Statistical tools for Astronomers

PHYS-788

Module 1: Introduction

Simone Paradiso

Module 1: Introduction

Overview

- Part I:
  - Quick recollection of statistical concepts: random variables, probability distributions, moments of a distribution.
  - Examples of probability distributions.
- Part II:
  - Normal distribution and χ^2 distribution.
  - Quick recollection of the concept of conditional probability and marginal probability. Bayes theorem.
  - Bayesian Approach.
  - Frequentist Approach.

Part I

“The practice or science of collecting and analyzing numerical data in large quantities, especially for the purpose of inferring proportions in a whole from those in a representative sample”

Oxford Dictionary

Random Variables

Definition:

A Random Variable (RV) is a mathematical formalization of quantity which depends on random events.

Mathematically, a RV X is a is a measurable function $X : \Omega \rightarrow E$ from a sample space Ω as a set of possible outcomes to a measurable space E. The probability that X takes on a value in a measurable set $S \subseteq E$ is written as:

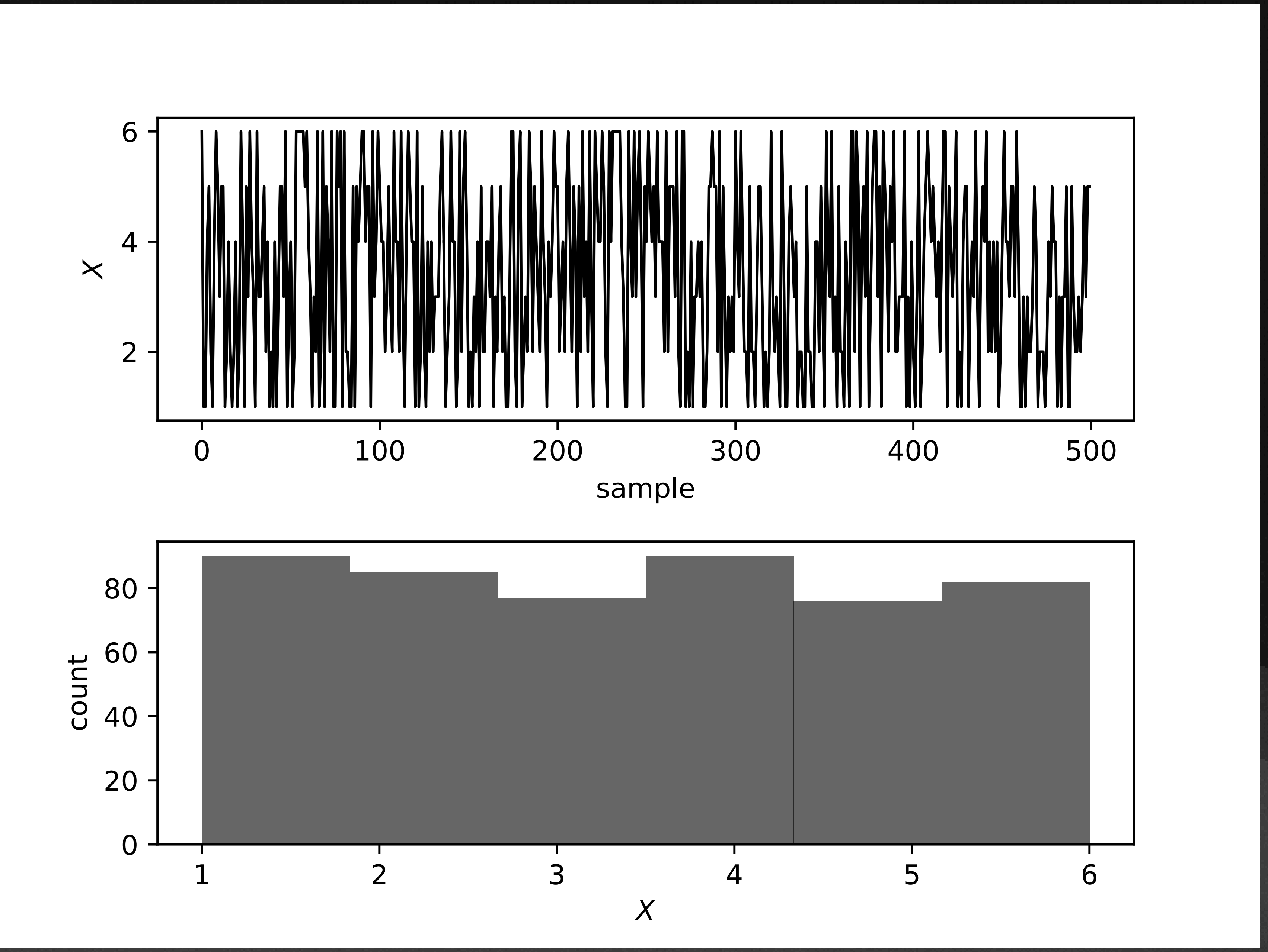
$$P(X \in S) = P(\{\omega \in \Omega \,|\, X(\omega) \in S\})$$

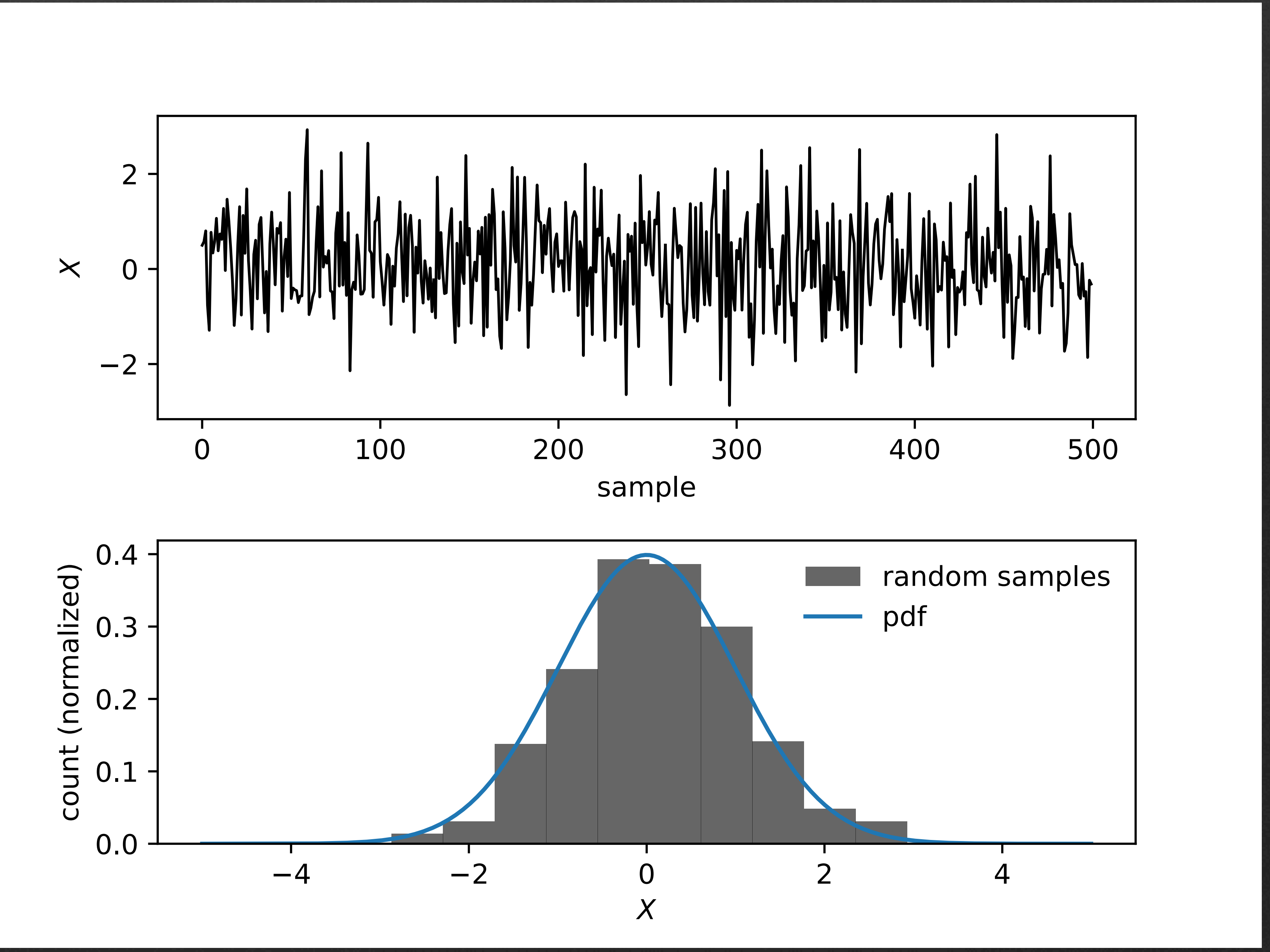
In many cases, X is Real valued, i.e. $E = \mathbb{R}$.

