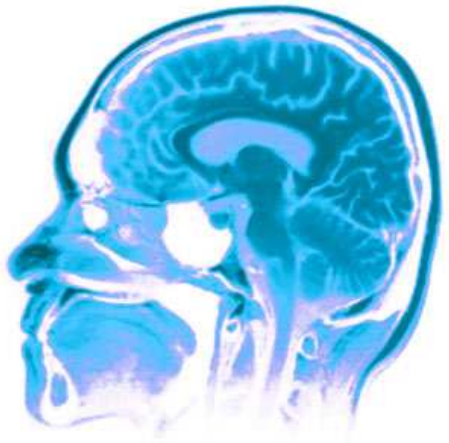
# CPSC340



## Probability

Nando de Freitas
*September, 2012*
*University of British Columbia*

# Outline of the lecture

This lecture is intended at revising probabilistic concepts that play an important role in the design of machine learning and data mining algorithms. It is expected that you will learn and master the following three topics:

1. Frequentist and axiomatic definition of probability.
2. Conditioning.
3. Marginalization.

# Probability as frequency

Consider the following questions:

1. What is the probability that when I flip a coin it is “heads”?

   1/2

2. Why?

   frequency

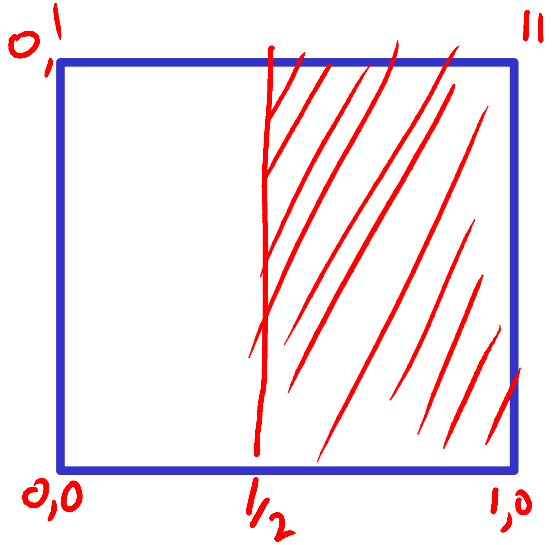
3. What is the probability that the Lion’s gate bridge will collapse before the term is over?

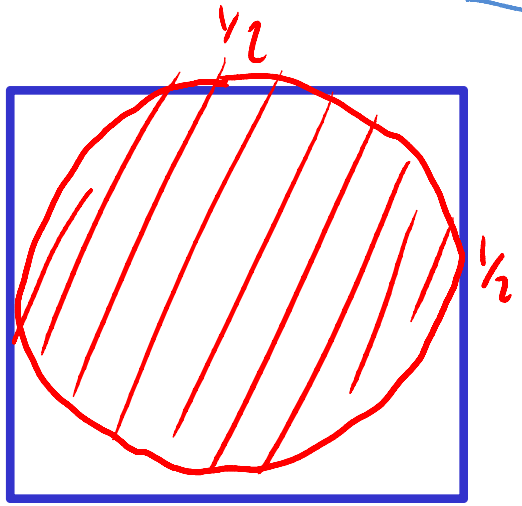
   Can't compute, 0.005, 0, 0.005 ...

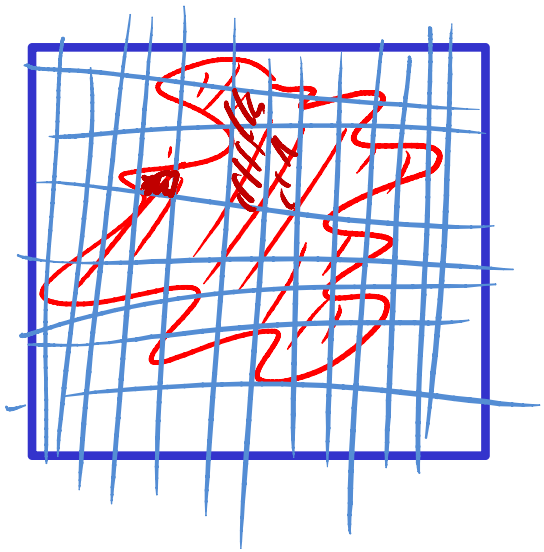
**Message:** *The frequentist view is very useful, but it seems that we also use domain knowledge to come up with probabilities. Moreover, it seems that probability can be subjective (different people have different probabilities for the same event).*

# Probability as measure

Imagine we are throwing darts at a wall of size 1x1 and that all darts are guaranteed to fall within this 1x1 wall. What is the probability that a dart will hit the shaded area?



