Probability

Lecturer: Stanley B. Gershwin

Trick Question

I flip a coin 100 times, and it shows heads every time.

Question: What is the probability that it will show heads on the next flip?

Probability and Statistics

Probability $\neq$ *Statistics*

Probability: mathematical theory that describes uncertainty.

Statistics: set of techniques for extracting useful information from data.

Interpretations of probability

Frequency

The probability that the outcome of an experiment is A *is* $P(A)$

if the experiment is performed a large number of times and the fraction of times that the observed outcome is A is $P(A)$.

Parallel universes

The probability that the outcome of an experiment is A *is* $P(A)$

if the experiment is performed in each parallel universe and the fraction of universes in which the observed outcome is A is $P(A)$.

Interpretations of probability

Betting odds

The probability that the outcome of an experiment is A *is* $P(A)$

if *before the experiment is performed* a risk-neutral observer would be willing to bet \$1 against more than \$ $\frac{1-P(A)}{P(A)}$.

Interpretations of probability

State of belief

The probability that the outcome of an experiment is A *is* $P(A)$

if that is the opinion (ie, belief or state of mind) of an observer *before* the experiment is performed.

Interpretations of probability

Abstract measure

The probability that the outcome of an experiment is A *is* $P(A)$

if $P()$ satisfies a certain set of conditions: *the axioms of probability.*

Interpretations of probability

Abstract measure

Axioms of probability

Let U be a set of *samples* . Let E_1, E_2, ... be subsets of U. Let ϕ be the *null set* (the set that has no elements).

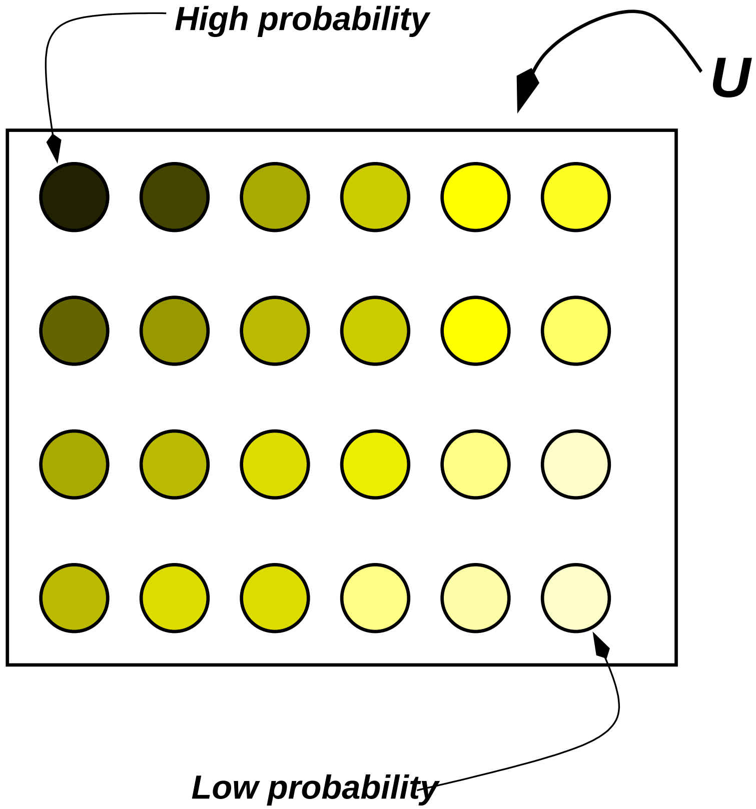
- $0 \leq P(E_i) \leq 1$
- $P(U) = 1$
- $P(\phi) = 0$
- If $E_i \cap E_j = \phi$, then $P(E_i \cup E_j) = P(E_i) + P(E_j)$

Discrete Sample Space

- Subsets of U are called *events.*
- $P(E)$ is the *probability* of E.

Probability Basics

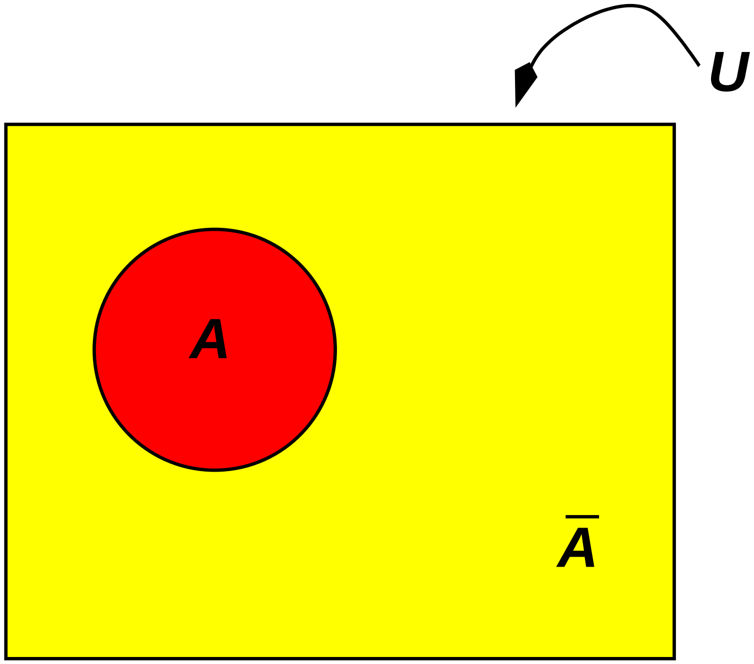
Discrete Sample Space



Probability Basics

Set Theory

Venn diagrams

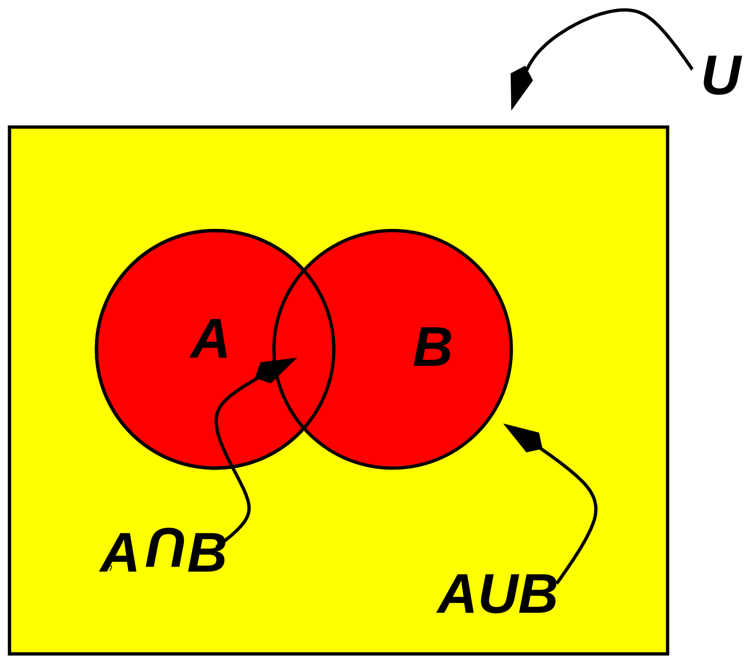


$$P(\bar{A}) = 1 - P(A)$$

Probability Basics

Set Theory

Venn diagrams



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

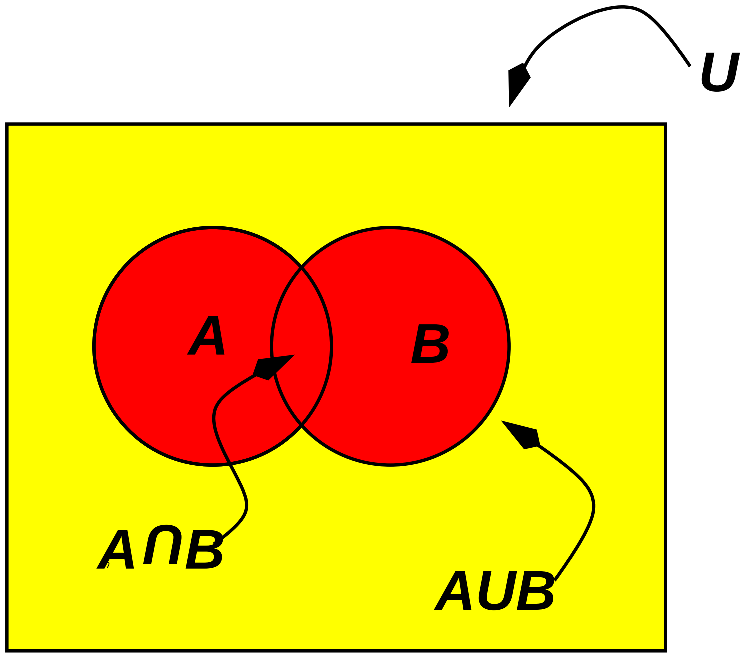
Independence

A and B are *independent* if

$$P(A \cap B) = P(A)P(B).$$

Probability Basics

Conditional Probability



If $P(B) \neq 0,$

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

We can also write $P(A \cap B) = P(A|B)P(B)$.

Probability Basics

Conditional Probability

Example

Throw a die.