Random Variables

- When the range of X is finitely or infinitely countable, X is called a **discrete random variable**, and its distribution is a **discrete probability distribution**. (E.G. the roll of a dice)

- If the range of X is uncountably infinite (usually an interval), then X is called a **continuous random variable**. In the special case that it is **absolutely continuous** (continuous and differentiable), then its distribution can be generalized to a **probability density function**, which assigns probability to intervals; in particular, each individual point must have zero probability for an absolutely continuous random variable.

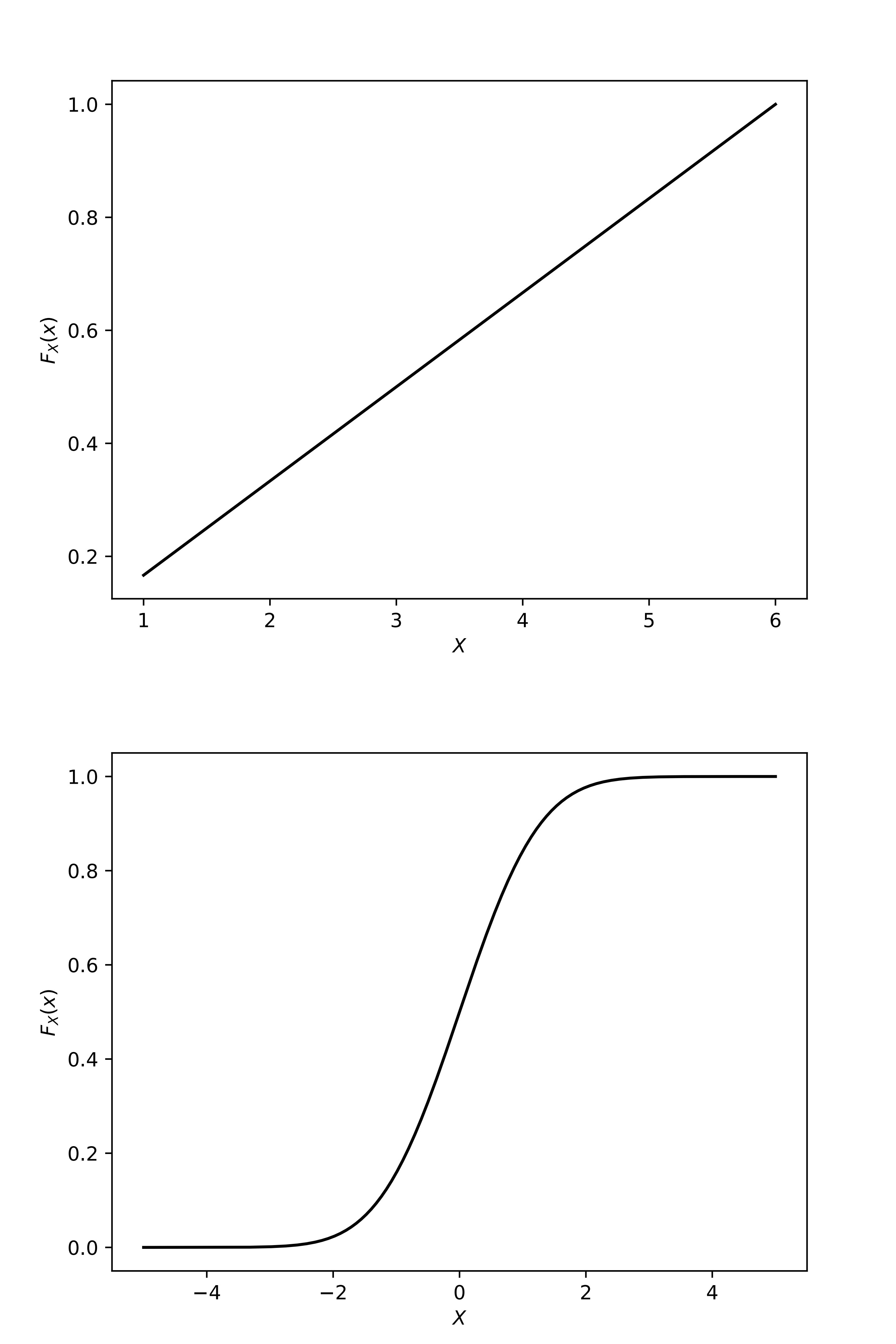




Random Variables

Any random variable X can be described by its cumulative distribution function, which describes the probability that X will be less or equal than a certain value:

$$F_X(x) = P(X \leq x)$$



Moments of a probability distribution

We can characterize the probability distribution of a random variable by a small number of parameters, which also have a practical interpretation.

- We often refer to the "average value". This is captured by the expected value of a random variable, $E[X]$, also called the first moment. In general this is non linear, i.e. $E[f(X)] \neq f(E[X])$;
- Once we know the "average value", one might be interested in how far from this value, the values of X typically are. This is given by the variance and the standard deviation of a random variable.

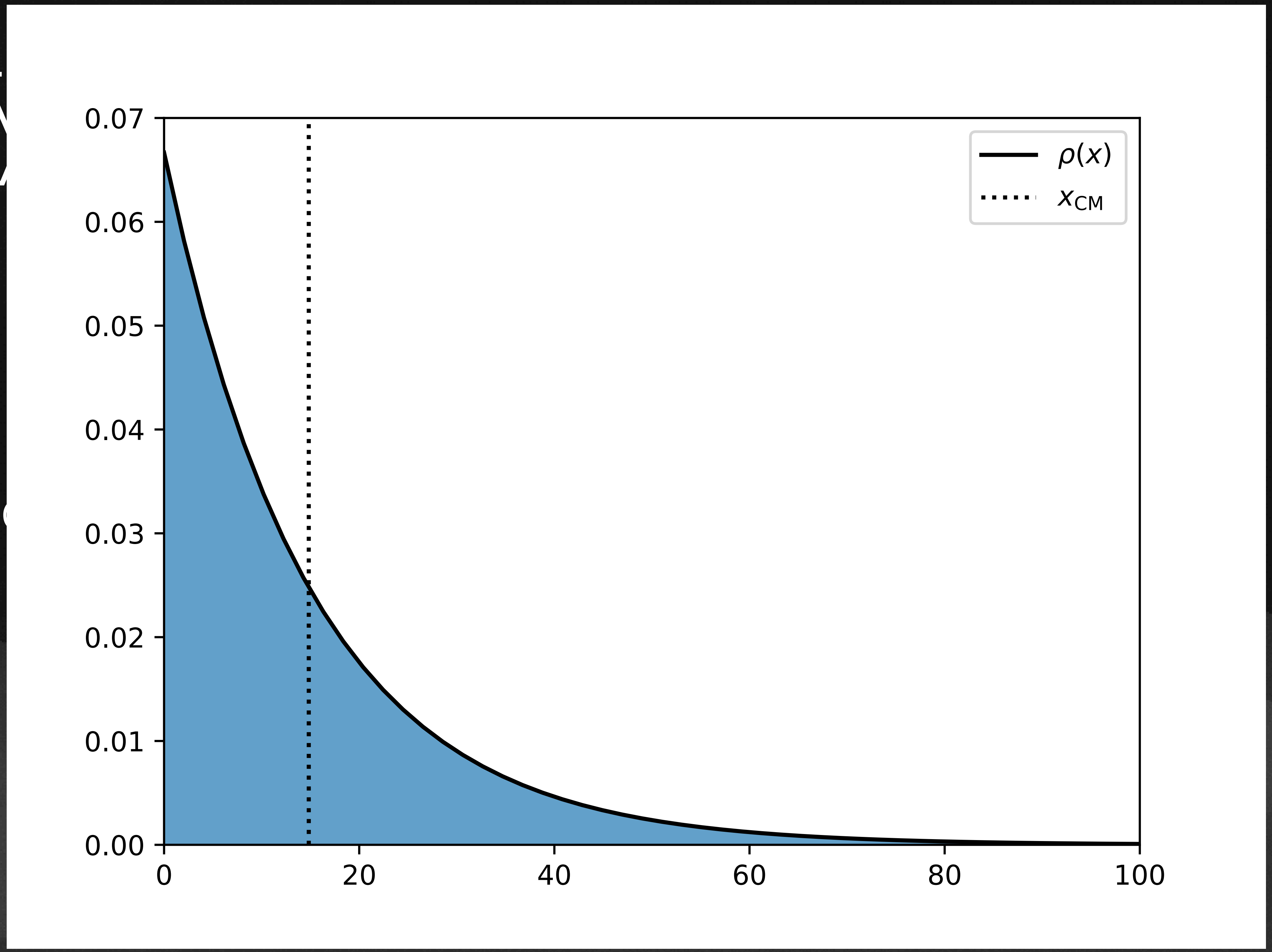
Moments of a probability distribution

Mathematically, the **moments** of a distribution are measures related to the functions graph. If the function represents mass density ρ, then:

- The 0th moment represents the total mass, $M(R) = \int_0^R \rho(r)dr$
- the 1st moment represents the center of mass, $x_{\text{CM}} = M^{-1} \cdot \int_0^R r\,\rho(r)dr$, and
- The 2nd moment represents the moment of inertia, $I = \int_0^R r^2 \rho(r)dr$.

