ML & Advanced Analytics For Biomedicine

Ishanu Chattopadhyay

Department of Medicine

University of Chicago

CCTS 40500 / CCTS 20500 / BIOS 29208

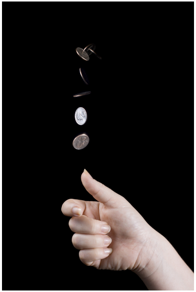
Winter 2020

Today

- Probability Theory
- Bayesian Statistics
- Naive Bayes Classifier

Probability Theory

- **Probability Theory is the cornerstone of ML. Almost all arguments are probabilistic.**
  - **when event A occurs, B generally follows**
  - **complicated model of a million possible events, and wish to simplify the model by allowing it to neglect the least likely events**

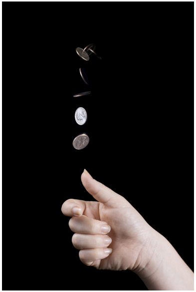


Possible Outcomes: H, T
Probabilities: $Pr(H) = Pr(T) = 0.5$

$$\frac{\#H}{\#Tosses} = \frac{\#H}{\#H + \#T} \longrightarrow 0.5 \tag{1}$$

Probability Theory

- **Probability Theory is the cornerstone of ML. Almost all arguments are probabilistic.**
  - **when event A occurs, B generally follows**
  - **complicated model of a million possible events, and wish to simplify the model by allowing it to neglect the least likely events**



Possible Outcomes: H, T **(Sample Space)**

Probabilities: $Pr(H) = Pr(T) = 0.5$

$$\frac{\#H}{\#Tosses} = \frac{\#H}{\#H + \#T} \longrightarrow 0.5 \tag{1}$$

- **Notion of Independence**
- **General Notion of Probability Measure**

Notion of Independence

Possible Outcomes: $1, 2, 3, 4, 5, 6$
Probabilities: $Pr(1) = Pr(2) = Pr(3) = Pr(4) = Pr(5) = Pr(6) = \frac{1}{6}$

$$\forall i, \frac{\#i}{\sum_{i=1}^{i=6} \#i} \longrightarrow \frac{1}{6} \tag{2}$$

Possible Outcomes: $(1, 1), (1, 2), (1, 3), \cdots, (6, 6)$
Probabilities: $\forall i, j, Pr((i, j)) = \frac{1}{36}$

$$\forall i, \frac{\#i}{\sum_{i=1}^{i=6} \#i} \longrightarrow \frac{1}{6} \times \frac{1}{6} \tag{3}$$

Notion of Independence

Possible Outcomes: $1, 2, 3, 4, 5, 6$ **(Sample Space)**
Probabilities: $Pr(1) = Pr(2) = Pr(3) = Pr(4) = Pr(5) = Pr(6) = \frac{1}{6}$

$$\forall i, \frac{\#i}{\sum_{i=1}^{i=6} \#i} \longrightarrow \frac{1}{6} \tag{2}$$

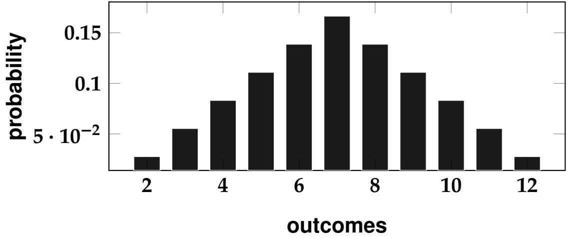
Possible Outcomes: $(1, 1), (1, 2), (1, 3), \cdots, (6, 6)$
Probabilities: $\forall i, j, Pr((i, j)) = \frac{1}{36}$

$$\forall i, \frac{\#i}{\sum_{i=1}^{i=6} \#i} \longrightarrow \frac{1}{6} \times \frac{1}{6} \tag{3}$$

Combining Events

What is the probability distribution of the sum of the rolls?