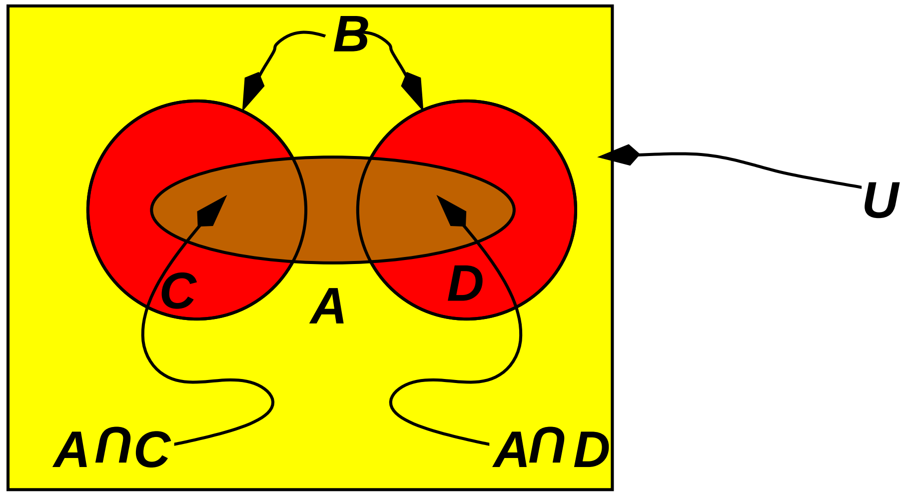
- $\boldsymbol{A}$ is the event of getting an odd number (1, 3, 5).
- $\boldsymbol{B}$ is the event of getting a number less than or equal to 3 (1, 2, 3).

Then $\boldsymbol{P(A) = P(B) = 1/2}$ and
$\boldsymbol{P(A \cap B) = P(1, 3) = 1/3}$.

Also, $\boldsymbol{P(A|B) = P(A \cap B)/P(B) = 2/3.}$

Probability Basics

Law of Total Probability



- Let $B = C \cup D$ and assume $C \cap D = \phi$. Then $P(A|C) = \dfrac{P(A \cap C)}{P(C)}$ and $P(A|D) = \dfrac{P(A \cap D)}{P(D)}$.

Probability Basics

Law of Total Probability

Also,

- $P(C|B) = \frac{P(C \cap B)}{P(B)} = \frac{P(C)}{P(B)}$ because $C \cap B = C$.

  Similarly, $P(D|B) = \frac{P(D)}{P(B)}$

- $P(A \cap B) = P(A \cap (C \cup D)) =$

  $P(A \cap C) + P(A \cap D) - P(A \cap (C \cap D)) =$

  or

  $P(A \cap B) = P(A \cap C) + P(A \cap D)$

Probability Basics

Law of Total Probability

- Or, $P(A|B)$ prob $(B) = P(A|C)P(C) + P(A|D)P(D)$

  or,

$$\frac{P(A|B) \text{ prob } (B)}{P(B)} = \frac{P(A|C)P(C)}{P(B)} + \frac{P(A|D)P(D)}{P(B)}$$

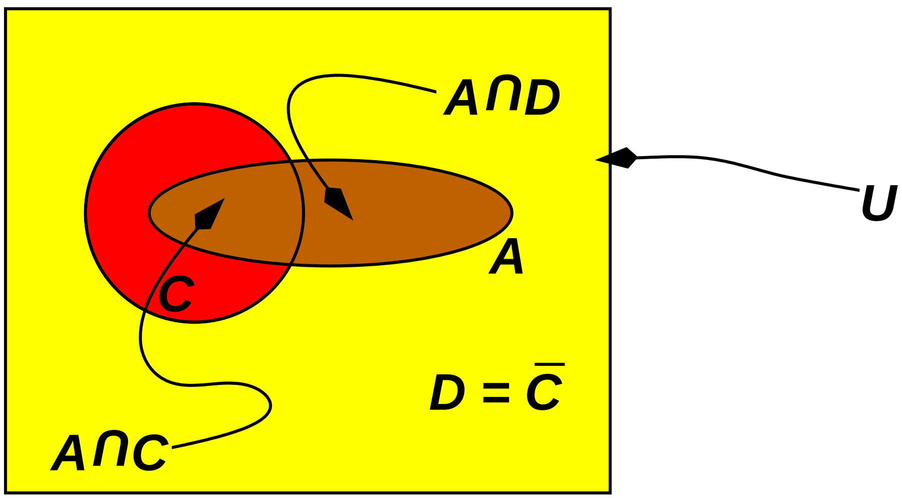
or,

$$P(A|B) = P(A|C)P(C|B) + P(A|D)P(D|B)$$

Law of Total Probability

An important case is when $C \cup D = B = U$, so that $A \cap B = A$. Then

$$P(A) = P(A \cap C) + P(A \cap D) = P(A|C)P(C) + P(A|D)P(D).$$



Law of Total Probability

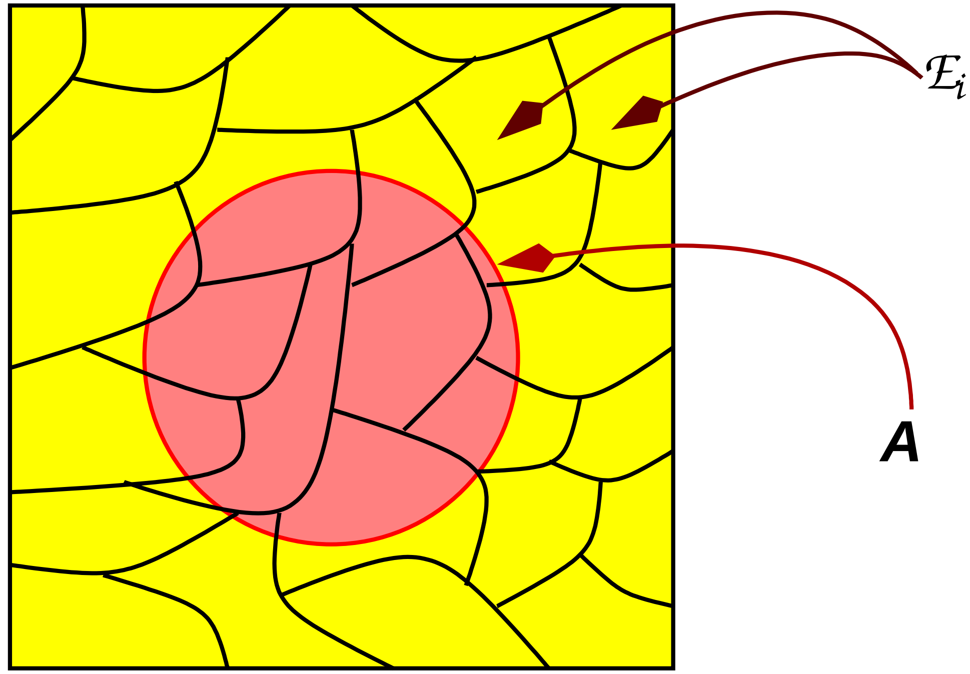
More generally, if A and $\mathcal{E}_1, \ldots \mathcal{E}_k$ are events and

$\mathcal{E}_i$ and $\mathcal{E}_j = \emptyset,$ for all $i \neq j$

and

$\bigcup_j \mathcal{E}_j =$ the universal set

(ie, the set of $\mathcal{E}_j$ sets is *mutually exclusive* and *collectively exhaustive*) then ...



Probability Basics

Law of Total Probability

$$\sum_j \text{prob}\,(\mathcal{E}_j) = 1$$

and

$$\text{prob}\,(A) = \sum_j \text{prob}\,(A|\mathcal{E}_j)\,\text{prob}\,(\mathcal{E}_j).$$

A = {I will have a cold tomorrow.}

B = {It is raining today.}

C = {It is snowing today.}

D = {It is sunny today.}

Assume $B \cup C \cup D = U$

Then $A \cap B$ = {I will have a cold tomorrow *and* it is raining today}.

And $P(A|B)$ is the probability I will have a cold tomorrow *given* that it is raining today.

etc.

Then

$\{$I will have a cold tomorrow.$\}$=

$\{$I will have a cold tomorrow *and* it is raining today$\}$ $\cup$

$\{$I will have a cold tomorrow *and* it is snowing today$\}$ $\cup$

$\{$I will have a cold tomorrow *and* it is sunny today$\}$

so

P($\{$I will have a cold tomorrow.$\}$)=

P($\{$I will have a cold tomorrow *and* it is raining today$\}$) +

P($\{$I will have a cold tomorrow *and* it is snowing today$\}$) +

P($\{$I will have a cold tomorrow *and* it is sunny today$\}$)

Probability Basics

Law of Total Probability

Example

$P(\{$I will have a cold tomorrow.$\})=$

$P(\{$I will have a cold tomorrow | it is raining today$\})P(\{$it is raining today$\})$ +

$P(\{$I will have a cold tomorrow | it is snowing today$\})P(\{$it is snowing today$\})$ +

$P(\{$I will have a cold tomorrow | it is sunny today$\})$ $P(\{$it is sunny today$\})$

Probability Basics

Random Variables

Let $\boldsymbol{V}$ be a vector space. Then a *random variable* $\boldsymbol{X}$ is a mapping (a function) from $\boldsymbol{U}$ to $\boldsymbol{V}$.

If $\boldsymbol{\omega} \in \boldsymbol{U}$ and $\boldsymbol{x} = \boldsymbol{X}(\boldsymbol{\omega}) \in \boldsymbol{V}$, then $\boldsymbol{X}$ is a random variable.

Probability Basics

Random Variables

Flip of a Coin

Let U=H,T. Let $\omega =$ H if we flip a coin and get heads; $\omega =$ T if we flip a coin and get tails.

