1 Probability Theory

A probability is a measure of how frequent or likely an event will take place.

Probability Space The probability space is a triplet space containing a sample/outcome space Ω (containing all possible atomic events), a collection of events S (containing a subset of Ω to which we want to assign probabilities) and the mapping P between Ω and S.

Axioms of Probability The mapping P must fulfill the axioms of probability:

1. $P(a) \geq 0$
2. $P(\Omega) = 1$
3. $a, b \in S$ and $a \cap b = \{\} \Rightarrow P(a \cup b) = P(a) + P(b)$

Random Variable A random variable is a function that maps points from the sample space Ω to some range (e.g. Real numbers or booleans). They are characterized by their distribution function. E.g. for a dice roll:

$$X(\omega) = \begin{cases} 0, & \text{if } \omega = heads \\ 1, & \text{if } \omega = tails. \end{cases}$$

Proposition A Proposition is a conclusion of a statistical inference that can be true or false (e.g. a classification of a datapoint). More formally: A disjunction of events where the logic model holds. An event can be written as a **propositional logic model**:
$A = true, B = false \Rightarrow a \wedge \neg b$. Propositions can be continuous, discrete or boolean.

1.1 Probability distributions

Probability distributions assign probabilities to to all possible points in Ω (e.g. $P(Weather) = \langle 0.3, 0.4, 0.2, 0.1 \rangle$, representing Rain, sunshine, clouds and snow). Joint probability distributions give you a probability for each atomic event of the random variables (e.g. $P(weather, accident)$ gives you a 2×4 matrix.)

Cumulative Distribution Function (CDF) The CDF is defined as $F_X(x) = P(X \leq x)$ (See figure 2).

Probability Density Function (PDF) For continuous functions the PDF is defined by

$$p(x) = \frac{d}{dx} p(X \leq x). \tag{1}$$

The probability of x being in a finite interval is

$$P(a < X \leq b) = \int_a^b p(x) dx \tag{2}$$

A PDF is shown in figure

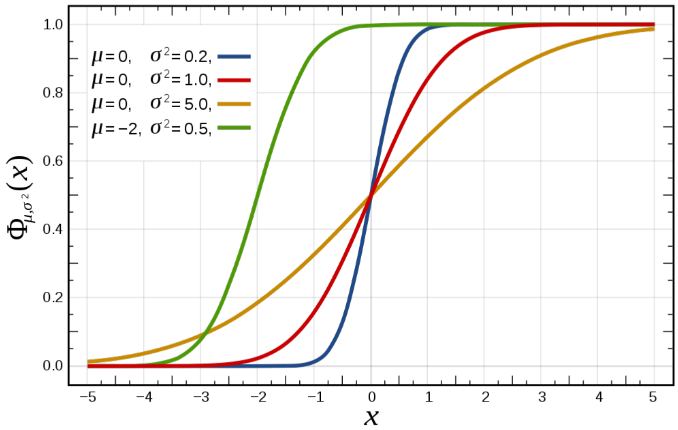

Fig. 1: Cumulative distribution function of a normal distribution for different mean (μ) and variance (σ). *Figure from user Inductiveload on wikimedia.org.*

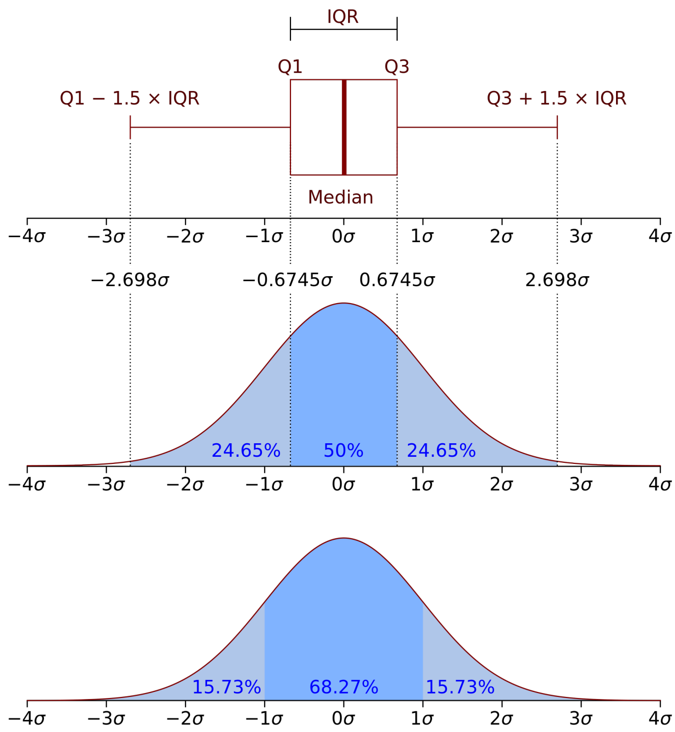

Fig. 2: Probability density function of a normal distribution with variance (σ). In red a range from a Box-plot is shown with quartiles (Q1, Q3) and interquartile range (IQR). For the cutoffs (borders to darker blue regions) the IQR (on top) and σ are chosen. Another common cutoff is the confidence interval with light blue regions having a probability mass of $2 * \alpha/2$. *Figure from user Jhguch on wikimedia.org.*

Uniform distribution The uniform distribution has the same probability throughout a specific interval and is defined as

$$\text{Unif}(a,b) = \frac{1}{b-a}\mathbb{I}(a < x \leq b) = \begin{cases} \frac{1}{b-a}, & \text{if } x \in [a,b] \\ 0, & \text{else} \end{cases}$$

Properties of Distributions