AID-521 Mathematics for Data Science

Module: Probability | Lecture: 3

RANDOM VARIABLES & PROB. DISTRIBUTIONS

random variables vs. probability, probability distribution, moments, moment generating funtion*

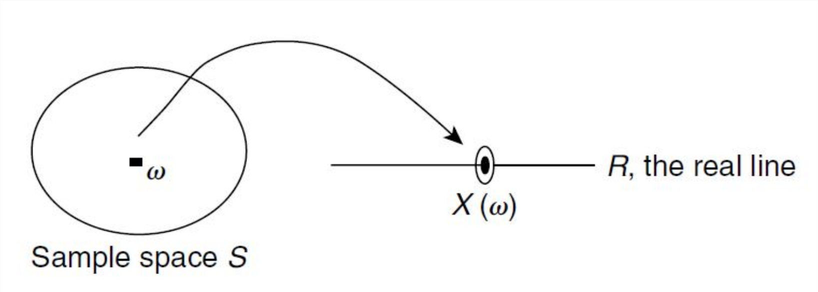
Random Variable

Given a sample space S, a random variable (r.v.) X is a function from S to $\mathbb{R} := (-\infty, \infty)$.

$$X : S \rightarrow \mathbb{R}$$

What's the difference between probability function and random variable? Remember the axiomatic definition of probability?

Random Variable



The simple coin toss - always helps!

Tossing of two coins one after another. (This is the process/experiment.) Sample space $S = \{HH, HT, TH, TT\}$

If $P(HH) = 1/4, P(HT) = 1/4, P(TH) = 1/4, P(TT) = 1/4$, then $P(\cdot)$ defined on S is a probability function.

If $P(HH) = 1/2, P(HT) = 0, P(TH) = 0, P(TT) = 1/2$, then $P(\cdot)$ defined on S is also another probability function.

Occurence of an event means one of the elements in the event set is an outcome of a particular trial. Now check if the two different $P(\cdot)$ functions defined above satisfy the axioms of probability function or not.

The simple coin toss - always helps!

Are the two different $P(\cdot)$ functions defined in previous slide random variabes?

What about the function X below, is it a random variable?

$X(HH) = 1, P(HT) = 0, P(TH) = 10, P(TT) = 1.89$

What about the function X defined on S below?

$X(\omega) = \#H$ in ω

The simple coin toss - always helps!

But, I don't see any probability! Indeed, because random variable is not the same as probability.

Try yourself

Assume both the coins are unbiased.

After tossing each coin sequentially, what is the probability of observing the same outcome for each coin?

$\rightarrow$ Can you define a random variable here?

Probability Distribution

Suppose X is a discrete r.v. with values $x_1, x_2, \ldots$

Probabilty function of a discrete random variable X is

$$p(x_i) := P(X = x_i), \; i = 1, 2, 3, \ldots$$

$\rightarrow$ $\sum_{i=1}^{\infty} p(x_i) \; = \; 1,$ and $p(a) \geq 0 \; \forall \; a$

The cumulative distribution function (cdf) F of the random variable X is

$$F(a) = P(X \leq a); \;\; a \in (-\infty, \infty)$$

$$= \sum_{x \leq a} p(x)$$

$\rightarrow$ Also called probability distribution function

Probability Distribution

Try yourself

A fair coin is tossed twice. Let X be the number of heads.

- $\rightarrow$ Write the probability function of X
- $\rightarrow$ Write the cdf of X

Probability Distribution

Try yourself – solution

The probability function..

$p(\cdot)$ is defined for all values of the r.v. X – Remember?
What is the sample space S here? What is the r.v. here?

The values taken by X are $0, 1, 2$.
So, $p(0) = P(X = 0) = P(event\{TT\} \subset S) = ?$

Probability Distribution

Try yourself – solution

The CDF of X..

$F(\cdot)$ is defined on the entire real line – Remember?

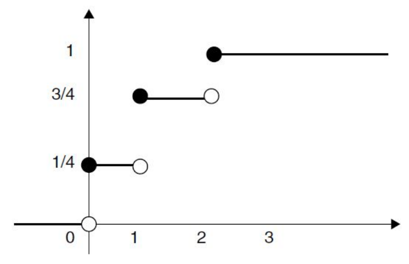
So what are $F(2)$ and $F(-2)$?