Let $X(\omega)$ be the number of times we get heads. Then $X(\omega)$ = 0 or 1.

$$P(\omega = \text{T}) = P(X = 0) = 1/2$$

$$P(\omega = \text{H}) = P(X = 1) = 1/2$$

Probability Basics

Random Variables

Flip of Three Coins

Let U=HHH, HHT, HTH, HTT, THH, THT, TTH, TTT.

Let $\omega =$ HHH if we flip 3 coins and get 3 heads; $\omega =$ HHT if we flip 3 coins and get 2 heads and *then* tails, etc. *The order matters!*

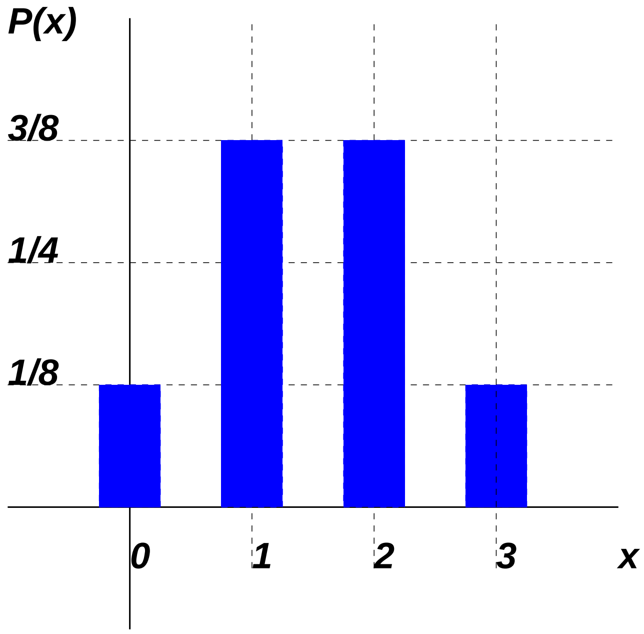
- $P(\omega)$ = 1/8 for all ω.

Let X be the number of heads. Then X = 0, 1, 2, or 3.

- $P(X = 0)$=1/8; $P(X = 1)$=3/8; $P(X = 2)$=3/8; $P(X = 3)$=1/8.

Probability Distributions

Let $X(\omega)$ be a random variable. Then $P(X(\omega) = x)$ is the *probability distribution* of X (usually written $P(x)$). For three coin flips:



Probability Basics

Probability Distributions

Mean and Variance

Mean (average): $\bar{x} = \mu_x = E(X) = \sum_x xP(x)$

Variance:

$$V_x = \sigma_x^2 = E(x - \mu_x)^2 = \sum_x (x - \mu_x)^2 P(x)$$

Standard deviation: $\sigma_x = \sqrt{V_x}$

Coefficient of variation (cv): σ_x / μ_x

Probability Basics

Example

For three coin flips:

$$\bar{x} = 1.5; V_x = 0.75; \sigma_x = 0.866; \text{cv} = 0.577.$$

A function of a random variable is a random variable.

For every ω, let $Y(\omega) = aX(\omega) + b$. Then

- $\bar{Y} = a\bar{X} + b$.
- $V_Y = a^2 V_X$; $\sigma_Y = |a|\sigma_X$.

Covariance

X and Y are random variables. Define the *covariance* of X and Y as:

$$\text{Cov}(X, Y) = E\left[(X - \mu_x)(Y - \mu_y)\right]$$

Facts:

- $\text{Var}(X + Y) = V_x + V_y + 2\text{Cov}(X, Y)$
- If X and Y are independent, $\text{Cov}(X, Y) = 0$.
- If X and Y vary in the same direction, $\text{Cov}(X, Y) > 0$.
- If X and Y vary in the opposite direction, $\text{Cov}(X, Y) < 0$.

Correlation

The *correlation* of X and Y is

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y}$$

$$-1 \leq \text{Corr}(X, Y) \leq 1$$

Discrete Random Variables

Bernoulli

Flip a biased coin. Assume all flips are independent.

X^B is 1 if outcome is heads; 0 if tails.

$P(X^B = 1)$ = p.

$P(X^B = 0)$ = $1 - p$.

X^B is Bernoulli.

Binomial

The sum of n Bernoulli random variables X_i^B with the same parameter p is a binomial random variable X^b.

$$X^b = \sum_{i=0}^{n} X_i^B$$

$$P(X^b = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{(n-x)}$$

Discrete Random Variables

Geometric

The number of Bernoulli random variables X_i^B with the same parameter p tested *until the first 1 appears* is a geometric random variable X^g.

$$X^g = \min_i \{X_i^B = 1\}$$

To calculate $P(X^g = x)$,

$$P(X^g = 1) = p; \quad P(X^g > 1) = 1 - p$$

$$P(X^g > x) = P(X^g > x | X^g > x - 1) P(X^g > x - 1)$$

$$= (1 - p) P(X^g > x - 1)$$, so

$P(X^g > x) = (1 - p)^x$ and $P(X^g = x) = (1 - p)^{x-1} p$

Discrete Random Variables

Poisson Distribution

$$P(X^P = x) = e^{-\lambda} \frac{\boldsymbol{\lambda}^x}{x!}$$

Discussion later.

Continuous random variables

Philosophical issues

1. *Mathematically* , continuous and discrete random variables are very different.

2. *Quantitatively* , however, some continuous models are very close to some discrete models.

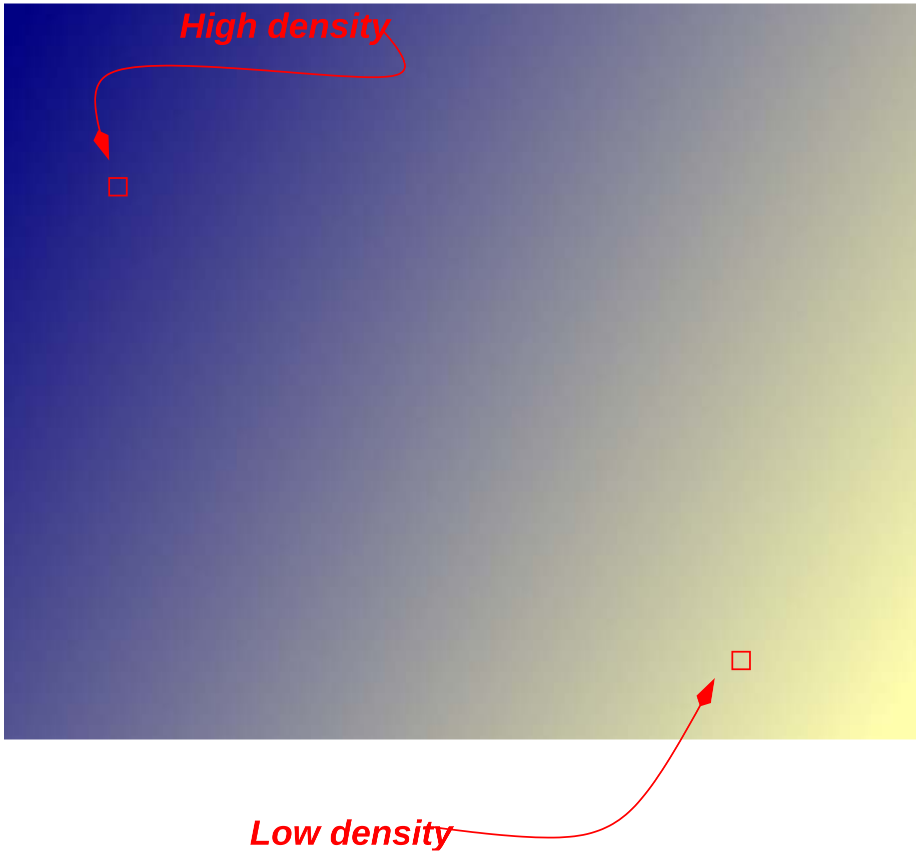
3. Therefore, which kind of model to use for a given system is a matter of *convenience* .

Philosophical issues

Example: The production process for small metal parts (nuts, bolts, washers, etc.) might better be modeled as a continuous flow than as a large number of discrete parts.