Moments of a probability

Mathematically, the **moments** of a distribution are measures relate
represents mass density ρ, then:



- The 0th moment represents the total mass, $M(R) = \int_0^R \rho(r)dr$
- the 1st moment represents the center of mass, $x_{\mathrm{CM}} = M^{-1} \cdot \int_0^R r\,\rho(r)dr$, and
- The 2nd moment represents the moment of inertia, $I = \int_0^R r^2\rho(r)dr$.

Moments of a probability distribution

If we are dealing with a probability distribution, then:

- the 1st moment is the **expected value**, and
- the 2nd central moment represents the **variance**;

Moreover, the 3rd standardized moment is the skewness and the 4th standardized moment is the kurtosis. Notice the analogy with moments in physics.

For a probability distribution on a bounded interval, the collection of all the moments (from 0th order to infinite) uniquely determines the distribution.

Moments of a probability distribution

The n-th raw moment (moment about 0) of a distribution is defined by:

$$\mu' = \langle x^n \rangle$$

Where:

$$\langle f(x) \rangle = \begin{cases} \sum f(x)P(x) & \text{discrete} \\ \int f(x)P(x)dx & \text{continuous} \end{cases}$$

The real moment of a continuous distribution about a value c is:

$$\mu_n = \int (x-c)^n f(x)dx$$

Moments of a probability distribution

If we have a cumulative probability distribution function $F(x)$, then the n-th order moment of the probability distribution $f(x)$ is given by the Riemann-Stieltjes integral:

$$\mu_n' = E[X^n] = \int_{-\infty}^{+\infty} x^n dF(x)$$

If $E[\,|X^n|\,] = \infty$ the n-th order moment does not exist, and so it is for all the lower order moments. The 0th order moments must be always 1 since it represents the area under the probability function.

The standardized central n-th moments are defined as the central n-th moments divided by σ^n. These are dimensionless quantities and are independent of any linear change of scale.

Moments of a probability distribution

- The 1st raw order moment is the **mean**: $\mu = E[X]$
- The 2nd central order moment is the **variance**: $\sigma = E[(x-\mu)^2]$. The positive square root of the variance is the **standard deviation**.
- The 3rd standardized central moment is the **skewness**. This quantity is 0 for any symmetric distribution, is >0 for a right skewed distribution (tail longer on the right), and is <0 for a left skewed distribution.
- The 4th standardized central moment is the **kurtosis**. For the normal distribution this is $3\sigma^4$. It measures the “fatness” of the tails. If a distribution has heavier tails the kurtosis will be higher.

Estimators

Since we have **observed data**, distributed according to some probability distribution (discrete or continuous), we define the **estimator** as a rule for calculating an estimate of a given quantity —based on observed data.

- The error of an estimator $\hat{\theta}$ is: $e(x) = \hat{\theta} - \theta$.
- The mean squared error MSE of $\hat{\theta}$ is then: $MSE(\hat{\theta}) = E[(\hat{\theta}(X) - \theta)^2]$

The latter indicates how far, on average, the estimator is far from the true value. In other word, the MSE represents the precision of an estimator.

Estimators: bias

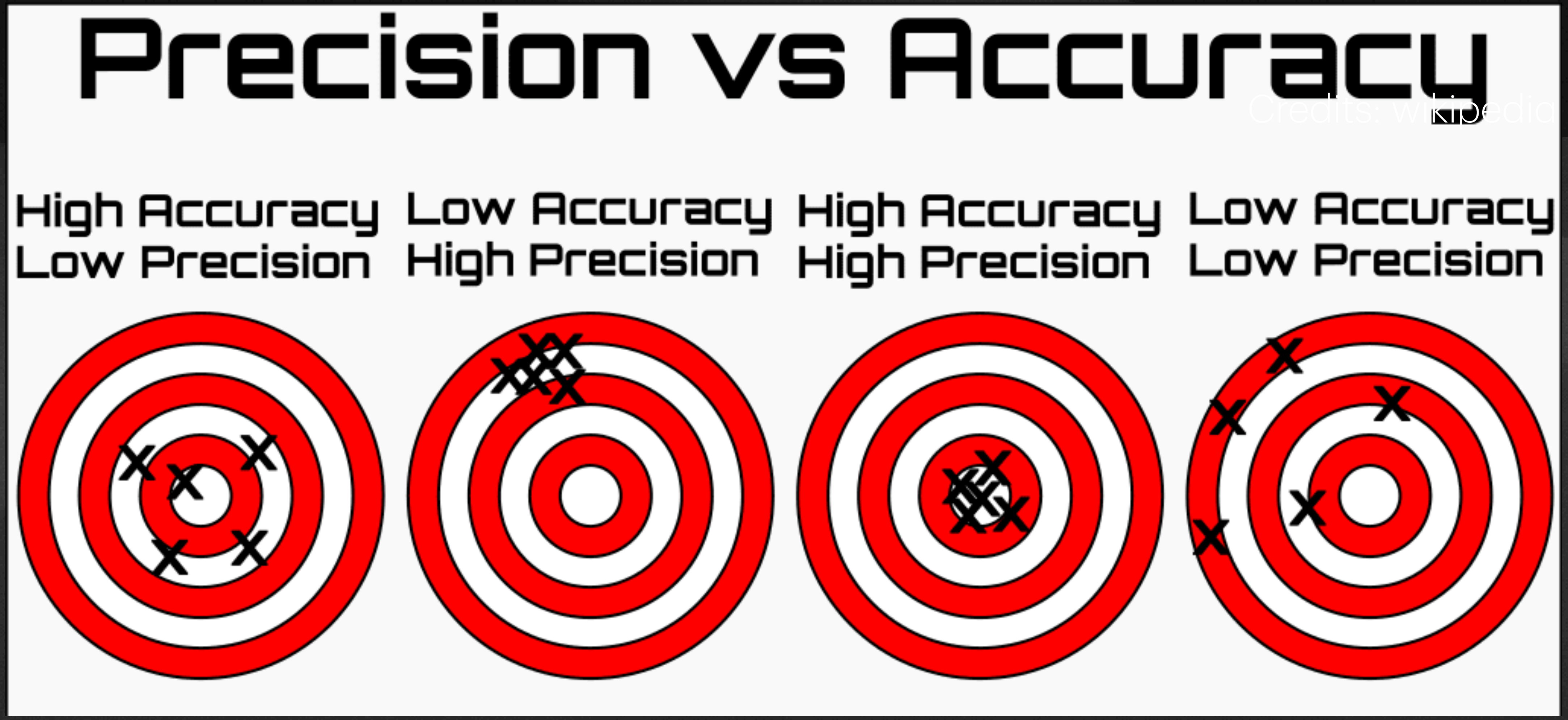
Introduction

- The **bias** of an estimator $\hat{\theta}$ of the quantity θ is:

$$B(\hat{\theta}) = E[\hat{\theta}] - \theta$$

If $B(\hat{\theta}) = 0$ then the estimator is **unbiased**. The bias represents the accuracy of an estimator.

Notice that the bias is a property of the estimator and not the estimate. Do not confuse the error of an estimate with the bias of the estimator. An unbiased estimator may have a large error, and the opposite is also true: the error of a biased estimate may be zero for some realizations.



Estimators: sampled moments

If we have n samples of random variable X, the **k-th sample moment** is $\frac{1}{n}\sum_{i=1}^{n} X_i^k$.

This is an unbiased estimator of any raw moment of the distribution.

This is not true for central moments, whose computation uses up a degrees of freedom by using the sample mean. For instance, an unbiased estimator of the variance is given by:

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Where $\bar{X}$ is the **sample mean** given by $\frac{1}{n}\sum_{i=1}^{n} X_i$. We refer to this quantity as the **sample variance**.

Discrete probability distributions

Binomial distribution

- Problem: A bent coin has probability p of coming up heads. The coin is tossed N times. What is the probability distribution of the number of heads, r? What are the mean and variance of r?

