Introduction to data science & artificial intelligence (INF7100)

Arthur Charpentier

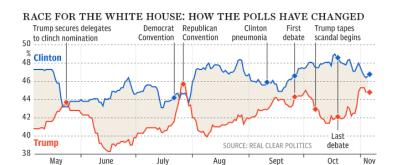
#131 Uncertainty and Randomness

été 2020

Uncertainty

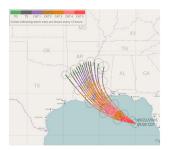


U.S. Elections



Uncertainty

Data from an experiment may appear rock solid. Upon further examination, the data may morph into something much less firm. A knee-jerk reaction to this conundrum may be to try and hide uncertain scientific results, which are unloved fellow travelers of science. After all, words can afford ambiguity, but with visuals, "we are damned to be concrete," says Bang Wong, who is the creative director of the Broad Institute of MIT and Harvard. The alternative is to face the ambiguity head-on through visual means.. via Data visualization: ambiguity as a fellow traveler (see also How data visualizations can clarify and confound uncertainty)





Probabilities

See visualizing uncertainty

Consider a (standard) dice, taking values $\{1, 2, 3, 4, 5, 6\}$ Here, lower case denotes specific values x, e.g. x = 3while upper case denotes a random variable X.

To describe X use can give its distribution,

$$\mathbb{P}(X=1) = \mathbb{P}(X=2) = \cdots = \mathbb{P}(X=5) = \mathbb{P}(X=6) = \frac{1}{6}$$

For any subset \mathcal{X} of $\{1,2,3,4,5,6\}$ we can compute $\mathbb{P}(X \in \mathcal{X})$



Birthday Paradox

Consider a set of n randomly chosen people. If $n \ge 23$, there is more than 50% chances that some pair of them will have the same birthday.

(assuming that each day of the year is equally probable for a birthday) "Class 1.pdf" p15

$$\mathbb{P}(A') = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \frac{362}{365} \times \dots \times \frac{343}{365} \simeq 49.2703\%$$

Poisson approximation,
$$\lambda=\frac{1}{365}\binom{23}{2}=\frac{253}{365}\simeq -0.6932$$
 so

$$\mathbb{P}(X > 0) = 1 - \mathbb{P}(X = 0) \simeq 1 - e^{-0.6932} \simeq 0.500002$$



Joint Distribution

Consider two dices.

 X_1 denotes the value of the first one,

 X_2 denotes the value of the second one.

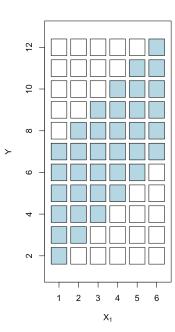
Let
$$Y = X_1 + X_2$$
.

$$\mathbb{P}(X_1 = x_1, Y = y)$$
 is the (joint) probability

$$x_1 \in \{1, 2, \dots, 6\}$$

▶
$$y \in \{2, 3, \dots, 12\}$$

E.g.
$$\mathbb{P}(X_1 = 4, Y = 3) = 0$$



Conditional Events

$$\mathbb{P}(Y = y | X_1 = x_1) \stackrel{def}{=} \frac{\mathbb{P}(X_1 = x_1, Y = y)}{\mathbb{P}(X_1 = x_1)}$$
or $\mathbb{P}(X_1 = x_1) \cdot \mathbb{P}(X_1 = x_1)$

$$= \mathbb{P}(X_2 = y - x_1)$$

$$\stackrel{\Sigma}{=} \mathbb{P}(X_2 = y - x_1)$$

$$\stackrel{\Sigma}{=} \mathbb{P}(X_2 = x_1) \stackrel{def}{=} \mathbb{P}(X_1 = x_1) \stackrel{def}{=} \mathbb{P}(X_2 = x_1)$$

$$\stackrel{\Sigma}{=} \mathbb{P}(X_2 = y - x_1)$$

$$\stackrel{\Sigma}{=} \mathbb{P}(X_2 = y - x_1)$$

$$\stackrel{\Sigma}{=} \mathbb{P}(X_2 = y - x_1)$$

Conditional Events

$$\mathbb{P}(Y = y | X_1 = x_1) \stackrel{def}{=} \frac{\mathbb{P}(X_1 = x_1, Y = y)}{\mathbb{P}(X_1 = x_1)}$$

and similarly

$$\mathbb{P}(X_1 = x_1 | Y = y) = \frac{\mathbb{P}(X_1 = x_1, Y = y)}{\mathbb{P}(Y = y)}$$

