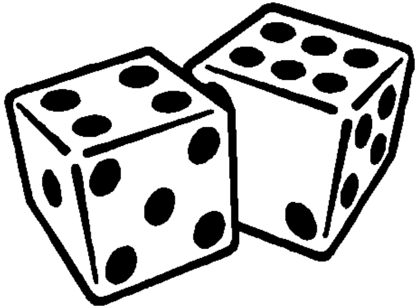
CS224W
Quick tour to Basic Probability Theory



Fall 2014

Outline

Just a gentle refresher on probability

You should have seen this before!

Outline

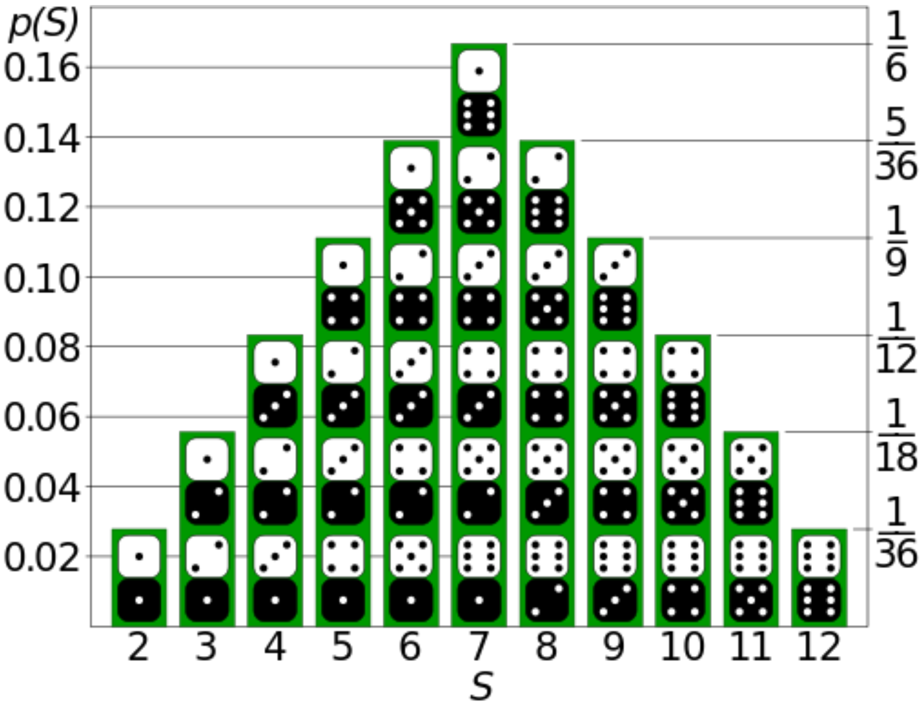
- Fundamentals
- Union/Intersection
- Conditional Probability
- Bayes Rule
- Random variables
- Distributions
- Expectation and Variance
- Indicator variables
- Inequalities

