Advanced Topics in Software Engineering: Data Analysis

Prof. Michele Loreti

Advanced Topics in Software Engineering
Corso di Laurea in Informatica (L31)
Scuola di Scienze e Tecnologie

Probability. . .

Recall. . .

A sample space Ω is the set of possible outcomes of an experiment.

Probability...

Recall...

A sample space Ω is the set of possible outcomes of an experiment.

A σ-algebra Σ on Ω is a family of subsets of Ω such that:

- $\Omega \in \Sigma$;
- if $A \in \Sigma$ then $\overline{A} = \Omega - A \in \Sigma$;
- for any $A_1, \ldots, A_n \in \Sigma$

$$\bigcup_i A_i \in \Sigma$$

A sample space Ω is the set of possible outcomes of an experiment.

A σ-algebra Σ on Ω is a family of subsets of Ω such that:

- $\Omega \in \Sigma$;
- if $A \in \Sigma$ then $\overline{A} = \Omega - A \in \Sigma$;
- for any $A_1, \ldots, A_n \in \Sigma$

$$\bigcup_i A_i \in \Sigma$$

An element $\omega \in \Omega$ is named a sample outcomes or realisation while $A \in \Sigma$ is an event.

Probability. . .

Recall. . .

Example: Tossing a coin twice

Probability. . .

Recall. . .

Example: Tossing a coin twice

$$\Omega = \{TT, TH, HT, HH\}$$

Example: Tossing a coin twice

$$\Omega = \{TT, TH, HT, HH\}$$

The event *“the first is head”* is

$$A = \{HT, HH\}$$

Probability...

Recall...

Example: Tossing a coin twice

$$\Omega = \{TT, TH, HT, HH\}$$

The event *"the first is head"* is

$$A = \{HT, HH\}$$

Example: Measurement of a physical experiment

Probability. . .

Recall. . .

Example: Tossing a coin twice

$$\Omega = \{TT, TH, HT, HH\}$$

The event *"the first is head"* is

$$A = \{HT, HH\}$$

Example: Measurement of a physical experiment

$$\Omega = \mathbb{R} = [-\infty, +\infty]$$

Probability. . .

Recall. . .

Example: Tossing a coin twice

$$\Omega = \{TT, TH, HT, HH\}$$

The event *"the first is head"* is

$$A = \{HT, HH\}$$

Example: Measurement of a physical experiment

$$\Omega = \mathbb{R} = [-\infty, +\infty]$$

The event *"measure is larger than* 10 *but less or equale to* 23*"* is

$$A = (10, 23]$$

Probability...

Recall...

A probability space is a tuple (Ω, Σ, Pr) where

Probability...

Recall...

A probability space is a tuple (Ω, Σ, Pr) where

- Ω is a sample space;

Probability...

Recall...

A probability space is a tuple (Ω, Σ, Pr) where

- Ω is a sample space;
- Σ is a σ-algebra on Ω;

Probability...

Recall...

A probability space is a tuple (Ω, Σ, Pr) where

- Ω is a sample space;
- Σ is a σ-algebra on Ω;
- $Pr : \Sigma \rightarrow [0, 1]$ such that:
  - $Pr(\Omega) = 1$
  - for any $A_1, \ldots, A_n$ ($A_i \cap A_j = \emptyset$ for any $i \neq j$):

$$Pr\left(\bigcup_i A_i\right) = \sum_i Pr(A_i)$$

Probability...

Recall...

A probability space is a tuple (Ω, Σ, Pr) where

- Ω is a sample space;
- Σ is a σ-algebra on Ω;
- $Pr : \Sigma \to [0,1]$ such that:
  - $Pr(\Omega) = 1$
  - for any $A_1, \ldots, A_n$ ($A_i \cap A_j = \emptyset$ for any $i \neq j$):

$$Pr\left(\bigcup_i A_i\right) = \sum_i Pr(A_i)$$

Remark: If Ω is finite, and if each outcome is equally likely, then

$$Pr(A) = \frac{|A|}{|\Omega|}$$

Probability...

Recall...

Let (Ω, Σ, Pr) be a probability space...

Probability. . .

Recall. . .

Let (Ω, Σ, Pr) be a probability space. . .

For any $A, B \in \Sigma$, $Pr(A \cup B) = Pr(A) \cup Pr(B) - Pr(A \cap B)$.

Probability...

Recall...

Let (Ω, Σ, Pr) be a probability space...

For any $A, B \in \Sigma$, $Pr(A \cup B) = Pr(A) \cup Pr(B) - Pr(A \cap B)$.

Two events $A, B \in \Sigma$ are independent if and only if

$$Pr(A \cap B) = Pr(A) \cdot Pr(B).$$

Probability...

Recall...

Let $A, B \in \Sigma$, if $Pr(B) > 0$ then the conditional probability of A *given* B is:

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

Probability. . .

Recall. . .

Let $A, B \in \Sigma$, if $Pr(B) > 0$ then the conditional probability of A *given* B is:

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

Remark: $Pr(A|B)$ is the fraction of times A occurs among those in which B occurs!

Let $A, B \in \Sigma$, if $Pr(B) > 0$ then the conditional probability of A *given* B is:

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

Remark: $Pr(A|B)$ is the fraction of times A occurs among those in which B occurs!

If A and B are independent...

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(A) \cdot Pr(B)}{Pr(B)} = Pr(A)$$

Random Variables. . .

A random variable is a mapping $X : \Omega \to \mathbb{R}$.

