

1. Probability

Data Analysis for Networks - DataNets'20
Anastasios Giovanidis

Sorbonne-LIP6



30 Septembre 2020

Bibliography

B.1 Christopher M. Bishop, "Pattern Recognition and Machine Learning", Springer 2006.

👉 Chapter 1.2

B.2 H. Pishro-Nik, "Introduction to probability, statistics, and random processes", available at <https://www.probabilitycourse.com>, Kappa Research LLC, 2014.

👉 Chapter 7.1

Basics

Distributions

Expectation

Important distributions: Discrete

Important distributions: Continuous

Limit Theorems

Uncertainty

The key concept we are dealing with is uncertainty.

Uncertainty arises through:

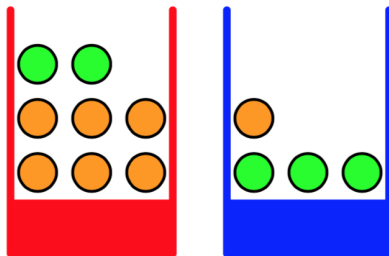
- ▶ noise on measurements
- ▶ finite size of data-sets

Probability theory provides a consistent framework for the quantification and manipulation of uncertainty.

Boxes, apples and oranges

A. Giovanidis 2020

Consider two boxes (red and blue), that contain apples and oranges.



☞ Pick one box at random, and then pick one fruit, note the result and put it back.

Event Probability

Probability of an event is the fraction of times that the event occurs out of the total number of trials, in the limit that the number of trials goes to infinity.

☞ This is the **frequentist** definition, in contrast to the **axiomatic** one.

Let us suppose for the sake of example:

- ▶ We pick the **red box** 40% of times and
- ▶ the **blue box** 60% of times.

Event Probability

We write:

$$P(B = r) = \frac{4}{10}$$

$$P(B = b) = \frac{6}{10}$$

- ▶ Each probability must lie in the interval $[0, 1]$.
- ▶ If the events are **mutually exclusive** and if they include all possible outcomes (e.g. **red**, or **blue** box) then these probabilities must sum to 1.

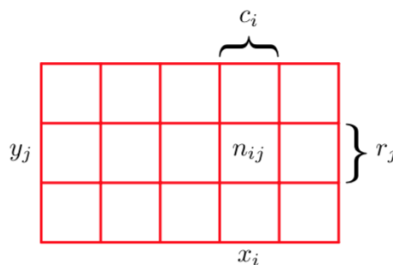
Questions

A. Giovanidis 2020

- ▶ “What is the overall probability that the selection procedure will pick an apple?”
- ▶ “Given that we have chosen an orange, what is the probability that the box chosen is the blue one?”

More general framework

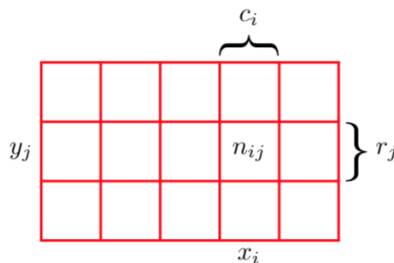
A. Giovanidis 2020



- ▶ Two random variables, X and Y .
- ▶ X takes any value x_i , $i = 1, \dots, M$
- ▶ Y takes any value y_j , $j = 1, \dots, L$.
- ▶ We sample both r.v.'s N times each.

Joint probability

A. Giovanidis 2020



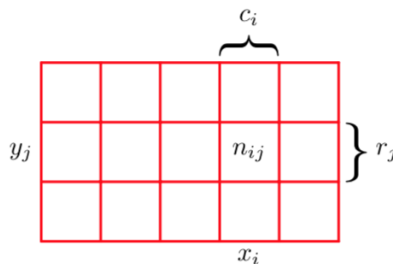
- ▶ The number of times we find $(X = x_i, Y = y_j)$ is n_{ij} .
- ▶ The joint probability reads

$$P(X = x_i, Y = y_j) = \frac{n_{ij}}{N}.$$

(we implicitly consider the limit $N \rightarrow \infty$)