Fundamentals

Sample space

Rolling two dice: $\{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$

Fundamentals



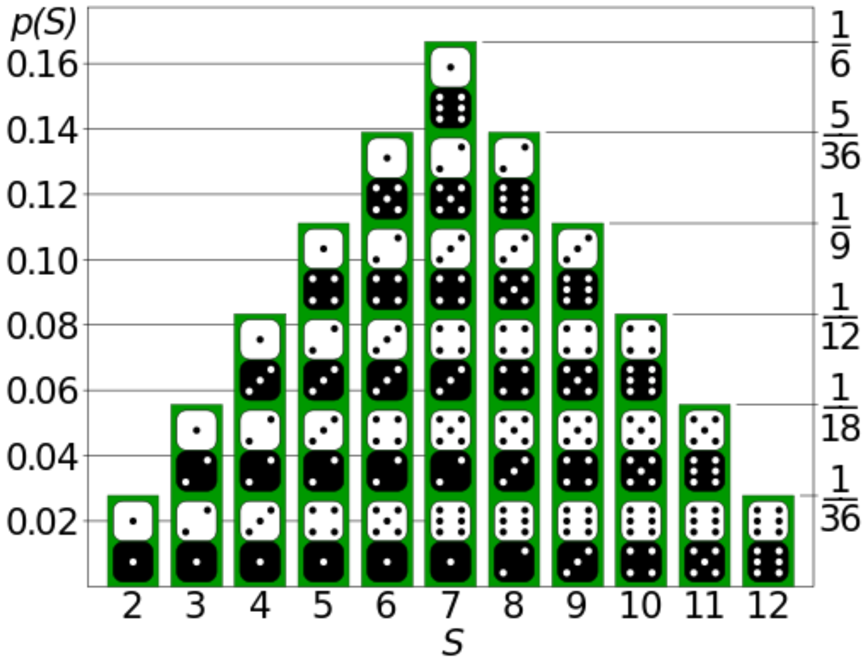
P(sum = 6)
5/36

P(sum = 12)
1/36

P(sum != 12)
1 - 1/36 = 35/36

P(both dice are odd)
1/2 * 1/2 = 1/4

Union / Intersection



P(sum = 6 AND sum = 12)
0

P(sum = 6 OR sum = 12)
5/36 + 1/36 = 6/36 = 1/6

P(sum = 6 AND both dice are odd)
3/36

P(sum = 6 OR both dice are odd)

Union[OR] / Intersection[AND]

For any two events A and B, the union of the two is:

P(A U B) = P(A) + P(B) - P(A n B)

e.g. P(sum = 6 U both dice are odd) =
5/36 + 9/36 - 3/36 = 11/36

Conditional Probability

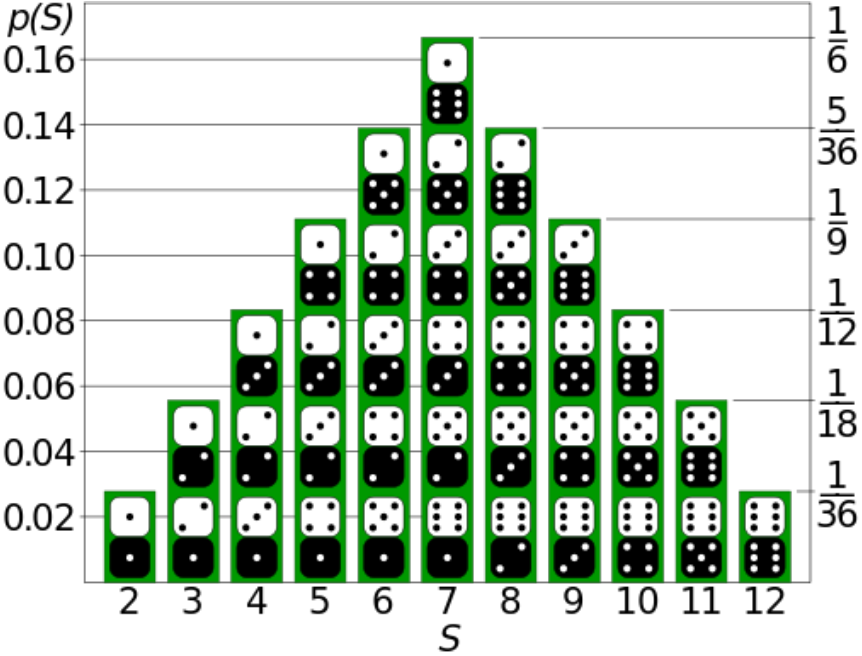
The conditional probability of A given B is:

P(A | B) = P(A n B) / P(B)

also gives **P(A n B) = P(A | B) P(B)**

“What’s the probability of A once we know B has happened?”

Conditional Probability



P(both dice are odd | sum = 6)
3/36

P(first die = 4 | sum = 7)
1/6

P(first die = 4 | sum = 6)
1/5

Note: P(first die = 4 | sum = 7) = P(first die = 4) = 1/6
P(first die = 4 | sum = 6) != P(first die = 4)

Independence

In general P(A|B) != P(A).

But there are special cases where **P(A|B) = P (A)**, which also implies **P(A n B) = P(A)P(B).**

In such cases, **A and B are independent events**.

Example: Events (sum = 7) and (first die = 4) are independent

Bayes Rule

From conditional probability we get,

P(A | B) = P(A n B) / P(B)

= P(B | A) P(A) / P(B)

Bayes Rule - Example

Your friend told you she had a great conversation with someone on the Caltrain. Not knowing anything else, your prior belief that her conversation partner was a woman is 50%. Let W denote this event. Let L denote the event that her conversation partner has long hair. If you learn L to be true, how should you update your beliefs about W?

P(W) = 0.5 and suppose **P(L) = 0.6**, and **P(L|W) = 0.75** are known.

Using Bayes Rule, **P(W | L) = P(L | W)P(W)/P(L) = 0.75 * 0.5 / 0.6 = 62.5%**

Random Variables

A random variable is a variable that can take on a set of different values, each with an associated probability.