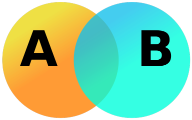
$1+1=2$	$2+1=3$	$3+1=4$	$4+1=5$	$5+1=6$	$6+1=7$
$1+2=3$	$2+2=4$	$3+2=5$	$4+2=6$	$5+2=7$	$6+2=8$
$1+3=4$	$2+3=5$	$3+3=6$	$4+3=7$	$5+3=8$	$6+3=9$
$1+4=5$	$2+4=6$	$3+4=7$	$4+4=8$	$5+4=9$	$6+4=10$
$1+5=6$	$2+5=7$	$3+5=8$	$4+5=9$	$5+5=10$	$6+5=11$
$1+6=7$	$2+6=8$	$3+6=9$	$4+6=10$	$5+6=11$	$6+6=12$



Combining Events

- **what is the probability that either a 1 or 6 come up on any particular toss?**
  - **Pr(1) + Pr(6) = (0+5)/36 = 5/36**

- **what is the probability that the outcome is prime?**
  - **Pr(2) + Pr(3)+Pr(5)+Pr(7)+Pr(11) = (1+2+4+6+2)/36 = 15/36 = 5/12**



$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B) \tag{4}$$

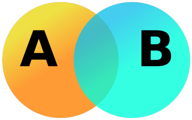
A **and** B **are independent** $\Leftrightarrow$**:**

$$Pr(A \cap B) = Pr(A)Pr(B) \tag{5}$$

Combining Events

- **what is the probability that either a 1 or 6 come up on any particular toss?**
  - **Pr(1) + Pr(6) = (0+5)/36 = 5/36**

- **what is the probability that the outcome is prime?**
  - **Pr(2) + Pr(3)+Pr(5)+Pr(7)+Pr(11) = (1+2+4+6+2)/36 = 15/36 = 5/12**



$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B) \tag{4}$$

A and B are independent $\Leftrightarrow$:

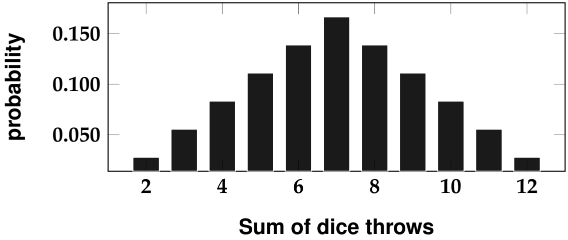
$$Pr(A \cap B) = Pr(A)Pr(B) \tag{5}$$

Q. What happens to the equality when A and B are not independent? It is possible that

$$Pr(A \cap B) < Pr(A)Pr(B)? \tag{6}$$

Can you give an example?

What is the connection between sample space, event probabilities, and random variables?



Define variable X : Sum of Dice Throws
Then X is a function that maps the sample space to the space of integers.

$$X : \Omega \rightarrow \mathbb{N} \tag{7}$$

Here $\mathscr{S} = (1,1), (1,2), (1,3), \cdots, (6,6)$

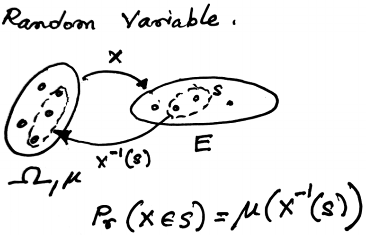
Then, we can ask:
What is the probability that X takes the value 4?
What is the probability that X takes the value between 4 and 6?

Random Variables & Probability Distributions

A random variable $X : \Omega \to E$ is a measurable function from a set of possible outcomes Ω to a measurable space E

Let Ω be the sample space. In the example of 2 dice, the sample space is the set of all possible outcomes $(1,1),(1,2),\cdots,(6,6)$.
Then $X : \Omega \to E$ is a function on the sample space.
For some subset $S \subset E$, we can then calculate:

$$\boxed{Pr(X \in S) = Pr\left(\{\omega \in \Omega : X(\omega) \in S\}\right)} \tag{8}$$

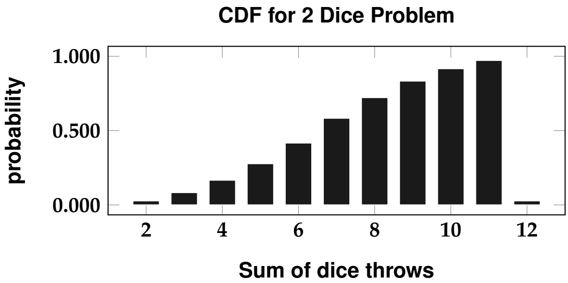


Example of Dice:

$$U = \{(1,1),(1,2),\cdots,(6,6)\} \tag{9}$$
$$\Omega = 2^{U} \tag{Sample space}$$
$$\mu : \Omega \to [0,1] \tag{Prob. Measure}$$
$$\mu(\emptyset) = 0 \tag{10}$$
$$\forall a \in U, \mu(\{a\}) = 1/36 \tag{11}$$
$$\forall a,b \in U, \mu(\{a,b\}) = \mu(a) + \mu(b) \tag{12}$$
$$X : \Omega \to \mathbb{N} \tag{13}$$
$$\forall (i,j) \in U, X((i,j)) = i + j \tag{14}$$