$F(1.0001) = P(X \leq 1.0001) = P(\{TT, HT, TH\})$ or $p(0) + p(1)$

Properties of Distribution Function

- $0 \le F(a) \le 1$
- $\lim_{a \to -\infty} F(a) = 0$, and $\lim_{a \to \infty} F(a) = 1$
- $F(\cdot)$ is non-decreasing, and right continuous

Look at the $F(\cdot)$ of tossing coin twice.



Expected Value

The expected value of a discrete random variable X with probability (mass) function $p(\cdot)$ is

$$\mu \quad = \quad E(X) \quad = \quad \sum_x x \cdot p(x)$$

$\rightarrow$ very intuitive indeed – measures the mean value basically

It is assumed that $\sum_x |x| \cdot p(x) < \infty$, so that $E(X)$ is well-defined.

Variance

The variance of a random variable X is

$$\sigma^2 \quad = \quad Var(X) \quad = \quad E\left[(X-\mu)^2\right]$$

→ measures the extent to which the values of X vary around its mean value

→ Standard deviation, denoted by σ, is simply $\sqrt{\sigma^2}$

Expected Value of a Function of R.V.

Let X be a discrete r.v. with probability function $p(\cdot)$, and $g(\cdot)$ be a function defined on X.

The expected value of $g(X)$ is

$$E[g(X)] = \sum_x g(x) \cdot p(x)$$

$$\mu = E(X) = \sum_x x \cdot p(x)$$

Properties of 1st Two Moments

- $\rightarrow$ $E(c) = c$, when c is a constant
- $\rightarrow$ $E[c \cdot g(X)] \; = \; c \cdot E[g(X)]$
- $\rightarrow$ $E[\sum_i g_i(X)] \; = \; \sum_i \; E[g_i(X)]$
- $\rightarrow$ $Var(aX + b) = a^2 \cdot Var(X)$
- $\rightarrow$ $Var(X) = E(X^2) - \mu^2$

You can try to prove all of these by yourself. You will have a better grasp.

The *k*th Moment

Given a r.v. X, its *k*th moment (about its mean) is denoted as μ_k and defined as

$$E[\,(X-\mu)^k\,],\ \ k \in \{2, 3, \ldots\}$$

whenever it exists.

The *k*th moment about origin is defined as $E[X^k]$.

The *k*th Moment

What is the use of these higher-order moments?

→ Just like mean and variance of a distribution describe the shape of the distribution, higher-order moments also describe other aspects of its shape.

→ Two distributions with the same mean and variance can have different higher-order moments, and hence different shapes.

→ Two frequently used higher-order moments are skewness and kurtosis.

Skewness & Kurtosis

Skewness of the distribution of X is the *standardized* third moment about the mean.

$$\alpha_3 = \frac{E[(X-\mu)^3]}{\sigma^3} = \frac{\mu_3}{\mu_2^{3/2}}$$

→ A measure of asymmetry of density function around its mean. The density function has a longer right tail when $\alpha_3 > 0$.

Kurtosis of the distribution of X is the *standardized* fourth moment about its mean.

$$\alpha_4 = \frac{E[(X-\mu)^4]}{\sigma^4}$$

→ A measure of the size of a distribution's tails. There are few observations in the tails (relative to normal distribution) when $\alpha_4 > 0$.

In all definitions so far,
replace summations $\sum$ by INTEGRALS $\int$ IF

the random variable *X* is a CONTINUOUS function
instead of being discrete.

That is, if *X* takes values in a range, for example $[5, 10] \subset \mathbb{R}$,
instead of values from $\{5, 5.5, 7, 10\}$.

Definitions for the Continuous Case

X is a random variable. If there exists a function $f(\cdot)$ such that

- $\rightarrow$ f is non-negative and real valued, i.e., $f : \mathbb{R} \rightarrow [0, \infty)$,
- $\rightarrow$ $P(X \in [a, b]) = \int_a^b f(t)dt$ for all intervalls $[a, b]$,

then X is a continuous r.v. and f is the probability density function (pdf) of X.

Why is $\int_{-\infty}^{\infty} f(t)\, dt = 1$?

The cumulative distribution function (cdf) of X is
$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt; \ \forall\, x \in \mathbb{R}$.

Why is $\frac{dF(x)}{dx} = f(x)$ when f is continuous?

PDF & CDF for a Continuous R.V.

An example is shown below