Lectures 3: Random variables

Statistical Methods for Natural Language Processing
Fredrik Engström

February 7, 2011

Summary of lecture 2

- P satisfies:
  - $0 \leq P(A) \leq 1$,
  - $P(\Omega) = 1$,
  - $P(A \cup B) = P(A) + P(B)$ whenever $A \cap B = \emptyset$.
- $P(A|B) = P(A \cap B)/P(B)$
- $P(A|B) = P(B|A)P(A)/P(B)$.
- A and B are independent if $P(A \cap B) = P(A)P(B)$.

Stochastic variables

Intuition: A variable depending on some random process.
Example: The sum of two dice. Number of letters in words.

Definition

A **stochastic variable** (or **random variable**) is a function from the sample space Ω to $\mathbb{R}$.

Given a stochastic variable X we can define the **frequency function** or the **probability mass function** (**pmf**) p as follows:

$$p(x) = P(A_x),$$

where A_x is the event of all a:s such that $X(a) = x$.

$$p_X(x) = P(X = x) = p(x).$$

Intuition: $p(x)$ is the probability that X has value x.
Example: Rolling two dice.

Expected value

Definition

The **expected value** of a random variable X is

$$E[X] = \sum_{x \in \mathbb{R}} x p(x).$$

Digress: The summation sign.
Examples: Dice.
Assume $Y = g(X)$ then

$$E[Y] = E[g(X)] = \sum_{y \in \mathbb{R}} y p_Y(y) = \sum_{x \in \mathbb{R}} g(x) p_X(x)$$

Thus, for example $E[5X] = 5E[X]$ and $E[X + 5] = E[X] + 5$.

Joint and conditional distributions

$(X, Y) : \Omega \to \mathbb{R}^2$, almost a random variable. (Two dimensional.)

Definition

$$p(x, y) = P(\{\omega \in \Omega \mid X(\omega) = x \text{ and } Y(\omega) = y\}).$$

Example: Two dice. pos-tagged words.

Definition

$$p(x|y) = \frac{p(x, y)}{p_Y(y)}$$

$$E[X + Y] = E[X] + E[Y]$$

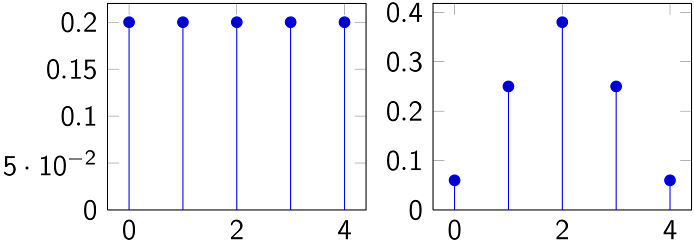
Example: Roll two dice.

Independence

Definition

X and Y are **independent** if $p(x, y) = p_X(x)p_Y(y)$.

I.e., iff $P(X = x, Y = y) = P(X = x)P(Y = y)$.
Example: The sum of two dice and the value of the first. Number of letters and noun.


0.2
0.15
0.1
$5 \cdot 10^{-2}$
0
0
2
4
0.4
0.3
0.2
0.1
0
0
2
4

Variance

Definition

The **variance** of a random variable X is

$$\mathrm{Var}(X) = E[(X - \mu)^2],$$

where $\mu = E[X]$.

Intuition: The expected value of the (squared) distance from mean.
Example: Rolling one n-die (uniform distribution): $E[X] = \frac{1+n}{2}$
$\mathrm{Var}(X) = \frac{n^2-1}{12}$
Example: The sum of two dice.
$\mathrm{Var}(X) = E[X^2] - E^2[X]$.

Binomial distribution

Let say we have an alphabet of five letters.
How many **words** of length three are there?

$$5^3$$

How many **words** of length three **without repetition** are there?

$$5 \cdot 4 \cdot 3 = \frac{5!}{2!}$$

How many **"bags"** consisting of **three different letters** are there?

$$\frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = \frac{5!}{2! \cdot 3!} = \binom{5}{3}$$

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

Binomial distribution cont.

Example: Coin tosses (number of heads). Fair/Unfair coin.

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$E[X] = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} = np$$

$$\mathrm{Var}[X] = \sum_{k=0}^{n} (k-np)^2 \cdot \binom{n}{k} p^k (1-p)^{n-k} = np(1-p)$$

Summary

- Random variable $X : \Omega \to \mathbb{R}$.
- Expected value/mean: $E[X] = \sum xp(x)$.
- Joint and conditional: $p(x|y) = p(x, y)/p_Y(y)$.
- Independence: $p(x, y) = p_X(x)p_Y(y)$.
- Variance: $\mathrm{Var}(X) = E[(X - \mu)^2]$.
- Binomial coefficients: $\binom{n}{k}$
- Distributions: Uniform / Binomial.