Random Variables...

A random variable is a mapping $X : \Omega \to \mathbb{R}$.

Let $(\Omega, \Sigma, \Pr)$ be a probability space, a random variable $X : \Omega \to \mathbb{R}$ is a measurable function from Ω to $\mathbb{R}$.

Random Variables...

A random variable is a mapping $X : \Omega \to \mathbb{R}$.

Let $(\Omega, \Sigma, \Pr)$ be a probability space, a random variable $X : \Omega \to \mathbb{R}$ is a measurable function from Ω to $\mathbb{R}$.

The probability that X takes value in a measurable set $S \subseteq \mathbb{R}$ is written as:

$$\Pr(X \in S) = \Pr(\{\omega \in \Omega | X(\omega) \in S\})$$

Random Variables...

Example...

The sample space of 3 coin flips is:

$$\Omega = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$$

Random Variables...
Example...

The sample space of 3 coin flips is:

$$\Omega = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$$

If the coin is *fair*, for each $\omega \in \Omega$ $Pr(\omega) = \frac{1}{8}$.

Random Variables...

Example...

The sample space of 3 coin flips is:

$$\Omega = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$$

If the coin is *fair*, for each $\omega \in \Omega$ $Pr(\omega) = \frac{1}{8}$.

Let $X(\omega)$ be the number of *heads* in the sequence ω.

Random Variables...
Example...

The sample space of 3 coin flips is:

$$\Omega = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$$

If the coin is *fair*, for each $\omega \in \Omega$ $Pr(\omega) = \frac{1}{8}$.

Let $X(\omega)$ be the number of *heads* in the sequence ω.

Let $k \in \{0, 1, 2, 3\}$:

$$Pr(X = k) = \binom{3}{k}\frac{1}{8} = \frac{3!}{k! \cdot (3-k)!}\frac{1}{8}$$

Distribution function

Let X a random variable on the probability space $(\Omega, \Sigma, \Pr)$, we define the distribution function F_X for each real $x \in \mathbb{R}$ by

$$F_X(x) = \Pr[X \leq x] = \Pr(\{\omega | X(\omega) \leq x\})$$

Distribution function

Let X a random variable on the probability space $(\Omega, \Sigma, \Pr)$, we define the distribution function F_X for each real $x \in \mathbb{R}$ by

$$F_X(x) = \Pr[X \leq x] = \Pr(\{\omega | X(\omega) \leq x\})$$

We associate another function $p_X(\cdot)$, called the probability mass function, with X (pmf), for each $x \in \mathbb{R}$:

$$p(x) = \Pr[X = x] = \Pr(\{\omega | X(\omega) \leq x\})$$

Distribution function

Let X a random variable on the probability space $(\Omega, \Sigma, \Pr)$, we define the distribution function F_X for each real $x \in \mathbb{R}$ by

$$F_X(x) = \Pr[X \leq x] = \Pr(\{\omega | X(\omega) \leq x\})$$

We associate another function $p_X(\cdot)$, called the probability mass function, with X (pmf), for each $x \in \mathbb{R}$:

$$p(x) = \Pr[X = x] = \Pr(\{\omega | X(\omega) \leq x\})$$

A random variable X is continuous if $p(x) = 0$ for all real x.

Distribution function

Let X a random variable on the probability space $(\Omega, \Sigma, \Pr)$, we define the distribution function F_X for each real $x \in \mathbb{R}$ by

$$F_X(x) = \Pr[X \leq x] = \Pr(\{\omega | X(\omega) \leq x\})$$

We associate another function $p_X(\cdot)$, called the probability mass function, with X (pmf), for each $x \in \mathbb{R}$:

$$p(x) = \Pr[X = x] = \Pr(\{\omega | X(\omega) \leq x\})$$

A random variable X is continuous if $p(x) = 0$ for all real x.

NB: If X is a continuous random variable, then X can assume infinitely many values, and so it is reasonable that the probability of its assuming any specific value we choose beforehand is zero.

Example: Dice Roll

A random variable can be used to describe the process of rolling two (fair) dice and the possible outcomes.

Example: Dice Roll

A random variable can be used to describe the process of rolling two (fair) dice and the possible outcomes.

We can consider the probability space $(\Omega_{2D}, \Sigma_{2D}, \mathrm{Pr}_{2D})$ such that:

$$\Omega_{2D} = \{(n_1, n_2) | 1 \leq n_1, n_2 \leq 6\} \qquad \Sigma_{2D} = 2^{\Omega_{2D}} \qquad Pr(A) = \frac{|A|}{36}$$

Example: Dice Roll

A random variable can be used to describe the process of rolling two (fair) dice and the possible outcomes.

We can consider the probability space $(\Omega_{2D}, \Sigma_{2D}, \mathrm{Pr}_{2D})$ such that:

$$\Omega_{2D} = \{(n_1, n_2) | 1 \leq n_1, n_2 \leq 6\} \qquad \Sigma_{2D} = 2^{\Omega_{2D}} \qquad Pr(A) = \frac{|A|}{36}$$

The total number rolled is then a random variable X given by the function that maps the pair to the sum: $X((n_1, n_2)) = n_1 + n_2$

Example: Dice Roll

A random variable can be used to describe the process of rolling two (fair) dice and the possible outcomes.

