

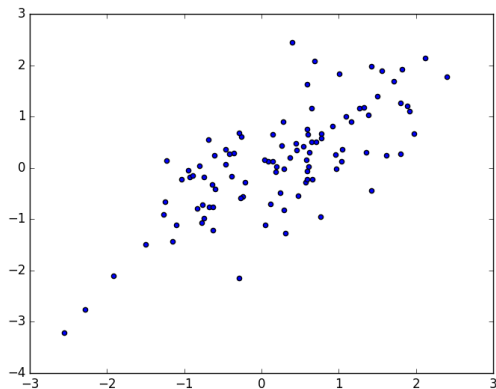


Big Data Analytics

– Week 2: Probability Theory –

Knowledge Discovery and Data Mining
Prof. Dr. Emmanuel Müller

Descriptive Statistics & Probability Theory

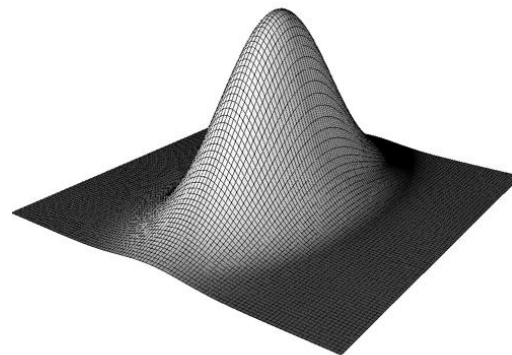


Descriptive Statistics

descriptions of
observed data (sample)

**Statistical
Inference** →

e.g. maximum
likelihood



Probability Theory

abstract models for
full population

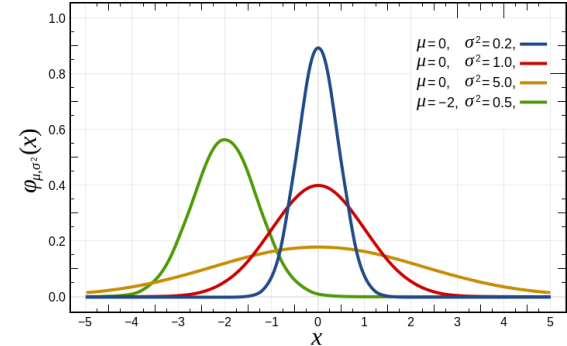
Continuous: Normal Distribution

- also known as **Gaussian distribution**
- fundamental probability density function for continuous X
 - data often assumed to be normally distributed
 - nice analytical properties (marginals, conditionals)

$X \sim N(\mu, \sigma^2)$ with mean μ and variance σ^2

$$N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}$$

$$P(X \in [a, b]) = \int_a^b N(\mu, \sigma^2) dx$$



Probability Theory (short intro)

- Probability theory is the mathematical study of uncertainty
 - Plays a central role in data analytics
 - Many algorithms rely on probabilistic assumption of data

Formally, we study the probabilities of events,
i.e. a probability space is defined by a triple (Ω, F, P)

- Ω is the space of possible outcomes
- $F \subseteq 2^\Omega$ (the power set of Ω) is the space of events
- P is the probability measure (or probability distribution) that maps each event in F to a real value between 0 and 1

Restrictions to F : *trivial events in F ($\Omega \subseteq F, \emptyset \subseteq F$),
 F is closed under union and under complement*

Example

Suppose we throw a six-sided dice

- The outcome space is

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

We may decide that the events of interest are odd or even

- The event space is

$$F = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$$

Probability Measure

Given an event space F , the probability measure P has to satisfy certain axioms:

- non-negative

$$\forall a \subseteq F: P(a) \geq 0$$

- trivial event Ω

$$P(\Omega) = 1$$

- additivity

$$\forall a, b \in F \wedge a \cap b = \emptyset: P(a \cup b) = P(a) + P(b)$$

Example:

$$P(\{1\}) = P(\{2\}) = \dots = P(\{6\}) = \frac{1}{6}$$

$$P(\{2, 4, 6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

Random Variables

We abstract from individual events and use *random variables* that capture the appropriate events

For example, we could define a random variable $X = 1$ (if outcome is odd) and $X = 0$ (otherwise)

Formally, we write $P(X = a)$ with $a \in \text{Val}(X)$ (the range of values)

Joint and Marginal Distributions

P	X = 1	X = 2	X = 3	X = 4	X = 5	X = 6
Y = 0	1/12	1/12	1/12	1/12	1/12	1/12
Y = 1	1/12	1/12	1/12	1/12	1/12	1/12

- $P(X = a, Y = b)$ is the joint probability
 $P(X, Y)$ the respective probability distribution
- $P(X)$ and $P(Y)$ are the marginal distributions with $P(X) = \sum_{b \in \text{Val}(Y)} P(X, Y = b)$

Conditional Distributions

Specify the distribution of a random variable
when the value of another random variable is known

Formally, conditional probability of
 $X = a$ given $Y = b$ is defined as:

$$P(X = a \mid Y = b) = \frac{P(X = a, Y = b)}{P(Y = b)}$$

Example: Suppose we know that a dice throw was odd ($Y = 1$), and want to compute the probability of a „three“ event ($X = 3$):

$$P(X = 3 \mid Y = 1) = \frac{P(X = 3, Y = 1)}{P(Y = 1)} = \frac{1/6}{1/2} = 1/3$$

Independence

Independence means that the distribution of a random variable does not change on learning the value of another random variable.

Mathematically, a random variable X is independent of Y when:

$$P(X) = P(X | Y)$$

or in an equivalent statement:

$$P(X, Y) = P(X)P(Y)$$

Expectation and Variance

Expectation of a random variable X :

$$E(X) = \sum_{a \in \text{Val}(X)} a P(X = a) \quad (\text{discrete})$$

$$E(X) = \int_{\text{Val}(X)} x f(x) dx \quad (\text{continuous})$$

Variance of a distribution is a measure of the “spread” of a distribution.
It is measured by the second moment:

$$\text{Var}(X) = E\left(\left(X - E(X)\right)^2\right)$$

or in an equivalent statement:

$$\text{Var}(X) = E(X^2) - (E(X))^2$$