



School of Business

BIA-652

Multivariate Data Analytics

Review of Probability & Random Variables

Prof. Feng Mai
School of Business

For academic use only.



First theme of the class

t.
Year

$$\text{Income} = f(\text{Age}, \text{Gender}, \text{edu} \dots)$$

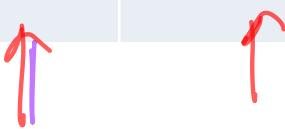
Individual	Age	Gender	Education	Weight	Height	Income

$$= \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Gender}$$

$$+ \dots$$

$$\begin{array}{l} P(F \mid C) \\ P(M \mid C) \end{array}$$

conditional



- Various practical methods of analyzing a “dataframe”
- Interpret the results

• Tabular dataset

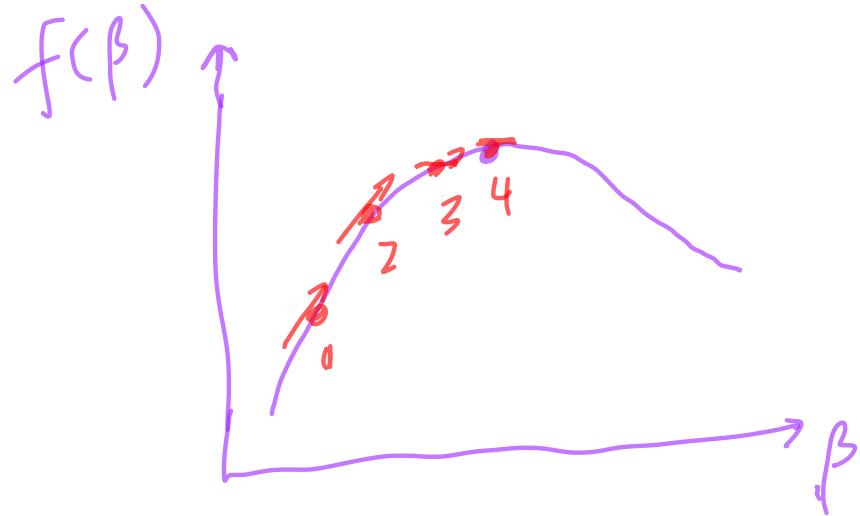
Not: Time - series.

Panel Data

	C	$<C$	$>C$	
M				
F				
Total				

Second theme of the class

- Almost every statistics problem is an optimization problem.
- Learn how the optimization problem is formulated
... and assume that we know how to solve the optimization problems.