We can consider the probability space $(\Omega_{2D}, \Sigma_{2D}, \Pr_{2D})$ such that:

$$\Omega_{2D} = \{(n_1, n_2) | 1 \leq n_1, n_2 \leq 6\} \qquad \Sigma_{2D} = 2^{\Omega_{2D}} \qquad Pr(A) = \frac{|A|}{36}$$

The total number rolled is then a random variable X given by the function that maps the pair to the sum: $X((n_1, n_2)) = n_1 + n_2$

The pms function p_X and the df F_X function can be defined as:

$$p_X(x) = \begin{cases} \frac{\min(x-1, 13-x)}{36} & 2 \leq x \leq 12 \\ 0 & \text{otherwise} \end{cases} \qquad F_X(x) = \sum_{y \leq x} p_X(y)$$

Mean, or expected value

If X is a discrete random variable with probability mass function $p(\cdot)$, we define the mean or expected value of $X \in S$, $\mu = E[X]$ by

$$E(X) = \sum_{x \in S} x \cdot p(x)$$

Mean, or expected value

If X is a discrete random variable with probability mass function $p(\cdot)$, we define the mean or expected value of $X \in S$, $\mu = E[X]$ by

$$E(X) = \sum_{x \in S} x \cdot p(x)$$

If X is a continuous random variable with density function $f(\cdot) = \frac{dF(\cdot)}{dx}$, we define the mean or expected value of X, $\mu = E[X]$ by

$$\mu = E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Variance

The expectation only gives us an idea of the average value assumed by a random variable, not how much individual values may differ from this average.

Variance

The expectation only gives us an idea of the average value assumed by a random variable, not how much individual values may differ from this average.

The variance, $Var(X)$, gives us an indication of the "spread" of values:

$$Var(X) = E\left[(X - E[X])^2\right] = E\left[X^2\right] - E\left[X\right]^2$$

Variance

The expectation only gives us an idea of the average value assumed by a random variable, not how much individual values may differ from this average.

The variance, $Var(X)$, gives us an indication of the "spread" of values:

$$Var(X) = E\left[(X - E[X])^2\right] = E\left[X^2\right] - E\left[X\right]^2$$

The standard deviation of X, $sd(X) = \sqrt{Var(X)}$.

Covariance. . .

Let X and Y be two random variables with means μ_X and μ_Y and standard deviations σ_X and σ_Y. The covariance between X and Y is defined as:

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

Covariance...

Let X and Y be two random variables with means μ_X and μ_Y and standard deviations σ_X and σ_Y. The covariance between X and Y is defined as:

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

The correlation is defined as:

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Normal Distribution

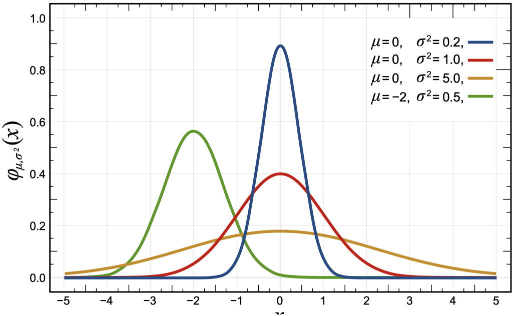
A random variable X has a Normal (or Gaussian) distribution with parameters μ and σ if and only if it has *probability density function*:

$$\phi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Normal Distribution

A random variable X has a Normal (or Gaussian) distribution with parameters μ and σ if and only if it has *probability density function*:

$$\phi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$



Normal Distribution

We say that X has a standard Normal distribution if $\mu = 0$ and $\sigma = 1$.

Normal Distribution

We say that X has a standard Normal distribution if $\mu = 0$ and $\sigma = 1$.
Random variables with standard Normal distribution are denoted by Z.

Normal Distribution

We say that X has a standard Normal distribution if $\mu = 0$ and $\sigma = 1$. Random variables with standard Normal distribution are denoted by Z.

Some facts about Normal Distribution:

- If X has distribution $N(\mu, \sigma^2$ then $Z = \frac{(X-\mu)}{\sigma}$ has distribution $N(0, 1)$

Normal Distribution

We say that X has a standard Normal distribution if $\mu = 0$ and $\sigma = 1$. Random variables with standard Normal distribution are denoted by Z.

Some facts about Normal Distribution:

- If X has distribution $N(\mu, \sigma^2$ then $Z = \frac{(X-\mu)}{\sigma}$ has distribution $N(0, 1)$
- If Z has distribution $N(0, 1)$ then $X = \mu + \sigma Z$ has distribution $N(\mu, \sigma^2)$.

Normal Distribution

We say that X has a standard Normal distribution if $\mu = 0$ and $\sigma = 1$. Random variables with standard Normal distribution are denoted by Z.

Some facts about Normal Distribution:

- If X has distribution $N(\mu, \sigma^2$ then $Z = \frac{(X-\mu)}{\sigma}$ has distribution $N(0, 1)$
- If Z has distribution $N(0, 1)$ then $X = \mu + \sigma Z$ has distribution $N(\mu, \sigma^2)$.
- $X_1, \ldots, X_n$ are independent and distributed with $N(\mu_i, \sigma_i^2)$ then $\sum_i X_i$ has distribution

$$N\left(\sum_i \mu_i, \sum_i \sigma_i^2\right)$$

Normal Distribution

Let X be a random variable distributed as $N(\mu, \sigma^2)$:

$$Pr(a < X < b) = Pr\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

where Φ is the distribution function of Z.

Normal Distribution

Let X be a random variable distributed as $N(\mu, \sigma^2)$:

$$Pr(a < X < b) = Pr\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

where Φ is the distribution function of Z.