Continuous random variables

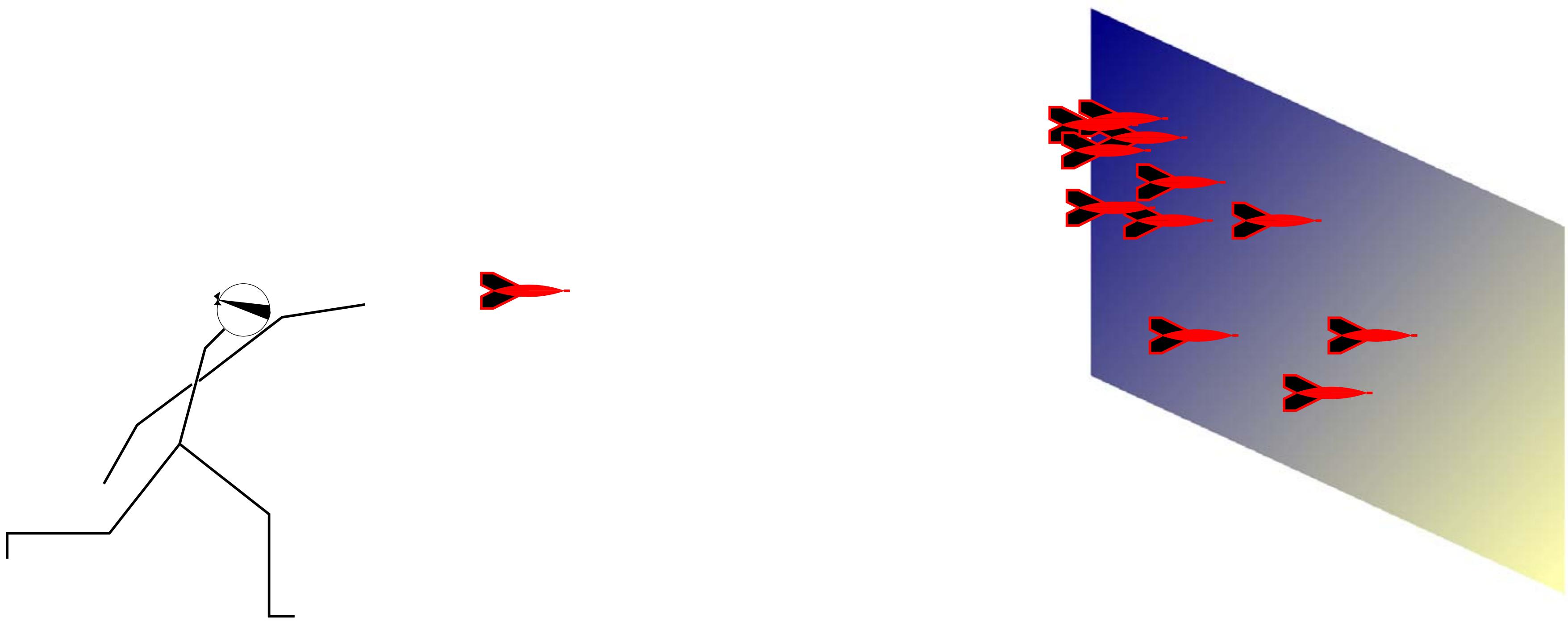
Probability density



The probability of a two-dimensional random variable being in a small square is the *probability density* times the area of the square. (Actually, it is more general than this.)

Continuous random variables

Probability density



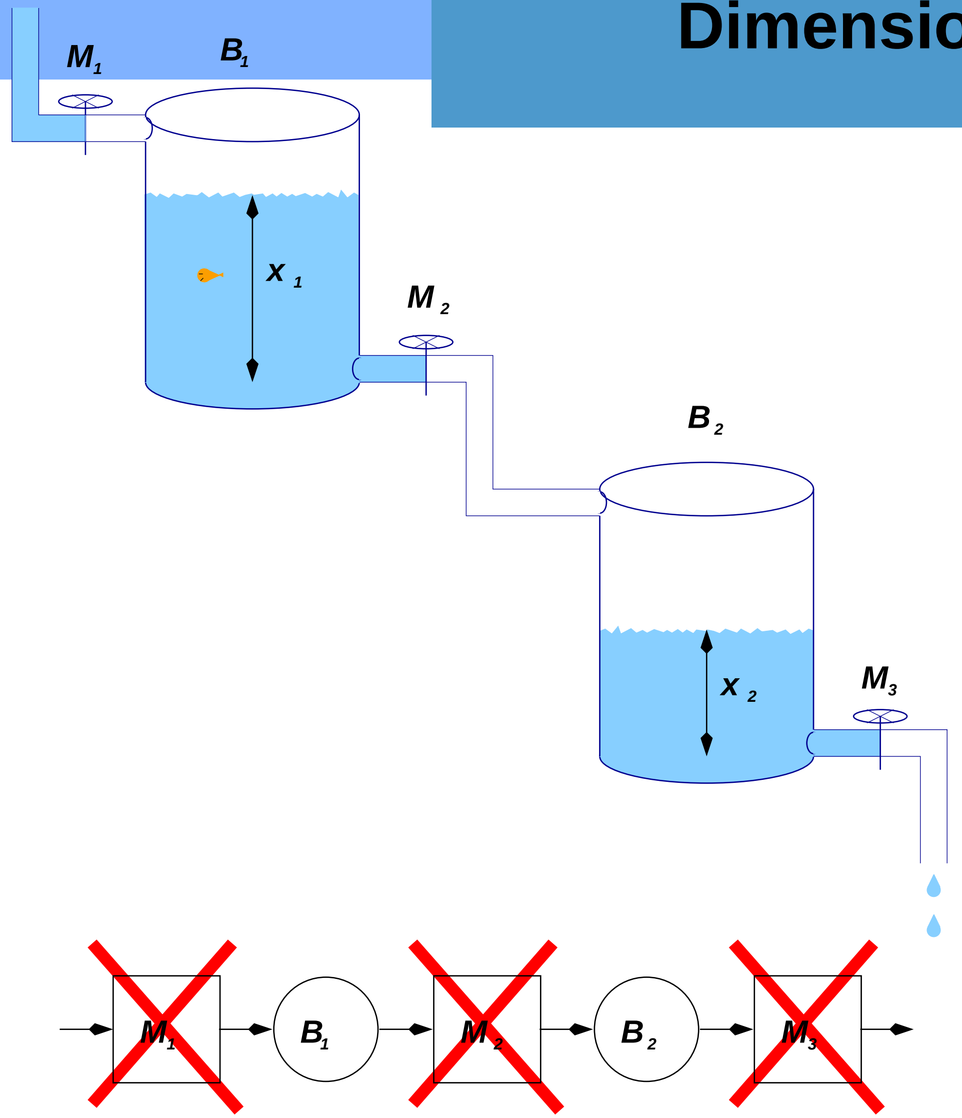
Continuous random variables

- Continuous random variables can be defined
  - ⋆ in one, two, three, ..., infinite dimensional spaces;
  - ⋆ in finite or infinite regions of the spaces.
- Continuous random variables can have
  - ⋆ probability measures with the same dimensionality as the space;
  - ⋆ lower dimensionality than the space;
  - ⋆ a mix of dimensions.