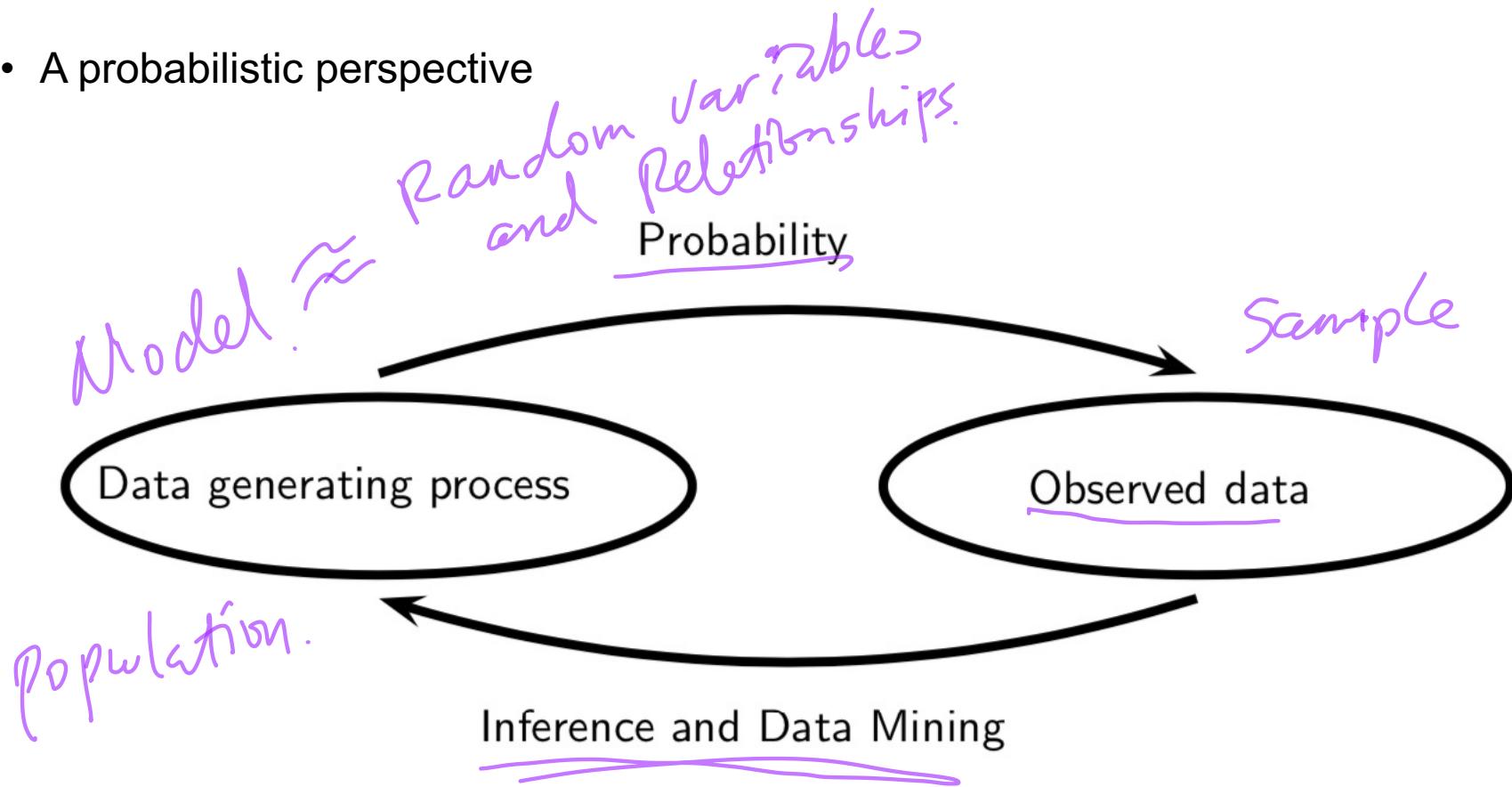


$$f'(\beta) = 0$$

solve for β .

Third theme of the class

- A probabilistic perspective





Why start with probability? What about deep learning?

- <https://www.youtube.com/watch?v=x7psGHgatGM>
- (starts from 11:05)



Goal of the class

- NOT to teach you every multivariate technique
- Provide you the knowledge to
 - Read more advanced statistics/ML books
 - Know how to read the manual of a statistical package/software



Prerequisite of the class

- Have taken a calculus class (derivative and integration)
- Know some basic Linear Algebra (vector and matrix operation)
- Prior Python knowledge is not required