Unfortunately, there is not any closed form for Φ!

Normal Distribution

Let X be a random variable distributed as $N(\mu, \sigma^2)$:

$$Pr(a < X < b) = Pr\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

where Φ is the distribution function of Z.

Unfortunately, there is not any closed form for Φ!

Tables are available!

Normal Distribution

Let X be distributed as $N(3, 5)$...

Normal Distribution

Let X be distributed as $N(3,5)$...
Compute $Pr(X > 1)$.

Normal Distribution

Let X be distributed as $N(3,5)$...
Compute $Pr(X > 1)$.

$$Pr(X > 1) = 1 - Pr(X < 1) = 1 - Pr\left(Z < \frac{1-3}{\sqrt{5}}\right)$$

Normal Distribution

Let X be distributed as $N(3,5)$...
Compute $Pr(X > 1)$.

$$Pr(X > 1) = 1 - Pr(X < 1) = 1 - Pr\left(Z < \frac{1-3}{\sqrt{5}}\right) = 1 - \Phi(-.8944)$$

Normal Distribution

Let X be distributed as $N(3,5)$...
Compute $Pr(X > 1)$.

$$Pr(X > 1) = 1 - Pr(X < 1) = 1 - Pr\left(Z < \frac{1-3}{\sqrt{5}}\right) = 1 - \Phi(-.8944) = 0.81$$

Normal Distribution

Let X be distributed as $N(3, 5)$...
Compute $Pr(X > 1)$.

$$Pr(X > 1) = 1 - Pr(X < 1) = 1 - Pr\left(Z < \frac{1-3}{\sqrt{5}}\right) = 1 - \Phi(-.8944) = 0.81$$

Find q such that $Pr(X < q) = .2$.

Normal Distribution

Let X be distributed as $N(3,5)$...
Compute $Pr(X > 1)$.

$$Pr(X > 1) = 1 - Pr(X < 1) = 1 - Pr\left(Z < \frac{1-3}{\sqrt{5}}\right) = 1 - \Phi(-.8944) = 0.81$$

Find q such that $Pr(X < q) = .2$.

$$0.2 = Pr(X < q)$$

Normal Distribution

Let X be distributed as $N(3,5)$...
Compute $Pr(X > 1)$.

$$Pr(X > 1) = 1 - Pr(X < 1) = 1 - Pr\left(Z < \frac{1-3}{\sqrt{5}}\right) = 1 - \Phi(-.8944) = 0.81$$

Find q such that $Pr(X < q) = .2$.

$$0.2 = Pr(X < q) = Pr\left(Z < \frac{q-3}{\sqrt{5}}\right) = \Phi\left(\frac{q-3}{\sqrt{5}}\right)$$

Normal Distribution

Let X be distributed as $N(3,5)$...
Compute $Pr(X > 1)$.

$$Pr(X > 1) = 1 - Pr(X < 1) = 1 - Pr\left(Z < \frac{1-3}{\sqrt{5}}\right) = 1 - \Phi(-.8944) = 0.81$$

Find q such that $Pr(X < q) = .2$.

$$0.2 = Pr(X < q) = Pr\left(Z < \frac{q-3}{\sqrt{5}}\right) = \Phi\left(\frac{q-3}{\sqrt{5}}\right)$$

From the Normal table, $\Phi(-.8416) = .2$. Hence:

$$-.8416 = \frac{q-3}{\sqrt{5}} \Rightarrow q = 1.1181$$

Inequialities: Markov and Chebyshev Inequialities

Markov's Inequiality: Let X be a non-negative random variable and suppose that $E[X]$ exists. For any $t > 0$:

$$Pr(X > t) \leq \frac{E[X]}{t}$$

Inequialities: Markov and Chebyshev Inequialities

Markov's Inequiality: Let X be a non-negative random variable and suppose that $E[X]$ exists. For any $t > 0$:

$$Pr(X > t) \leq \frac{E[X]}{t}$$

Chebyshev Inequiality: Let $\mu = E[X]$ and $\sigma^2 = Var[X]$. The,

$$Pr(|X - \mu| > t) \leq \frac{\sigma^2}{t} \qquad Pr(|Z| \geq k) \leq \frac{1}{k^2}$$

where $Z = \frac{X-\mu}{\sigma}$.

Statistics. . .

Probability provides *a priori* information about a *random phenomena*.

Statistics. . .

Probability provides *a priori* information about a *random phenomena*.

Unfortunately, often we don't know the exact probability distribution of a *random variable* X.

Statistics. . .

Probability provides *a priori* information about a *random phenomena*.

Unfortunately, often we don't know the exact probability distribution of a *random variable* X.

In this case we can try to reconstruct the properties of X by using a number of observation.

Statistics...

Probability provides *a priori* information about a *random phenomena*.

Unfortunately, often we don't know the exact probability distribution of a *random variable* X.

In this case we can try to reconstruct the properties of X by using a number of observation.

We can consider two approaches:

Statistics. . .

Probability provides *a priori* information about a *random phenomena*.

Unfortunately, often we don't know the exact probability distribution of a *random variable* X.

In this case we can try to reconstruct the properties of X by using a number of observation.

We can consider two approaches:

- Descriptive Statistics, that is used to say something about a set of information that has been collected only.

Statistics...

Probability provides *a priori* information about a *random phenomena*.

Unfortunately, often we don't know the exact probability distribution of a *random variable* X.