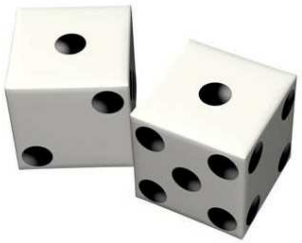
Prob = 1/2

Prob: $\frac{\pi}{4}$

Prob. = $\frac{\#\text{ red boxes}}{\#\text{ boxes}}$

**Message:** *Probability is a measure of certainty of an event taking place. i.e. in the example above we were measuring the chances of hitting the shaded area.*

# Probability



***Probability*** *is the formal study of the laws of chance. Probability allows us to* ***manage uncertainty****.*

*The* ***sample space*** *is the set of all* ***outcomes****. For example, for a die we have 6 outcomes:*

$\Omega_{die} = \{1, 2, 3, 4, 5, 6\}$ $\Omega_{coin} = \{H, T\}$

*Probability allows us to measure many* ***events****.* ***The events are subsets of the sample space****. For example, for a die we may consider the following events:*

$Even = \{2, 4, 6\}$ $Odd = \{1, 3, 5\}$

$greaterThanFour = \{5, 6\}$

*We assign probabilities to these events:*

$P(Even) = 1/2$ $P(gTF) = \frac{2}{6} = \frac{1}{3}$

# The axioms

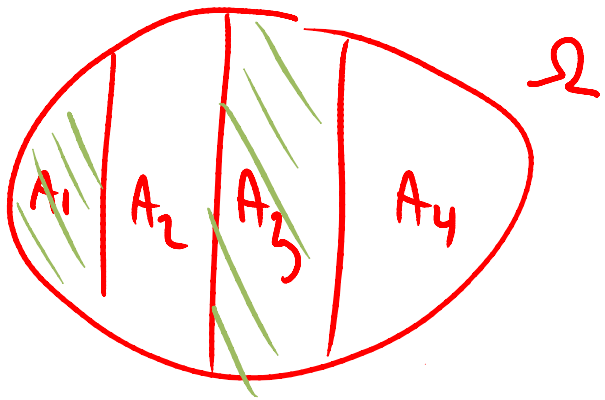
*The following two laws are the key axioms of probability:*

1. $P(\emptyset) = 0 \leq p(A) \leq 1 = P(\Omega)$

2. For **disjoint sets** $A_n$, $n \geq 1$, we have

$$P\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

null event




$\Omega = A_1 \cup A_2 \cup A_3 \cup A_4$

$P(\Omega) = 1$

$P(A_1 \cup A_2 \cup A_3 \cup A_4) = 1$

$= P(A_1) + P(A_2) + P(A_3) + P(A_4)$

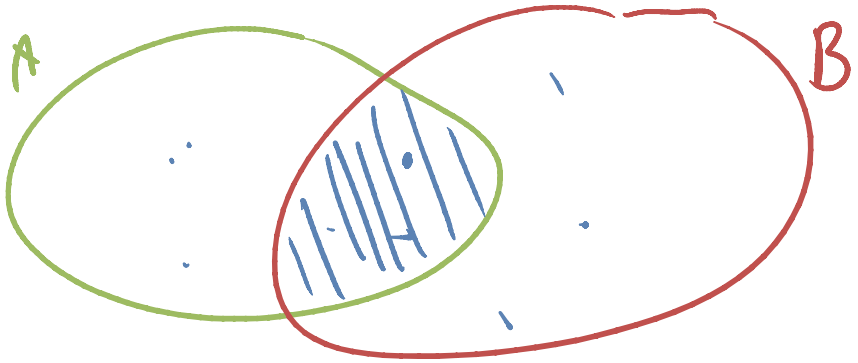
*We can visualize all these sets with* ***Venn Diagrams***

