REVIEW OF MAIN CONCEPTS AND FORMULAS

Boolean algebra of `events` (subsets of a sample space)
DeMorgan's formula:

$$\begin{aligned} \overline{A \cup B} &= \bar{A} \cap \bar{B} \\ \overline{A \cap B} &= \bar{A} \cup \bar{B} \end{aligned}$$

The notion of `conditional probability`, and of `mutual independence` of two or more events. The latter implies:

$$\begin{aligned} \Pr(A \mid B) &= \Pr(A) \\ \Pr(A \mid B \cap C) &= \Pr(A) \end{aligned}$$

Probability rules
Product rule:

$$\Pr(A \cap B \cap C) = \Pr(A) \cdot \Pr(B \mid A) \cdot \Pr(C \mid A \cap B)$$

and

$$\Pr(A \cup B \cup C) = \Pr(A) + \Pr(B) + \Pr(C)$$
$$- \Pr(A \cap B) - \Pr(A \cap C) - \Pr(B \cap C) + \Pr(A \cap B \cap C)$$

`Partition` of sample space and total-probability formula:

$$\begin{aligned} \Pr(B) &= \Pr(B|A_1) \cdot \Pr(A_1) + \Pr(B|A_2) \cdot \Pr(A_2) + ... \\ &\qquad ... + \Pr(B|A_k) \cdot \Pr(A_k) \end{aligned}$$

and extensions of this (shortly).

Random variables are either of the `discrete` (non-negative integers) or `continuous` (real numbers) type.

For the former, their `distribution` is most often specified by spelling out the corresponding `probability function`

$$\Pr(X = i) \equiv f_X(i)$$

for the latter, we need their `probability density function` (pdf for short), defined by

$$\lim_{\varepsilon \to 0} \frac{\Pr(x \le X < x + \varepsilon)}{\varepsilon} \equiv f_X(x)$$

We like to look at their graphs.

An alternate way is to specify the corresponding (cumulative) `distribution function`, defined as

$$\Pr(X \le x) \equiv F_X(x)$$

The `expected value` (also called the **mean**) of a RV is computed by

$$\mathbb{E}(X) = \sum_{\text{All } i} i \cdot f_X(i)$$

(discrete type), or

$$\mathbb{E}(X) = \int\limits_{\text{All } x} x \cdot f_X(x)\, dx$$

Since $X = i$ is an event, and a collection of these, for all possible values of i, a partition, a special case of total-probability formula reads:

$$\Pr(B) = \sum_{\text{All } i} \Pr(B \mid X = i) \cdot f(i)$$

While at it, let me mention that this can be extended to X continuous, thus:

$$\Pr(B) = \int\limits_{\text{All } x} \Pr(B \mid X = x) f(x)\, dx$$

The concept of probability function and, particularly, of pdf (called joint pdf) can be extended to bi-variate and **multivariate distributions**.

We should then be able to find the `marginal` pdf of X by

$$f_X(x) = \int\limits_{\text{All } y|x} f_{XY}(x, y)\, dy$$

and `conditional` pdf of Y given $X = \mathbf{x}$ by

$$f_{Y|\mathbf{x}}(y) = \frac{f_{XY}(\mathbf{x}, y)}{f_X(\mathbf{x})}$$

and the corresponding `conditional mean`:

$$\mathbb{E}\,(Y \mid X = \mathbf{x}) = \frac{\int_{\text{All } y|\mathbf{x}} y \cdot f_{XY}(\mathbf{x}, y) dy}{f_X(\mathbf{x})}$$

In this context, we will need yet another extension of the total-probability formula, namely

$$\mathbb{E}(Y) = \int\limits_{\text{All } x} \mathbb{E}(Y \mid X = x) f(x)\; dx$$

which I'll call total-mean formula.

Moments of a RV (or its distribution) are either `simple`

$$\mathbb{E}(X^k) = \int\limits_{\text{All } x} x^k \cdot f(x)\; dx = \sum_{\text{All } i} i^k \cdot f(i)$$

(when $k = 1$, this is called the `mean`, or expected value, of X, and denoted μ_X), or `central`

$$\mathbb{E}[(X - \mu)^k] = \int\limits_{\text{All } x} (x - \mu)^k \cdot f(x)\; dx = \sum_{\text{All } i} (i - \mu)^k \cdot f(i)$$

The most important of these is the `variance` ($k = 2$). The corresponding `standard deviation` is $\sigma \equiv \sqrt{\text{Var}(X)}$.