Syllabus

BIA 652, Multivariate Data Analytics



1- Probability

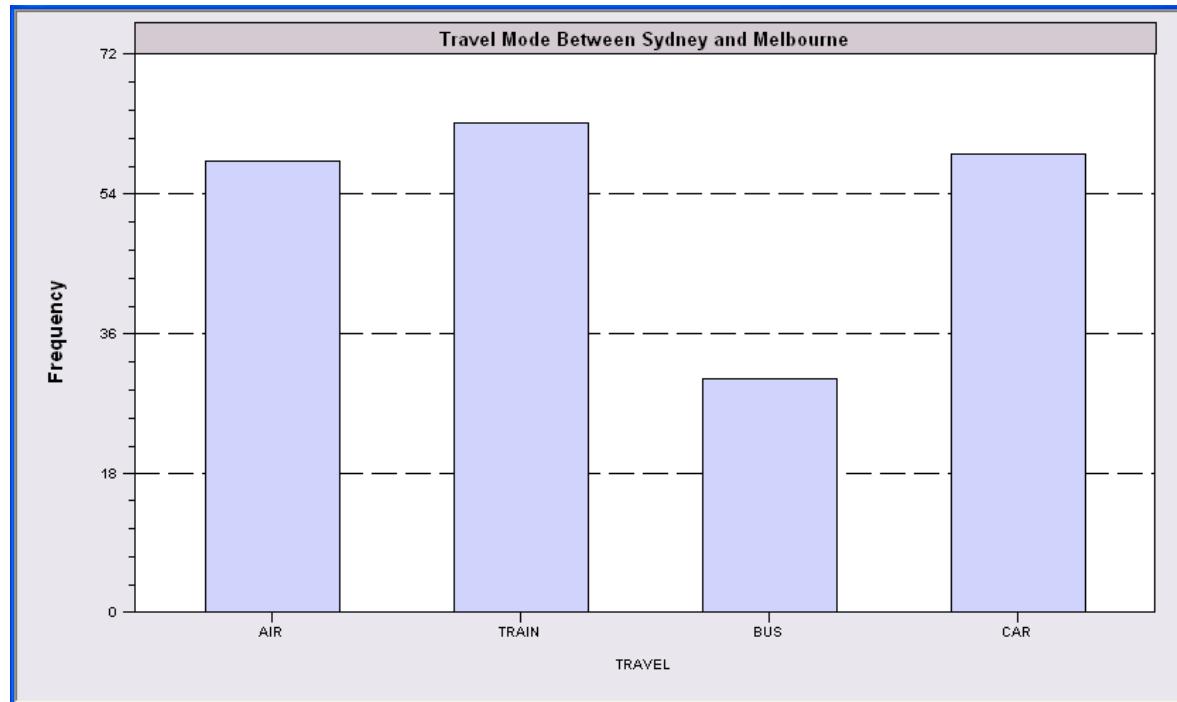


The Sample Space

- Collection of all possible outcomes
 - Exclusive
 - Exhaustive
 - $\Omega = \{\text{the set of possible outcomes}\}$
- Random outcomes: The result of a process
 - Sequence of events,
 - Number of events,
 - Measurement of a length of time, space, etc.
- Experiments, outcomes, and sample spaces

Consumer Choice:

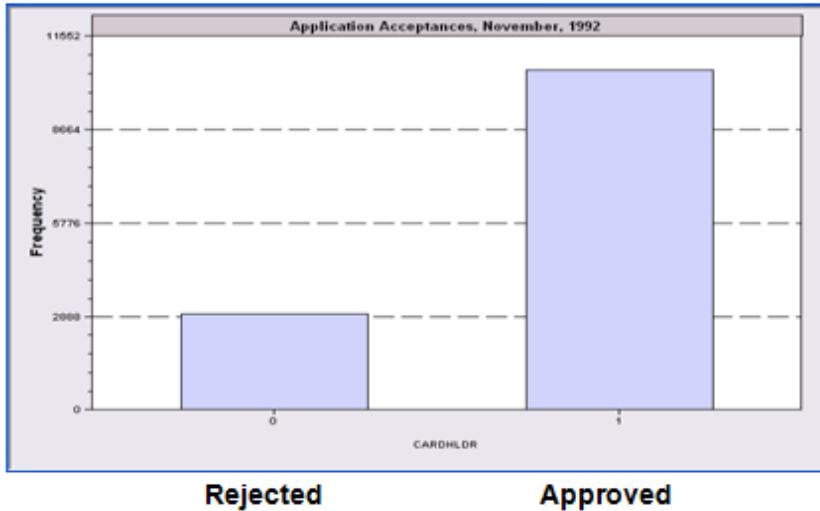
4 possible ways a randomly chosen traveler might have traveled between Sydney and Melbourne



$$\Omega = \{\text{Air, Train, Bus, Car}\}$$

Market Behavior: outcome of credit card application

13,444 Applicants for a Credit Card (November, 1992)



Experiment = A randomly picked application.

Let $X = 0$ if Rejected

Let $X = 1$ if Accepted

X is DISCRETE (Binary). This is called a Bernoulli random variable.

$$\Omega = \{\text{Reject, Accept}\}$$



Lifetimes of light bulbs

- A box of light bulbs states “Average life is 1500 hours”
- Outcome = length of time until failure (lifetime) of a randomly chosen light bulb

$$\Omega = \{\text{lifetime} \mid \text{lifetime} \geq 0\}$$

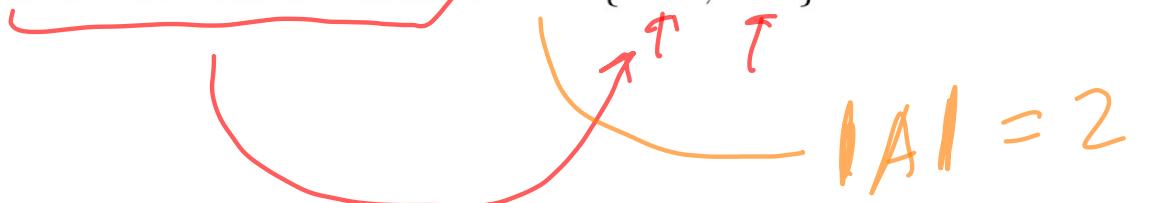
Events

Prob (Events)