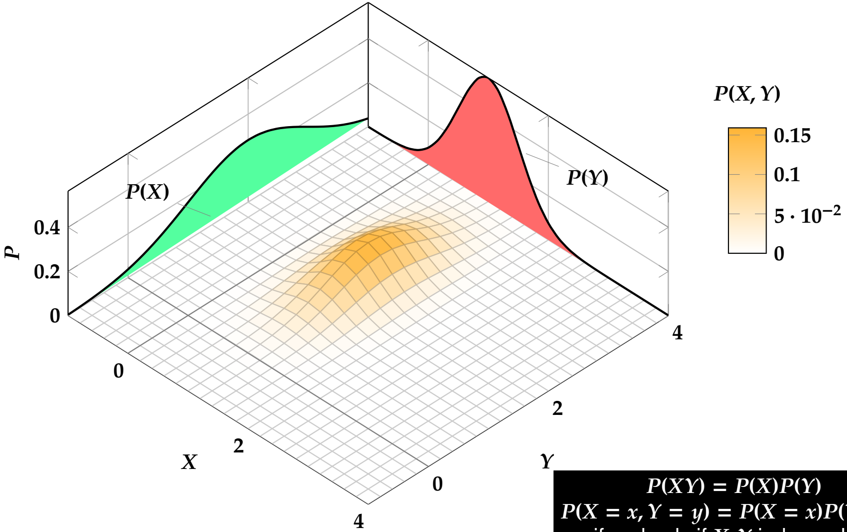
Random Variables & Probability Distributions

- Random variables can be discrete or continuous.
- For any discrete random variable, we can define a corresponding discrete distribution.
- We also can define the cumulative distribution, which is $Pr(X \leqq x)$



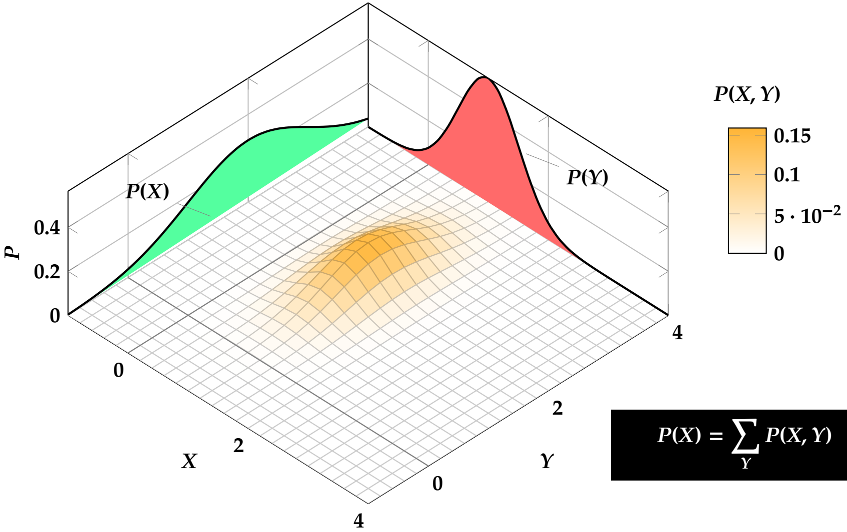
- CDFs are always monotonicaly non-decreasing
- In the continuous domain, the analogous notion to "probability distribution" in the discrete case is "probability density function".
- Loosely speaking the density function is the derivative of the distribution. You have to be careful about the differentiation.

Joint Distributions



$$P(XY) = P(X)P(Y)$$
$$P(X = x, Y = y) = P(X = x)P(Y = y)$$
if and only if X, Y independent

Marginal Distributions

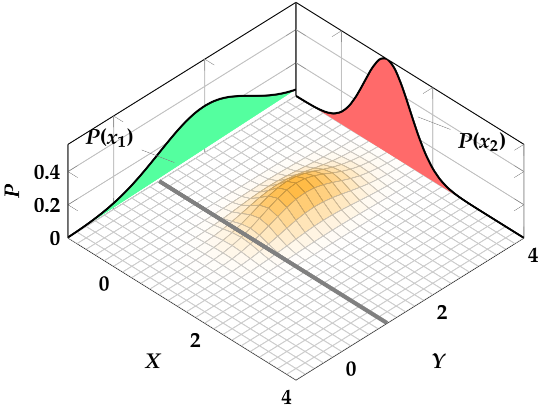


$$P(X) = \sum_{Y} P(X, Y) \qquad (15)$$

Conditional Probability

Incorporating knowledge of events

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}} \tag{16}$$



	y_1	$y_\star$	y_m
x_1	$\cdots$	p_1	
$\vdots$	$\cdots$	p_i	
x_n		p_n	

$$\begin{aligned} &Pr(X|Y = y_\star) \\ &= Pr(X, Y = y_\star) / \sum_{Y=Y_\star} Pr(X, Y = y_\star) \end{aligned}$$