We can write

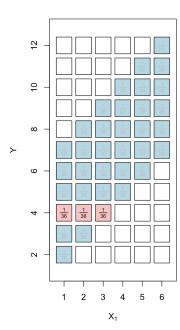
$$\mathbb{P}(Y=y\big|X_1=x_1)=\frac{\mathbb{P}(Y=y)}{\mathbb{P}(X_1=x_1)}\cdot\mathbb{P}(X_1=x_1\big|Y=y)$$

Conditional does not mean causal!



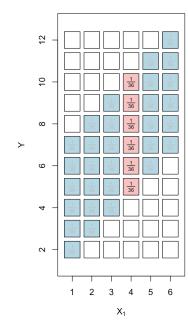
Marginal Events

$$\mathbb{P}(Y=y)=\sum_{x_1}\mathbb{P}(X_1=x_1,Y=y)$$
 $=\sum_{x_1}\mathbb{P}(Y=yig|X_1=x_1)\cdot\mathbb{P}(X_1=x_1)$ e.g. $\mathbb{P}(Y=4)=rac{3}{36}=rac{1}{12}$



Marginal Events

$$\mathbb{P}(X_1 = x_1) = \sum_{y} \mathbb{P}(X_1 = x_1, Y = y)$$
 $= \sum_{y} \mathbb{P}(X_1 = x_1 | Y = y) \cdot \mathbb{P}(Y = y)$
e.g. $\mathbb{P}(X_1 = 4) = \frac{6}{36} = \frac{1}{6}$



(no surprise here...)

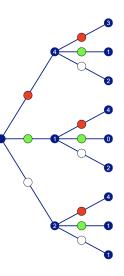
Trees & Sets

Consider an urn with 10 balls, 5 red ● 2 green ● and 3 white \circ .

The conditional probability of event A occurring given that event B occurred is defined as

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \text{ and } B]}{\mathbb{P}[B]}$$

or $\mathbb{P}[A \text{ and } B] = \mathbb{P}[A|B] \cdot \mathbb{P}[B]$ (so called chain rule) We draw two balls, without replacement what is the probability to have (at least) one green?



Trees & Sets

We draw two balls, without replacement what is the probability to have (at least) one green

$$p = \mathbb{P}\big[X_1 = ullet \text{ or } X_2 = ullet \big]$$

$$p = \mathbb{P}[X_1 = ullet] + \mathbb{P}[X_2 = ullet$$
 and $X_1
eq ullet]$

$$p = \mathbb{P}[X_1 = \bullet] + \mathbb{P}[X_2 = \bullet | X_1 = \bullet] \cdot \mathbb{P}[|X_1 = \bullet] + \mathbb{P}[X_2 = \bullet | X_1 = \circ] \cdot \mathbb{P}[|X_1 = \circ]$$

$$p = \frac{1}{10} + \frac{1}{9} \cdot \frac{5}{10} + \frac{1}{9} \cdot \frac{3}{10} = \dots = \frac{17}{45}$$

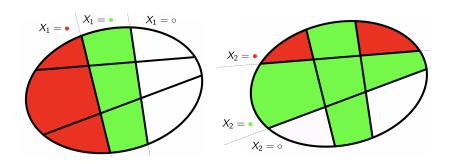
or more simple

$$p=1-\mathbb{P}\big[X_1\in\{ullet,\circ\} \text{ and } X_2\in\{ullet,\circ\}\big]=1-rac{8}{10}\cdotrac{7}{9}=rac{17}{45}$$



Trees & Sets

One can also consider sets,



 $\mathbb{P}[A \text{ and } B] = \mathbb{P}[A \cap B] \text{ and } \mathbb{P}[A \text{ or } B] = \mathbb{P}[A \cup B]$