Marginal probability

A. Giovanidis 2020



- ▶ The number of times we find $X = x_i$ (for any Y) is c_i .
- ▶ The number of times we find $Y = y_j$ (for any X) is r_j .
- ▶ The probability for $X = x_i$ reads (**marginal probability**)

$$P(X = x_i) = \frac{c_i}{N} = \frac{1}{N} \sum_{j=1}^L n_{ij} = \sum_{j=1}^L P(X = x_i, Y = y_j).$$

Conditional probability

By considering only those instances for which $X = x_i$:

- ▶ The fraction of instances for which $Y = y_j$ is

$$P(Y = y_j \mid X = x_i) = \frac{n_{ij}}{c_i}.$$

- ▶ We can derive the following relation

$$\begin{aligned} P(X = x_i, Y = y_j) &= \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} \\ &= P(Y = y_j \mid X = x_i) \cdot P(X = x_i) \end{aligned}$$

- ☞ We denote by $P(X)$ the **distribution** over the random variable X .

The probability rules

sum rule $P(X) = \sum_Y P(X, Y)$

product rule $P(X, Y) = P(Y | X)P(X)$

The probability rules

sum rule $P(X) = \sum_Y P(X, Y)$

product rule $P(X, Y) = P(Y | X)P(X)$

☞ By symmetry we get

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

which gives

Bayes' theorem $P(Y | X) = \frac{P(X | Y)P(Y)}{P(X)}.$

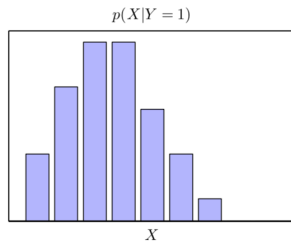
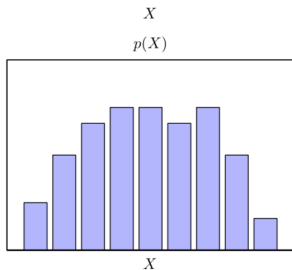
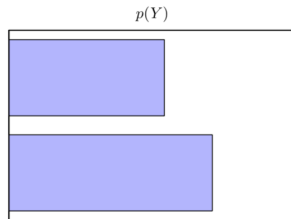
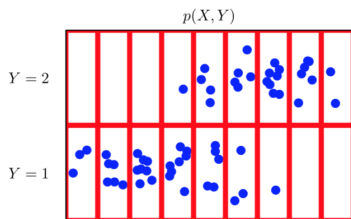
Back to apples and oranges*

Given we picked an orange, what is the probability that it came from the **red** box?

$$\begin{aligned}P(B = \text{red} \mid F = \text{orange}) &= \frac{P(F = \text{orange} \mid B = \text{red}) \cdot P(B = \text{red})}{P(F = \text{orange})} \\&= \frac{\frac{6}{8} \cdot \frac{4}{10}}{\frac{6}{8} \cdot \frac{4}{10} + \frac{1}{4} \cdot \frac{6}{10}} \\&= \frac{24/80}{36/80} \\&= \frac{2}{3}.\end{aligned}$$

Distribution over 2 variables

A. Giovanidis 2020



Elementary properties of probability*

Consider events A, B as ensembles of realisations for the r.v. X , e.g.

$A = \{x_1, x_4, x_{M-1}\}$, and $B = \{x_2\}$.

1. **Complementarity rule** $P(A^c) = 1 - P(A)$,
where A^c is the complementary event of A , i.e. $A^c = \Omega/A$.
2. **Empty set** $P(\emptyset) = 0$.
3. **Inclusion property**: $B \subseteq A \Rightarrow P(B) \leq P(A)$ and
 $P(A - B) = P(A) - P(B)$.
4. **Additivity**: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
5. **Exclusivity**: $P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$, if the events A_k are mutually exclusive, i.e. $A_i \cap A_j = \emptyset$, for $i \neq j$.

