Applications

- Probability estimation

Let A be an event (set of outcomes of a random phenomenon) of which we want to estimate the probability $P(A)$.

Examples

- $A = \{$the outcome of flipping a coin is head$\}$
- $A = \{$a person who tested positive to COVID-19 is asymptomatic$\}$
- $A = \{X(t) = x\}$, where $X(t)$ is the state of a stochastic timed automaton at time t.

Consider N independent observations of the random phenomenon, and let N_A be the number of times that A was observed. Then, an estimate of $P(A)$ can be computed as:

$$\boxed{\hat{P}(A) = \frac{N_A}{N}}$$

The properties of this estimator can be studied with the Law of Large Numbers.

Let ω_i be the i-th outcome of the random phenomenon, and define the random variable:

$$\mathbb{1}_A(\omega_i) \triangleq \begin{cases} 1 & \text{if event } A \text{ is observed in } \omega_i \\ 0 & \text{otherwise} \end{cases}$$

Indicator function of event A

$\mathbb{1}_A$ is a random variable such that:

- $E[\mathbb{1}_A] = 0 \cdot [1-P(A)] + 1 \cdot P(A) = P(A)$
- $Var(\mathbb{1}_A) = [0-P(A)]^2[1-P(A)] + [1-P(A)]^2 P(A)$

$$= P(A)[1-P(A)][\cancel{P(A)} + 1 - \cancel{P(A)}]$$

$$= P(A)[1-P(A)]$$

$\Rightarrow$ Since $E[\mathbb{1}_A] = P(A)$, an estimate of $E[\mathbb{1}_A]$ is an estimate of $P(A)$.

$\Rightarrow$ $E[\mathbb{1}_A]$ can be estimated using the Law of Large Numbers, and with the theoretical guarantees thereof.

In this respect, notice that:

$$\frac{\sum_{i=1}^{N} \mathbb{1}_A(\omega_i)}{N} = \frac{N_A}{N}$$

Arithmetic mean of the observations $\mathbb{1}_A(\omega_i)$

Hence, according to the Law of Large Numbers,

$\frac{N_A}{N}$ is an unbiased, consistent estimate of $P(A)$.

For the choice of N (whenever this is possible), consider the following.

Assume that a desired accuracy $\Delta > 0$ of the estimate is given. The problem is to choose N such that

$$|\hat{P}(A) - P(A)| \leq \Delta$$

For the Central Limit Theorem, we know that

$$\hat{P}(A) \sim N\left(\mu, \frac{\sigma^2}{N}\right),$$

where

- $\mu = E[\mathbb{1}_A] = P(A)$
- $\sigma^2 = Var(\mathbb{1}_A) = P(A)[1-P(A)]$

for N sufficiently large.

Hence,

$$P\left(|\hat{P}(A) - P(A)| \leq \frac{3\sigma}{\sqrt{N}} \right) \simeq 0.9973$$

From the tables of the Normal distribution:

if $X \sim N(\mu, \sigma^2)$,

$P(|X-\mu| \leq \sigma) \simeq 0.6827$

$P(|X-\mu| \leq 2\sigma) \simeq 0.9545$

$P(|X-\mu| \leq 3\sigma) \simeq 0.9973$

If we accept that 3 times out of 1000 (on average) the estimate differs from the true value more than $\frac{3\sigma}{\sqrt{N}}$, we can set

$$\Delta = \frac{3\sigma}{\sqrt{N}} \Rightarrow \boxed{N = \frac{9\sigma^2}{\Delta^2}} \quad (*)$$

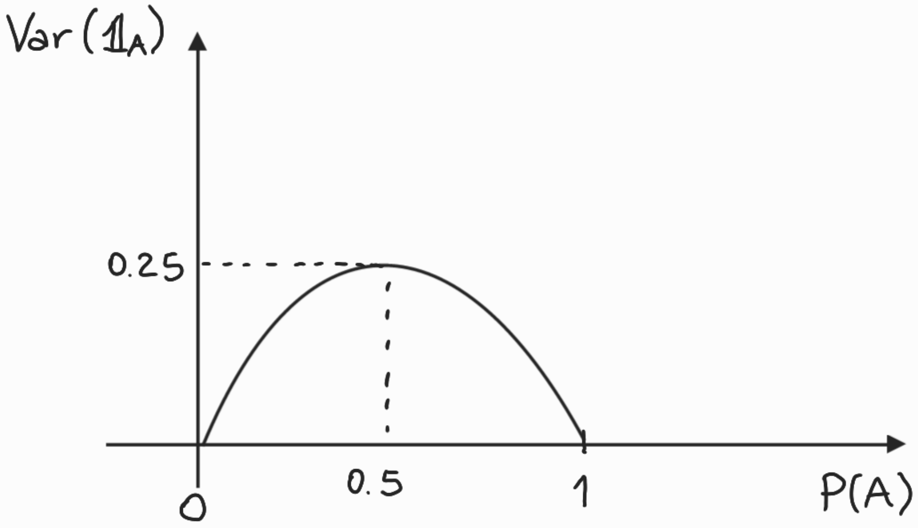
The problem with this formula is that

$\sigma^2 = P(A)[1-P(A)] \Rightarrow$ It depends on the quantity we want to estimate

However, if we plot this function versus $P(A)$, we observe that

$$P(A)[1-P(A)] \leq 0.25 \quad \forall\, 0 \leq P(A) \leq 1$$

The maximum is achieved for $P(A) = 0.5$.



This makes it possible to replace σ^2 in (*) with its upper bound 0.25, thus obtaining the following formula for the choice of N:

$$\boxed{N = \frac{2.25}{\Delta^2}}$$

We used that $9\sigma^2 \leq 9 \cdot 0.25 = 2.25$.

EXAMPLE

We have an unfair coin.

The probability to get head is $p = 0.7$.

Assume that we do not know p, and we want to estimate it from the results of flipping the coin repeatedly.

Let $\Delta = 0.01$ be the required accuracy for estimating p.

Using the approximated formula for N, we have

$$N = \frac{2.25}{(0.01)^2} = 22500$$

With this choice of N, we expect that, on average, no more than 3 times out of 1000 the estimate $\hat{p}$ should differ from the true value p more than Δ.

We compute $M = 1000$ estimates $\hat{p}$ of p.

Each estimate is computed using $N = 22500$ observations of the random experiment.

One observation consists of flipping the coin, and recording the result (head or tail).

The figure shows the histogram of the M estimates.

All the estimates except one are within the interval $[p-\Delta, p+\Delta] = [0.69, 0.71]$.