Baves Rule

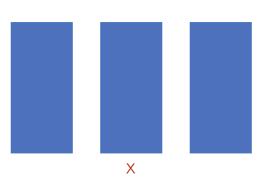
$$\mathbb{P}(X_{1} = x_{1} | Y = y) = \frac{\mathbb{P}(X_{1} = x_{1}) \cdot \mathbb{P}(Y = y | X_{1} = x_{1})}{\mathbb{P}(Y = y)}$$

$$\mathbb{P}(X_{1} = x_{1} | Y = y) = \frac{\mathbb{P}(X_{1} = x_{1}) \cdot \mathbb{P}(Y = y | X_{1} = x_{1})}{\sum \mathbb{P}(X_{1} = x) \cdot \mathbb{P}(Y = y | X_{1} = x)}$$



Monty Hall

Three doors: one has a treasure chest behind it and the other two have goats. You pick a door and indicate it to Monty. He opens one of the other two doors to reveal a goat. Now, should you stick to your initial choice, or switch to the other unopened door? see wikipedia



Letters to the Editor

<u> Leries io il le Luiloi</u>										
	A PROBLEM IN PROBABILITY	SECT. THE	contract to	own what he is	delay					
his Let Ref.	's Main a Deal'—a famous TV above staying Monte				States Sub-in- Sub-oil of Phili	in Heels				
Monte Hall:	One of the three boson inhelpt A, B, and C contains the lays to that now 2015 Lincoln Continued. The				Daily, of Calif Berhaley, Cit.	94700 94700				
	other two are amply. If you shows the box containing the keys, you win the car.		"A Problem is							
Contentant		Cotable	Consisty Moste Hell knows which has in the winner and							
Monte Tail:	Select one of these boson.	therefore,	therefore, would not open the box containing the large to the car Consider all consider automes.							
	Fil take how M.	CONSTRUCT S	a possion russ							
Monto Fall:	Now how A and how C are on the table and how is hos B positionism gripe how B rightly). It is possible the nex keps are in that had I'll give you \$100 for the hos.	in box	Conteston choses ben	Moote Hall spens box	Contestant switches A for Har C	Famil				
-	No, there you	*		8=0	A fire Marc	***				
Mosta Volt	Non about \$2007		ě.		Chr					
Contestant		ii ii		e	Abel	when				
Antierry			ъ	Apr	Miler A or C	Science William				
	Economics that the excludibly of your less	B	. 0		ABrC	100				
	containing the large to the car is 1/2 and the		ñ			whee				
	probability of your has being empty in 2/3. I'll give you \$600.	ē	ē	Aug	Chraus	leen.				
Antieren		Reservation	as above prob	ability of win	ning is 619 - 1	US. OF the				
	Me, I think I'll keep this box.	contestant.	contestant does not ewitch boxes, then his probability of winning							
Stante Hall	I'll do you a favor and open one of the remaining boson on the table the opens box Al. It's energic (buillence applasse). Now other box C or your box is containe the our lays. Since there are two losses	tional hos.	the car remains unchanged (1/9) after Moste Hall open an additional loss.							
	ieft, the probability of your ten considering the keys in new 1/2. I'd give you E1000 reals for your ton.									
	Narrin		PARCAL	GENERALIZA	LATTON'S					
beam on the thin beyond containing a has to be aboving his probability Contamination	right? The contention knows that at least one of the continue to table is empty. He new known is use he A. Done stage change this probability of lawing the length of the large length is 10.70 met of the beautiful the empty. Each Monte show the continues a four by empty. Each Monte show the continues a four by empty. The Monte show the continues a four by empty. The Monte show the continues as four by empty. The law is the law is the continue of the law is of continue flowers. Life 1.27. If the law is my less B for the less C on the table. That's writed?	methons of it desiration labeled (L- sets, Thee, in which as of problem of problem of into 2 + 2 N is sub- sement the a school left in the state left in the state left in the state left in the state of the state left in the state left in the state of the state left in the state left in the state of the state left in the state left in the state of the state left in the state left in the state left in the state of the state left in the state left in the state left in the state of the state left in the state left	Noted [] has resimbed used the value and department of general makins of Plant's Unsign but the millionide state. A strategies the millionist plant of the plant of the strategies of the plant behavior [1] as a fund principal strategies of the strategies of the latest the plant of the strategies of the strategies of the strategies of its behavior of the plant of the strategies of the strategies of the strategies and strategies on table the strategies of the strategies of the partition of strategies of the stra							
	THERN METHODEST UNIVERSITY levies application for the protion of PROFESSION OF STATISTICS could applicate matches bear a application for special applications of superiors bearing fairly and other medican of superiors requisite. Mall it am equal appendicularly in action employee. Prepared with 16.	which is such. Let us soles a contribut is which the content enter to the quantion in I_{c} , I_{c} , the case, I_{c} is two, I_{c} is two, I_{c} . It is the I_{c} is two, I_{c} is two, I_{c} . If I_{c} is I_{c} is I_{c} is I_{c} is I_{c} . If I_{c} is I_{c} is I_{c} is I_{c} is I_{c} is I_{c} . If I_{c} is I_{c} is I_{c} is I_{c} is I_{c} in I_{c} in I_{c} in I_{c} in I_{c} is I_{c} in I_{c}								
De A. Over, Confermen Department of Emission Department of Emission Developes Methodate University Developes Methodate University Developes Service of the Conferment of the American Service of Servi										
	D140 852-3445									