The last property is due to the fact that for two **exclusive events**:

$$P(A_i \cap A_j) = P(\emptyset) = 0.$$

Conditional Probability and Independence

Two r.v.'s are **independent** when the probability of the one does not depend on the realization of the other

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

so that conditioning does not add information. This definition has the following consequence. From the **product rule**

$$\begin{aligned} P(X, Y) &= P(X|Y) \cdot P(Y) \\ &\stackrel{\text{indep.}}{=} P(X) \cdot P(Y) \end{aligned}$$

☞ In our example, $P(X|Y) = P(X)$, when $\frac{n_{ij}}{r_j} = \frac{c_i}{N}$.

Basics

Distributions

Expectation

Important distributions: Discrete

Important distributions: Continuous

Limit Theorems

Mass and Density

- **Discrete** r.v. \rightarrow probability mass function (p.m.f)

$$p_i := P(X = x_i), \quad \text{for } i = 1, \dots, M$$

$$p_i \geq 0, \quad \text{and} \quad \sum_{i=1}^M p_i = 1.$$

Mass and Density

- **Discrete** r.v. \rightarrow probability mass function (p.m.f)

$$p_i := P(X = x_i), \quad \text{for } i = 1, \dots, M$$

$$p_i \geq 0, \quad \text{and} \quad \sum_{i=1}^M p_i = 1.$$

- **Continuous** r.v. \rightarrow probability density function (p.d.f)

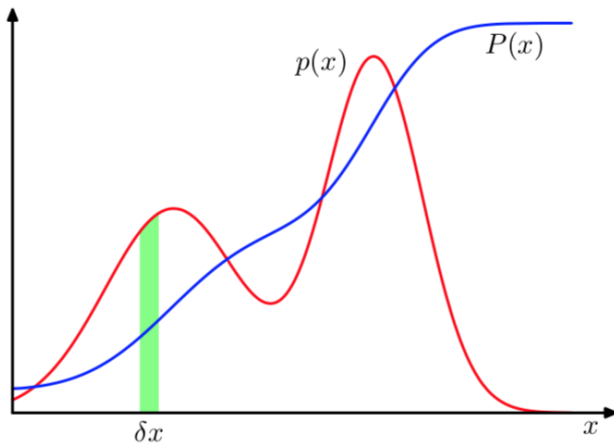
$$p(x)\delta x \quad \text{for } (x, x + \delta x)$$

and when $\delta x \rightarrow 0$ then $p(x)$ is the **probability density over x** .

$$p(x) \geq 0, \quad \text{and} \quad \int_0^{\infty} p(x)dx = 1.$$

PDF

A. Giovanidis 2020



PDF and CDF

☞ Using the PDF we can derive the probability that the r.v. lies within a certain interval

$$P(X \in (a, b)) = \int_a^b p(x) dx$$

The probability that x lies in the interval $(-\infty, z)$ is given by the **cumulative distribution function** defined by

$$P(z) = \int_{-\infty}^z p(x) dx$$

which satisfies $dP(z)/dz|_{z=x} = p(x)$.

Sum and product rule

These rules apply also in the case of continuous r.v.'s for the PDF.

Sum Rule $p(x) = \int p(x, y) dy$

Product Rule $p(x, y) = p(y|x)p(x).$

Bayes' Theorem $p(y|x) = \frac{p(x|y)p(y)}{p(x)}$

Change of variables*

Suppose we know the pdf $p_X(x)$ and we would like to find the pdf of a function of X , say $Y = f(X)$. So, we are looking for the pdf

$$p_X(x) \rightarrow p_Y(y)$$

Practically, since we know $p_X(x)$ we will need the inverse of f

$$X = f^{-1}(Y) := g(Y)$$

BUT WE CANNOT JUST WRITE $p_X(g(y))$!!

