[From O. Häggström, Finite Markov Chains and Algorithms]

RANDOM VARIABLES (R.V.)

A random variable is a quantity which depends on chance.
Usually a random variable is real valued, then it is a function

$$X : \Omega \to \mathbb{R}$$

but it can be also integer value (a counting) and then $X : \Omega \to \mathbb{N}$.
To a r.v. is associated a family of events, called σ-algebra, of which we can compute the probability from the value of X.

Examples:

$$A = \{X \leq 4.7\} = \{X(\omega) : X(\omega) \leq 4.7\}$$

$$B = \{X \text{ is an even integer}\}$$

Two r.v. X and Y are independent if any event A_X defined in terms of X is independent from any event B_Y defined in terms of Y
(remember that A_X ind. from B_Y if
$$\mathbb{P}(A_X \cap B_Y) = \mathbb{P}(A_X)\,\mathbb{P}(B_Y) \qquad).$$

Similarly $X_1, X_2, \dots, X_k$ are independent if, called A_i any event defined in terms of X_i, $A_1, A_2, \dots, A_k$ are independent.

DISTRIBUTIONS

A <u>distribution</u> of a r.v. X is defined in terms of a probability measure.

Def: If X is a real valued r.v. then the <u>distribution</u> μ_X of X is the probability measure on $\mathbb{R}$ satisfying

$$\mu_X(A) = \mathbb{P}(X \in A)$$

for all (appropriate) $A \subseteq \mathbb{R}$.

Each distribution μ_X is characterized in terms of its <u>cumulative distribution function (cdf)</u>

$$F_X : \mathbb{R} \to [0, 1]$$

defined by

$$F_X(x) = \mathbb{P}(X \leq x)$$

A distribution μ on a finite set $S = \{s_1, \dots, s_k\}$ is often represented as a vector

$$(\mu_1, \dots, \mu_k)$$

where $\mu_i = \mu(s_i) = \mathbb{P}(X = s_i) \in [0, 1]$

and $\sum_{i=1}^{k} \mu_i = 1$, by definition of probability measure.

Def: A sequence of random variables $X_1, X_2, \dots$ is said to be <u>i.i.d (independent and identically distributed)</u> if

i) they are independent

ii) they have the same cdf, i.e., $\mathbb{P}(X_i \leq x) = \mathbb{P}(X_j \leq x)$ $\forall i, j \in \mathbb{N}$ $\forall x \in \mathbb{R}$

Very often a sequence $(X_1, X_2, \dots)$ of r.v.'s is interpreted as the evolution in time of some quantity: X_n is the (random) quantity at time n.

Such sequence is then called stochastic process or random process.

Markov chains (we will see later) are a special class of stochastic processes.

We will consider two classes of random variables:

a) DISCRETE R.V.'S: they take values in a countable or finite subset of $\mathbb{R}$. We will identify the subset with $\{0, 1, 2, \dots\}$
$\Rightarrow$ we will use nonnegative integer-valued random variables

Examples: tossing coins or dice; counting clients to a service

b) CONTINUOUS R.V.'S: they take values in all $\mathbb{R}$, and there exists a density function

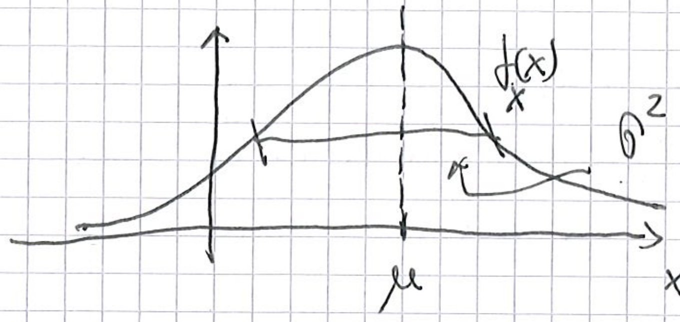
$$f_X : \mathbb{R} \to [0, +\infty) \quad \text{such that}$$

$$\int_{-\infty}^{x} f_X(y)\,dy = F_X(x) = \mathbb{P}(X \leq x) \qquad \forall x \in \mathbb{R}$$

Examples:

a) X is a Gaussian or Normal r.v. if its density function is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{with parameters } \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}_+$$

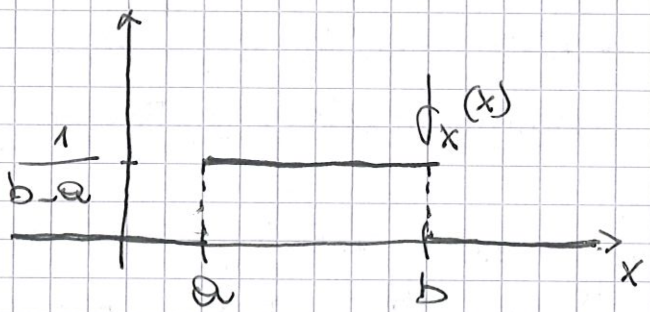


σ^2 is related to the width of the density

symmetric with respect to μ

b) X is uniformly distributed in $[a,b]$
$X \sim U(a,b)$
if its density function is