Monty Hall

Three doors: one has a treasure chest behind it and the other two have goats. You pick a door and indicate it to Monty. He opens one of the other two doors to reveal a goat. Now, should you stick to your initial choice, or switch to the other unopened door? see wikipedia







Letters to the Editor

	A PROBLEM IN PROBABILITY	SINT D	outstant in	own what he is	doing			
It is "Let's Male a Deal"—a famous TV show storying Monte Hadi.					States Sub-lin School of Pull Date of Out I			
Monte Hall:	One of the three boson labeled A, R, and C contains the lays to that new 2015 Liscoln Continuental. The other two are semply. If you choose the box		*A Problem is		Behaley, CA			
	containing the keys, you win the can.	Cesardy Moste Hell house which has is the winner an						
Continuent		therefore, would not open the box containing the large to the car Consider all consider subcomes.						
	Select one of these boson.							
	Fli take bee M.							
Monto Fail:	Now how A and how C are on the table and have in box B (contentant pripe how B tightip). It is possible the new large are in than hard I'll give you \$100 for	Kon are in box	Consessor chooses ber	Moote Hall spare box	Contactent reliables	Famili		
	the box.							
	No, there yes.			. c	B for A	whe		
	How about \$2007	6	ō	8	C for A	- 22		
Contestant		8.		April	Abr Aur C	home		
			ě.					
Moute Fall					Abre			
	containing the large to the car is 1/2 and the							
	probability of your hos being empty is 2/2. I'll give you \$600.	ē	ē	Aug	Chekerk	-		
	Mill.	Responsible shows probability of winning is 649 - 545. (f th						
	Ma, I think I'E keep this box.	contested does not evitab house, then his probability of winels the our remains unchanged (1/2) after Monte Half open on seld						
Monte Hall	I'll de you a fever and open one of the remaining beam an the halle the open box Al. I'v stopp blassimers applicated. Nor either her Cor year her K contine the our lays. Since there are two boxes lift, the probability of year less considering the Kays in new 1/E. I'll give you E000 reads for your box.	tional Jun.	_		_	e uz edi		
	Natrin	PARCAL GENERALIZATIONS						
beam on the Shin knowle containing of has to be showing his	right? The contentant because that at least one of the a table is empty. He now because it was been A. Daws edge change his probability of having the been be keys from 1.78 is 1.97 flow of the because the table have by men to be a content of the because the table may be a flow to the because when the down by me which of the two because was empty? In the of whentier flows 1.25 is 1.27.	Exits [1] has reminded us of the value and degener of praces actions of Panel? Unsage to the millimental uses. A standar- derivation of such formatic offers the sation of startlant labeled (i.e. distinguishable) objects to include pigeneshote, sets. Then, Shard formatic [1] easily follows these consideration in which set the ultimate object in planel. There is mentione due of problems quicky from the combility task uniter or lack.						