There are bi-variate analogs of these, of which the most important is the `covariance` (the first-first central moment):

$$\mathbb{E}[(X - \mu_X) \cdot (X - \mu_Y)] = \int\limits_{\text{All } x,y} (x - \mu)(y - \mu) f_{XY}(x, y)\; dx$$

Computationally, it's easier to do

$$\text{Cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mu_X \cdot \mu_Y$$

Independence of X and Y implies zero covariance (but not necessarily reverse).

And, based on these, we can define the `correlation coefficient` between X and Y as

$$\rho_{XY} \equiv \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

(must be between -1 and 1).

We know that, in general

$$\mathbb{E}\,[g(X)] \neq g(\mu_X)$$

unless the **transformation** is **linear:**

$$\mathbb{E}(aX + b) = a\mu_X + b$$

This can be extended to any number of RVs, independent or NOT:

$$\mathbb{E}(aX + bY + c) = a\mu_X + b\mu_y + c$$

There is also a similar important formula for the variance of a linear combination of RVs:

$$\text{Var}(aX + bY + c) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$

When X, Y, ... are independent, the formula simplifies (no covariances).

And two more important concepts:

Probability generating function (PGF) of a discrete-type RV:

$$P(z) = \mathbb{E}\left[z^X\right] = \sum_{\text{All } i} z^i \cdot f(i)$$

and **moment generating function** (MGF) of a continuous-type RV:

$$M(u) = \mathbb{E}\left[e^{uX}\right] = \int_{\text{All } x} e^{ux} \cdot f(x)\ dx$$

Based on each, we can easily compute the mean and standard deviation of X, by either

$$\begin{aligned} \mu &= P'(z)\big|_{z=1} \\ \mathbb{E}\left[X(X-1)\right] &= \sigma^2 + \mu^2 - \mu = P''(z)\big|_{z=1} \end{aligned}$$

or

$$\begin{aligned} \mu &= M'(u)\big|_{u=0} \\ \mathbb{E}\left[X^2\right] &= \sigma^2 + \mu^2 = M''(u)\big|_{u=0} \end{aligned}$$

The PGF can also yield

$$f(i) = \frac{P^{(i)}(z)\big|_{z=0}}{i!}$$

inverting the MGF to get $f(x)$ is more difficult (and we won't try it).

To get a moment generating function of a linear transformation of X, we do

$$M_{aX+b}(u) = e^{bu} \cdot M(au)$$

Sum of two independent RVs, say $V = X + Y$

We can instantly find its PGF (MGF) by simply multiplying the individual PGFs (MGFs).

To get its pdf, we need to compute the so called `convolution` of $f_X(x)$ and $f_Y(y)$, thus

$$f_V(v) = \int_{\text{All } x} f_X(x) \cdot f_Y(v-x)\, dx$$

The concept of random independent **sample** of size n, and the corresponding **sample mean**

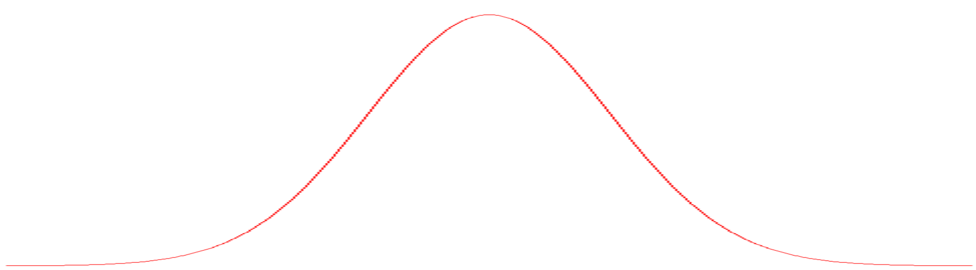
$$\bar{X} \equiv \frac{\sum_{i=1}^{n} X_i}{n}$$

also a *random variable* whose distribution has the same mean as the distribution we are sampling, but its standard deviation is reduced by a factor of $\sqrt{n}$, thus

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

(it clearly tends to 0 in the $n \to \infty$ limit - this implies the Law of Large Numbers).

Central Limit Theorem tells us that the *shape* of the $\bar{X}$ distribution tends (as n increases) to the Normal 'bell-shaped' curve



regardless of the shape of the original distribution (this is true even when sampling from a *discrete* distribution - know how to do `continuity correction`)!

Sum of N IIDs (independent, identically distributed RVs), where N is a random variable itself, i.e.

$$S_N \equiv X_1 + X_2 + ... + X_N$$

has a PGF given by the following `composition` of the PGF of N, and of the X_is.

$$P_N[P_X(z)]$$

(mathematically, a very simple operation).