Continuous random variables

Spaces

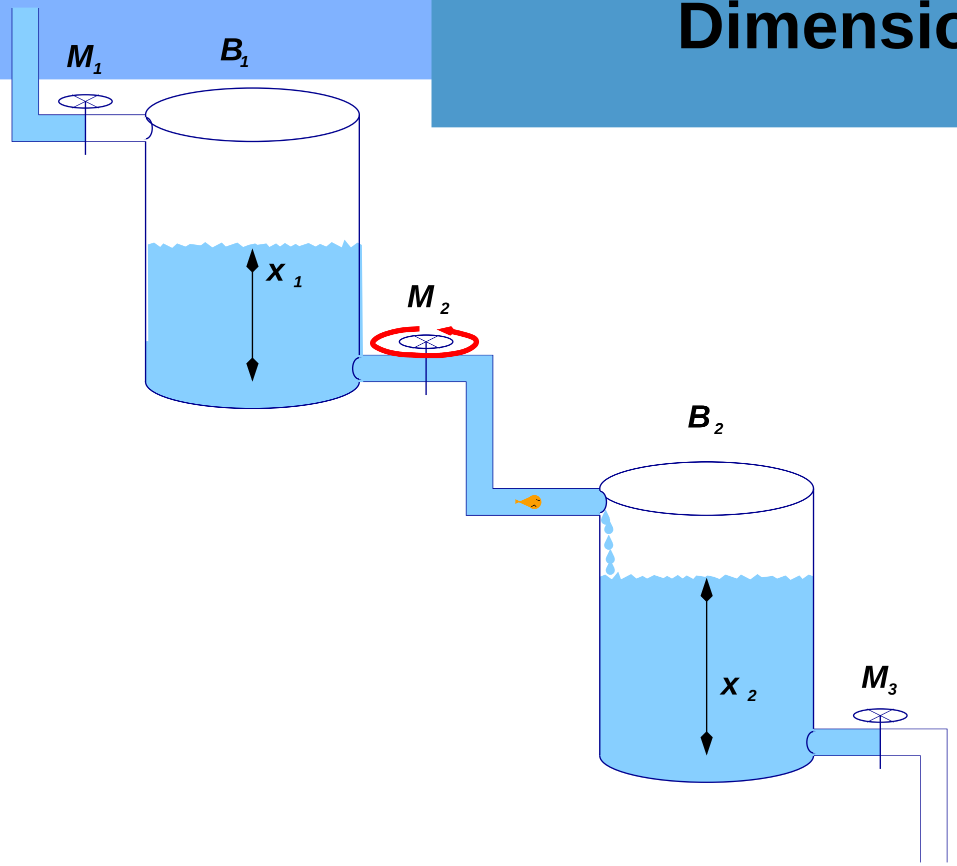
Dimensionality

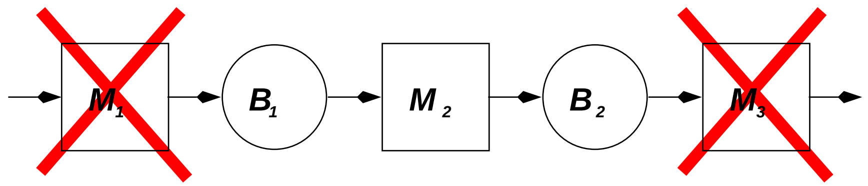


Continuous random variables

Spaces

Dimensionality

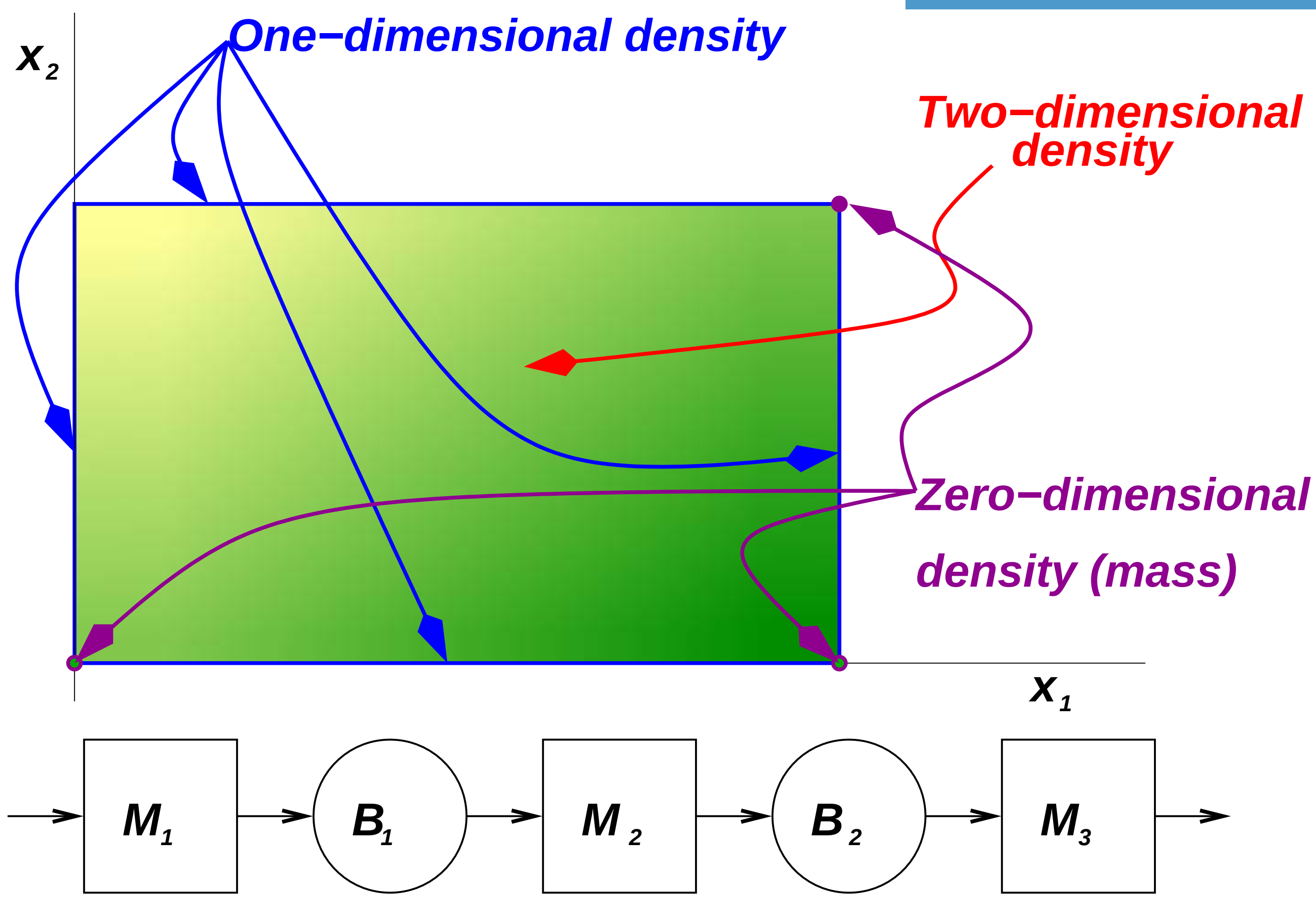




Continuous random variables

Spaces

Dimensionality

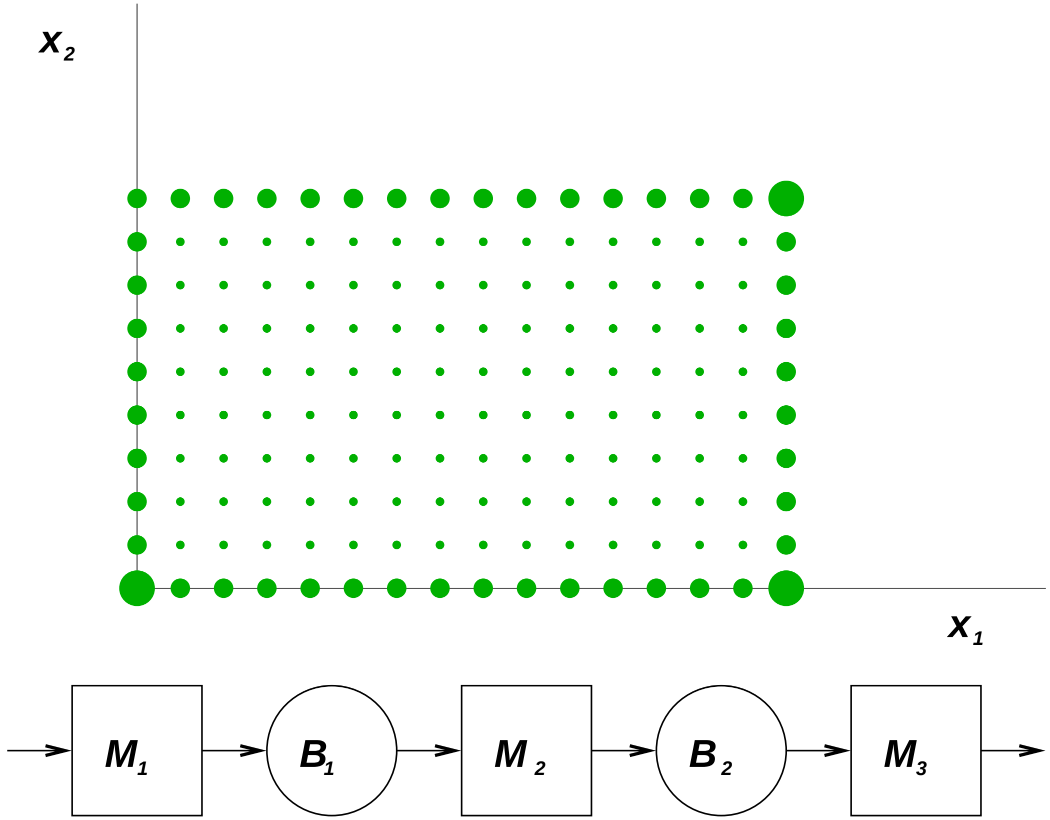


Probability distribution of the amount of material in each of the two buffers.

Continuous random variables

Spaces

Discrete approximation



Probability distribution of the amount of material in each of the two buffers.

Continuous random variables

In one dimension, $\boldsymbol{F}()$ is the *cumulative probability distribution of* $\boldsymbol{X}$ if

$$\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{P}(\boldsymbol{X} \leq \boldsymbol{x})$$

$\boldsymbol{f}()$ is the *density function of* $\boldsymbol{X}$ if

$$\boldsymbol{F}(\boldsymbol{x}) = \int_{-\infty}^{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{t}) \boldsymbol{dt}$$

or

$$\boldsymbol{f}(\boldsymbol{x}) = \frac{\boldsymbol{dF}}{\boldsymbol{dx}}$$

wherever $\boldsymbol{F}$ is differentiable.

Continuous random variables

Fact: $F(b) - F(a) = \int_a^b f(t)dt$

Fact: $f(x)\delta x \approx P(x \leq X \leq x + \delta x)$ for sufficiently small δx.

Definition: $\bar{x} = \int_{-\infty}^{\infty} t f(t)dt$

Normal Distribution

The density function of the *normal* (or *gaussian*) distribution with mean 0 and variance 1 (the *standard normal*) is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

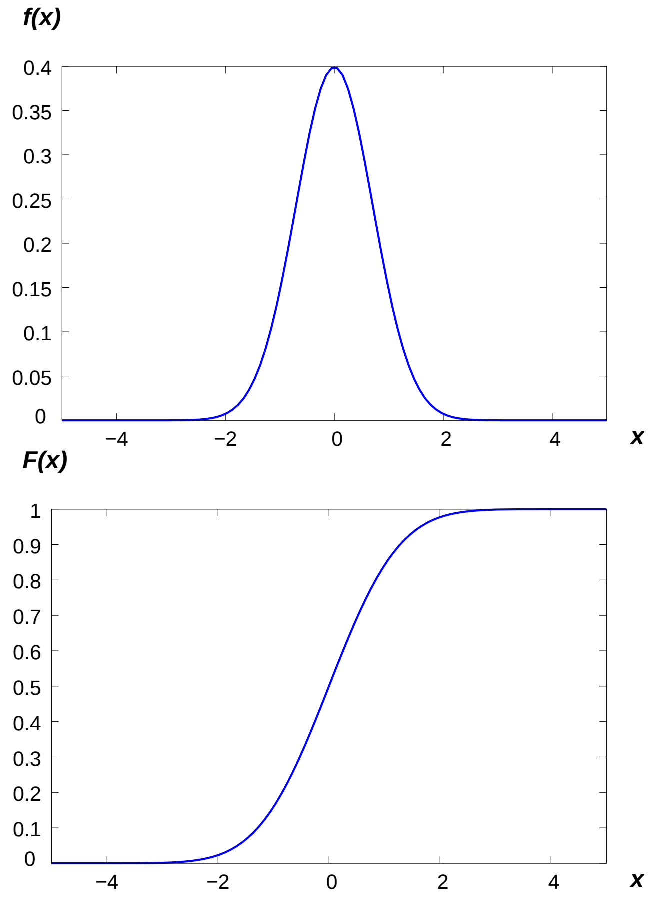
The *normal distribution function* is

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

(There is no closed form expression for $F(x)$.)

