Random variables - probability distribution

Exercise 1 (Binomial distribution)

During university registrations, each student completes a registration file. All checks carried out indicate that the probability that any registration is well filled in is equal to $p = 0.94$.

1) Introduce a random variable X which describes the two possible states for each file.

2) Provide its probability distribution

3) Calculate its expectation and its variance.

We now consider a batch of n files and we are interested in number of well-filled files among the n files.

4) Introduce a random variable X which represents the number of well-filled files.

5) What is its probability distribution ? Provide its expectation and its variance.

6) If $n = 5$, calculate the probability of the following events : $\{$ no file is well filled$\}$, $\{$all files are well filled$\}$, $\{X > 3\}$, $\{ 2 < X < 4\}$.

7) If $n = 100$, what probability distribution can we use to approximate the distribution of X ?

Solution

6) We compute the following probabilities :

— $\mathbb{P}[\{\text{no file is well filled}\}] = \mathbb{P}[\{X = 0\}] = \binom{5}{0} p^0 (1-p)^5 = 7.8 \times 10^{-7}$;

— $\mathbb{P}[\{X > 3\}] = \binom{5}{4} p^4 (1-p)^1 + \binom{5}{5} p^5 (1-p)^0 = 0.97$;

— $\mathbb{P}[\{2 < X < 4\}] = \mathbb{P}[\{X = 3\}] = \binom{5}{3} p^3 (1-p)^2 = 0.18$.

7) For $n = 100$, we can approximate the binomial distribution by the normal distribution with mean $\mu = np$ and standard deviation $\sigma = \sqrt{npq}$. This approximation is valid if $np > 5$ and $nq > 5$, which is true here.

Exercise 2 (Distribution function)
Atmospheric ozone concentration (in $\mu g/m^3$ = microgram (one millionth of a gram) per cubic meter) is modeled by a Gaussian random variable X of mean m and variance σ^2, denoted by $\mathcal{N}(m, \sigma^2)$, where $m = 178$ and $\sigma^2 = 3.1$.

1) What are the units of measure of m and σ ? What do they represent ?

An ozone concentration greater than $180\mu g/m^3$ is considered dangerous for humans.

2) a) What is the probability that the concentration exceeds 180 ?
2) b) Assuming $m = 180$, find a real number δ such that the probability

$$\mathbb{P}(180 - \delta \leq X \leq 180 + \delta)$$

is larger than 95%.

3) Now assume that m and σ are unknown. For a fixed value x, calculate the probability that X is less than or equal to x. Deduce the distribution function of X.

Solution

1) m and σ are in $\mu g/m^3$. m represents the real atmospheric concentration and σ the imprecision of the measure.
2) (a) For this exercise, you can use the table of the cdf of the standard normal distribution. This probability is

$$\begin{aligned} P_{\mathcal{N}(178,3.1)}(X \geq 180) &= P_{\mathcal{N}(178,3.1)}\left(\frac{X-178}{\sqrt{3.1}} \geq \frac{180-178}{\sqrt{3.1}}\right) \\ &= P_{U\sim\mathcal{N}(0,1)}\left(U \geq \frac{2}{\sqrt{3.1}}\right) = 1 - F_{\mathcal{N}(0,1)}\left(\frac{2}{\sqrt{3.1}}\right) \sim 12.74\%, \end{aligned}$$

where $F_{\mathcal{N}(0,1)}$ is the cdf of a standard normal distribution.

(b)

$$\begin{aligned} P_{\mathcal{N}(180,3.1)}(180 - \delta \leq X \leq 180 + \delta) &= P_{\mathcal{N}(180,3.1)}\left(\frac{180-\delta-180}{\sqrt{3.1}} \leq \frac{X-180}{\sqrt{3.1}} \leq \frac{180+\delta-180}{\sqrt{3.1}}\right) \\ &= P_{U\sim\mathcal{N}(0,1)}\left(-\frac{\delta}{\sqrt{3.1}} \leq U \leq \frac{\delta}{\sqrt{3.1}}\right) \\ &= 1 - 2P_{U\sim\mathcal{N}(0,1)}\left(U \geq \frac{\delta}{\sqrt{3.1}}\right) \leq 0.95 \\ &\Rightarrow 2F_{\mathcal{N}(0,1)}\left(\frac{\delta}{\sqrt{3.1}}\right) - 1 \leq 0.95 \\ &\iff F_{\mathcal{N}(0,1)}\left(\frac{\delta}{\sqrt{3.1}}\right) \leq 0.975. \end{aligned}$$

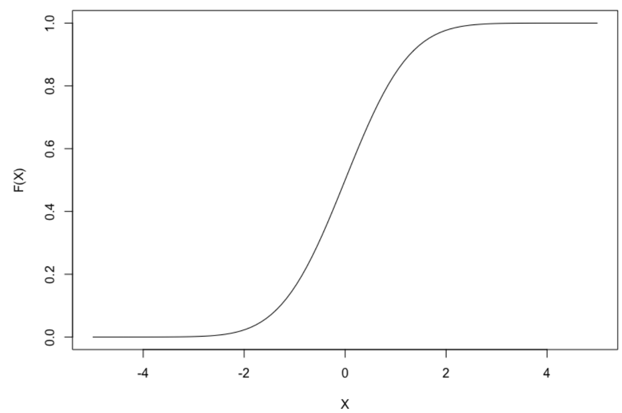

FIGURE 1 – CDF of the standard Gaussian for $m = 0$ and $\sigma = 1$.

hence, using the table of the cdf of the standard normal distribution, we see that $\delta/\sqrt{3.1} = 1.96 \implies \delta = 3.45$ is a suitable choice.

4) See Figure 1.