In this case we can try to reconstruct the properties of X by using a number of observation.

We can consider two approaches:

- Descriptive Statistics, that is used to say something about a set of information that has been collected only.
- Inferential Statistics, that is used to make prediction or comparisons about a larger group (a population) using information gathered about a small part of that population.

Independent and Identically Distributed Random Variables. . .

Let us consider a set of data $\mathcal{X}$ collected by observing a random phenomenon:

$$\mathcal{X} = (v_1, \ldots, v_n)$$

Independent and Identically Distributed Random Variables...

Let us consider a set of data $\mathcal{X}$ collected by observing a random phenomenon:

$$\mathcal{X} = (v_1, \ldots, v_n)$$

We can say that $X = (X_1, \ldots, X_n)$ is a random vector and that $X_1, \ldots, X_n$ are Independent and Identically Distributed Random Variables with a Cumulative Distribution Function F.

Independent and Identically Distributed Random Variables...

Let us consider a set of data $\mathcal{X}$ collected by observing a random phenomenon:

$$\mathcal{X} = (v_1, \ldots, v_n)$$

We can say that $X = (X_1, \ldots, X_n)$ is a random vector and that $X_1, \ldots, X_n$ are Independent and Identically Distributed Random Variables with a Cumulative Distribution Function F.

We call $(v_1, \ldots, v_n)$ a random sample from F.

Medians. . .

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

Medians...

Let $\mathcal{X} = (v_1, \dots, v_n)$ be a sequence of data.

The median of $\mathcal{X}$ is the middle number of a set of numbers arranged in numerical order. If the number of values in a set is even, then the median is the sum of the two middle values, divided by 2.

Medians. . .

Let $\mathcal{X} = (v_1, \dots, v_n)$ be a sequence of data.

The median of $\mathcal{X}$ is the middle number of a set of numbers arranged in numerical order. If the number of values in a set is even, then the median is the sum of the two middle values, divided by 2.

Example:

$$1, 3, 3, 6, 7, 8, 9$$

Medians...

Let $\mathcal{X} = (v_1, \dots, v_n)$ be a sequence of data.

The median of $\mathcal{X}$ is the middle number of a set of numbers arranged in numerical order. If the number of values in a set is even, then the median is the sum of the two middle values, divided by 2.

Example:

$$1, 3, 3, 6, 7, 8, 9$$

Medians...

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The median of $\mathcal{X}$ is the middle number of a set of numbers arranged in numerical order. If the number of values in a set is even, then the median is the sum of the two middle values, divided by 2.

Example:

$$1, 3, 3, 6, 7, 8, 9 \qquad \Rightarrow \qquad \text{Median} = 6$$

Medians. . .

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The median of $\mathcal{X}$ is the middle number of a set of numbers arranged in numerical order. If the number of values in a set is even, then the median is the sum of the two middle values, divided by 2.

Example:

$$1, 3, 3, 6, 7, 8, 9 \qquad \Rightarrow \qquad \text{Median} = 6$$

$$1, 2, 3, 4, 5, 6, 8, 9$$

Medians...

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The median of $\mathcal{X}$ is the middle number of a set of numbers arranged in numerical order. If the number of values in a set is even, then the median is the sum of the two middle values, divided by 2.

Example:

$$1, 3, 3, 6, 7, 8, 9 \qquad \Rightarrow \qquad \text{Median} = 6$$

$$1, 2, 3, 4, 5, 6, 8, 9$$

Medians. . .

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The median of $\mathcal{X}$ is the middle number of a set of numbers arranged in numerical order. If the number of values in a set is even, then the median is the sum of the two middle values, divided by 2.

Example:

$$1, 3, 3, 6, 7, 8, 9 \qquad \Rightarrow \qquad \text{Median} = 6$$

$$1, 2, 3, 4, 5, 6, 8, 9 \qquad \Rightarrow \qquad \text{Median} = \frac{4+5}{2} = 4.5$$

Mode. . .

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

Mode. . .

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The mode is the most frequent value in the set. A set can have more than one mode; if it has two, it is said to be bimodal, or in general multimodal.

Mode...

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The mode is the most frequent value in the set. A set can have more than one mode; if it has two, it is said to be bimodal, or in general multimodal.

Example:

$$1, 1, 2, 3, 5, 8$$

Mode...

Let $\mathcal{X} = (v_1, \dots, v_n)$ be a sequence of data.

The mode is the most frequent value in the set. A set can have more than one mode; if it has two, it is said to be bimodal, or in general multimodal.

Example:

$$1, 1, 2, 3, 5, 8 \qquad \Rightarrow \qquad \text{mode is} = 1$$

Mode...

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The mode is the most frequent value in the set. A set can have more than one mode; if it has two, it is said to be bimodal, or in general multimodal.

Example:

$$1, 1, 2, 3, 5, 8 \qquad \Rightarrow \qquad \text{mode is} = 1$$

$$1, 3, 5, 7, 9, 9, 21, 25, 25, 31$$

Mode. . .

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The mode is the most frequent value in the set. A set can have more than one mode; if it has two, it is said to be bimodal, or in general multimodal.

Example:

$$1, 1, 2, 3, 5, 8 \qquad \Rightarrow \qquad \text{mode is} = 1$$

$$1, 3, 5, 7, 9, 9, 21, 25, 25, 31 \qquad \Rightarrow \qquad \text{modes are} = 9 \text{ and } 25$$

Mean...