Discrete probability distributions

Binomial distribution

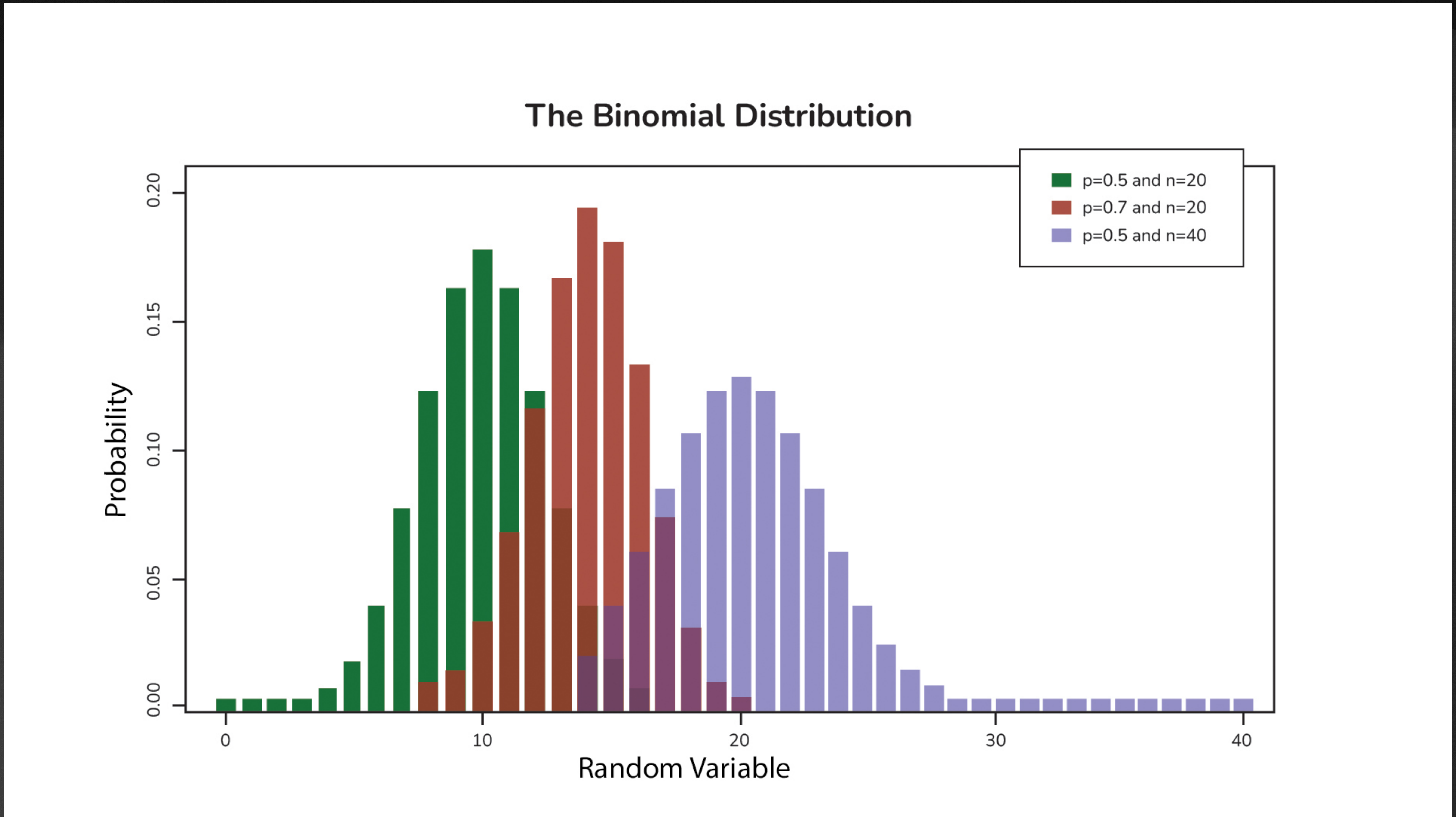
- Problem: A bent coin has probability p of coming up heads. The coin is tossed N times. What is the probability distribution of the number of heads, r? What are the mean and variance of r?

- Solution: The number of heads has a **binomial distribution**.

$$P(r \mid p, N) = \binom{N}{r} p^r (1-p)^{N-r}$$

Discrete probability distributions

Binomial distribution



Discrete probability distributions

Binomial distribution

The mean and the variance are given by

$$E[r] = \sum_{r=0}^{N} P(r \,|\, p, N)$$

$$var[r^2] = E[(r - E[r])^2] = E[r^2] - E[r]^2 = \sum_{r=0}^{N} P(r \,|\, p, N) r^2 - E[r]^2$$

Discrete probability distributions

Binomial distribution

Rather than evaluating the sums over r, let's notice that mean and variance are the sum over N independent random variables (number of heads in each toss). In general:

$$E[X+Z] = E[X] + E[Z] \text{ and } var[X+Z] = var[X] + var[Z]$$

The mean number of heads in a single toss is:

$$p \times 1 + (1-p) \times 0 = p$$

And the variance of number of heads in a single toss is:

$$[p \times 1^2 + (1-p) \times 0^2] - p^2 = p - p^2 = p(1-p)$$

So mean and variance for the Binomial distribution are:

$$E[r] = Np \text{ and } var[r] = Np(1-p)$$

Discrete probability distributions

Poisson distribution

The binomial distribution works for a finite number of events, N, with a fixed probability of success p. What if we did not know either N or p, but the average number of success rate per time period, i.e., the quantity:

$$\lambda = Np$$

This is average success rate from the binomial distribution. The success rate is then $p = \lambda/N$.

Let's therefore take the limit of the binomial distribution for $N \to \infty$, and replace the success rate definition by the quantity above:

$$\lim_{N\to\infty} P(X=k) = \lim_{N\to\infty} \frac{N!}{k!(N-k)!}(\frac{\lambda}{N})^k(1-\frac{\lambda}{N})^{N-k}$$

Discrete probability distributions

Poisson distribution

Let's do some math:

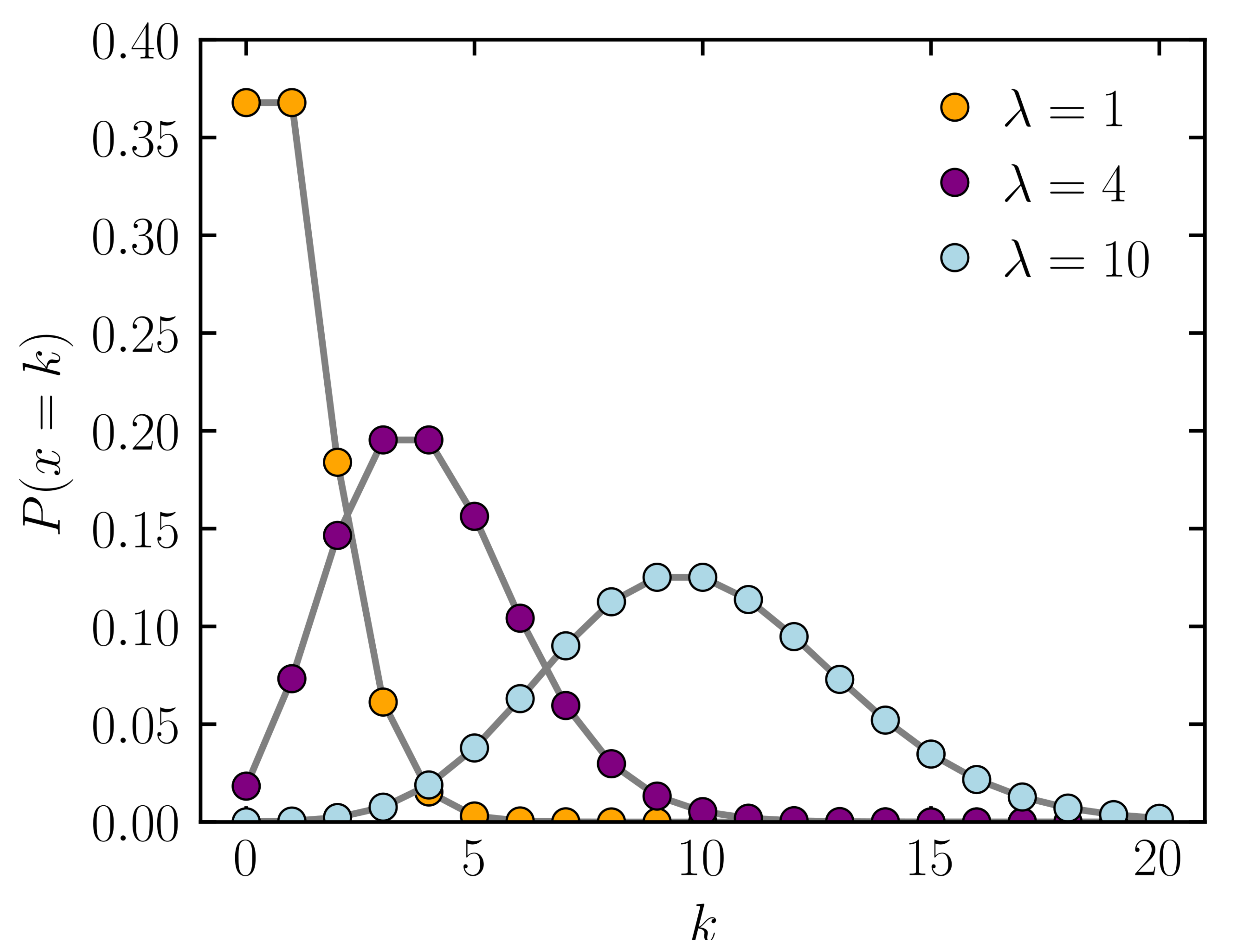
$$\left(\frac{\lambda^k}{k!}\right) \lim_{N\to\infty} \overset{\to 1}{\boxed{\frac{N!}{(N-k)!}\left(\frac{1}{N^k}\right)}} \overset{\to e^{-\lambda}}{\boxed{(1-\lambda/N)^n}} \overset{\to 1}{\boxed{(1-\lambda/N)^{-k}}}$$