Monty Hall

 $\mathbb{P}(\text{treasure behind the other door})$

- $= \mathbb{P}(\text{treasure behind the other door}|X \text{ was correct}) \cdot \mathbb{P}(X \text{ was correct})$
- $+\mathbb{P}(\text{treasure behind the other door}|X \text{ was wrong}) \cdot \mathbb{P}(X \text{ was wrong})$

$$= 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}$$

so yes, we should switch...



Exercise

The probability that a woman has breast cancer is 1% If a woman has breast cancer, probability to test positive is 90% If a woman does not have breast cancer, the probability that she nevertheless tests positive is 9%

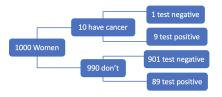
A 50-year-old woman, no symptoms, participates in routine mammography screening. She tests positive, is alarmed, and wants to know from you whether she has breast cancer for certain or what the chances are. Apart from the screening results, you know nothing else about this woman. How many women who test positive actually have breast cancer? What is the best answer?

- A) nine in 10
- B) eight in 10
- C) one in 10
- D) one in 100



Exercise & Probability Trees

The probability that a woman has breast cancer is 1% If a woman has breast cancer, probability to test positive is 90% If a woman does not have breast cancer, the probability that she nevertheless tests positive is 9%



$$\mathbb{P}[\mathsf{have\ cancer}|\mathsf{test\ positive}] = \frac{9}{9+89} \simeq \frac{1}{10}$$

See Do doctors understand test results? half the group of 160 gynaecologists responded that the woman's chance of having cancer was nine in 10

Independence

Definition: Two events A and B are independent if the probability of B occurring is the same whether or not A occurs.

Example: $A = \{ \text{ first coin is heads } \} \text{ and } B = \{ \text{ second coin is } \} \}$ heads }

Formally, $\mathbb{P}[B|A] = \mathbb{P}[B]$ or $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$

Quizz: with two dices, $A = \{ \text{ first dice is } 6 \} \text{ and } B = \{ \text{ sum } > 6 \}$

Quizz: with two cards (deck of 52),

 $A = \{ \text{ first is heart } \} \text{ and } B = \{ \text{ second is club } \}$

Quizz: with two cards (deck of 52), $A = \{ \text{ first is heart } \} \text{ and } B = \{ \text{ second is } 10 \} \}$

Quizz: $A = \{ \text{ first child boy} \} \text{ and } B = \{ \text{ second child boy} \}$

Independence

Definition: Two events A and B are independent if the probability of B occurring is the same whether or not A occurs.

Example: $A = \{ \text{ first coin is heads } \} \text{ and } B = \{ \text{ second coin is } \} \}$ heads }

Formally, $\mathbb{P}[B|A] = \mathbb{P}[B]$ or $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$

Quizz: with two dices, $A = \{ \text{ first dice is } 6 \} \text{ and } B = \{ \text{ sum } > 6 \}$ Not independent $(A \subset B)$

Quizz: with two cards (deck of 52),

$$A = \{ \text{ first is heart } \} \text{ and } B = \{ \text{ second is club } \}$$

Not independent
$$\mathbb{P}[A \cap B] = \frac{1}{4} \cdot \frac{13}{51} > \frac{1}{4} \cdot \frac{13}{52} = \mathbb{P}[A] \cdot \mathbb{P}[B]$$

Quizz: with two cards (deck of 52),

$$A = \{ \text{ first is heart } \} \text{ and } B = \{ \text{ second is } 10 \}$$

Independent
$$\mathbb{P}[A \cap B] = \frac{12}{52} \cdot \frac{4}{51} + \frac{1}{52} \cdot \frac{3}{51} = \frac{1}{52} = \mathbb{P}[A] \cdot \mathbb{P}[B]$$

Quizz: $A = \{ \text{ first child boy} \} \text{ and } B = \{ \text{ second child boy} \}$