$P(C \mid A) = \frac{P(A \mid C)\,P(C)}{P(A)}$$

Predictor Prior

Predictor Prior is the probability of observing the attribute A.

Algorithm: Bayes Theorem

Prior

Likelihood

$$P(C \mid A) = \frac{P(A \mid C)\, P(C)}{P(A)}$$

Predictor Prior

Posterior

And on LHS is the posterior probability we are interested in.

Sample Problem 1: Fund managers

- Suppose skillful fund managers are 1 in 100.
- A test is 99% accurate in identifying a fund manager is skillful.
- For skillful fund managers, test is 99% correct. Similarly, for those who are not skillful, test correctly indicates they are not skillful with 99% of accuracy.
- If a fund manager takes test and test is positive, what is probability fund manager is actually skillful?

Answer to Sample Problem 1

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

$$P(p) = P(p \cap \text{skillful}) + P(p \cap \text{unskillful})$$

$$= P(p\,|\,\text{skillful})\, P(\text{skillful}) + P(p\,|\,\text{unskillful})\, P(\text{unskillful})$$

$$= 99\% \times 1\% + 1\% \times 99\%$$

$$= 2\% \times 99\%.$$

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

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

Results are **extremely sensitive** to the prior (that skillful managers are 1 in 100)

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.
- 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, new manager has beaten the market every year.

  1. What was probability that manager was a star when she was first added?
  2. What is probability that manager is a star now?
  3. After observing manager beat market over past three years, what is probability that manager will beat 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%:

$$P(B|star) = 75\% = \frac{3}{4}$$

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

$$P(B|nonstar) = 50\% = \frac{1}{2}$$

- 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.

$$P(star) = 16\% = \frac{4}{25}$$

Answer to Sample Problem 2-2

- We need to find P(star | 3B), the probability that the manager is a star, given that the manager has beaten the market three years in a row (3B).

$$P(star|3B) \quad = \frac{P(3B|star)P(star)}{P(3B)}$$

- Because outperformance is independent from one year to the next, P(3B | star) is just the probability that a star beats the market in any given year to the third power.

$$P(3B|star) \ = (x+a)^n = \left(\frac{3}{4}\right)^3 = \frac{27}{64}$$

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

- What is 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:

P(3B) = P (3B | star) P (star) + P (3B | nonstar) P (nonstar)

$$= \left(\frac{3}{4}\right)^3 \frac{4}{25} + \left(\frac{1}{2}\right)^3 \frac{21}{25} = \frac{69}{400}.$$

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

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

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.
- No substitute for time in building a track record? [maybe]

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:

  P(B) = P (B|star) P (star) + P(B|nonstar) P (nonstar) $= \frac{3}{4}\frac{9}{23} + \frac{1}{2}\ \frac{14}{23} = 60\%$

- 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, P (3B|star)= 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.

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
  - Underperformers: Beat the market 25% of the time
  - In-line performers: beat the market 50% of the time
  - 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.

P(p=0.25) 20%

P(p=0.50) 60%

P(p=0.75) 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:

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

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

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

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

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

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

1. Underperformer=

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

Answer to Many-State Problem (cont'd)

2. In-line performer

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

3. Outperformer

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

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, P (2B) = 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

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

Bayes Theorem

- 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 for 6 of 10 years?

- Answer:
  For each possible type of manager, the likelihood of beating market 6 times out of 10 can be determined using **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 shortcut method, we first calculate posterior probabilities in terms of an arbitrary constant, *c*.
If manager is an underperformer,

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

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

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

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

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

Answer to Sample Problem 3 (cont'd)

- In this case, probability that manager is an in-line performer has increased from 60% to 79.13%.
- Probability 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 probability 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 frequentist approach. The conclusion is based only on observed frequency of positive results.

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

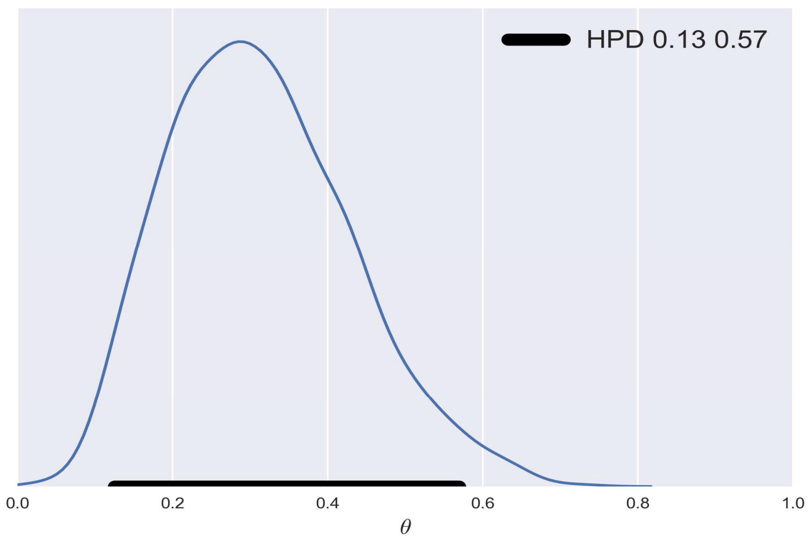
Bayes versus Frequentists (cont'd)

- 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 analysts would probably use 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.

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.

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.