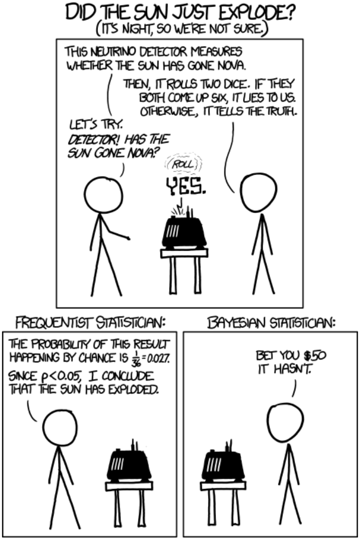
The Chain Rule

$$P(A,B) = \frac{P(A,B)}{P(B)}P(B) = P(A|B)P(B) \tag{17}$$

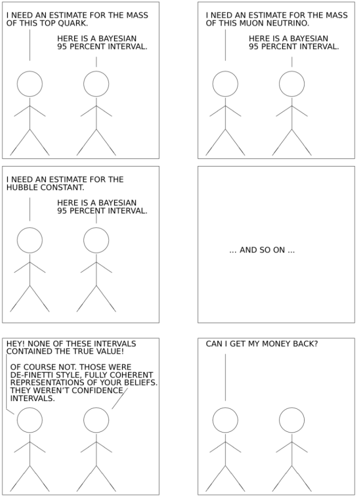
This generalizes to n variables:

$$P(X_1, \cdots, X_n) = \prod_{i=1}^{N} P(X_n|X_1, \cdots, X_{i-1}) \tag{18}$$

Bayes Rule



DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
ROLL
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS $\frac{1}{36}=0.027$.
SINCE $p<0.05$, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU $50 IT HASN'T.



I NEED AN ESTIMATE FOR THE MASS OF THIS TOP QUARK.
HERE IS A BAYESIAN 95 PERCENT INTERVAL.
I NEED AN ESTIMATE FOR THE MASS OF THIS MUON NEUTRINO.
HERE IS A BAYESIAN 95 PERCENT INTERVAL.
I NEED AN ESTIMATE FOR THE HUBBLE CONSTANT.
HERE IS A BAYESIAN 95 PERCENT INTERVAL.
... AND SO ON ...
HEY! NONE OF THESE INTERVALS CONTAINED THE TRUE VALUE!
OF COURSE NOT. THOSE WERE DE-FINETTI STYLE, FULLY COHERENT REPRESENTATIONS OF YOUR BELIEFS. THEY WEREN'T CONFIDENCE INTERVALS.
CAN I GET MY MONEY BACK?

Bayes Rule

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{\sum P(B|A)P(A)}} \tag{19}$$

- **Why is this advantageous?**

Example (Chance of rare disease given positive test)

A blood test for a rare disease correctly returns positive 99% of the time if the subject indeed has the disease. And falsely indicates a positive result in 1% of subjects who do not have the disease. The prevalence of the disease in the population is 0.1%.
What is the probability that you have the disease if you test positive?

$$Pr(H|E) = \frac{Pr(E|H)Pr(H)}{Pr(E|H)Pr(H) + Pr(E|\neg H)Pr(\neg H)} \tag{20}$$

$$= \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.01 \times 0.999} = 0.0902 \approx 9\% \tag{21}$$

Independence and Conditional Independence

- **Independence ($X \perp Y$)**

$$P(XY) = P(X)P(Y) \tag{22}$$

$$P(X|Y) = \frac{P(XY)}{P(Y)} = P(X) \tag{23}$$

$$P(Y|X) = P(Y) \tag{24}$$

- **Conditional Independence ($(X \perp Y)|Z$)**

$$P(X|ZY) = P(X|Z) \tag{25}$$

$$P(Y|ZX) = P(Y|Z) \tag{26}$$

$$P(XY|Z) = P(X|Z)P(Y|Z) \tag{27}$$

Example 2

Example (Chance of rare disease given mutiple positive tests)