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

Mean...

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The mean is the sum of all the values in a set, divided by the number of values. The mean of a sample $\mathcal{X}$ is usually denoted by $\overline{\mathcal{X}}$.

Mean...

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The mean is the sum of all the values in a set, divided by the number of values. The mean of a sample $\mathcal{X}$ is usually denoted by $\overline{\mathcal{X}}$.

The mean is sensitive to any change in value, unlike the median and mode, where a change to an extreme or uncommon value usually has no effect.

Mean...

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The mean is the sum of all the values in a set, divided by the number of values. The mean of a sample $\mathcal{X}$ is usually denoted by $\overline{\mathcal{X}}$.

The mean is sensitive to any change in value, unlike the median and mode, where a change to an extreme or uncommon value usually has no effect.

One disadvantage of the mean is that a small number of extreme values can distort its value:

$$1, 1, 1, 2, 2, 3, 3, 3, 200$$

Mean...

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The mean is the sum of all the values in a set, divided by the number of values. The mean of a sample $\mathcal{X}$ is usually denoted by $\overline{\mathcal{X}}$.

The mean is sensitive to any change in value, unlike the median and mode, where a change to an extreme or uncommon value usually has no effect.

One disadvantage of the mean is that a small number of extreme values can distort its value:

$$1, 1, 1, 2, 2, 3, 3, 3, 200$$

The trimmed mean, where the smallest and largest quarters of the values are removed before the mean is taken, can solve this problem.

Variability

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

Variability

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The range of $\mathcal{X}$ is the difference between the largest and smallest values of $\mathcal{X}$.

The range of a set is simple to calculate, but is not very useful because it depends on the extreme values, which may be distorted.

Variability

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The range of $\mathcal{X}$ is the difference between the largest and smallest values of $\mathcal{X}$.

The range of a set is simple to calculate, but is not very useful because it depends on the extreme values, which may be distorted.

Example:

$$1, 1, 1, 2, 2, 3, 3, 3, 200$$

Interquartile range

The Interquartile Range (IRQ) is computed as the range of the set with smallest and largest quarters removed.

Interquartile range

The Interquartile Range (IRQ) is computed as the range of the set with smallest and largest quarters removed.

Algorithm:

1. Quartiles are calculated recursively, by using median;
2. If the number of entries is an even number $2n$:
   - first quartile $Q1$ is defined as median of the n smallest entries;
   - the third quartile $Q3$ is the median of the n largest entries.
3. If the number of entries is an odd number $2n + 1$:
   - first quartile $Q1$ is defined as median of the n smallest entries;
   - the third quartile $Q3$ is the median of the n largest entries;
   - the second quartile $Q2$ is the the same as the ordinary median.

Interquartile range

Example. . .

i	x[i]	Median	Quartile
1	7		
2	7		
3	31		Q_1=31
4	31		(median of upper half, from row 1 to 6)
5	47		
6	75		
7	87	Q_2=87	
8	115	(median of whole table)	
9	116		
10	119		Q_3=119
11	119		(median of lower half, from row 8 to 13)
12	155		
13	177		

Outliers. . .

The IQR is useful for determining outliers, or extreme values such as the element 200 in the following dataset:

$$1, 1, 1, 2, 2, 3, 3, 3, 200$$

Outliers...

The IQR is useful for determining outliers, or extreme values such as the element 200 in the following dataset:

$$1, 1, 1, 2, 2, 3, 3, 3, 200$$

If $Q1$ and $Q3$ are the lower and the upper quartiles respectively, then one could define an outlier to be any observation outside the range:

$$[Q1 - k(Q3 - Q1), Q3 + k(Q3 - Q1)]$$

where k is a nonnegative constant.

Outliers...

The IQR is useful for determining outliers, or extreme values such as the element 200 in the following dataset:

$$1, 1, 1, 2, 2, 3, 3, 3, 200$$

If $Q1$ and $Q3$ are the lower and the upper quartiles respectively, then one could define an outlier to be any observation outside the range:

$$[Q1 - k(Q3 - Q1), Q3 + k(Q3 - Q1)]$$

where k is a nonnegative constant.

This method has been proposed by John Tukey and suggested $k = 1.5$ to indicate an outlier and $k = 3$ for far out.

Variance and standard deviation

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

Variance and standard deviation

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The variance s^2 of $\mathcal{X}$ is a measure of how items are dispersed about their mean. It can be calculated as:

$$s^2 = \frac{\sum(v_i - \overline{\mathcal{X}})}{n-1}$$

Variance and standard deviation

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a sequence of data.

The variance s^2 of $\mathcal{X}$ is a measure of how items are dispersed about their mean. It can be calculated as:

$$s^2 = \frac{\sum(v_i - \overline{\mathcal{X}})}{n-1}$$

The standard deviation s of $\mathcal{X}$ is the square root of the variance.

The relative variability of $\mathcal{X}$ is the standard deviation of $\mathcal{X}$ divided by its mean.

Position

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a set of data we are interested study how each v_i is positioned (or ranked) in $\mathcal{X}$.

Position

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a set of data we are interested study how each v_i is positioned (or ranked) in $\mathcal{X}$.

A simple ranking is used when an element is ranked as its position in the order.

Position

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a set of data we are interested study how each v_i is positioned (or ranked) in $\mathcal{X}$.

A simple ranking is used when an element is ranked as its position in the order.

The percentile ranking of a value v_i is the percent of values that are below it.