So the distribution of infinite trials with fixed probability of success is the Poinsson distribution with probability function:

$$P(\lambda, k) = \lambda^k \frac{e^{\lambda}}{k!}$$

Discrete probability distributions

Poisson distribution



Discrete probability distributions

Poisson distribution - Moment Generating Function

To compute the moments of the Poisson distribution we define the **moment-generating function** (MGF):

$$M_X(t) = E[e^{tX}]$$

The name is self-explanatory. The series expansion of e^{tX} is:

$$e^{tX} = 1 + tX + \frac{t^2X^2}{2!} + \ldots + \frac{t^nX^n}{n!} + \ldots$$

Hence we have:

$$M_X(t) = E[e^{tX}] = 1 + tE[X] + t^2\frac{E[X^2]}{2!} + \ldots + t^n\frac{E[X^n]}{n!} + \ldots$$

Where we identify the n-th orders moments. We only need to differentiate the MGF n times and set t=0 to gent the n-th order moment.

Discrete probability distributions

Poisson distribution

The MGF for the Poisson distribution is:

$$E[e^{tX}] = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}$$

This is related to the exponential function $\sum_{x=0}^{\infty} \frac{a^x}{x!} = e^a$. Hence we have:

$$E[e^{tX}] = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

The mean and variance of the Poisson distribution hence are:

$E[X] = dE[e^{tX}]/dt\,|_{t=0} = \lambda$ and $var[X] = d^2E[e^{tX}]/dt^2\,|_{t=0} = \lambda$

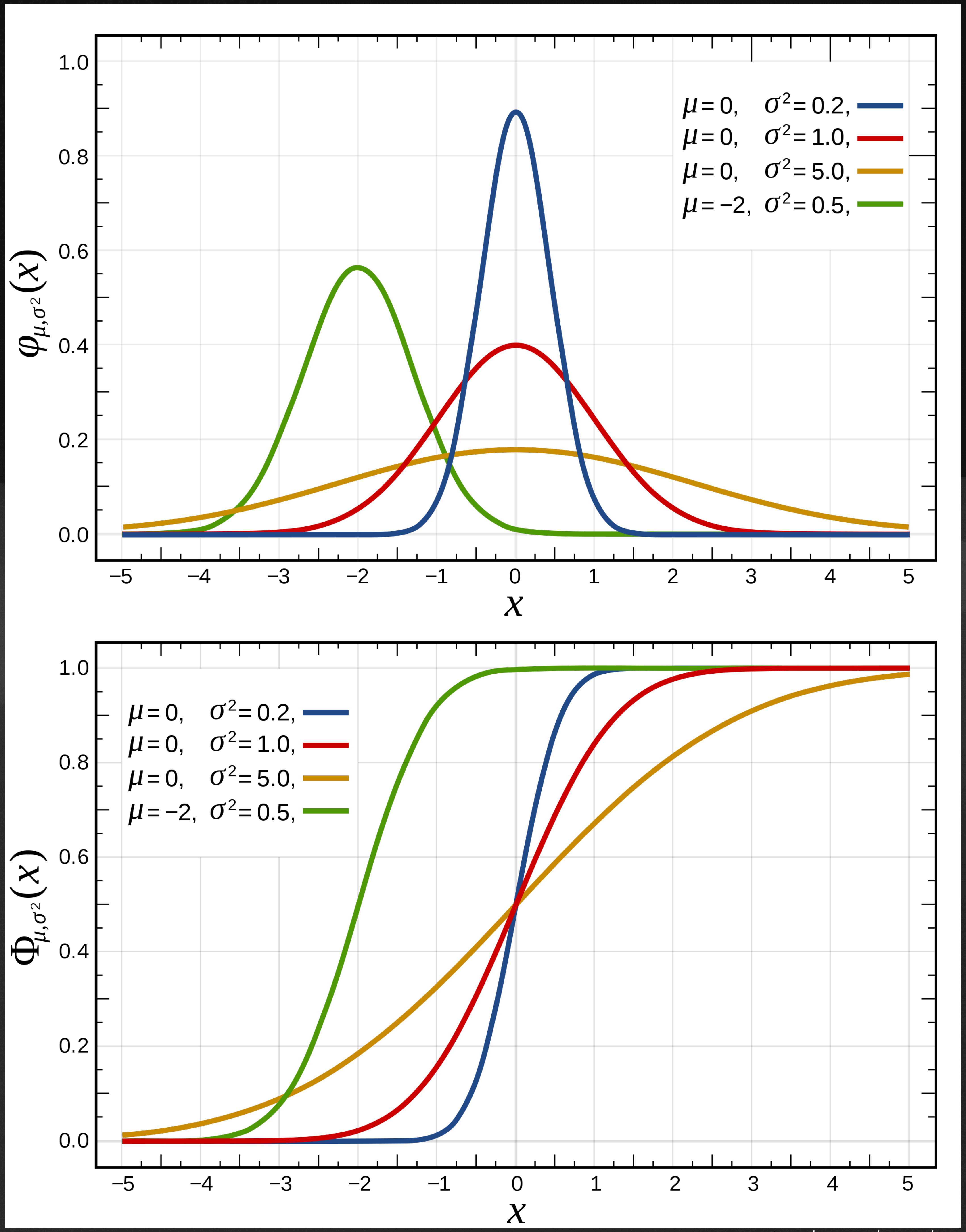
Part II

Normal distribution

Definition

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

$$\Phi(x) = \frac{1}{2\pi} \int_{-\infty}^{x} e^{-t^2/2} dt$$



Credits: wikipedia

Normal distribution

Properties

- It is symmetric around $x = \mu$;
- The probability distribution os completely defined by its mean and variance (first two order moments).

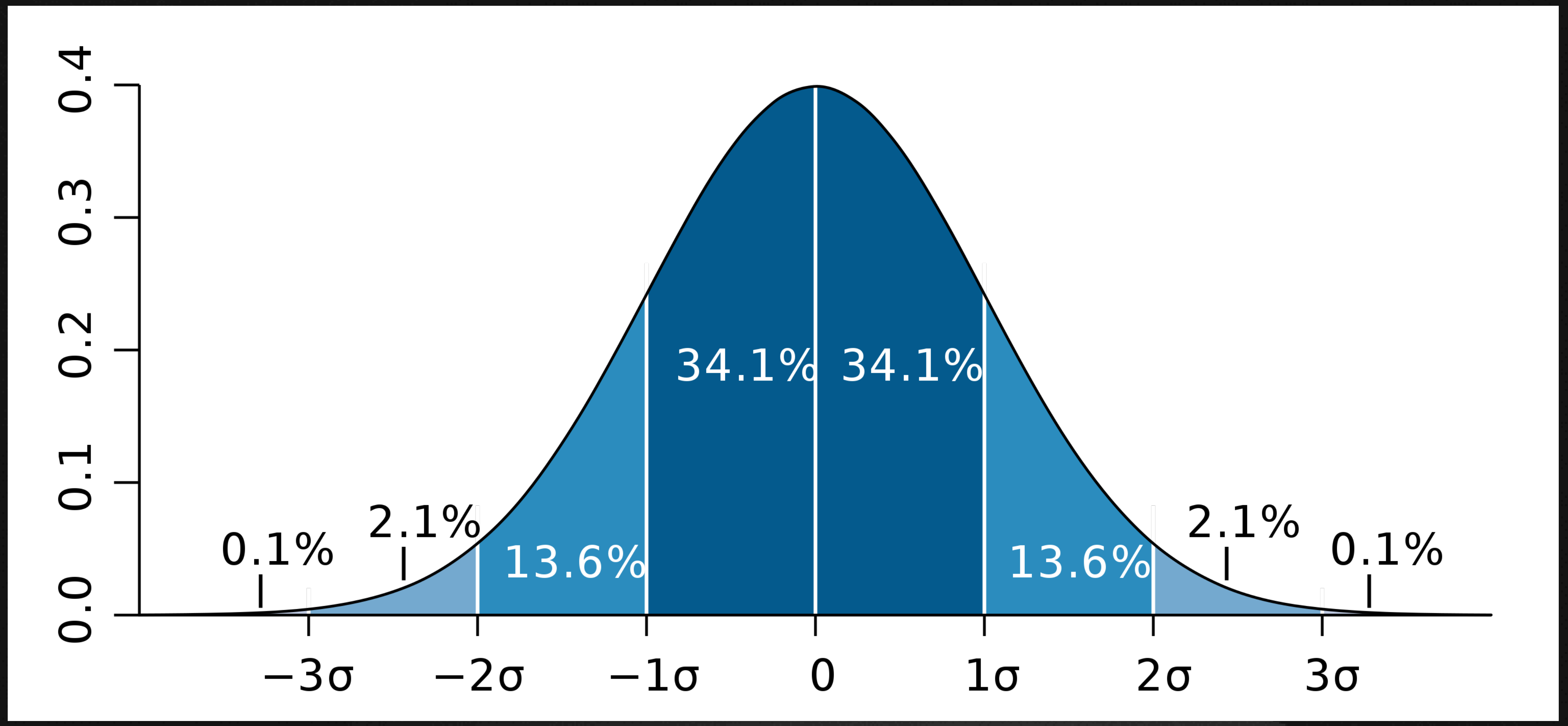
For any non negative value of p, the plain central moments are:

$$E[(x-\mu)^p] = \begin{cases} 0 & \text{if p is odd} \\ \sigma^p (p-1)!! & \text{if p is even} \end{cases}$$

Where $n!!$ is the double factorial (the product of all numbers with same parity as n).