☞ Intuition: for small δx , δy the range $(x + \delta x)$ transforms to $(y + \delta y)$ and it holds $p_X(x)\delta x \approx p_Y(y)\delta y$, or

$$p_Y(y) = p_X(x) \frac{\delta x}{\delta y}.$$

Change of variables cont'd*

$$\begin{aligned} p_Y(y) &= p_X(x) \left| \frac{dx}{dy} \right| \\ &= p_X(g(y)) |g'(y)|. \end{aligned}$$

Basics

Distributions

Expectation

Important distributions: Discrete

Important distributions: Continuous

Limit Theorems

Expectation

The expectation is the weighted average value of some function of X , say $f(X)$ under a distribution $p(X)$ and is defined as

$$\text{Discrete} \quad \mathbb{E}[f] = \sum_x p(x)f(x)$$

$$\text{Continuous} \quad \mathbb{E}[f] = \int p(x)f(x)dx.$$

Note that

$$\mathbb{E}_x[f(X, Y)] = g(Y)$$

$$\mathbb{E}_X[f(X, Y) \mid Y] = \sum_x p(x|Y)f(x)$$

for discrete variables, and using integrals for continuous.

Estimating Expectation from Samples

Suppose we are given N samples drawn from the $p(X)$, then the expectation can be approximated as a finite sum over these points:

$$\mathbb{E}[f] \approx \frac{1}{N} \sum_{n=1}^N f(x_n).$$

When $N \rightarrow \infty$ this converges to the true expectation, for both continuous and discrete variables.

Variance

The variance is a measure of variability of $f(X)$ around its mean $\mathbb{E}[f(X)]$.

Definition

$$\begin{aligned}\text{Var}(f(X)) &= \mathbb{E}[(f(X) - \mathbb{E}[f(X)])^2] \\ &= \mathbb{E}[f(X)^2] - \mathbb{E}[f(X)]^2.\end{aligned}$$

The variance of the variable X itself, is given as

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Covariance

The **covariance** expresses the extent to which X and Y vary together.

Definition

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}_{X,Y}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}_{X,Y}[XY] - \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

Note here that, when X, Y are **independent** r.v.'s, $P(X, Y) = P(X)P(Y)$, and hence

$$\text{Cov}(X, Y) \stackrel{\text{indep.}}{=} 0.$$

Basics

Distributions

Expectation

Important distributions: Discrete

Important distributions: Continuous

Limit Theorems

Bernoulli(p)

The result of an experiment is either positive or negative

$$X = \begin{cases} 0, & \text{with probability } 1 - p \\ 1, & \text{with probability } p \end{cases}.$$

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1 - p) = p$$

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 1^2 \cdot p - p^2 = p(1 - p).$$

Binomial(N, p)

We repeat the same experiment N times, where the result of each experiment is an independent Bernoulli(p) distribution. Here, X counts the number of times among the N efforts that the experiment was successful.

$$P(X = k) = \binom{N}{k} p^k (1 - p)^{N-k}, \quad k = 0, \dots, N$$

where,

$$\binom{N}{k} = \frac{N!}{k!(N-k)!}.$$

counts the different combinations of k successes among the N efforts.
From the [binomial formula](#)

$$\sum_{k=0}^N P(X = k) = (p + (1 - p))^N = 1.$$

Binomial(N, p) cont'd

A. Giovanidis 2020

$$\mathbb{E}(X) = \sum_{k=0}^N k \binom{N}{k} p^k (1-p)^{N-k} = N \cdot p$$

$$\text{Var}(X) = Np(1-p).$$

Geometric(p)*

Repeat independent Bernoulli(p) trials until observing the first success.
Here, X is the number of efforts until success.

$$P(X = k) = (1 - p)^{k-1}p, \quad k = 1, 2, \dots$$

The mean and variance are equal to

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p = \frac{1}{p}, \\ \text{Var}(X) &= \frac{1-p}{p^2}.\end{aligned}$$

Applications: ARQ retransmission, Wifi ALOHA, CSMA protocols, etc.