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$



and the cdf is

$$F_X(x) = \int_{-\infty}^{x} f(y)\,dy = \begin{cases} 0 & \text{if } x \leq a \\ \dfrac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } x \geq b \end{cases}$$

$U(a,b)$ distribution corresponds to the concept of "equally likely" in $[a,b]$

The distribution of a r.v. is characterized by its expected value and variance (first two moments).

Def: The expected value or mean of a r.v. X is

a) if X is discrete, taking values in $M \subseteq \mathbb{N}$,
$$E(X) = \sum_{i \in M} x_i \, \mathbb{P}(X = x_i) = \sum_{i \in M} x_i \, \mu_i$$

b) if X is continuous with pdf $f_X(x)$
$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)\, dx$$

If we have an iid sample for X, $X_1, \dots, X_n$, a good estimator of $E(X)$ is the sample mean
$$\boxed{\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i}$$

As much the sample size n increases, as much the approximation improves, because of the

Law of large numbers ("$\overline{X}_n$ is consistent for $E(X)$")
If $X_1, \dots, X_n, \dots$ is an i.i.d. sequence of r.v.'s distributed like X, having $E(X) < +\infty$, then
$$\mathbb{P}\left(\overline{X}_n \xrightarrow[n \to \infty]{} E(X)\right) = 1$$
i.e. $\overline{X}_n \xrightarrow[n \to \infty]{} E(X)$ almost surely (a.s.)

Def: the variance of a r.v. X is $V(X) = E\left[(X - E(X))^2\right]$

Thus:

a) if X is discrete, taking values in $M \subseteq \mathbb{N}$

$$V(X) = \sum_{i \in M} (x_i - E(X))^2 \, \mathbb{P}(X = x_i)$$

b) if X is continuous, with density $f_X(x)$

$$V(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f_X(x) \, dx$$

If we have an iid sample $X_1 \ldots X_n$ for a r.v. X, a good estimator of $V(X)$ is the (unbiased) sample variance

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

Also this estimator is consistent, i.e.

$$S_n^2 \xrightarrow[n \to \infty]{} V(X) \quad \text{a.s.}$$

Properties of E(X) and V(X)

$E(\cdot)$ is linear, i.e.

a) $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$

b) $E(cX) = c\,E(X)$ $\quad c$ = constant, deterministic

$$\Rightarrow E(c_1 X_1 + \cdots + c_n X_n) = c_1 E(X_1) + \cdots + c_n E(X_n)$$

Variance is non linear:

a) $Var(cX) = c^2 Var(X)$

b) In general

$$Var(X_1 + \dots + X_n) \neq Var(X_1) + \dots + Var(X_n)$$

but if $X_1, \dots, X_n$ are <u>independent</u>, then

$$Var(X_1 + \dots + X_n) = Var(X_1) + \dots + Var(X_n)$$

<u>Example 1</u>

$X \in \{0, 1\}$ $\qquad X = \begin{cases} 0 & prob = 1-p \\ 1 & prob = p \end{cases}$

result of tossing a coin,
or outcome of a dicotomic experiment

$X \sim Bernoulli(p)$ $\qquad p =$ probability of a "success"

$\boxed{E(X) =} \; 0 \cdot (1-p) + 1 \cdot p \; \boxed{= p}$

$\boxed{V(X) =} \; (0-p)^2 (1-p) + (1-p)^2 p =$

$= p^2(1-p) + (1-p)^2 p =$

$= p(1-p)(\cancel{p} + 1 - \cancel{p}) \; \boxed{= p(1-p)}$

Example 2

Let Y be the sum of n independent Bernoulli(p) r.v.'s $X_1, \dots, X_n$ (ex: the number of heads in n tosses of a coin; the number of females in a classroom; ...)

$Y \sim$ Binomial(n, p) $\qquad Y = X_1 + \dots + X_n$

Using the properties of E and V we have

$$\boxed{E(Y)} = E(X_1 + \dots + X_n) = \underbrace{E(X_1)}_{=p} + \dots + \underbrace{E(X_n)}_{=p} = \boxed{np}$$

$$\boxed{V(Y)} = V(X_1 + \dots + X_n) \underset{X_i \text{ independent}}{=} \underbrace{V(X_1)}_{=p(1-p)} + \dots + \underbrace{V(X_n)}_{=p(1-p)} = \boxed{np(1-p)}$$