Probability Basics in Computer Science

Ulises Tirado Zatarain [1]
(ulises.tirado@cimat.mx)

[1] Algorists Group

February, 2016

Outline

1. Introduction
   - Definitions
   - Properties and relations
   - Probability in continuous space $\mathbb{R}^n$ and random variables
2. Probability measures
   - Centrality measures
   - Dispersion measures
3. Applications
   - Linear median
   - More applications

Definitions

Experiments and Events

An event is a set of posible results in an experiment execution. For example, taking a card from a deck or rolling a dice. We can denote all posible results with Ω and an event with uppercase letter such that $A \subseteq \Omega$ or $A \in 2^{\Omega}$, then A is a subset of Ω.

Probability

Is an indicator that describes the frecuency of an event in one universal set of posibilities. Daily, we express that indicator as a percentage value or value between 0 and 1. Then we can define the probability as:

$$p : 2^{\Omega} \mapsto [0, 1]$$

Definitions

- So, when Ω is a discrete and finite set, then:

$$p(A) = \frac{\#A}{\#\Omega}$$

- We name this as uniform distribution or counting distribution.
- However, counting elements in A and Ω isn't always trivial. Maybe we need to use operations like factorial, combinations, permutations, etcetera.

Examples

- Key pressing random letter of the english keyboard such that the letter be a vowel.
  - Let $A = \{a, e, i, o, u\}$ and $\Omega = \{a, ..., z\}$,then $p(A) = \frac{5}{26}$.
- Rolling a dice such that the result be greater than 2.
  - Let $A = \{3, 4, 5, 6\}$ and $\Omega = \{1, 2, 3, 4, 5, 6\}$,then $p(A) = \frac{2}{3}$.,
- Taking a card from a deck such that getting a red card.
  - Let $A = \{A\heartsuit, 2\heartsuit, ..., 10\heartsuit, J\heartsuit, Q\heartsuit, K\heartsuit, A\diamondsuit, 2\diamondsuit, ..., 10\diamondsuit, J\diamondsuit, Q\diamondsuit, K\diamondsuit\}$ and $\Omega = \left\{ \begin{array}{cc} A\heartsuit, 2\heartsuit, ..., 10\heartsuit, J\heartsuit, Q\heartsuit, K\heartsuit, & A\diamondsuit, 2\diamondsuit, ..., 10\diamondsuit, J\diamondsuit, Q\diamondsuit, K\diamondsuit, \\ A\spadesuit, 2\spadesuit, ..., 10\spadesuit, J\spadesuit, Q\spadesuit, K\spadesuit, & A\clubsuit, 2\clubsuit, ..., 10\clubsuit, J\clubsuit, Q\clubsuit, K\clubsuit \end{array} \right\}$, then $p(A) = \frac{1}{2}$.

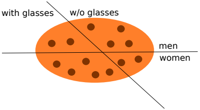
Properties and relations

Let A and B events in Ω, then

- $p(\Omega) = 1$ and $p(\emptyset) = 0$.
- $p(A \cap B)$ means the probability of a value that is in both sets, A and B.
- $p(A \cup B) = p(A) - p(A \cap B) + p(B)$ represents the probability of a value that is at least in one set, A or B.
- $p(A^c) = p(\Omega \setminus A) = 1 - p(A)$ is the probability of a value that isn't in A.
- If $A \cap B = \emptyset$ then $p(A \cup B) = p(A) + p(B)$.

Conditional probability

- Starting with following example: we have a set tokens whose represent the people whose are showing in the figure. So, closing our eyes we take a random token then the probability to take a woman is $\frac{\#\text{women}}{\#\text{people}} = \frac{7}{13}$.



- However, if someone tells us that (before we open our eyes) the chosen person wears glasses, then the probability will be conditioned to the fact that person wears glasses, then will be $\frac{\#\text{women with glasses}}{\#\text{people with glasses}} = \frac{5}{7}$.