Poisson(λ)*

Here, X is the number of efforts until success.

$$P(X = k) = \exp(-\lambda) \frac{\lambda^k}{k!} \quad k = 0, 1, \dots$$

The mean and variance are equal to

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=1}^{\infty} k \exp(-\lambda) \frac{\lambda^k}{k!} = \lambda, \\ \text{Var}(X) &= \lambda.\end{aligned}$$

Applications: models telephone calls, or packets within some time-interval.

☞ The Poisson distribution is the limit of a Binomial(N, p), as $N \rightarrow \infty$ and $p \rightarrow 0$, with $Np = \lambda$ fixed, for all pairs of values.

Basics

Distributions

Expectation

Important distributions: Discrete

Important distributions: Continuous

Limit Theorems

Uniform(a, b)

A. Giovanidis 2020

The Uniform r.v. X has a PDF which takes a constant value in a predefined interval, else 0. Hence the probability to find X in a small continuous subset of the interval, is proportional to the length of the subset,

$$p_X(x) = \begin{cases} \frac{1}{b-a}, & \text{for } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}.$$

It is indeed a PDF because $1/(b-a) \geq 0$ and

$$\int_{-\infty}^{+\infty} \frac{1}{b-a} dx = \int_a^b \frac{1}{b-a} dx = 1.$$

and $\mathbb{E}[X] = \frac{a+b}{2}$, and $\text{Var}[X] = \frac{(b-a)^2}{12}$.

Exponential(λ)

A continuous random variable X is said to have an exponential distribution with parameter $\lambda > 0$, shown as $X \sim \text{Exponential}(\lambda)$, if its PDF is given by

$$p_X(x) = \begin{cases} \lambda \exp(-\lambda x), & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

The mean and variance are equal to

$$\mathbb{E}[X] = \lambda^{-1} \quad \text{Var}(X) = \lambda^{-2}.$$

Applications: Inter-arrival time between consecutive telephone calls, or packets. ***Memoryless Property!**

☞ The $\text{Poisson}(\lambda)$ distribution in 1-D has the property that the distance between any two consecutive points is $\text{Exp}(\lambda)$ distributed.

Standard Normal(0, 1)

A continuous random variable Z is said to be a **standard normal** (**standard Gaussian**) r.v., shown as $Z \sim \mathcal{N}(0, 1)$, if its PDF is given by

$$p_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right).$$

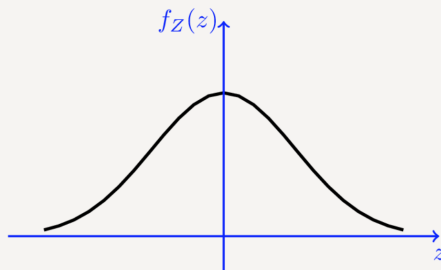


Fig.4.6 - PDF of the standard normal random variable.

Standard Normal(0, 1) cont'd

A continuous random variable Z is said to be a **standard normal** (**standard Gaussian**) r.v., shown as $Z \sim \mathcal{N}(0, 1)$, if its PDF is given by

$$p_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right).$$

- ▶ The $1/\sqrt{2\pi}$ guarantees that the area under the PDF is equal to 1.
- ▶ It holds $\mathbb{E}[Z] = 0$ and $\text{Var}(Z) = 1$.

The CDF of the standard normal distribution is denoted by the Φ function:

$$\Phi(x) = P(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{u^2}{2}\right) du.$$

Standard Normal(0, 1) ccont'd

☞ Because of the importance of the normal distribution, the values of the Φ function have been tabulated and many calculators and software packages have this function.

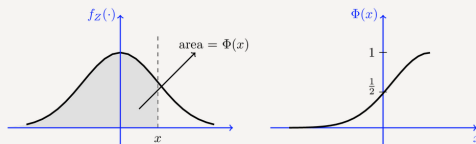


Fig.4.7 - The Φ function (CDF of standard normal).

Here are some properties of the Φ function that can be shown from its definition.