- Events are defined as
 - Subsets of sample space, such as empty set
- It can be
 - Empty
 - Intersection of related events
 - Complements such as “A” and “not A”
 - Disjoint sets such as (train, bus),(air, car)

$$|\mathcal{N}| = 4$$

1.1 Example. If we toss a coin twice then $\Omega = \{HH, HT, TH, TT\}$. The event that the first toss is heads is $A = \{HH, HT\}$. ■



$$|\mathcal{A}| = 2$$

$$P(A) = \frac{2}{4} = 0.5$$

Summary of Terminology

Ω	sample space
ω	outcome (point or element)
A	event (subset of Ω)
A^c	complement of A (not A) <i>not happening</i>
$A \cup B$	union (A or B)
$A \cap B$ or AB	intersection (A and B)
$A - B$	set difference (ω in A but not in B)
$A \subset B$	set inclusion
\emptyset	null event (always false)
Ω	true event (always true)



Mutually Exclusive (Disjoint) Events and Partition

We say that A_1, A_2, \dots are **disjoint** or are **mutually exclusive** if $A_i \cap A_j = \emptyset$ whenever $i \neq j$. For example, $A_1 = [0, 1), A_2 = [1, 2), A_3 = [2, 3), \dots$ are disjoint. A **partition** of Ω is a sequence of disjoint sets A_1, A_2, \dots such that $\bigcup_{i=1}^{\infty} A_i = \Omega$. Given an event A , define the **indicator function of A** by

$$I_A(\omega) = I(\omega \in A) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A. \end{cases}$$

Probability

- Probability is a measure defined on all subsets of Ω
- Probability is a function P that assigns a real number $P(A)$ to each event A
- The three Axioms of Probability
 - $A \subset \Omega \Rightarrow P(A) \geq 0$
 - $P(\Omega) = 1$
 - If $A \cap B = \{\emptyset\}$, $P(A \cup B) = P(A) + P(B)$
- Interpretation
 - Frequency
 - Degree-of-belief

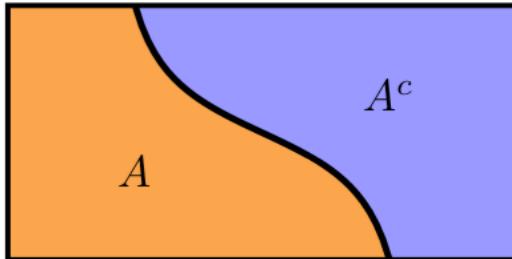
Mutually exclusive

*A and B cannot happen
at the same time*

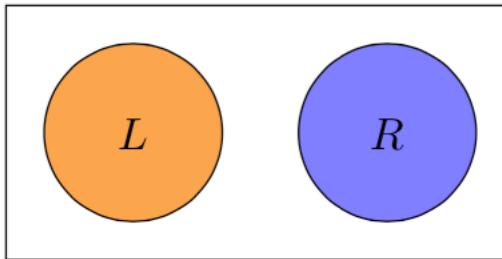
Implications of the Axioms

- $P(A^c) = 1 - P(A)$ as $A \cup A^c = \Omega$
- $P(\emptyset) = 0$
- $A \subset B \Rightarrow P(A) \leq P(B)$
- **Addition rule:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

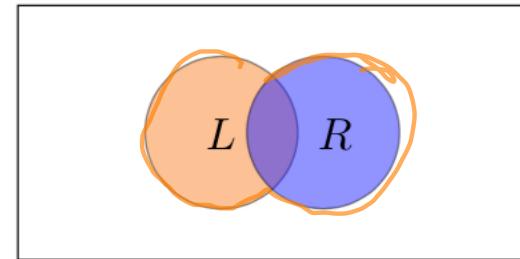
A^c \bar{A} $\sim A$
 A not happening



$\Omega = A \cup A^c$, no overlap



$L \cup R$, no overlap



$L \cup R$, overlap = $L \cap R$

Calculating Probability: Counting Rules

- If Ω is finite and if each outcome is equally likely, then assigning probability: '**Size** of an event relative to size of sample space.'