# Or and And operations

*Given two events, A and B, that are not disjoint, we have:*

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

intersection



# Conditional probability

*Assuming (given that) B has been observed (i.e. there is no uncertainty about B), the following tells us the probability that A will take place:*

$$P(A \text{ given } B) = P(A \text{ and } B) / P(B)$$

*That is, in the frequentist interpretation, we calculate the ratio of the number of times both A and B occurred and divide it by the number of times B occurred.*

*For short we write:* $P(A|B) = P(AB)/P(B)$*; or* $P(AB)=P(A|B)P(B)$*, where P(A|B) is the* ***conditional*** *probability, P(AB) is the* ***joint****, and P(B) is the* ***marginal****.*

*If we have more events, we use the chain rule:*

$$P(ABC) = P(A|BC)\, P(B|C)\, P(C)$$

# Conditional probability

*Can we write the joint probability,* ***P(AB)****, in more than one way in terms of conditional and marginal probabilities?*

$$P(AB) = P(A|B)\,P(B)$$

$$P(AB) = P(B|A)\,P(A)$$

# Independence

*We know that:*

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

*But what happens if **A** does not **depend** on **B**? That is, the value of **B** does not affect the chances of **A** taking place. How does the above expression simplify?*

$$P(AB) = P(A)P(B)$$

*How does the expression below simplify?*

$P(ABCD) =$ $P(A)P(B)P(C)P(D)$

# Conditional probability example

*Assume we have a dark box with 3 red balls and 1 blue ball. That is, we have the* ***set {r,r,r,b}****. What is the probability of drawing 2 red balls in the first 2 tries?*

$P(B_1 = r, B_2 = r) = P(B_2 = r \mid B_1 = r)\, P(B_1 = r) = \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{2}$

AND

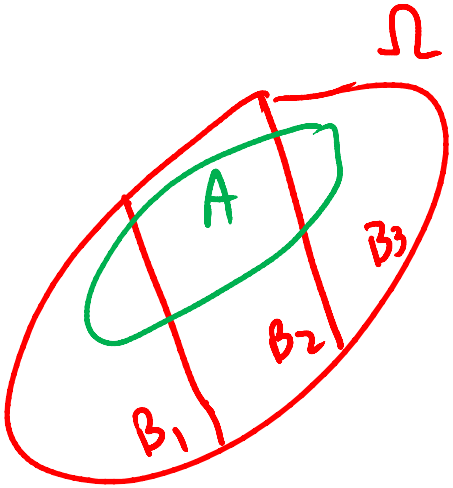
given

# Marginalization

Let the sets $B_{1:n}$ be disjoint and $\bigcup_{i=1}^{n} B_i = \Omega$. Then

$$P(A) = \sum_{i=1}^{n} P(A, B_i) = P(AB_1) + P(AB_2) + \cdots + P(AB_n)$$

*Proof sketch:*




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

$$= P[(A \cap B_1) \cup (A \cap B_2) \cup (AB_3)]$$

$$= P(AB_1) + P(AB_2) + P(AB_3)$$

# Conditional probability example

*What is the probability that the 2nd ball drawn from the* ***set {r,r,r,b}*** *will be red?*

*Using marginalization,* $\boldsymbol{P(B_2 = r)}$ $= P(B_2 = r, B_1 = r) + P(B_2 = r, B_1 = b)$

$= \frac{1}{2} + 1 \times \frac{1}{3}$

$= \frac{3}{4}$

# Matrix notation

*Assume that* $X$ *can be 0 or 1. We use the math notation:* $X \in \{0,1\}$.