1. $\lim_{x \rightarrow \infty} \Phi(x) = 1$, $\lim_{x \rightarrow -\infty} \Phi(x) = 0$;
2. $\Phi(0) = \frac{1}{2}$;
3. $\Phi(-x) = 1 - \Phi(x)$, for all $x \in \mathbb{R}$.

Normal(μ, σ^2)

A. Giovanidis 2020

We can obtain any normal random variable X by **shifting and scaling** a standard normal random variable Z .

$$X = \sigma Z + \mu,$$

where $\sigma > 0$. Then

$$\mathbb{E}[X] = \sigma \mathbb{E}[Z] + \mu = \mu$$

$$\text{Var}(X) = \sigma^2 \text{Var}(Z) = \sigma^2.$$

Then X is a normal random variable with mean μ and variance σ^2 .
We write $X \sim (\mu, \sigma^2)$.

Normal(μ, σ^2) cont'd

A. Giovanidis 2020

Conversely, if $X \sim \mathcal{N}(\mu, \sigma^2)$, then

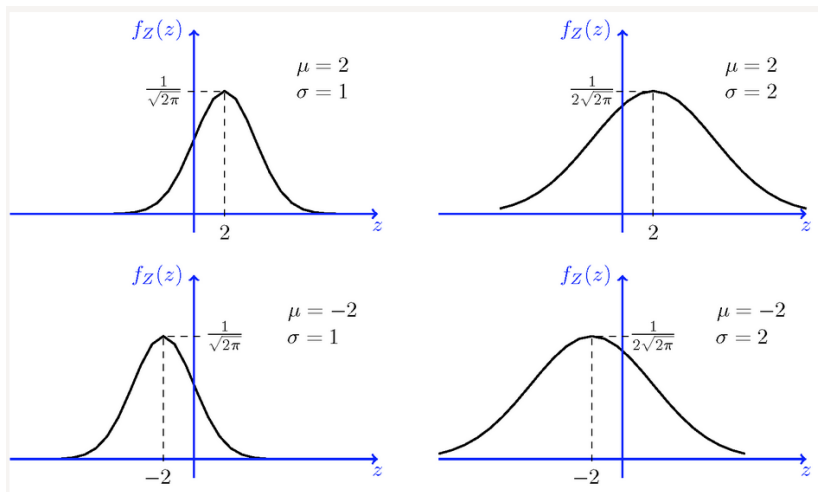
$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$F_X(x) = P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

$$P(a < X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Normal(μ, σ^2) Examples

A. Giovanidis 2020



Addition of 2 Normal(μ_k, σ_k^2) r.v.'s*

Suppose we would like to find the distribution of the sum

$$Y = X_1 + X_2,$$

where each $X_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$, and they are **independent**. Then,

$$Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

If they are not independent, but they are **jointly normal** then

$$Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho(X_1, X_2)\sigma_1\sigma_2).$$

where

$$\rho(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} = \frac{\text{Cov}(X_1, X_2)}{\sigma_1\sigma_2}.$$

Basics

Distributions

Expectation

Important distributions: Discrete

Important distributions: Continuous

Limit Theorems

1st important limit Theorem

- ▶ Law of Large Numbers (LLN),

LLN: the average of a large number of i.i.d. random variables converges to the expected value.

Sample mean

Let X_1, X_2, \dots, X_n be i.i.d. random variables.

Sample Mean:

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}.$$

- ▶ The sample mean is a random variable.
- ▶ $\mathbb{E}[\bar{X}] = \mathbb{E}[X_1]$.
- ▶ $\text{Var}(\bar{X}) = \frac{\text{Var}(X_1)}{n}$.

Law of Large Numbers

☞ If you repeat an experiment independently a large number of times and average the result, what you obtain should be close to the expected value.

Let X_1, X_2, \dots, X_n be i.i.d. random variables with a finite expected value $E[X_i] = \mu < \infty$.