Continuous random variables

Normal Distribution



Normal Distribution

Notation: $N(\mu, \sigma)$ is the normal distribution with mean μ and variance σ^2.

Note: Some people write $N(\mu, \sigma^2)$ for the normal distribution with mean μ and variance σ^2.

Fact: If X and Y are normal, then $aX + bY + c$ is normal.

Fact: If X is $N(\mu, \sigma)$, then $\frac{X-\mu}{\sigma}$ is $N(0, 1)$, the standard normal.

This is why $N(0, 1)$ is tabulated in books and why $N(\mu, \sigma)$ is easy to compute.

Law of Large Numbers

Let $\{X_k\}$ be a sequence of independent identically distributed *(i.i.d.)* random variables that have the same finite mean μ. Let S_n be the sum of the first n X_ks, so

$$S_n = X_1 + ... + X_n$$

Then for every $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) = 0$$

That is, the average approaches the mean.

Let $\{X_k\}$ be a sequence of i.i.d. random variables with finite mean μ and finite variance σ^2.

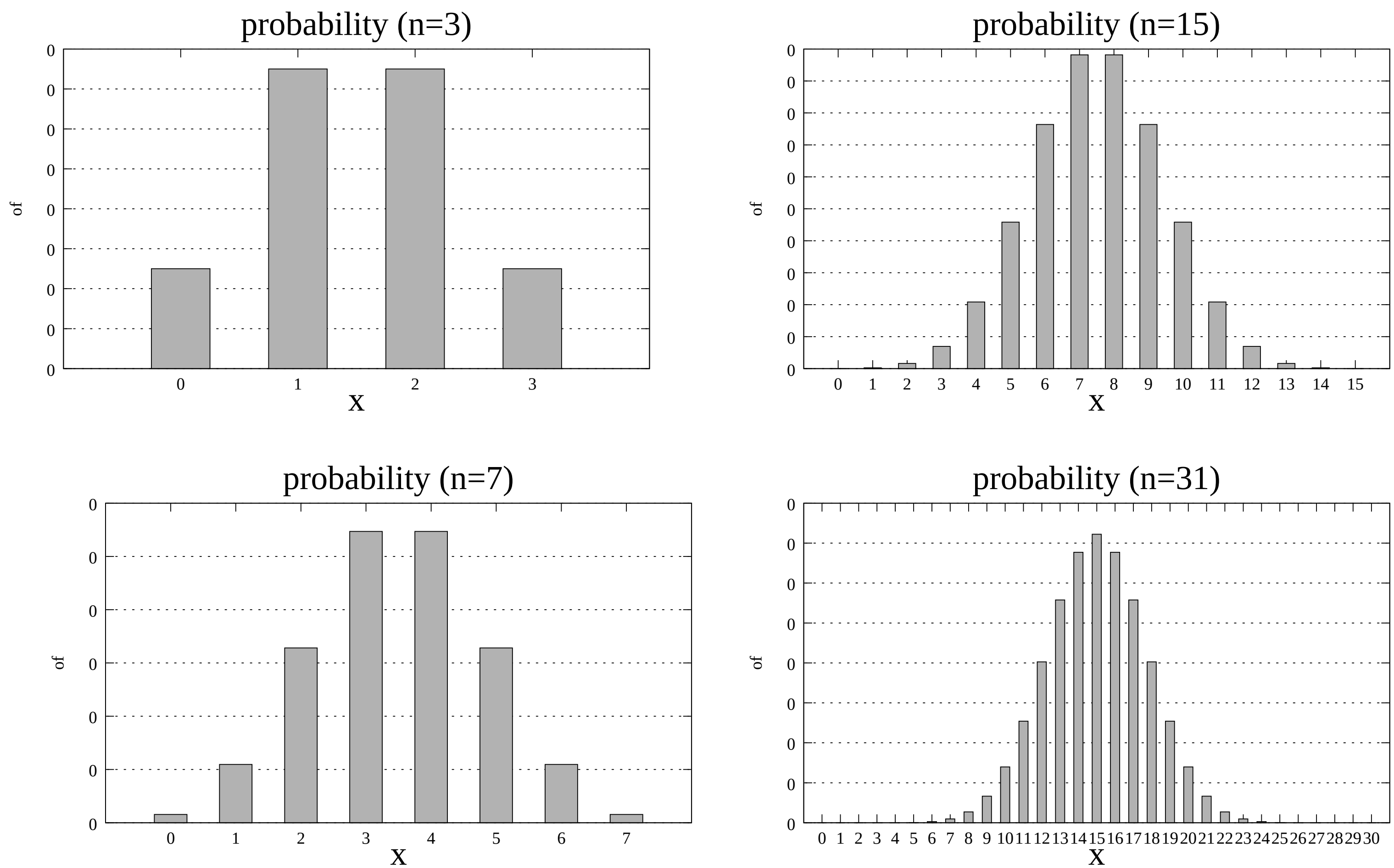
Then as $n \to \infty$, $P(\frac{S_n - n\mu}{\sqrt{n}\sigma}) \to N(0,1)$.

If we define A_n as S_n/n, the average of the first n X_ks, then this is equivalent to:

As $n \to \infty$, $P(A_n) \to N(\mu, \sigma/\sqrt{n})$.

Coin flip examples

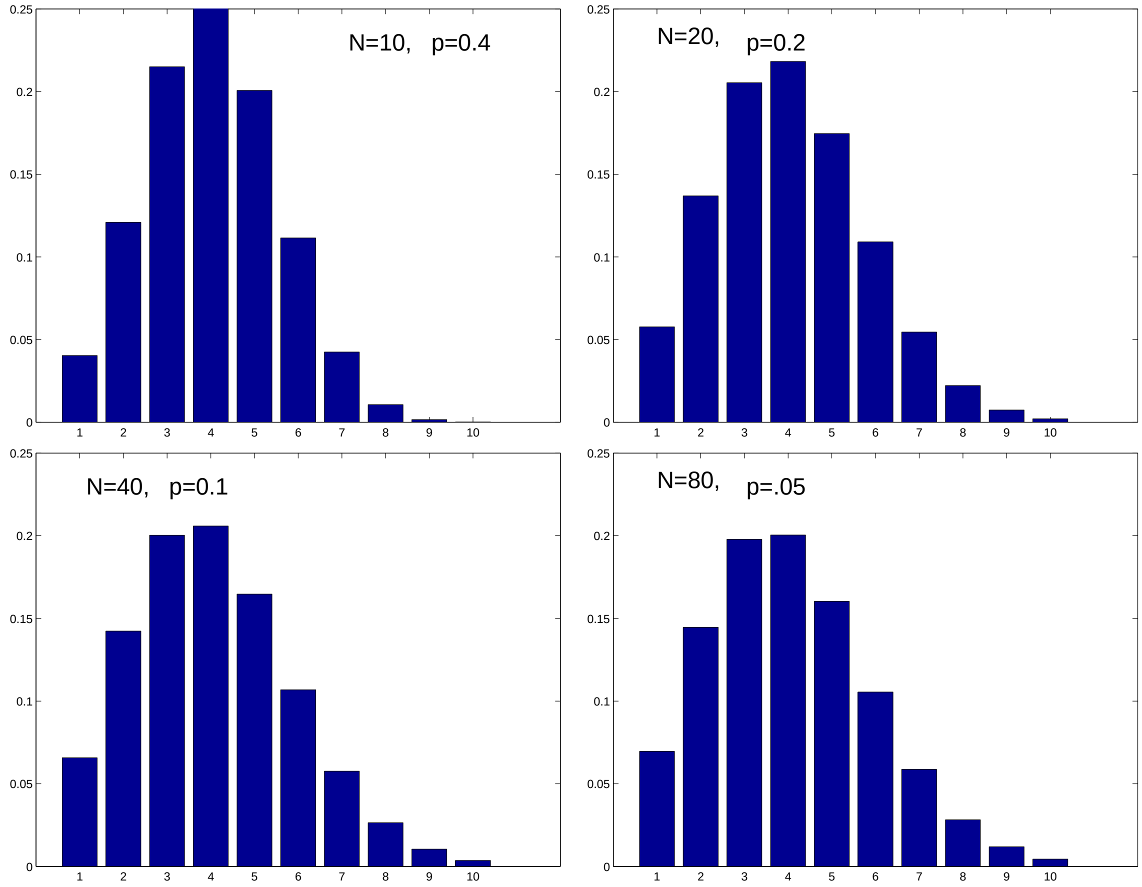
Probability of x heads in n flips of a fair coin