Normal distribution

Useful results



From the Normal CDF we define the Error function:

$$erf(z) = \frac{2}{\pi^{1/2}} \int_0^z e^{-t^2} dt$$

Represents the probability of a normal distributed continuous variable X to fall in the interval $[-z, z]$. Given a normal distributed variable, the expected fraction of population within x standard deviations, $\mu \pm x\sigma$ are $erf(\frac{x}{\sqrt{w}})$, and the occurrences is therefore 1 every $\frac{1}{1 - erf(x/\sqrt{2})}$.

This useful result leads to the so called 68-95-99.7 rule (or 3σ rule), giving the probability fraction within 1, 2 and 3 standard deviations respectively.

Figure By M. W. Toews - Own work, based (in concept) on figure by Jeremy Kemp, on 2005-02-09, CC BY 2.5, https://commons.wikimedia.org/w/index.php?curid=1903871

Normal distribution

Useful results

Range	Expected fraction of population inside range	Expected fraction of population outside range	Approx. expected frequency outside range		Approx. frequency for daily event
$\mu \pm 0.5\sigma$	0.38	61.71%	3 in 5	5	Four or five times a week
$\mu \pm \sigma$	0.68	31.73%	1 in 3	3	Twice or thrice a week
$\mu \pm 1.5\sigma$	0.87	13.36%	2 in 15	15	Weekly
$\mu \pm 2\sigma$	0.95	4.550%	1 in 22	22	Every three weeks
$\mu \pm 2.5\sigma$	0.988	1.242%	1 in 81	81	Quarterly
$\mu \pm 3\sigma$	0.997	2.700 ‰	1 in 370	370	Yearly
$\mu \pm 3.5\sigma$	0.999534741841929	0.04653 % = 465.3 ppm	1 in 2149	2149	Every 6 years
$\mu \pm 4\sigma$	0.999936657516334	63.34 ppm	1 in 15787	15787	Every 43 years (twice in a lifetime)
$\mu \pm 4.5\sigma$	0.999993204653751	6.795 ppm	1 in 147160	147160	Every 403 years (once in the modern era)
$\mu \pm 5\sigma$	0.999999426696856	0.5733 ppm = 573.3 ppb	1 in 1744278	1744278	Every 4776 years (once in recorded history)

https://en.wikipedia.org/wiki/68–95–99.7_rule

Normal distribution

Operations with normal variables

If X is distributed with zero mean, and unity standard deviation:

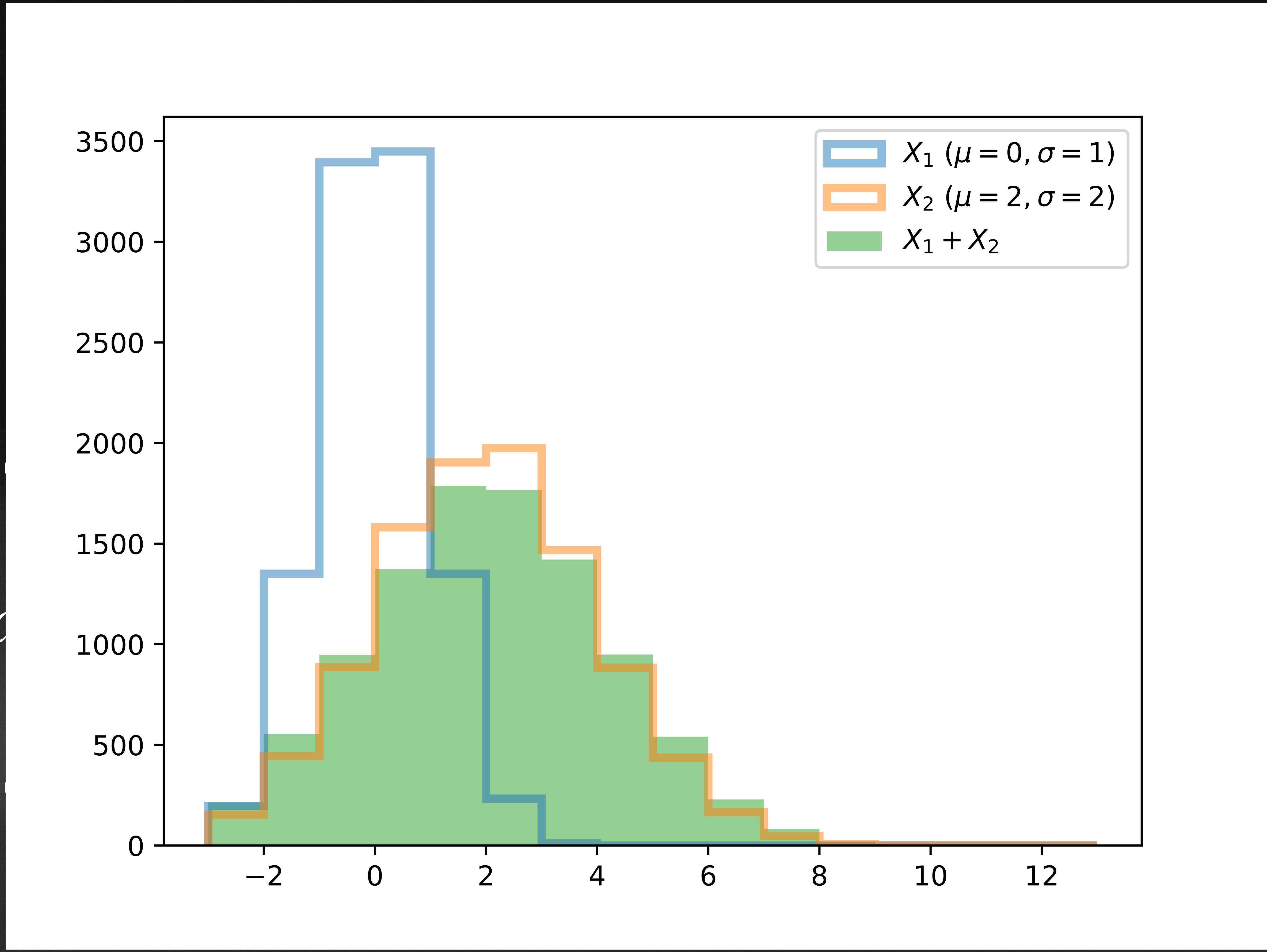
- $aX + b$ is also normally distributed, with mean $a\mu + b$ and standard deviation $|a|\sigma$.

Moreover, if X_1 and X_2 are two independent normal variables with means μ_1,μ_2 and variances σ_1^2, σ_2^2 :

- $X_1 + X_2$ is also normally distributed with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$

Normal distribution

Operations with normal variables



If X is distributed with zero mean, and unity stand

- $aX + b$ is also normally distributed, with mean a

Moreover, if X_1 and X_2 are two independent norm
σ_1^2, σ_2^2 :