Weak Law of Large Numbers:

For any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|\bar{X} - \mu| \geq \epsilon) = 0$$

○ Convergence *in probability* (weak).

A Proof for WLLN

We assume for the samples X_1, \dots, X_n finite mean and variance:

$$\begin{aligned} P(|\bar{X} - \mu| \geq \epsilon) &\leq \frac{\text{Var}(\bar{X})}{\epsilon^2} \quad (\text{Chebyshev}) \\ &= \frac{\text{Var}(X_1)}{n\epsilon^2} \\ &\rightarrow 0 \end{aligned}$$

2nd important limit Theorem

- ▶ Central Limit Theorem (CLT).

CLT: the sum of a large number of random variables has an approximately normal distribution (under some conditions).

Normalisation

Suppose X_1, \dots, X_n are i.i.d. random variables with expected values $\mathbb{E}[X_i] = \mu < \infty$ and variance $\text{Var}(X_i) = \sigma^2 < \infty$.

- ▶ Sample mean has $\mathbb{E}[\bar{X}] = \mu$ and $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$.
- ▶ Thus, the **normalised** random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

has mean $\mathbb{E}[Z_n] = 0$ and variance $\text{Var}(Z_n) = 1$.

Remark: If r.v. Y has mean $M = 0$ and variance S^2 , then with high probability all values fall within the interval $[0 - kS, 0 + kS]$. The new variable Y/S will have mean $M' = 0$ and variance $S' = 1$, and a range of values $[0 - k, 0 + k]$ with high probability. It is a **shrunked** version of Y .

CLT

The Central Limit Theorem (CLT)

Under the above assumptions (i.i.d. samples, finite mean, variance) the random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

converges in distribution to the standard normal random variable as n goes to infinity, that is

$$\lim_{n \rightarrow \infty} P(Z_n \leq x) = \Phi(x), \text{ for all } x \in \mathbb{R},$$

where $\Phi(x)$ is the standard normal CDF.

CLT for all distributions!

The CLT does not care what the distribution of the X_i 's is!
The X_i 's can be discrete, continuous, or mixed r.v.'s.

☞ Example 1: X_i 's are Bernoulli(p) (discrete with PMF).

- ▶ $\mathbb{E}[X_i] = p$, and $\text{Var}(X_i) = p(1 - p)$.
- ▶ $Y_n = X_1 + X_2 + \dots + X_n$ has Binomial(n, p) distribution.
- ▶ Normalisation

$$Z_n = \frac{Y_n - np}{\sqrt{np(1 - p)}},$$

where $Y_n \sim \text{Binomial}(n, p)$.

CLT for Bernoulli

Assumptions:

- X_1, X_2, \dots are iid $\text{Bernoulli}(p)$.
- $Z_n = \frac{X_1 + X_2 + \dots + X_n - np}{\sqrt{np(1-p)}}$.

We choose $p = \frac{1}{3}$.

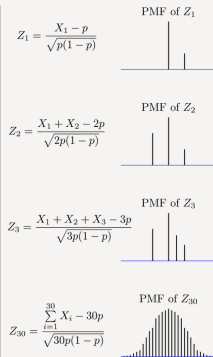


Fig.7.1 - Z_n is the normalized sum of n independent $\text{Bernoulli}(p)$ random variables. The shape of its PMF, $P_{Z_n}(z)$, resembles the normal curve as n increases.

Another example

☞ Example 2: X_i 's are Uniform(0, 1) (continuous with PDF).