A blood test for a rare disease correctly returns positive 99% of the time if the subject indeed has the disease. And falsely indicates a positive result in 1% of subjects who do not have the disease. The prevalence of the disease in the population is 0.1%.
What is the probability that you have the disease if you test positive twice?

$$
\begin{aligned}
Pr(H|E) &= \frac{Pr(E|H)Pr(H)}{Pr(E|H)Pr(H) + Pr(E|\neg H)Pr(\neg H)} && (28)\\
&= \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.01 \times 0.999} = 0.0902 \approx 9\% && (29)
\end{aligned}
$$

$$
\begin{aligned}
Pr(H|E) &= \frac{Pr(E|H)Pr(H)}{Pr(E|H)Pr(H) + Pr(E|\neg H)Pr(\neg H)} && (30)\\
&= \frac{0.99 \times 0.0902}{0.99 \times 0.0902 + 0.01 \times (1 - 0.0902)} = 0.908 \approx 91\% && (31)
\end{aligned}
$$

Bayesian Framework

$$P(H|E) = \frac{P(E|H)\,P(H)}{P(E)}$$

Prior Probability

Model Evidence or Marginal likelihood

Likelihood

Posterior Probability.

$$P(E) = P(E|H)\,P(H) + P(E|\neg H)\,P(\neg H)$$

$$P(H|E) = \frac{P(E|H)\,P(H)}{P(E|H)\,P(H) + P(E|\neg H)\,P(\neg H)}$$

$$P(H|E) \propto P(E|H)\,P(H)$$

Posterior $\propto$ Likelihood $\times$ Prior

Bayesian Inference

$$x \sim P(x|\theta)$$

$$\theta \sim P(\theta|\alpha)$$

Sampling distribution or Likelihood $P(X|\theta)$ $L(\theta|X)$

Marginal Likelihood $P(X|\alpha)$

Posterior $P(\theta|X,\alpha)$

Bayesian Inference

$$P(x|\alpha) = \int P(x|\theta)\,P(\theta|\alpha)\,d\theta$$

$$P(\theta|x,\alpha) = \frac{P(\theta,x,\alpha)}{P(x,\alpha)} = \frac{P(x|\theta,\alpha)\,P(\theta,\alpha)}{P(x|\alpha)\,P(\alpha)}$$

$$= \frac{P(x|\theta,\alpha)\,P(\theta|\alpha)}{P(x|\alpha)}$$

Bayesian Prediction

$$P(\tilde{x} \mid X, \alpha) = \int P(\tilde{x} \mid \theta) P(\theta \mid X, \alpha) \, d\theta$$

$$P(\tilde{x} \mid \alpha) = \int P(\tilde{x} \mid \theta) P(\theta \mid \alpha) \, d\theta$$

Conjugate Priors

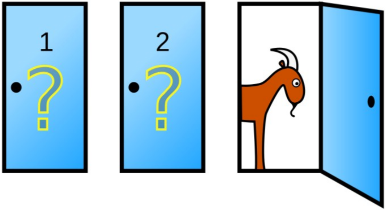
Conjugate Priors.

$$P(\theta|\alpha) \longrightarrow P(\theta|X,\alpha)$$

$$f(\alpha_1,\ldots,\alpha_k) \longrightarrow f(\alpha_1',\ldots,\alpha_k')$$

where $\alpha_i' = f(\alpha_1,\ldots,\alpha_k, X)$

Example 2: The Monty Hall Problem



In a gameshow, contestants try to guess which of 3 closed doors contain a cash prize (goats are behind the other two doors).

Odds of choosing the correct door are 1 in 3.

As a twist, the host of the show occasionally opens a door after a contestant makes his or her choice. This door is always one of the two the contestant did not pick, and is also always one of the goat doors (note that it is always possible to do this, since there are two goat doors).

At this point, the contestant has the option of keeping his or her original choice, or swtiching to the other unopened door.

The question is: is there any benefit to switching doors?

Example (Two Child Problem)

The two child problem asks the following question: I have two children, and at least one of them is a son. What is the probability that both children are boys?
Hint: It is not 1/2.
Here is a table of the possible family makeups:

First child	Second child	
Boy	Boy	
Boy	Girl	
Girl	Boy	
Girl	Girl	Not possible

The probability that both children are boys is actually **1/3**.

Q. Show this using Bayes Theorem

Q. I have two children, and at least one is a son born on a Tuesday. What is the probability that I have two boys?