Position

Let $\mathcal{X} = (v_1, \ldots, v_n)$ be a set of data we are interested study how each v_i is positioned (or ranked) in $\mathcal{X}$.

A simple ranking is used when an element is ranked as its position in the order.

The percentile ranking of a value v_i is the percent of values that are below it.

The z-score of a value v_i is the number of standard deviations it is from the mean:

$$z = \frac{v_i - \overline{\mathcal{X}}}{s}$$

Position

Example

Let $\mathcal{X} = \{1.1, 2.34, 2.9, 3.14, 3.29, 3.57, 4.0\}$, we have that:

- $\overline{\mathcal{X}} = 2.91$
- $s = 0.88$

Position

Example

Let $\mathcal{X} = \{1.1, 2.34, 2.9, 3.14, 3.29, 3.57, 4.0\}$, we have that:

- $\overline{\mathcal{X}} = 2.91$
- $s = 0.88$

Let us consider value 3.57:

- Its simple ranking is 2 out of 7;
- Its percentile ranking is $\frac{5}{7} = 71,43\%$;
- Its z-score is $\frac{3.57-2.91}{0.88} = 0.75$.

Five-number summary

The five-number summary is a set of descriptive statistics that provide information about a dataset.

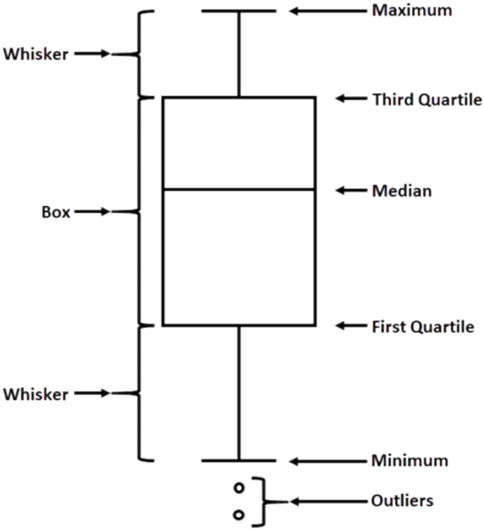
Five-number summary

The five-number summary is a set of descriptive statistics that provide information about a dataset.

It consists of the five most important sample percentiles:

- the sample minimum (smallest observation);
- the lower quartile or first quartile;
- the median (the middle value);
- the upper quartile or third quartile;
- the sample maximum (largest observation).

Box plot. . .


Maximum
Whisker
Third Quartile
Median
Box
First Quartile
Whisker
Minimum
Outliers

Convergence of Random Variables...

Let X_1, X_2,... be a sequence of random variables, and let X be another random variable. Let F_n denote the CDF of X_n and let F denote the CDF of X.

Convergence of Random Variables...

Let X_1, X_2,... be a sequence of random variables, and let X be another random variable. Let F_n denote the CDF of X_n and let F denote the CDF of X.

X_n converges to X in probability, written $X_n \xrightarrow{P} X$, if for every $\varepsilon > 0$,

$$Pr(|X_n - X| > 0) \to 0$$

as $n \to \infty$.

Convergence of Random Variables. . .

Let X_1, X_2,. . . be a sequence of random variables, and let X be another random variable. Let F_n denote the CDF of X_n and let F denote the CDF of X.

X_n converges to X in probability, written $X_n \xrightarrow{P} X$, if for every $\varepsilon > 0$,

$$Pr(|X_n - X| > 0) \to 0$$

as $n \to \infty$.

X_n converges to X in distribution, written $X_n \rightsquigarrow X$, if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} F_n(t) = F(t)$$

at all t for which F is continuous.

The Weak Law of Large Numbers. . .

Let X_1, X_2,. . . be an IID sample and let $\mu = E[X_1]$ and $\sigma^2 = Var[X_1]$ then:

$$\overline{X_n} \xrightarrow{P} \mu$$

where $\overline{X_n} = \frac{1}{n} \sum X_n$ and $Var[\overline{X_n}] = \frac{\sigma^n}{n}$.

The Weak Law of Large Numbers...

Let X_1, X_2,... be an IID sample and let $\mu = E[X_1]$ and $\sigma^2 = Var[X_1]$ then:

$$\overline{X_n} \xrightarrow{P} \mu$$

where $\overline{X_n} = \frac{1}{n} \sum X_n$ and $Var[\overline{X_n}] = \frac{\sigma^n}{n}$.

The Weak Law of Large Numbers guarantee that the distribution of $\overline{X_n}$ becomes more concentrated around μ as n gets large!

The Weak Law of Large Numbers...

Let X_1, X_2,... be an IID sample and let $\mu = E[X_1]$ and $\sigma^2 = Var[X_1]$ then:

$$\overline{X_n} \xrightarrow{P} \mu$$

where $\overline{X_n} = \frac{1}{n}\sum X_n$ and $Var[\overline{X_n}] = \frac{\sigma^n}{n}$.

The Weak Law of Large Numbers guarantee that the distribution of $\overline{X_n}$ becomes more concentrated around μ as n gets large!

X_1, X_2,... must be IID!

The Weak Law of Large Numbers...

Example

Consider flipping a coin for which the probability of *heads* is p. Let X_i denote the outcome of a single toss (0 or 1). Hence, $p = Pr(X_i = 1) = E[X_i]$.

The Weak Law of Large Numbers...