Continuous random variables

Binomial distributions

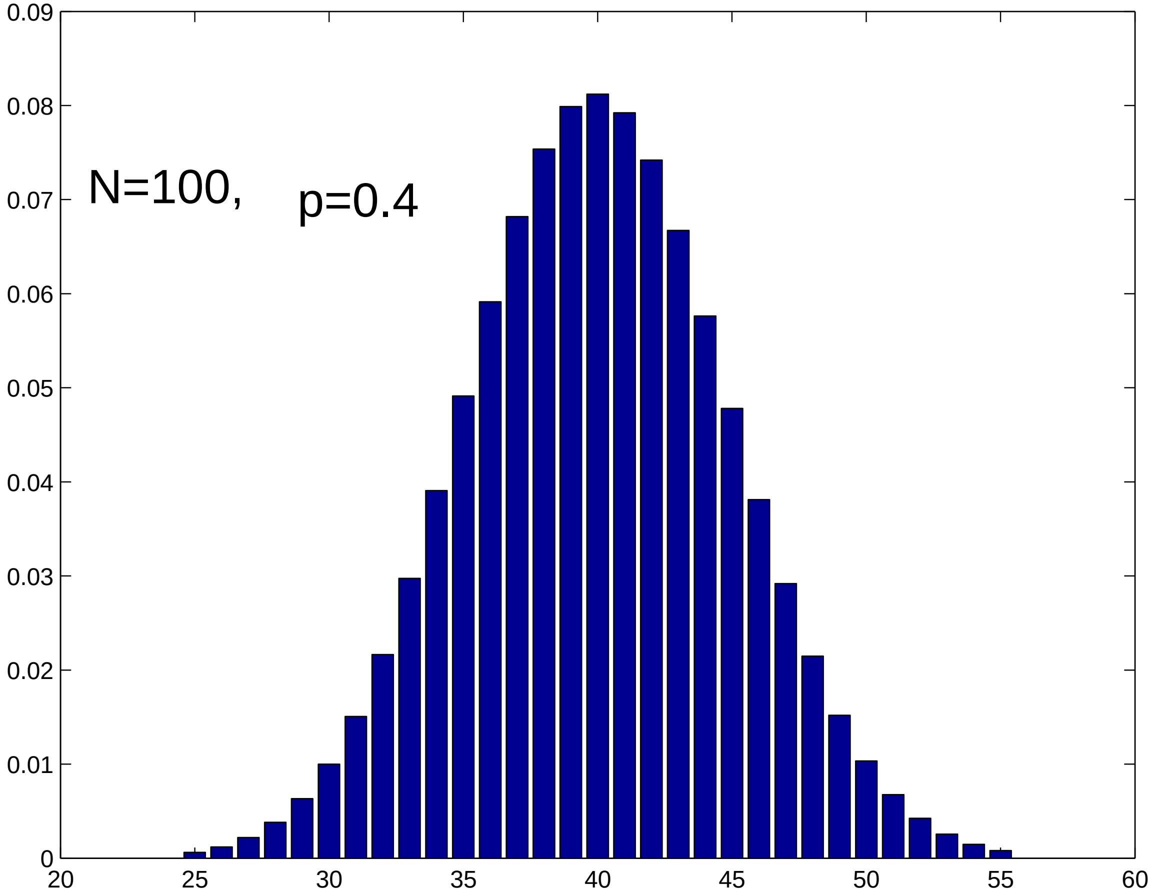
Why are these distributions so similar?



Continuous random variables

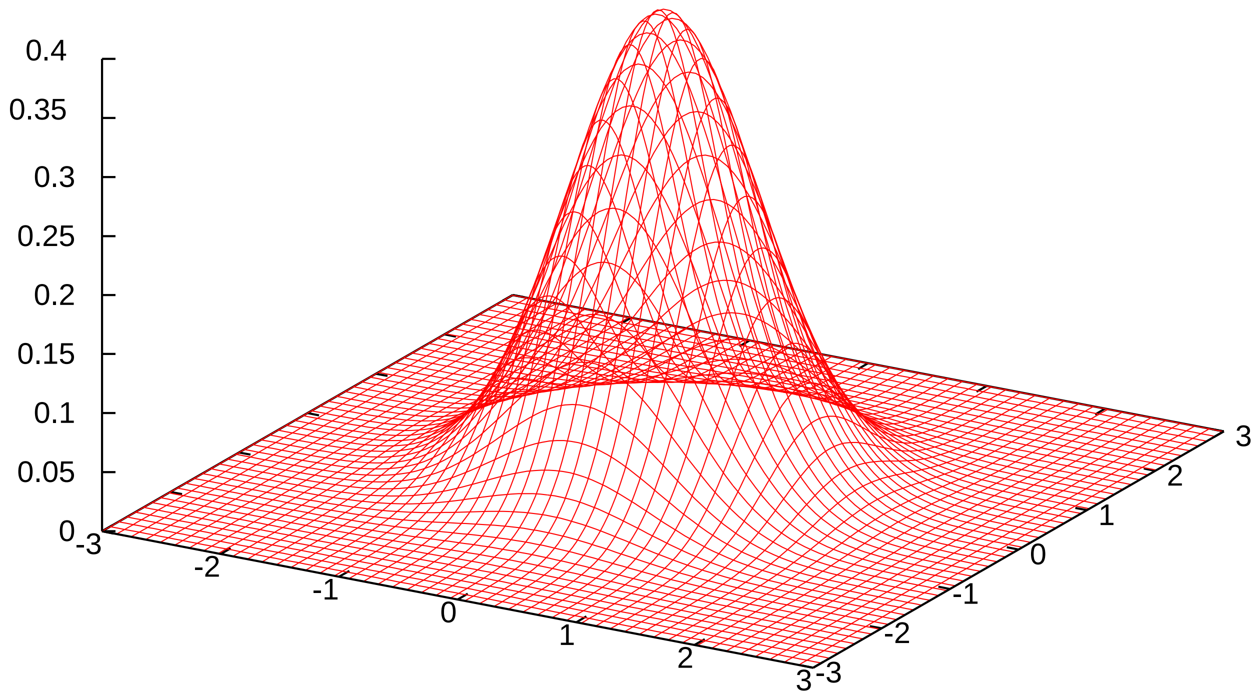
Binomial distributions

Binomial for large N approaches normal.



