12.2: Continuous random variables: Probability distribution functions

Given a sequence of data points $a_1, \ldots, a_n$, its *cumulative distribution function* $F(x)$ is defined by

$$F(A) := \frac{\text{number of } i \text{ with } a_i \leq A}{n}$$

That is, $F(A)$ is the relative proportion of the data points taking value less than or equal to A.

Properties of cumulative distribution functions

- The cumulative distribution function F for the data points $a_1, \ldots, a_n$ may be computed from the corresponding random variable X via the formula

$$F(A) = \sum_{v \leq A} vX(v)$$

- $\lim_{A \to -\infty} F(A) = 0$ and $\lim_{A \to \infty} F(A) = 1$
- $A \leq B \Rightarrow F(A) \leq F(B)$

Example

Given the data points $5, 3, 6, 2, 5, 2, 1, -4, 0, 4, 9, 10, 3, 3, 6, 8$, compute $F(4)$ where $F(x)$ is the corresponding cumulative distribution function.

Solution

There are a total of sixteen data points of which nine have a value less than or equal to four. Thus, $F(4) = \frac{9}{16}$.

Computing probabilities with cumulative distributions

One may regard the cumulative distribution function $F(x)$ as describing the probability that a randomly chosen data point will have value less than or equal to x.

If X is the correponding random variable, one often writes

$$\Pr(X \leq x) = F(x)$$

From F we may compute other probabilities. For instance, the probability of obtaining a value greater than A but less than or equal to B is

$$\Pr(A < X \leq B) = F(B) - F(A)$$

Continuous random variables

We may wish to express the probability that a numerical value of a particular experiment lie with a certain range even though infinitely many such values are possible.

- Express the probability that if a coin is flipped repeatedly, the first result of heads will occur by the n^{th} flip.
- What is the probability that a major earthquake will occur on the North Hayward fault within the next five years?
- What is the probability that a randomly selected high school senior will score at least 600 on the SAT?

General cumulative distribution functions

A cumulative distribution function (in general) is a function $F(x)$ defined for all real numbers for which

- $A \leq B \Rightarrow F(A) \leq F(B)$
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$

We write X for the corresponding random variable and treat F as expressing $F(A) =$ the probability that $X \leq A = \Pr(X \leq A)$.

Probability densities

If the cumulative distribution function $F(x)$ (for the random variable X) is differentiable and have derivative $f(x) = F'(x)$, then we say that $f(x)$ is the *probability density function for X.*

For numbers $A \leq B$ we have

$$\begin{aligned} \Pr(A < X \leq B) &= F(B) - F(A) \\ &= \int_A^B f(x)dx \end{aligned}$$

Properties of probability densities

- $0 \leq f(x)$ for all values of x since F is non-decreasing.
- $F(A) = \int_{-\infty}^{A} f(x)dx$
- $\lim_{x \to -\infty} f(x) = 0 = \lim_{x \to \infty} f(x)$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

Conversely, any function satisfying the above properties is a probability density.

Example

The function

$$f(x) = \begin{cases} e^{-x} \text{ if } x \geq 0 \\ 0 \text{ if } x < 0 \end{cases}$$

is a probability density (for the random variable X).

Compute $\Pr(-10 \leq X \leq 10)$.

Solution

We know

$$\begin{aligned}
\Pr(-10 \leq X \leq 10) &= \int_{-10}^{10} f(x)dx \\
&= (\int_{-10}^{0} 0dx) + (\int_{0}^{10} e^{-x}dx) \\
&= 0 + (-e^{-x}|_{x=0}^{x=10}) \\
&= 1 - e^{-10}
\end{aligned}$$