Example

Consider flipping a coin for which the probability of *heads* is p. Let X_i denote the outcome of a single toss (0 or 1). Hence, $p = Pr(X_i = 1) = E[X_i]$.

The fraction of heads after n tosses is $\overline{X_n}$. According to the WLLN $\overline{X_n}$ converges to p in probability.

The Weak Law of Large Numbers...

Example

Consider flipping a coin for which the probability of *heads* is p. Let X_i denote the outcome of a single toss (0 or 1). Hence, $p = Pr(X_i = 1) = E[X_i]$.

The fraction of heads after n tosses is $\overline{X_n}$. According to the WLLN $\overline{X_n}$ converges to p in probability.

This does not mean that $\overline{X_n}$ will numerically equal p!

The Weak Law of Large Numbers. . .
Example

Consider flipping a coin for which the probability of *heads* is p. Let X_i denote the outcome of a single toss (0 or 1). Hence, $p = Pr(X_i = 1) = E[X_i]$.

The fraction of heads after n tosses is $\overline{X_n}$. According to the WLLN $\overline{X_n}$ converges to p in probability.

This does not mean that $\overline{X_n}$ will numerically equal p!

We only know that when n is large, $\overline{X_n}$ is tightly concentrated around p.

The Weak Law of Large Numbers...

Example

Consider flipping a coin for which the probability of *heads* is p. Let X_i denote the outcome of a single toss (0 or 1). Hence, $p = Pr(X_i = 1) = E[X_i]$.

The fraction of heads after n tosses is $\overline{X_n}$. According to the WLLN $\overline{X_n}$ converges to p in probability.

This does not mean that $\overline{X_n}$ will numerically equal p!

We only know that when n is large, $\overline{X_n}$ is tightly concentrated around p.

Question: How large should be n so that

$$Pr(|\overline{X_n} - p| < 0.1) \geq \overline{p}?$$

The Weak Law of Large Numbers. . .
Example

Answer: From Chebyshev's inequality we know that:

$$Pr(|\overline{X} - p| > 0.1) \leq \frac{\sigma^2}{n \cdot (0.1)^2}$$

The Weak Law of Large Numbers. . .

Example

Answer: From Chebyshev's inequality we know that:

$$Pr(|\overline{X} - p| > 0.1) \leq \frac{\sigma^2}{n \cdot (0.1)^2}$$

Hence:

$$Pr(|\overline{X_n} - p| \leq 0.1) = 1 - Pr(|\overline{X} - p| > 0.1) \geq 1 - \frac{\sigma^2}{n \cdot (0.1)^2}$$

The Weak Law of Large Numbers...

Example

Answer: From Chebyshev's inequality we know that:

$$Pr(|\overline{X} - p| > 0.1) \leq \frac{\sigma^2}{n \cdot (0.1)^2}$$

Hence:

$$Pr(|\overline{X_n} - p| \leq 0.1) = 1 - Pr(|\overline{X} - p| > 0.1) \geq 1 - \frac{\sigma^2}{n \cdot (0.1)^2}$$

Warning: In the general case σ^2 is unknown!

The Weak Law of Large Numbers. . .

Example

Answer: From Chebyshev's inequality we know that:

$$Pr(|\overline{X} - p| > 0.1) \leq \frac{\sigma^2}{n \cdot (0.1)^2}$$

Hence:

$$Pr(|\overline{X_n} - p| \leq 0.1) = 1 - Pr(|\overline{X} - p| > 0.1) \geq 1 - \frac{\sigma^2}{n \cdot (0.1)^2}$$

Warning: In the general case σ^2 is unknown!

Solution: We can use s^2!

Central Limit Theorem

Let X_1, X_2,... be an IID sample and let $\mu = E[X_1]$ and $\sigma^2 = Var[X_1]$ then:

$$Z_n \equiv \frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma} \rightsquigarrow Z$$

where Z is distributed as $N(0, 1)$.

Central Limit Theorem

Let X_1, X_2,... be an IID sample and let $\mu = E[X_1]$ and $\sigma^2 = Var[X_1]$ then:

$$Z_n \equiv \frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma} \rightsquigarrow Z$$

where Z is distributed as $N(0, 1)$.

Probability statements about X_n can be approximated using a Normal distribution. It's the probability statements that we are approximating, not the random variable itself.

After a reasonable number of observations we can estimate how good is the average value we have computed!

Central Limit Theorem

Let X_1, X_2,... be an IID sample and let $\mu = E[X_1]$ and $\sigma^2 = Var[X_1]$ then following notations are all equivalent:

- $Z_n \approx N(0,1)$
- $\overline{X_n} \approx N(\mu, \frac{\sigma^2}{n})$
- $\overline{X_n} - \mu = N(0, \frac{\sigma^2}{n})$
- $\sqrt{n}(\overline{X_n} - \mu) = N(0, \sigma^2)$
- $\frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma} = N(0,1)$

Central Limit Theorem

Let X_1, X_2,... be an IID sample and let $\mu = E[X_1]$ and $\sigma^2 = Var[X_1]$ then following notations are all equivalent:

- $Z_n \approx N(0,1)$
- $\overline{X_n} \approx N(\mu, \frac{\sigma^2}{n})$
- $\overline{X_n} - \mu = N(0, \frac{\sigma^2}{n})$
- $\sqrt{n}(\overline{X_n} - \mu) = N(0, \sigma^2)$
- $\frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma} = N(0,1)$

Remark: When μ and σ^2 are unknown we can use their estimations!

To be continued...