- ▶ $\mathbb{E}[X_i] = 1/2$, and $\text{Var}(X_i) = 1/12$.
- ▶ Normalisation

$$Z_n = \frac{X_1 + \dots + X_n - n/2}{\sqrt{n/12}}.$$

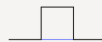
CLT for Uniform

Assumptions:

- X_1, X_2, \dots are iid $\text{Uniform}(0,1)$.
- $Z_n = \frac{X_1 + X_2 + \dots + X_n - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$.

$$Z_1 = \frac{X_1 - \frac{1}{2}}{\sqrt{\frac{1}{12}}}$$

PDF of Z_1



$$Z_2 = \frac{X_1 + X_2 - 1}{\sqrt{\frac{2}{12}}}$$

PDF of Z_2



$$Z_3 = \frac{X_1 + X_2 + X_3 - \frac{3}{2}}{\sqrt{\frac{3}{12}}}$$

PDF of Z_3



$$Z_{30} = \frac{\sum_{i=1}^{30} X_i - \frac{30}{2}}{\sqrt{\frac{30}{12}}}$$

PDF of Z_{30}

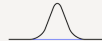


Fig. 7.2 - Z_n is the normalized sum of n independent $\text{Uniform}(0, 1)$ random variables. The shape of its PDF, $f_{Z_n}(z)$, gets closer to the normal curve as n increases.

CLT application 1

□ **Application 1:** A data-centre serves requests of service that arrive over time. Suppose the service time X_i for request i has mean $\mathbb{E}[X_i] = 1$ [msec] and variance $\text{Var}[X_i] = 1$ [msec] (e.g. we can consider exponential service time).

Let Y be the total time necessary to serve $n = 50$ requests.
What is the probability that this time ranges between $40 \leq Y \leq 60$ [msec]?

Solve application 1

Solution: We write the sum

$$Y_{50} = X_1 + \dots + X_{50}$$

for the total time needed to serve 50 requests. This is a r.v..
We are looking for:

$$P(40 \leq Y_{50} \leq 60) = P(40 \leq X_1 + \dots + X_{50} \leq 60)$$

The mean of Y is $50\mathbb{E}X_i = 50$ [msec] and the variance is $50\text{Var}[X_i] = 50$ [msec].

From the CLT, the normalised variable follows the Normal distribution

$$\frac{Y_{50} - 50\mathbb{E}X_i}{\sqrt{50\text{Var}(X_i)}} = \frac{Y_{50} - 50}{5\sqrt{2}} \sim \mathcal{N}(0, 1).$$

Solve application 1 cont'd

Hence, we can use the Normal CDF

$$\begin{aligned} P\left(\frac{40-50}{5\sqrt{2}} \leq \frac{Y_{50}-50}{5\sqrt{2}} \leq \frac{60-50}{5\sqrt{2}}\right) &= \\ P\left(-\sqrt{2} \leq \frac{Y_{50}-50}{5\sqrt{2}} \leq \sqrt{2}\right) &= \\ \Phi(\sqrt{2}) - \Phi(-\sqrt{2}) &= 0.8427. \end{aligned}$$

We conclude that the service period is between 40 and 60 [msec] with probability 84.27%.

CLT application 2

□ **Application 2:** A data packet consists of $n = 1000$ bits. Due to noise, bit errors occur independently, with probability $p = 0.1$.

What is the probability to have more that 120 errors within the same packet?

Solve application 2

Solution: We write the sum

$$Y_{1000} = X_1 + \dots + X_{1000}$$

where X_1 takes 1 if an error occurs, otherwise 0. The Y_{1000} is a r.v..
We are looking for:

$$P(Y_{1000} > 120)$$

We know that $X_i \sim \text{Bernoulli}(p = 0.1)$. The mean of Y is $1000p = 100$ [errors] and the variance is $1000p(1 - p) = 90$ [errors].

CLT: the normalised variable **approximately** follows the Normal distribution

$$\frac{Y_{1000} - 100}{\sqrt{90}} \sim \mathcal{N}(0, 1).$$

Solve application 2 cont'd

Hence, we can use the Normal CDF

$$\begin{aligned} P\left(\frac{Y_{1000} - 100}{\sqrt{90}} > \frac{120 - 100}{\sqrt{90}}\right) &= \\ 1 - \Phi\left(\frac{20}{\sqrt{90}}\right) &= 0.0175. \end{aligned}$$

We conclude that the probability to exceed 120 errors per packet is **approximately** 1.75%.

END