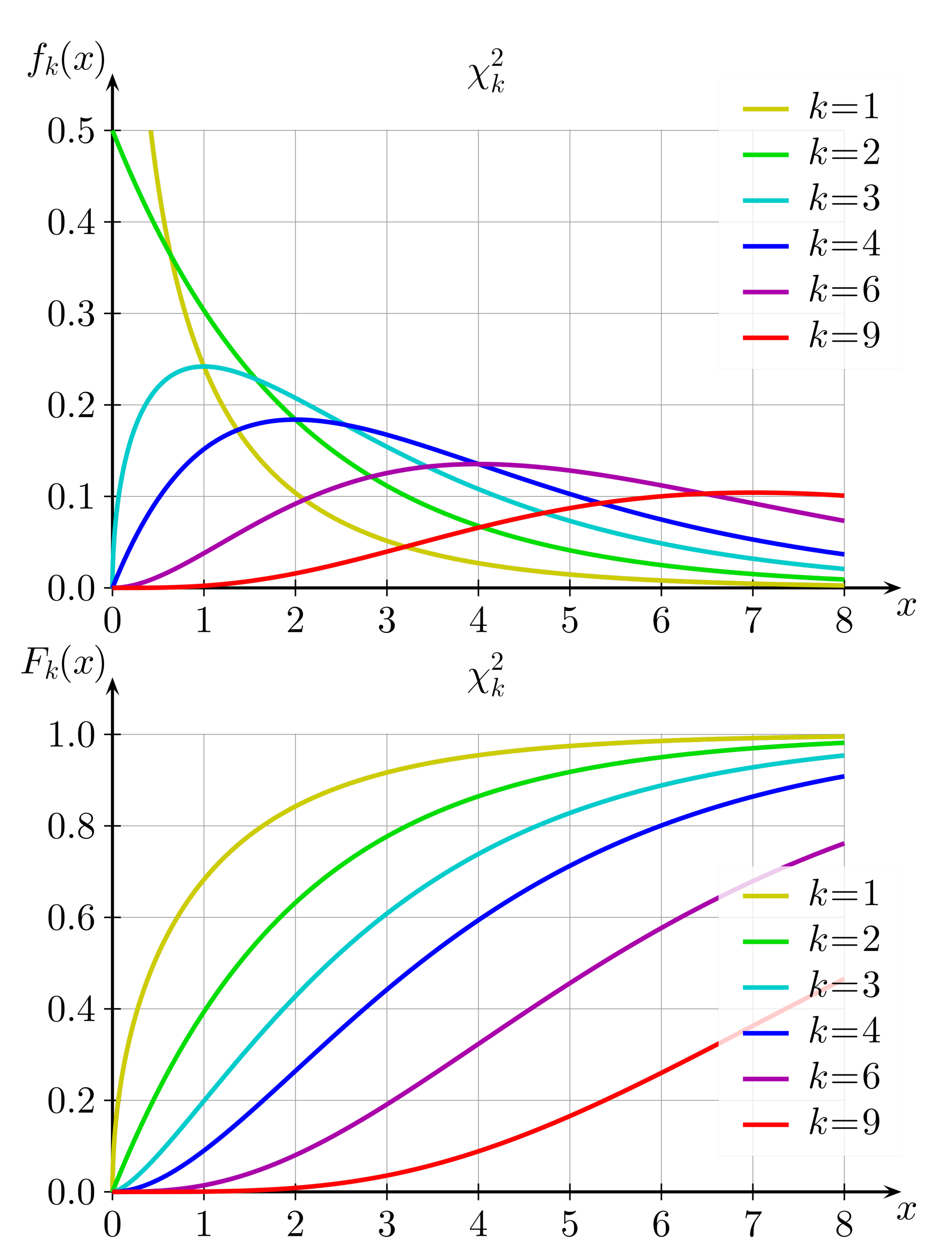
- $X_1 + X_2$ is also normally distributed with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$

χ^2 distribution

Suppose we have $X_1, X_2, \ldots, X_n$ normal random variables. The sum of their squares has the χ^2 distribution with n degrees of freedom:

$$X_1^2/\sigma_1^2 + X_2^2/\sigma_2^2 + \ldots + X_n^2/\sigma_n^2 \sim \chi_n^2 = \sum_{i=1}^{n} \frac{(X_i - \mu_i)^2}{\sigma_i^2}$$

The χ^2 distribution only depends on the degrees of freedom as a free parameter



Multivariate Normal distribution

Suppose we have a set of possibly non-independent random variables $X_1, X_2, \ldots, X_k$. We can define the k-dimensional random vector $\mathbf{X} = (X_1, X_2, \ldots, X_k)^T$.

The random vector $\mathbf{X}$ is k-variate normally distributed if any linear combination of its k components follows a univariate normal distribution.

$$\mathbf{X} \sim \mathcal{N}(\mu, \mathbf{\Sigma}) = (2\pi)^{-k/2} |\mathbf{\Sigma}|^{-1/2} \exp[-\frac{1}{2}(\mathbf{X} - \mu)^T \mathbf{\Sigma}^{-1} (\mathbf{X} - \mu)]$$

Where $\mu = (\mu_1, \mu_2, \ldots, \mu_k)$ and $\Sigma_{i,j} = E[(X_i - \mu_i)(X_j - \mu_j)] = Cov(X_i, X_j)$

Central Limit Theorem

- Suppose we have $X_1, X_2, \ldots, X_n$ **independent and identically distributed random variables**, with the **same arbitrary distribution**, mean μ and **finite** variance σ^2.

- Let $\bar{X}_n$ be the sample mean.

The Central Limit Theorem (CLT) states that in the limit of $n \to \infty$, the distribution of $\dfrac{\bar{X} - \mu}{\sigma_{\bar{X}_n}}$, with $\sigma_{\bar{X}_n} = \sigma/\sqrt{N}$, is a standard normal distribution.

- In other words, provided a sufficiently large sample set of independent random observations with its sample mean; if this procedure is repeated many times, leading to a collection of samples means, the distribution of such sample means follows a normal distribution.

- Galton's board online.

- This guy on YouTube is incredible. He is much better than me in explaining math, check his channel out.

A quick recollection on probabilities

Joint Probability

Problem: Each of two urns contains twice as many red balls as blue balls, and no others, and one ball is randomly selected from each urn, with the two draws independent of each other. Let X and Y be discrete random variables associated with the outcomes of the draw from the first urn and second urn respectively. **What is the probability of drawing a red ball from the first urn and a blue ball from the second urn?**

A quick recollection on probabilities

Joint Probability

Problem: Each of two urns contains twice as many red balls as blue balls, and no others, and one ball is randomly selected from each urn, with the two draws independent of each other. Let X and Y be discrete random variables associated with the outcomes of the draw from the first urn and second urn respectively. **What is the probability of drawing a red ball from the first urn and a blue ball from the second urn?**

Solution: The probability of drawing a red ball from either of the urns is 2/3, and the probability of drawing a blue one is 1/3. In this simple case we can count and see that, for two independent draws:

A quick recollection on probabilities

Joint Probability

Problem: Each of two urns contains twice as many red balls as blue balls, and no others, and one ball is randomly selected from each urn, with the two draws independent of each other. Let X and Y be discrete random variables associated with the outcomes of the draw from the first urn and second urn respectively. **What is the probability of drawing a red ball from the first urn and a blue ball from the second urn?**

Solution: The probability of drawing a red ball from either of the urns is 2/3, and the probability of drawing a blue one is 1/3. In this simple case we can count and see that, for two independent draws:

$$P(X = \text{blue}, Y = \text{red}) = 2/3 \cdot 1/3 = 2/9$$

A quick recollection on probabilities

Joint Probability

From a mathematical point of view, we were interested in the **joint probability** of X and Y:

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

The corresponding **joint probability distribution** is the corresponding probability distribution of all the possible combination of the random variables. In our case:

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

A quick recollection on probabilities

Independent random variables

- If two random variables X and Y are **independent**, the joint cumulative distribution satisfies: $F(x, y) = F(x) \cdot F(y)$ and

- the the joint probability distribution satisfies: $p(x, y) = p(x) \cdot p(y)$, for all x, y.

A quick recollection on probabilities

Marginal probability

Problem: Let's consider once more the previous problem. What is now the probability of drawing a red ball from the first urn, regardless of the color of the ball extracted in the second urn?

A quick recollection on probabilities

Marginal probability

Problem: Let's consider once more the previous problem. What is now the probability of drawing a red ball from the first urn, regardless of the color of the ball extracted in the second urn?

Solution: We need to consider the two possible cases: the first draw to give a red ball and the second a blue one (P(X=red,Y=blue)=4/9), and that both the draws give a red ball (P(X=red, Y=red)=2/9). The probability of getting a red ball in the first draw regardless of the second draw is then $P(\text{red,blue}) + P(\text{red,red}) = 4/9 + 2/9 = 2/3$

A quick recollection on probabilities

Marginal probability

We have computed the **Marginal probability** of the variable X, i.e. the probability of X regardless of Y. Y is said to be **marginalized out**.

The corresponding distribution is called the **Marginal probability distribution**. In the marginal probability distribution does not depend on the variable we marginalized on.

For a discrete 2-dimensional probability distribution:

$$p(x) = \sum_j P(x_i, y_j)$$

The analogous for the continuous case is:

$$p(x) = \int P(x, y)dy$$

The N dimension generalization is straightforward.

A quick recollection on probabilities

Conditional probability

Let's now consider the only case left. What is the probability of an event X to occur, given an event Y has already occurred?

A quick recollection on probabilities

Conditional probability

Let's now consider the only case left. What is the probability of an event X to occur, given an event Y has already occurred?