*Let* $P(X_1=0) = 3/4$ *and* $P(X_1=1) = 1/4$. *Assume too that* $P(X_2=1|X_1=0) = 1/3$, $P(X_2=1|X_1=1) = 0$. *Then,* $P(X_2=0|X_1=0) = 2/3$ , $P(X_2=0|X_1=1) = 1$

*We can obtain an expression for* $P(X_2)$ *easily using matrix notation:*

$$\begin{matrix} & \begin{matrix} X_1 & \\ 0 & 1 \end{matrix} \\ \begin{matrix} \\ X_1 \\ 0 \\ 1 \end{matrix} & \begin{bmatrix} 3/4 & 1/4 \end{bmatrix} \end{matrix} \begin{matrix} & \begin{matrix} X_2 & \\ 0 & 1 \end{matrix} \\ \begin{matrix} \\ X_1 \\ 0 \\ 1 \end{matrix} & \begin{bmatrix} 2/3 & 1/3 \\ 1 & 0 \end{bmatrix} \end{matrix} = \begin{matrix} & \begin{matrix} X_2 & \\ 0 & 1 \end{matrix} \\ & \begin{bmatrix} 3/4 & 1/4 \end{bmatrix} \end{matrix}$$

which agrees with the previous slide i.e. $P(B_2=r) = 3/4$

Componentwise matrix vector product is:

$$\sum_{x_1 \in \{0,1\}} P(x_1) P(x_2|x_1) = P(X_2)$$

# Matrix notation

$[1\ 2]\begin{bmatrix}1 & 0\\ 3 & 2\end{bmatrix} = [7\ 4]$

*We have:*

$$\sum_{X_1 \in \{0,1\}} P(X_1)\, P(X_2|X_1) = P(X_2)$$

$7 = 1\times 1 + 2\times 3$

$4 = 1\times 0 + 2\times 2$

*For short, we write this using vectors and a* ***stochastic matrix****:*

$$\underset{1\times 2}{\pi_1^T}\ \underset{2\times 2}{G} = \underset{1\times 2}{\pi_2^T} \quad \equiv \quad \pi_2(j) = \sum_{i=1}^{2} \pi_1(i)\, G(i,j)$$

*Imagine we kept on multiplying by* ***G****.*

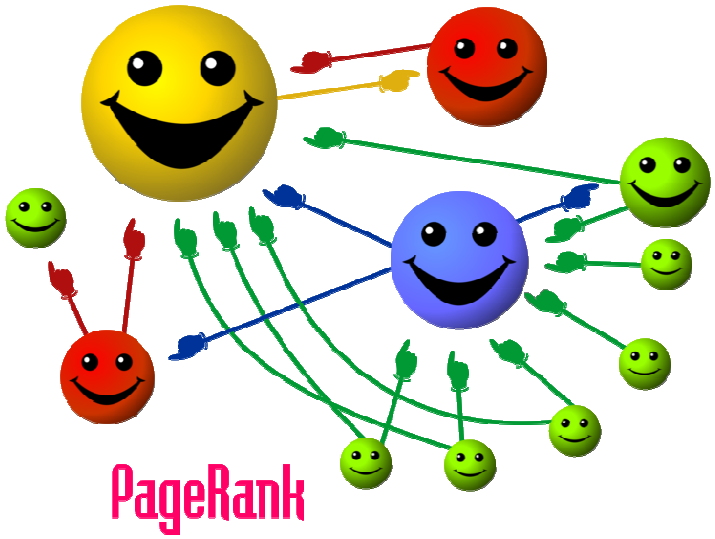
$$\pi_1^T G = \pi_2^T,\ \pi_2^T G = \pi_3^T,\ \pi_3^T G = \pi_4^T,\ \ldots\ \pi_{k-1}^T G = \pi_k^T \quad \equiv \quad \pi_1^T G^{k-1} = \pi_k^T$$

*Claim: For a very large number k, after k iterations, the value of* $\boldsymbol{\pi}$ *stabilizes.* if $\pi_k = \pi$, then $\pi^T G = \pi^T$

*That is,* $\boldsymbol{\pi_k}$ *is an* ***eigenvector*** *of* **G** *with* ***eigenvalue*** *1.*

# Google's website search algorithm

*The previous observation is key to understand how Google's* ***pageRank*** *algorithm works.*




Webpages

| | $P_1$ | $P_2$ | | | | | | | | $P_{109}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $P_1$ | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| $P_2$ | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| $P_{109}$ | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

Homework:

→ Convert to a valid stochastic matrix

→ Then the largest eigenvector $\pi$ is the probability of each page, or as Google says: its page rank