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|},$$

- Counting rules for equally likely discrete outcomes
 - How many ways for an outcome to be included in A?
 - Using **multi-step experiment**, **combinations** and **permutations** to count elements

Counting Rule for Multiple-Step Experiments

- If an experiment consists of a sequence of k steps in which there are n_1 possible results on the first step, n_2 possible results on the second step, and so on, then the total number of experimental outcomes is given by $(n_1)(n_2) \dots (n_k)$.
- Example:** Flipping 3 coins. How many outcomes are possible?

Step 1 Toss coin 1

$$n_1 = 2 \quad H, T$$

Step 2 Toss coin 2

$$n_2 = 2 \quad H, T$$

Step 3 Toss coin 3

$$n_3 = 2 \quad H, T$$

Total Number of Outcomes: $n_1 n_2 n_3 = 2 \cdot 2 \cdot 2 = 8$

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Counting Rule for Combinations

Another useful counting rule enables us to count the number of different experimental outcomes when k objects are to be selected from a set of n objects.

- The number of combinations of n objects taken k at a time is $(n \text{ choose } k)$

$$C_k^n = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

where $n! = n(n - 1)(n - 2) \dots (2)(1)$ (n factorial)

$$k! = k(k - 1)(k - 2) \dots (2)(1)$$

$$0! = 1$$



Counting Rule for Permutations

k objects are to be selected from a set of n objects, where the *order* of selection is important.

Number of Permutations of n Objects Taken k at a Time

$$P_k^n = k! \binom{n}{k} = \frac{n!}{(n-k)!}$$

- where: $n! = n(n - 1)(n - 2) \dots (2)(1)$
- $k! = k(k - 1)(k - 2) \dots (2)(1)$
- $0! = 1$

Bivariate Probabilities: Marginal, Joint, and Conditional

- Outcomes for bivariate events:

	B_1	B_2	...	B_k
A_1	$P(A_1 \cap B_1)$	$P(A_1 \cap B_2)$...	$P(A_1 \cap B_k)$
A_2	$P(A_2 \cap B_1)$	$P(A_2 \cap B_2)$...	$P(A_2 \cap B_k)$
.
.
A_h	$P(A_h \cap B_1)$	$P(A_h \cap B_2)$...	$P(A_h \cap B_k)$

Joint and Marginal Probabilities

- The probability of a joint event, $A \cap B$:

$$P(A \cap B) = \frac{\text{number of outcomes satisfying } A \text{ and } B}{\text{total number of elementary outcomes}}$$

- Computing a marginal probability:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \dots + P(A \cap B_k)$$



- Where B_1, B_2, \dots, B_k are k mutually exclusive and collectively exhaustive events

Bivariate Probabilities: Coffee preference and gender

$$P(y | x_1, x_2, \dots)$$

$$P(\text{Latte} | M) = \frac{7}{50}$$

$$P(M) = 50/100$$

$$P(\text{Latte}) = \frac{16}{100}$$

		Coffee Preference					Total
		Coffee	Mocha	Espresso	Latte	Tea	
Gender	Male	15	9	13	7	6	50
	Female	11	14	5	9	11	50
	Total	26	23	18	16	17	100

Joint Frequency

Marginal Frequency

$$P(\text{Latte} | \text{F}) = \frac{P(\text{Latte} \cap \text{F})}{P(\text{F})} = \frac{9/100}{50/100} = \frac{9}{50}$$

Given / conditioned on

$$\begin{aligned}
 P(\text{Latte}) &= \frac{7+9}{100} = \frac{7}{100} + \frac{9}{100} \\
 &= P(M \wedge L) + P(F \wedge L) \\
 &= \underbrace{P(L|M) \cdot P(M) + P(L|F) \cdot P(F)}_{\text{Law of total prob}}
 \end{aligned}$$

$$\begin{aligned}
 P(M \wedge L) &= P(L|M) \cdot P(M) \\
 P(L|M) &= \frac{P(M \wedge L)}{P(M)}
 \end{aligned}$$

Conditional Probability

Joint / Marginal

- Probability of event A given that event B occurs.