Example: Let X be a random variable that counts the number of 6's we roll in 2 rolls of a die

$P(X = 2) = P(\{6, 6\}) = 1/36$

$P(X = 1) = P(\{1,6\}) + \ldots + P(\{5,6\}) * 2 = 10/36$

$P(X = 0) = 1 - 11/36 = 25/36$

Common random variables

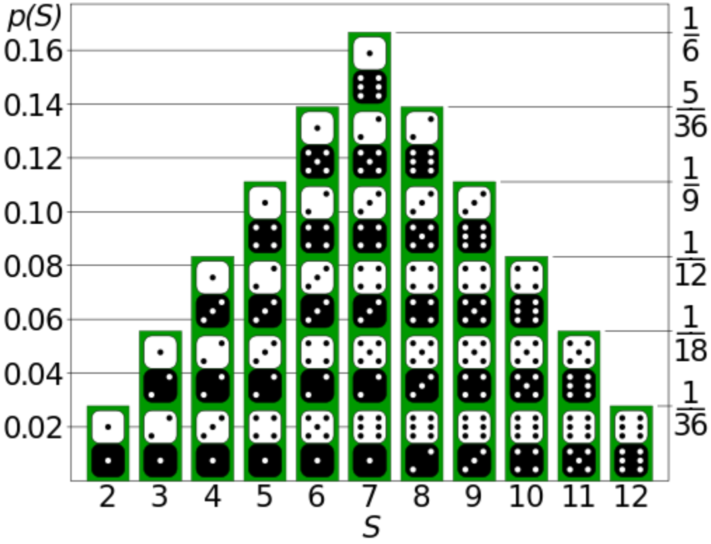
- $X \sim \mathit{Bernoulli}(p)$ $(0 \leq p \leq 1)$: $p_X(x) = \begin{cases} p & x{=}1, \\ 1-p & x{=}0. \end{cases}$
- $X \sim \mathit{Geometric}(p)$ $(0 \leq p \leq 1)$: $p_X(x) = p(1-p)^{x-1}$
- $X \sim \mathit{Uniform}(a, b)$ $(a < b)$: $f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases}$
- $X \sim \mathit{Normal}(\mu, \sigma^2)$: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

Outline

- Fundamentals
- Union/Intersection
- Conditional Probability
- Bayes Rule
- Random variables
- Distributions
- Expectation and Variance
- Indicator variables
- Inequalities

Distributions (pmf)

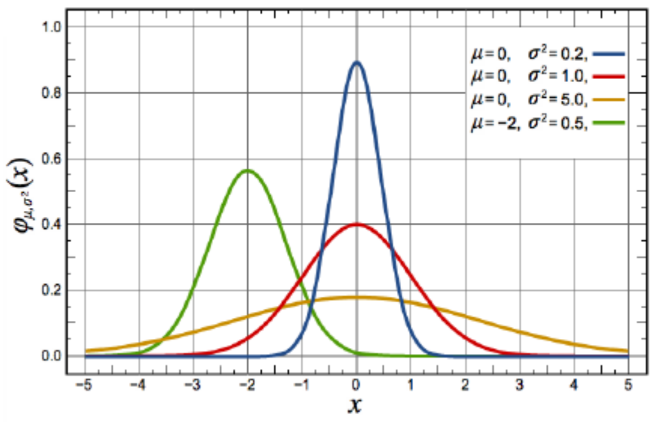
A probability mass function(pmf) assigns a probability to each possible value of a random variable.

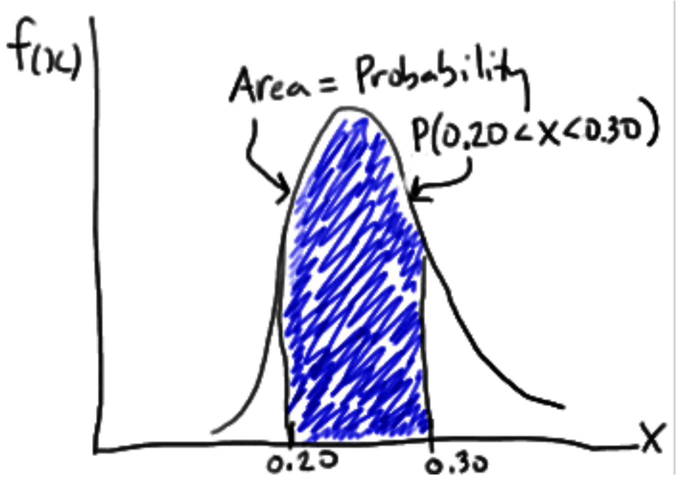


Distributions (pdf)