This is the **conditional probability** $P(X = x \mid Y = y) = p(x \mid y)$, with its **conditional probability distribution**:

$$p(x \mid y) = \frac{p(x, y)}{p(y)}$$

Where p is the point mass function for a discrete random variable, and the probability density function for a continuous one

A quick recollection on probabilities

Bayes' theorem

Let's consider a 2-dimensional distribution $p(x, y)$. We can write down the two conditional distributions for x and y:

$$p(x|y) = \frac{p(x, y)}{p(y)} \text{ and } p(y|x) = \frac{p(x, y)}{p(x)}$$

From the either of the two, we can have a formulation for the joint probability $p(x, y) = p(y|x) \cdot p(x)$ that, plugged into the other conditional definition gives:

$$p(x|y) = \frac{p(y|x) \cdot p(x)}{p(y)}$$

This is known as the **Bayes theorem**, and provides a useful relationship between conditional distributions.

Bayes' Theorem

$$P(x \mid y) = \frac{p(y \mid x) p(x)}{p(y)}$$

Bayes' Theorem

$$\boxed{P(x \mid y)} = \frac{p(y \mid x)p(x)}{p(y)}$$

The conditional probability of x given that y is true. Also known as **Posterior probability**

Bayes' Theorem

$$\boxed{P(x \mid y)} = \frac{\boxed{p(y \mid x)}\, p(x)}{p(y)}$$

The conditional probability of x given that y is true. Also known as **Posterior probability**

This is also a conditional probability. This is the probability of y given that x is true; it can also be interpreted as the **Likelihood** of x given y

Bayes' Theorem

$$\boxed{P(x \mid y)} = \frac{\boxed{p(y \mid x)}\,\boxed{p(x)}}{p(y)}$$

The conditional probability of x given that y is true. Also known as **Posterior probability**

This is also a conditional probability. This is the probability of y given that x is true; it can also be interpreted as the **Likelihood** of x given y

This is the probability of observing x without further condition. It is also known as the **prior probability**.

Bayes' Theorem

$$\boxed{P(x \mid y)} = \frac{\boxed{p(y \mid x)}\boxed{p(x)}}{\boxed{p(y)}}$$

The conditional probability of x given that y is true. Also known as **Posterior probability**

This is also a conditional probability. This is the probability of y given that x is true; it can also be interpreted as the **Likelihood** of x given y

This is the probability of observing x without further condition. It is also known as the **prior probability**.

This is the probability of y without any given conditions; this is known as the **evidence**. It can be written as:

$$P(y) = \int p(y \mid x)p(x)dx$$

Bayes' Theorem

An example

- A factory produces items using three different machines —A, B and C— accounting for 20%, 30% and 50% of the total output respectively. However, these machines have a percentage of defective products of 5%, 3% and 1% respectively.

Bayes' Theorem

An example

- A factory produces items using three different machines —A, B and C— accounting for 20%, 30% and 50% of the total output respectively. However, these machines have a percentage of defective products of 5%, 3% and 1% respectively.

- Question: If a random product is defective, what is the probability it was produced by machine C?

Bayes' Theorem

An example

- A factory produces items using three different machines —A, B and C— accounting for 20%, 30% and 50% of the total output respectively. However, these machines have a percentage of defective products of 5%, 3% and 1% respectively.
- Question: If a random product is defective, what is the probability it was produced by machine C?
- Solution: We can try applying the Bayes' theorem!

Bayes' Theorem

An example

Let's be X_i the event that a randomly chosen item was made by i-th machine (i = A, B or C). Let be Y the event that a random item is defective. Hence we have:

$$P(X_A) = 0.2,\ P(X_B) = 0.3 \text{ and } P(X_C) = 0.5$$

We also have that:

$$P(Y|X_A) = 0.05,\ P(Y|X_B) = 0.03 \text{ and } P(Y|X_C) = 0.01$$

Bayes' Theorem

An example

To use Bayes' theorem let's first compute the evidence:

$$P(Y) = \sum_i P(Y|X_i)P(X_i) = 5/24 \approx 0.024.$$

This tells us that ~2,4% of the products are defective. We are interested in the probability of a random defective product to be made by machine C. This is nothing but:

$$P(X_C|Y) = \frac{P(Y|X_C)P(X_C)}{P(Y)} = \frac{0.01 \cdot 0.5}{0.024} = \frac{5}{24} \approx 21\,\%$$

In other words, despite C produces 50% of the products, the a-priori knowledge of only 1% of it being defective gives us a conditional probability of ~21%.

Bayes' theorem $\neq$ Bayesian inference

Bayesian vs Frequentist

The difference between **Bayesian inference** and **Frequentist inference** is the **goal**:

Bayesian vs Frequentist

The difference between **Bayesian inference** and **Frequentist inference** is the **goal**:

- The Bayesian goal is to quantify and analyze subjective degrees of belief.

Bayesian vs Frequentist

The difference between **Bayesian inference** and **Frequentist inference** is the **goal**:

- The Bayesian goal is to quantify and analyze subjective degrees of belief.
- The frequentist goal is to create procedures that have frequency guarantees

Bayesian vs Frequentist

The difference between **Bayesian inference** and **Frequentist inference** is the **goal**:

- The Bayesian goal is to quantify and analyze subjective degrees of belief.
- The frequentist goal is to create procedures that have frequency guarantees

Neither method of inference is right or wrong. Which one you use depends on your goal. If your goal is to quantify and analyze your subjective degrees of belief, you should use Bayesian inference. If our goal create procedures that have frequency guarantees then you should use frequentist procedures.

Bayesian vs Frequentist

The difference between **Bayesian inference** and **Frequentist inference** is the **goal**:

- The Bayesian goal is to quantify and analyze subjective degrees of belief.
- The frequentist goal is to create procedures that have frequency guarantees

Neither method of inference is right or wrong. Which one you use depends on your goal. If your goal is to quantify and analyze your subjective degrees of belief, you should use Bayesian inference. If our goal create procedures that have frequency guarantees then you should use frequentist procedures.

Sometimes you can do both. That is, sometimes a Bayesian method will also have good frequentist properties. Sometimes it won't.

The Doomsday Argument

Hypothesis: The total number of human beings that will ever exist is fixed (N).

Question: Is there a way to estimate, from a statistical point of view, what N is based on how many human beings there are on our planet now, n?

The Doomsday Argument

[1] Oliver, Jonathan; Korb, Kevin (1998). "A Bayesian Analysis of the Doomsday Argument"

The Doomsday Argument

Any human should be equally likely to find himself in any position n of the total population N, so that we can define a fractional position $f = n/N$ uniformly distributed in $[0,1]$.

[1] Oliver, Jonathan; Korb, Kevin (1998). "A Bayesian Analysis of the Doomsday Argument"

The Doomsday Argument

Any human should be equally likely to find himself in any position n of the total population N, so that we can define a fractional position $f = n/N$ uniformly distributed in $[0,1]$.

Given the prior assumption on the distribution of f, we can say that there is a $95\ \%$ probability that any human is the fraction $[0.05,1]$, so $f > 0.05$. In other words, there is a $95\ \%$ probability for a human to be born in the last $95\ \%$ of all humans ever to be born.

[1] Oliver, Jonathan; Korb, Kevin (1998). "A Bayesian Analysis of the Doomsday Argument"

The Doomsday Argument

Any human should be equally likely to find himself in any position n of the total population N, so that we can define a fractional position $f = n/N$ uniformly distributed in $[0,1]$.

Given the prior assumption on the distribution of f, we can say that there is a $95\ \%$ probability that any human is the fraction $[0.05,1]$, so $f > 0.05$. In other words, there is a $95\ \%$ probability for a human to be born in the last $95\ \%$ of all humans ever to be born.

The knowledge of n implies a $95\ \%$ upper bound on N which gives $N = 20n$.

[1] Oliver, Jonathan; Korb, Kevin (1998). "A Bayesian Analysis of the Doomsday Argument"

The Doomsday Argument

Any human should be equally likely to find himself in any position n of the total population N, so that we can define a fractional position $f = n/N$ uniformly distributed in $[0,1]$.

Given the prior assumption on the distribution of f, we can say that there is a $95\ \%$ probability that any human is the fraction $[0.05,1]$, so $f > 0.05$. In other words, there is a $95\ \%$ probability for a human to be born in the last $95\ \%$ of all humans ever to be born.

The knowledge of n implies a $95\ \%$ upper bound on N which gives $N = 20n$.

If Leslie's figure [1] is used, then approximately 60 billion humans have been born so far, so it can be estimated that there is a $95\ \%$ chance that the total number of humans N will be less than 20x60 billion = 1.2 trillion. Assuming that the world population stabilizes at 10 billion and a life expectancy of 80 yrs it can be estimated that the remaining 1140 billion humans will be born in 9120 years. Depending on the projection of the world population in the forthcoming centuries, estimates may vary, but the argument states that it is unlikely that more than 1.2 trillion humans will ever live.

[1] Oliver, Jonathan; Korb, Kevin (1998). "A Bayesian Analysis of the Doomsday Argument"