Normal Density Function

... in Two Dimensions



More Continuous Distributions

Uniform

$$f(x) = \frac{1}{b-a} \quad \text{for } a \leq x \leq b$$

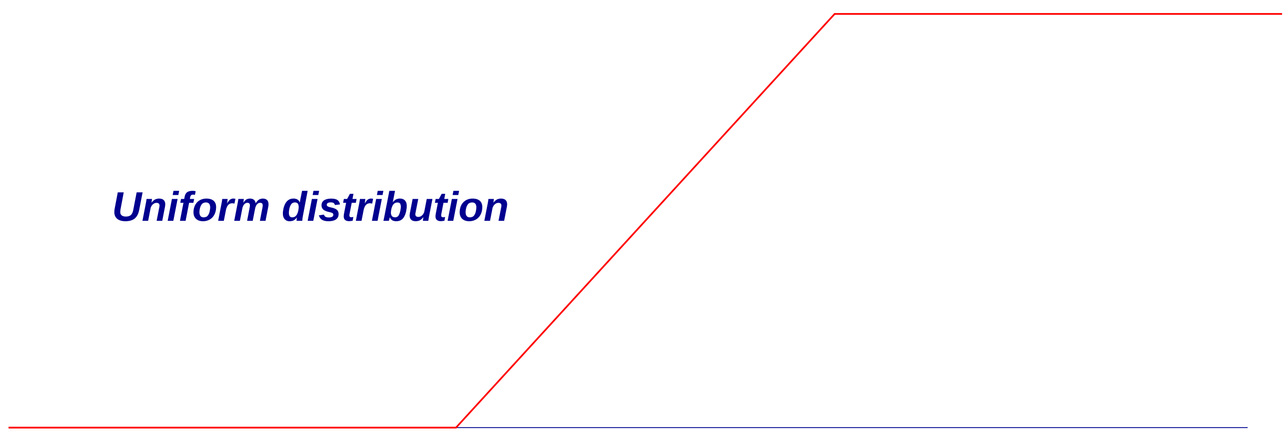
$$f(x) = 0 \qquad \text{otherwise}$$

More Continuous Distributions

Uniform

Uniform density

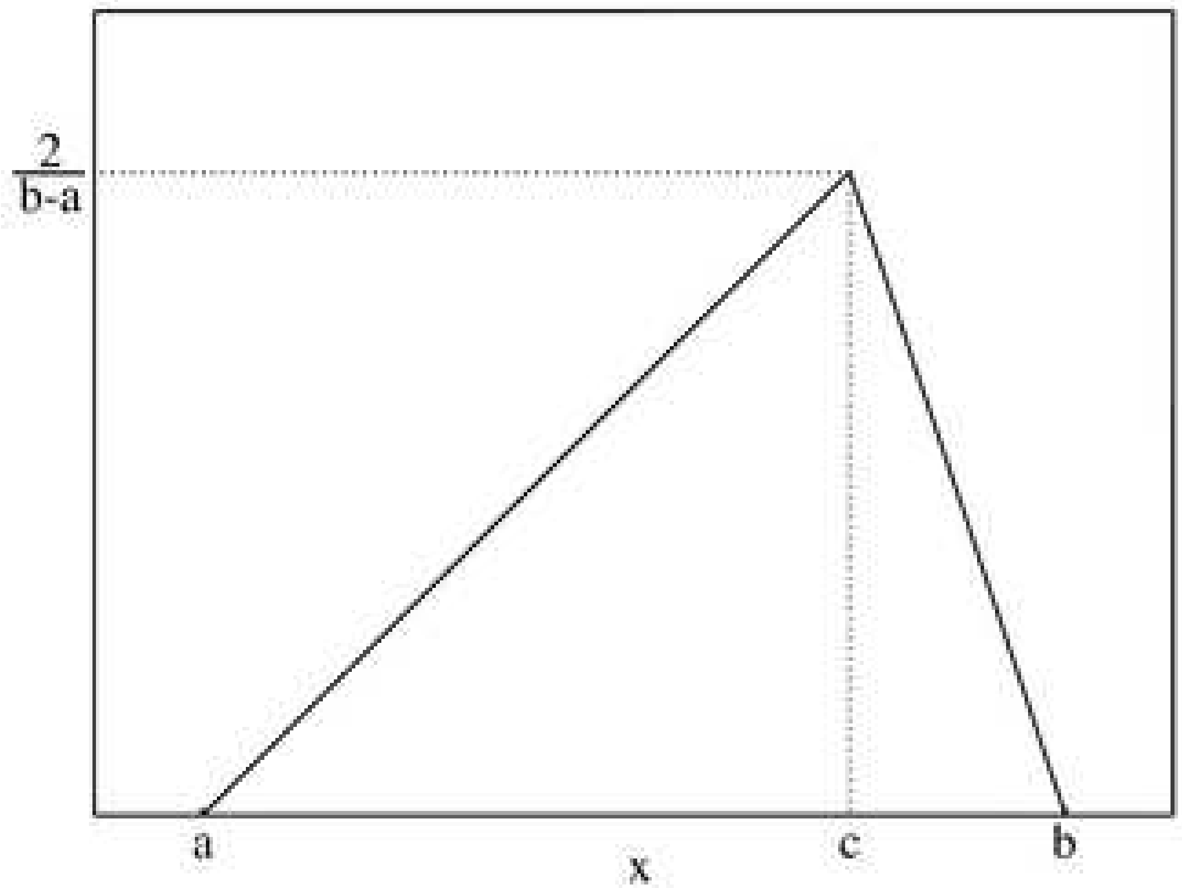
Uniform distribution



More Continuous Distributions

Triangular

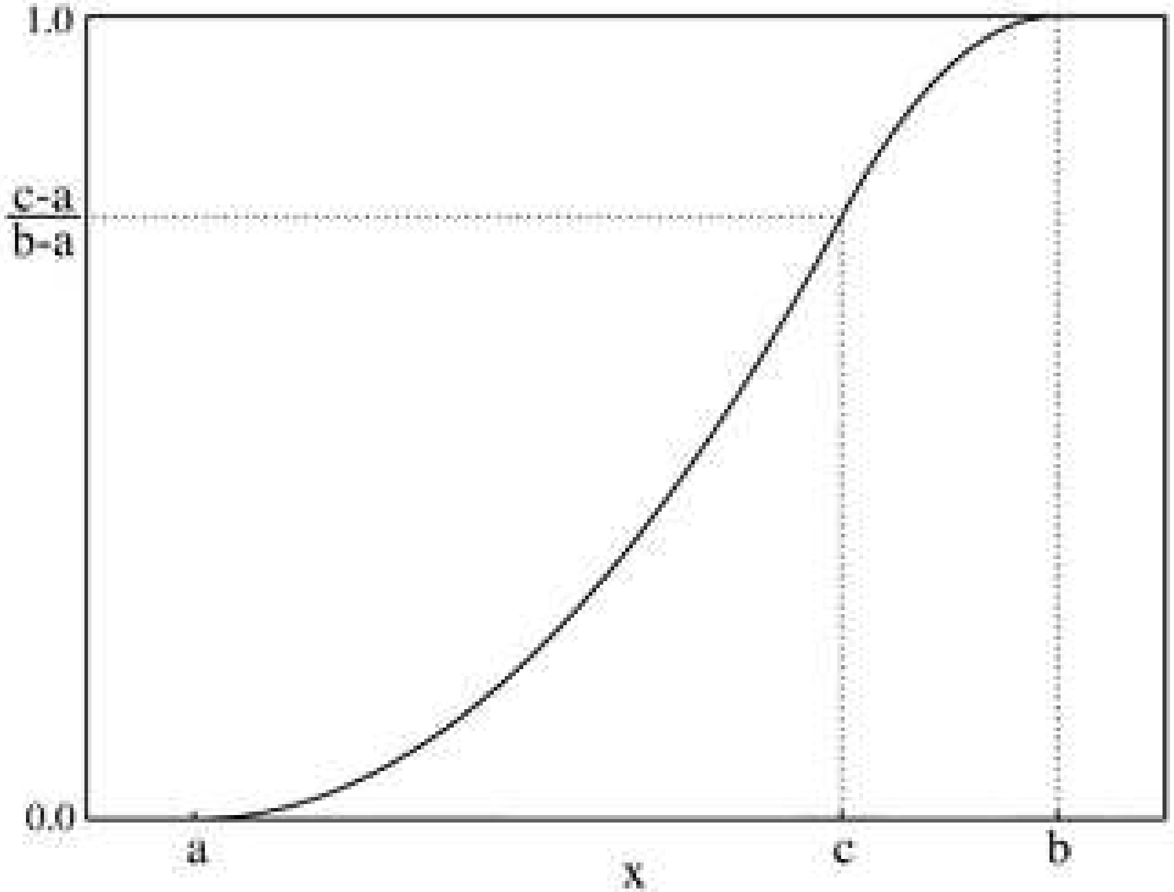
Probability density function



More Continuous Distributions

Triangular

Cumulative distribution function



More Continuous Distributions

Exponential

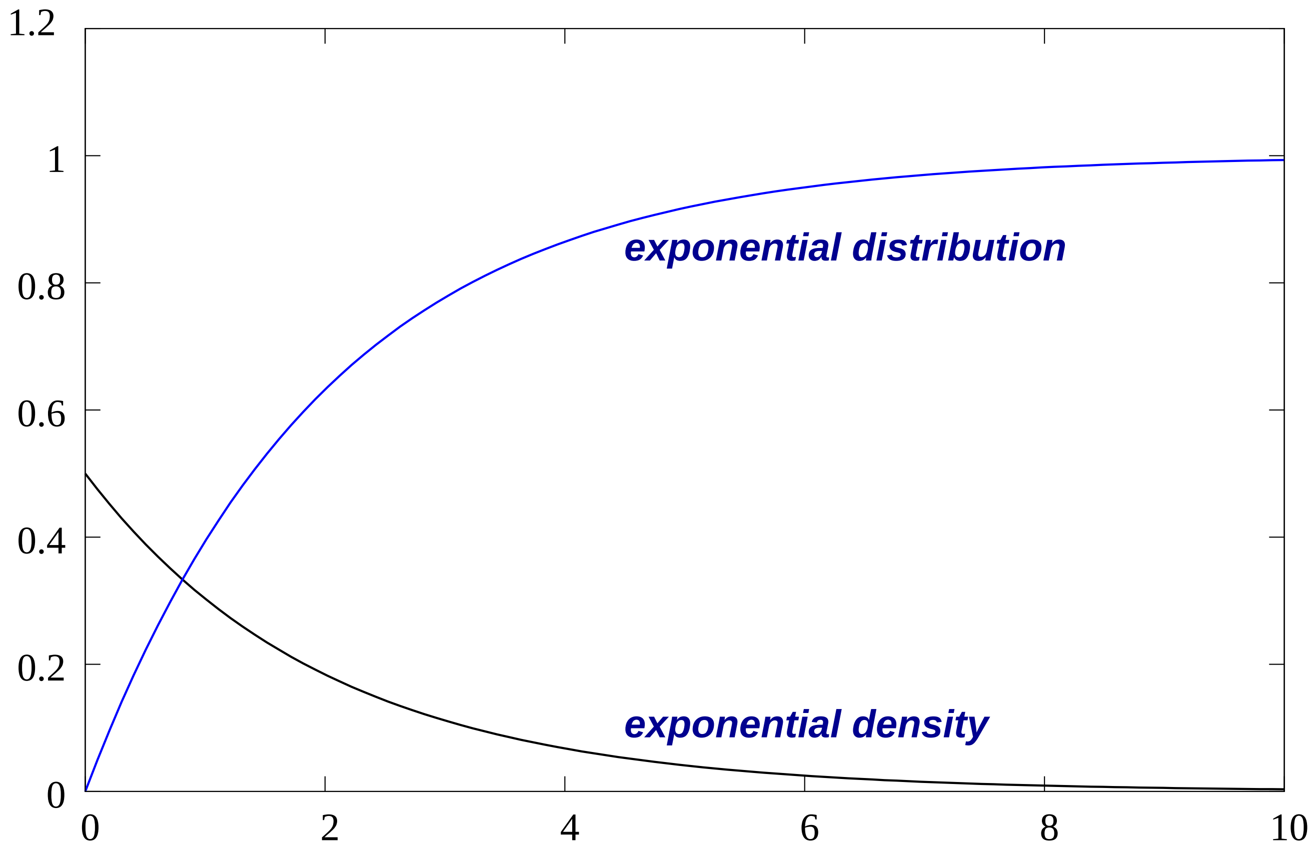
- $f(t) = \lambda e^{-\lambda t}$ for $t \geq 0$; $f(t) = 0$ otherwise;
  $P(T > t) = e^{-\lambda t}$ for $t \geq 0$; $P(T > t) = 1$ otherwise;
- Same as the geometric distribution but for continuous time.
- *Very* mathematically convenient. Often used as model for the first time until an event occurs.
- Memorylessness:
  $P(T > t + x | T > x) = P(T > t)$

The probability distribution

$F(t) = 1 - P(T > t) = 1 - e^{-\lambda t}$ for $t \geq 0$; $F(t) = 0$ otherwise;

More Continuous Distributions

Exponential



More Continuous Distributions

Exponential

Poisson Distribution

$$P(X^P = x) = e^{-\lambda t}\frac{(\lambda t)^x}{x!}$$

is the probability that x events happen in $[0, t]$ if the events are independent and the times between them are exponentially distributed with parameter λ.

Typical examples: arrivals and services at queues.
(Next lecture!)