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

= Size of A relative to B (a subset of Ω)

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

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

- Multiplication rule:

$$P(A \cap B) = P(A|B) P(B) \text{ (follows from the definition)}$$

$$= P(B|A) P(A)$$

- Factorization (chain rule)

$$P(a \cap b \cap c \dots y \cap z) = P(a | b, c, \dots, y, z) P(b | c, \dots, y, z) P(c | \dots, y, z) \dots P(y | z) P(z)$$

Bayes Theorem = def of $P(A|B)$
 Classification. + Law of Total Prob

1.16 Theorem (The Law of Total Probability). Let A_1, \dots, A_k be a partition of Ω . Then, for any event B ,

$y \in \{\mathcal{M}, \mathcal{F}\}$

$$\begin{aligned} \mathbb{P}(B) &= \sum_{i=1}^k \mathbb{P}(B|A_i) \mathbb{P}(A_i). \\ &= \underbrace{P(L|M) \cdot P(M)}_{\hookrightarrow \mathcal{M}, \mathcal{F}} + P(L|P) \cdot P(P) \end{aligned}$$

1.17 Theorem (Bayes' Theorem). Let A_1, \dots, A_k be a partition of Ω such that $\mathbb{P}(A_i) > 0$ for each i . If $\mathbb{P}(B) > 0$ then, for each $i = 1, \dots, k$,

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i) \mathbb{P}(A_i)}{\sum_j \mathbb{P}(B|A_j) \mathbb{P}(A_j)}. \quad (1.5)$$

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

Target

$$= \frac{P(B | A)P(A)}{P(B)} \leftarrow \text{Law of Total prob.}$$

Theorem

$$= \frac{P(B | A)P(A)}{P(A, B) + P(\text{not } A, B)}$$

Definition

$$= \frac{P(B | A)P(A)}{P(B | A)P(A) + P(B | \text{not } A)P(\text{not } A)}$$

Computation



Conditional Prob and Bayes Theorem Example: Color Blindness

- Inherited color blindness has different prevalence rates in men and women. Women usually carry the defective gene and men usually inherit it.

- Experiment: pick an individual at random from the population.**

CB = has inherited color blindness

MALE = gender, Not-Male = FEMALE

- Marginal:** $P(CB)$ = 2.75%
 $P(MALE)$ = 50.0%

- Joint:** $P(CB \cap MALE)$ = 2.5%
 $P(CB \cap FEMALE)$ = 0.25%

- Conditional:** $P(CB|MALE)$ = 5.0% (1 in 20 men)
 $P(CB|FEMALE)$ = 0.5% (1 in 200 women)



- $P(CB|Male) = .05$
- What is $P(Male|CB)$?
- $$\begin{aligned} P(M|CB) &= P(M \cap CB)/P(CB) \\ &= P(CB|M)P(M)/P(CB) = .05 \cdot .5 / P(CB) \end{aligned}$$
- $$\begin{aligned} P(CB) &= P(CB \cap M) + P(CB \cap F) \\ &= P(CB|M)P(M) + P(CB|F)P(F) \\ &= .05(.5) + .005(.5) = .0275 \text{ (as we knew)} \end{aligned}$$
- $P(M|CB) = .025 / .0275 = .909$ (i.e., 91% of colorblind people are male.)



A drilling company has estimated a 40% chance of striking oil for their new well. A detailed test has been scheduled for more information. Historically, 60% of successful wells have had detailed tests, and 20% of unsuccessful wells have had detailed tests.

Given that this well has been scheduled for a detailed test, what is the probability that the well will be successful?



Independent events

- Definition: $P(A|B) = P(A)$
- Multiplication rule: $P(A \cap B) = P(A)P(B)$ if A and B are independent
- Two random men are both color blind =
 $.05 \text{ (Men 1)} \times .05 \text{ (Men 2)} = .0025$

• Addition Rule

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

• Mul Rule

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

• Bayes Rule

• Axiom's

• Conditional Prob def.