A probability density function(pdf) of a continuous random variable describes the relative likelihood for X to take on a given value:

$$P[a \leq| X \leq b] = \int_a^b f(x)dx$$

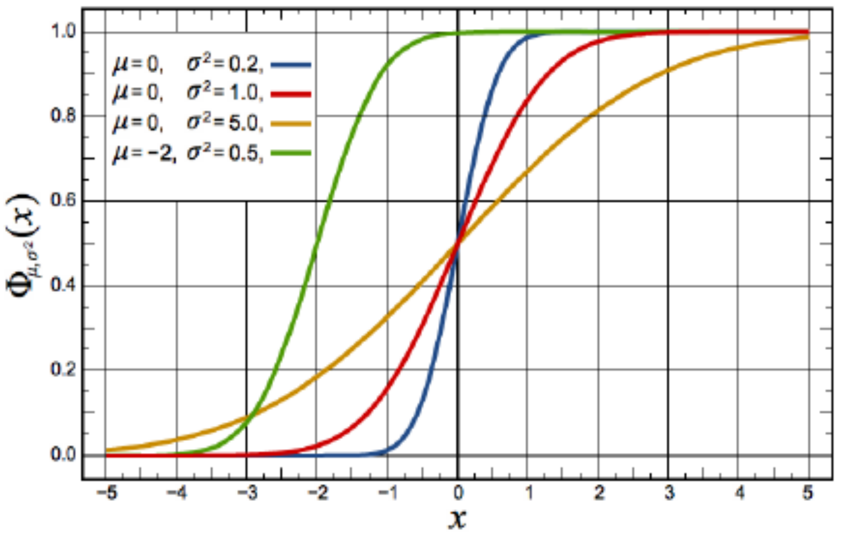




Distributions (cdf)

Cumulative distribution function of random variable X is

$$F(x) = P(X \leq x)$$



Distributions (general properties)

- CDF (cumulative distribution function):
  - $0 \leq F_X(x) \leq 1$
  - F_X monotone increasing, with $lim_{x \to -\infty} F_X(x) = 0$, $lim_{x \to \infty} F_X(x) = 1$
- pmf:
  - $0 \leq p_X(x) \leq 1$
  - $\sum_x p_X(x) = 1$
  - $\sum_{x \in A} p_X(x) = p_X(A)$
- pdf:
  - $f_X(x) \geq 0$
  - $\int_{-\infty}^{\infty} f_X(x) dx = 1$
  - $\int_{x \in A} f_X(x) dx = P(X \in A)$

Expectation and Variance

- If the discrete random variable X has pmf $p(x)$, then the expectation is $E[X] = \sum_x x \cdot p(x)$
- Continuous case is similar: $E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)dx$
- Expectation is linear:
  - for any constant $a \in \mathbb{R}$, $E[a] = a$
  - $E[a \cdot g(X) + b \cdot h(X)] = aE[g(X)] + bE[h(X)]$
- $Var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$
- Variance is **not** linear

Expectation of a die roll: 1 * 1/6 + 2 * 1/6 + … + 6 * 1/6 = 3.5

Indicator Variables

An indicator variable just indicates whether an event occurs or not:

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

They have a very useful property:

$$\begin{aligned} E[I_A] &= 1 \cdot P(I_A = 1) + 0 \cdot P(I_A = 0) \\ &= P(I_A = 1) \\ &= P(A) \end{aligned}$$

Method of indicators

Goal: find expected number of successes out of N trials

Method: define an indicator (Bernoulli) random variable for each trial, find expected value of the sum

Example: N professors are at dinner and take a random coat when they leave. Expected number of profs with the right coat?

Method of indicators(continued)

$$
\begin{aligned}
E[G] &= E[G_1 + G_2 + \ldots + G_n] \\
&= E[G_1] + E[G_2] + \ldots + E[G_n] \\
&= 1/n + 1/n + \ldots 1/n = 1
\end{aligned}
$$

Note: linearity of expectation does not assume independence!

Inequalities

Markov's inequality: $P(X \geq a) \leq \frac{E[X]}{a}$

Chernoff bound: Let $X_1, \ldots, X_n$ independent Bernoulli with $P(X_i = 1) = p_i$. Denoting $\mu = E[\sum_{i=1}^n X_i] = \sum_{i=1}^n p_i$,

$$P(\sum_{i=1}^{n} X_i \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{\mu}$$

for any δ. Multiple variants of Chernoff-type bounds exist, which can be useful in different settings

Inequalities

Example: die roll

$P(X >= 5) <= E[X] / 5$

$P(X >= 5) <= 3.5 / 5 = 0.7$

That’s all!

Have a nice weekend!