Conditional probability

- This probability is denoted as $p(A|B)$, where
  - A is the event of taking a women.
  - B represents the event of taking a person who wears glasses.
- In general way, if $p(B) > 0$ then $p(A|B)$ is defined as:

$$p(A|B) = \frac{p(A \cap B)}{p(B)} = \tfrac{5}{13} / \tfrac{7}{13} = \frac{5}{7}$$

- Two events A and B are independents iff

$$p(A|B) = p(A)$$

or well

$$p(A \cap B) = p(A)\,p(B)$$

Probability in $\mathbb{R}^n$ and random variables

- Let region $\Omega \subset \mathbb{R}^n$, then the probability of $A \subset \Omega$ will be:

$$p(A) = \frac{1}{c} \int_A f(\mathbf{x})\, d\mathbf{x}$$

where:

  - $\mathbf{x} \in A$ is the value of a random variable named $\mathbf{X} \in \mathbb{R}^n$ and $c = \int_\Omega f(\mathbf{x})\, d\mathbf{x}$ is knew as normalization factor. Random variables are denoted with uppercase letter and their values with lowercase letters.
  - $f : \Omega \mapsto \mathbb{R}$ is a function knew as probability density or simply density and for all $\mathbf{x} \in \Omega$ we have $f(\mathbf{x}) \geqslant 0$.
  - $p(A)$ is knew as probability distribution, cumulative distribution or simply distribution and is denoted with upper case letter, frequently as F.

Probability in $\mathbb{R}^n$ and random variables

- So, random variable can be discrete or continuous.
- We can denote the probability of a event where $\mathbf{X} = \mathbf{x}$ as $p(\mathbf{X} = \mathbf{x})$.
- Sometimes (always), the probability notation is very nasty. Beware!
- There are many probability distributions:
  - Uniform (discrete and continuous): $X \sim \mathcal{U}(a, b)$
  - Bernoulli (discrete): $X \sim Be(p)$
  - Binomial (discrete): $X \sim B(n; p)$
  - Geometric (discrete): $X \sim Geom(k; p)$
  - Poisson (discrete): $X \sim Pois(\lambda)$
  - Normal (continuous): $X \sim \mathcal{N}(\mu; \sigma^2)$
  - Chi-squared (continuous): $X \sim \chi^2(q)$
  - ...

Centrality

Let $X \in \mathbb{R}$ a random variable, then:

Expectation, expected value simply mean

Is the mean of the data based on their frequency:

$$\mathbb{E}X = \begin{cases} \sum_{x\in\Omega} xp\left(X = x\right) & X \text{ is discrete} \\ \frac{1}{c}\int_{\Omega} xf\left(x\right)dx & X \text{ is continuous} \end{cases}$$

Mode

Is most frequent value, then in the discrete case is $\arg\max_{x\in\Omega} p\left(X = x\right)$ and in the for continuos random variable is x such that $\frac{df}{dx} = 0$ and $\frac{d^2 f}{dx^2} < 0$.

Centrality

Median

Is the value separating the higher half of a data sample, a population, or a probability distribution, from the lower half.

Function composition

The mean for composition functions is defined as following:

$$\mathbb{E}\left[g\left(X\right)\right] = \begin{cases} \sum_{x \in \Omega} g\left(x\right) p\left(X = x\right) & X \text{ is discrete} \\ \frac{1}{c} \int_{\Omega} g\left(x\right) f\left(x\right) dx & X \text{ is continuous} \end{cases}$$

Dispersion

Let $X \in \mathbb{R}$ a random variable, then:

Variance

Is the mean square distance between the mean point and the data.

$$\operatorname{var}(X) = \mathbb{E}\left[X^2\right] - \left[\mathbb{E}X\right]^2$$

Entropy

Is the expected value of the information provided by a random variable X, we mean:

$$\mathbb{H}(X) = \mathbb{E}\left[\mathbb{I}(X)\right] = \mathbb{E}\left[-\log p(X)\right]$$

Linear median

function $\mathtt{median}(\boldsymbol{x} \in \mathbb{R}^n)$ **begin**

Parameters: $k \in [n]$

$p \leftarrow \mathtt{random}(n)$;

Split $\boldsymbol{x}$ into subarrays $\boldsymbol{l}$ and $\boldsymbol{g}$ by comparing each element to p-th elemnent. While we are at it, count the number K of elements going in to $\boldsymbol{l}$.

if $K = k - 1$ **then:**

return x_p;

else if $K > k - 1$ **then:**

return $\mathtt{median}(\boldsymbol{l}, k)$;

else:

return $\mathtt{median}(\boldsymbol{g}, k - K - 1)$;

end

end

More applications

- Skip lists
- Random trees
- Random graphs
- Complexity analysis
- Page rank
- Simulation
- Visualizations
- Comunications

References I

- Johan van Horebeek - Métodos Estocásticos en Computación
- Christopher Bishop - Pattern Recognition
- Wikipedia
- College of Computing - Georgia Tech
- A consistent adjacency spectral embedding for stochastic blockmodel graphs