Random Variables

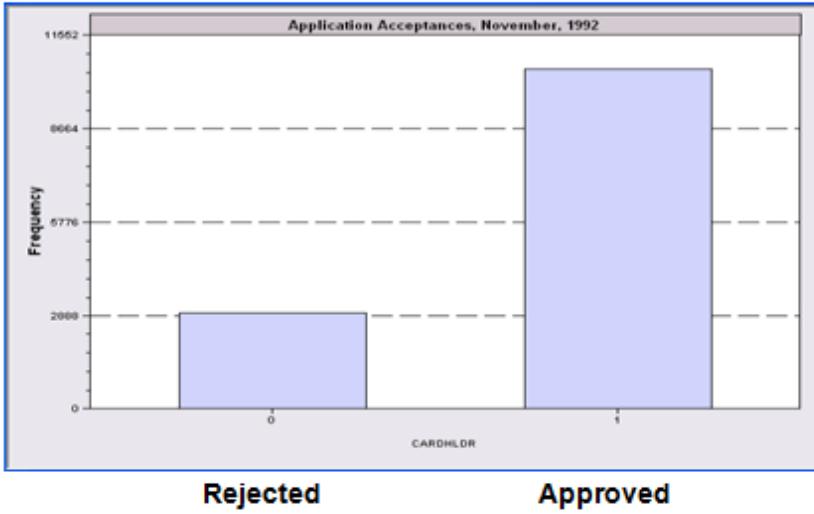


Random Variable

- Definition: Maps elements of the sample space to a variable.
 - Assigns a value to $\omega \in \Omega$
- Types of random variables
 - Discrete: Coin toss
 - Continuous: Lightbulb lifetimes

Market Behavior: outcome of credit card application

13,444 Applicants for a Credit Card (November, 1992)



Experiment = A randomly picked application.

Let $X = 0$ if Rejected

Let $X = 1$ if Accepted

X is DISCRETE (Binary). This is called a Bernoulli random variable.

$$\Omega = \{\text{Reject, Accept}\}$$

$$X = 0 = \text{reject}, 1 = \text{accept}$$



2.9 Definition. X is discrete if it takes countably³ many values $\{x_1, x_2, \dots\}$. We define the **probability function** or **probability mass function** for X by $f_X(x) = \mathbb{P}(X = x)$.

- A set is countable if it is finite or it can be put in a one-to-one correspondence with the integers. The even numbers, the odd numbers, and the rationals are countable; the set of real numbers between 0 and 1 is not countable.



Features of Random Variables

- Probability mass function (PMF)
 $f(x) = \text{Prob}(X=x)$ -- For discrete random variables
- Cumulative distribution (density) function (CDF)
 $F(x) = \text{Prob}(X \leq x)$
- Quantiles: x such that $F(x) = Q$
 - Median: $x = \text{median}$, $Q = 0.5$.



Properties of CDF

1. F is **non-decreasing**. That is, its graph never goes down, or symbolically if $a \leq b$ then $F(a) \leq F(b)$.
2. $0 \leq F(a) \leq 1$.
3. $\lim_{a \rightarrow \infty} F(a) = 1$, $\lim_{a \rightarrow -\infty} F(a) = 0$.



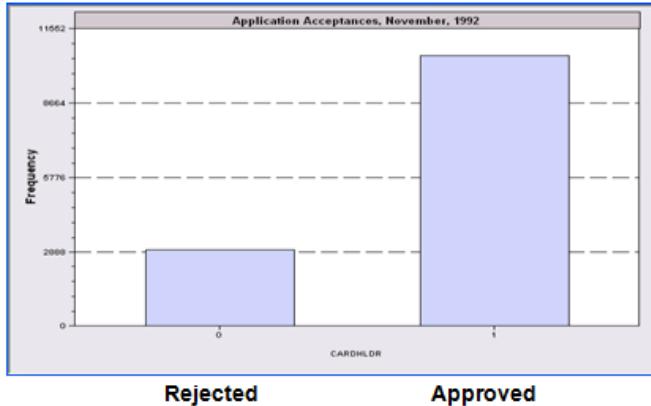
Discrete Random Variables

Bernoulli

← coin flip (can be biased unfair)



13,444 Applicants for a Credit Card (November, 1992)



Experiment = A randomly picked application.

Let $X = 0$ if Rejected

Let $X = 1$ if Accepted

X is DISCRETE (Binary). This is called a **Bernoulli random variable**.

$$X = \{0=\text{reject}, 1=\text{accept}\} \text{ or } X = 1[\text{Accepted}]$$

THE BERNOULLI DISTRIBUTION. Let X represent a binary coin flip. Then $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$ for some $p \in [0, 1]$. We say that \boxed{X} has a Bernoulli distribution written $\boxed{X \sim \text{Bernoulli}(p)}$. The probability function is $f(x) = \underline{p^x(1-p)^{1-x}}$ for $x \in \{0, 1\}$.

$$x=0 \quad f(x) = 1 - p$$

Type \uparrow parameters

BIA 652, Multivariate Data Analytics

$$x=1 \quad f(x) = p$$



Binomial: $X = \text{Number of successes in } n \text{ trials}$

THE BINOMIAL DISTRIBUTION. Suppose we have a coin which falls heads up with probability p for some $0 \leq p \leq 1$. Flip the coin n times and let X be the number of heads. Assume that the tosses are independent. Let $f(x) = \mathbb{P}(X = x)$ be the mass function. It can be shown that

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{for } x = 0, \dots, n \\ 0 & \text{otherwise.} \end{cases}$$

Sum of n Bernoulli trials

Family has 4 children.

$$\text{Prob}[4 \text{ daughters}] = \binom{4}{4} \left(\frac{1}{2}\right)^4 \left(1 - \frac{1}{2}\right)^0 = \frac{1}{16} \quad \text{What is P(2 daughters)?}$$

Weapons system has 20 components. It will fail if 2 or more break down.

Prob any component fails is 0.2. Component failures are independent.

$$\begin{aligned} \text{Prob(system breaks down)} &= \text{Prob}[X \geq 2] = \sum_{x=2}^{20} \binom{20}{x} \cdot 2^x \cdot 0.8^{20-x} \\ &= 1 - \text{Prob}[X < 2] \\ &= 1 - \text{Prob}(X=0) - \text{Prob}(X=1) \\ &= 1 - (1) \cdot 2^0 \cdot 0.8^{20} - (20) \cdot 2^1 \cdot 0.8^{19} = 1 - 0.011529 - 0.57646 \\ &= 1 - 8^{19} (0.8 + (20) \cdot 0.2) = 0.93082 \end{aligned}$$



Notation Reminder

- X is a random variable
- x denotes a particular value of the random variable
- n and p are parameters, that is, fixed real numbers.
- The parameter p is usually unknown and must be estimated from data.
- Be aware of the sample space and the support.



Poisson

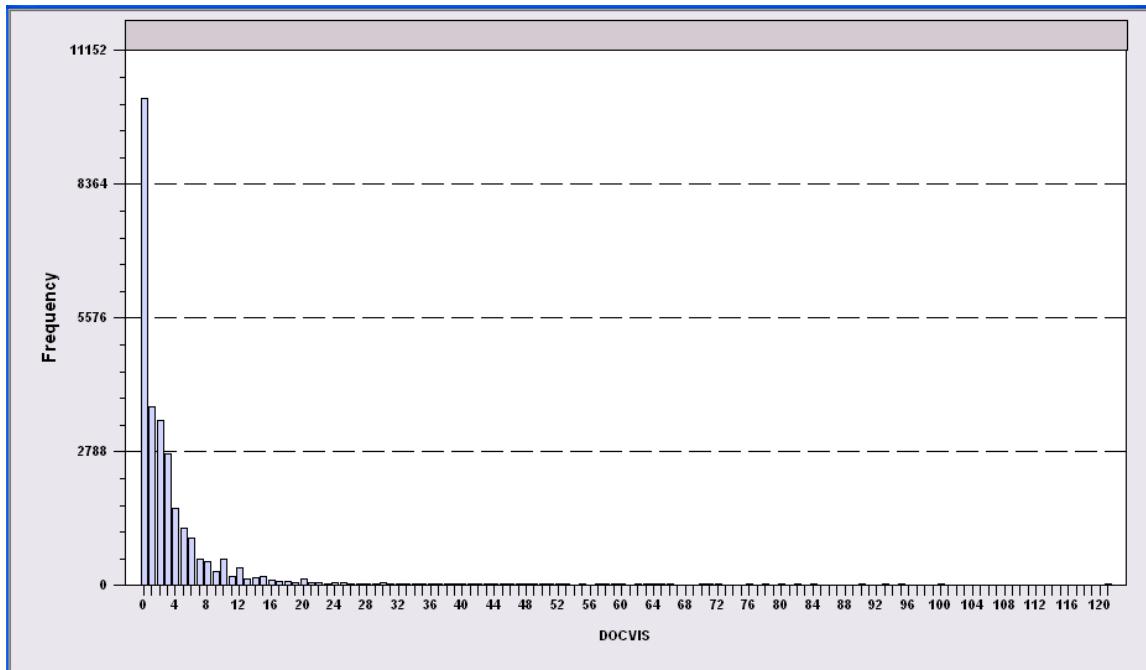
The **Poisson frequency function** with parameter λ ($\lambda > 0$) is

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

- **Two Interpretations:**

- Can be considered as an approximation to binomial
- General model for a type of process
(number of arrivals in a time period)

A Poisson Process: Doctor visits in the survey year by people in a sample of 27,326. $\lambda = .8$





Thank you!

Prof. Feng Mai
School of Business

For academic use only.