Statistical Methods in Particle Physics

Lecture 2

October 22, 2012

Silvia Masciocchi, GSI Darmstadt
s.masciocchi@gsi.de

Winter Semester 2012 / 13

Outline

- Uncertainty
- Probability
  - Definition and interpretation
  - Kolmogorov's axioms
- Bayes' theorem

- Random variables
- Probability density functions (pdf)
- Cumulative distribution function (cdf)

Uncertainty

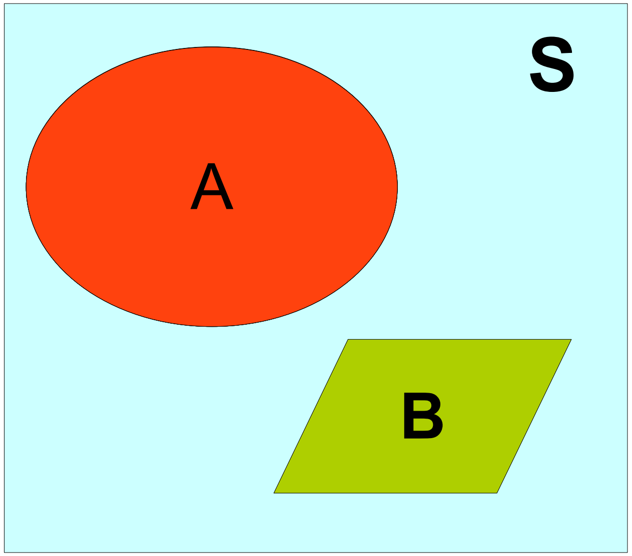
In particle physics there are various elements of uncertainty:

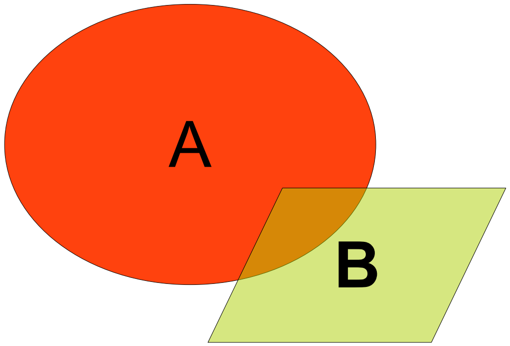
- theory is not deterministic
  - quantum mechanics
- random measurement errors
  - present even without quantum effects
- things we could know in principle but don't
  - e.g. from limitations of cost, time, ...

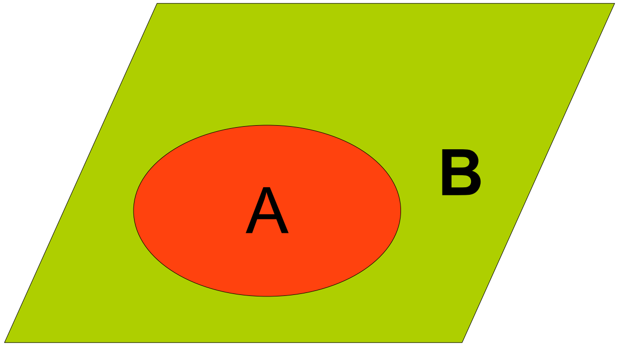
We can quantify the uncertainty using **PROBABILITY**

Some ingredients

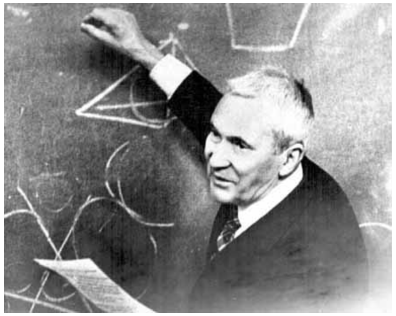
- Set of elements S (sample space)
- Subsets A, B, … of set S
- $A \cup B$: A or B (union of A and B)
- $A \cap B$: A and B (intersection of A and B)
- $\bar{A}$: not A (complementary of A)
- $A \subset B$: A contained in B
- P(A): probability of A (real number)

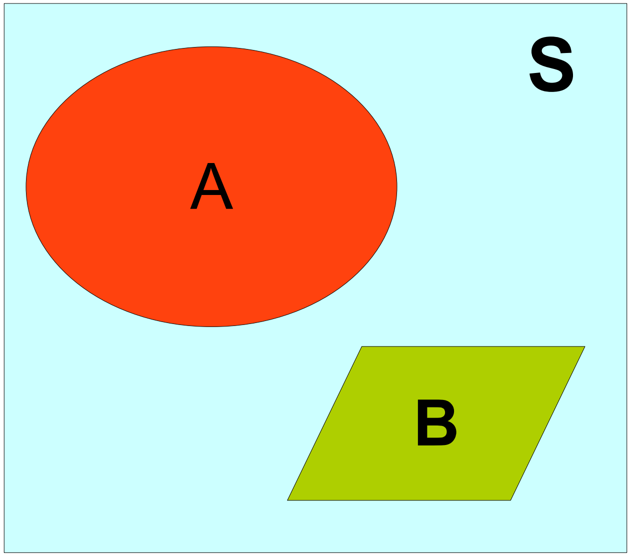






A definition of probability





Kolmogorov's axioms (1933)

1. For all $A \subset S, \;\; P(A) \geq 0$ — Positive definite
2. P(S) = 1 — Normalized
3. If $A \cap B = 0$, then $P(A \cup B) = P(A) + P(B)$ — Additive

   (A and B are *disjoint*)

Further properties, independence

From Kolmogorov's axioms we can derive further properties:

- $P(\bar{A}) = 1 - P(A)$
- $P(A \cup \bar{A}) = 1$
- $P(\emptyset) = 0$
- If $A \subset B$, then $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Subsets A and B are said independent if $P(A \cap B) = P(A)P(B)$

N.B. Do not confuse with disjoint subsets i.e. $A \cap B = 0$

Conditional probability

Define **conditional probability** of A, given B (with $P(B) \neq 0$)

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

Example: rolling dice:

$$P(n<3|n\,\text{even}) = \frac{P((n<3)\cap(n\,\text{even}))}{P(\text{even})} = \frac{1/6}{3/6} = \frac{1}{3}$$

If A and B *independent* (see previous page):

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

Bayes' theorem

From the definition of conditional probability we have:

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

But $P(A \cap B) = P(B \cap A)$, so:

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

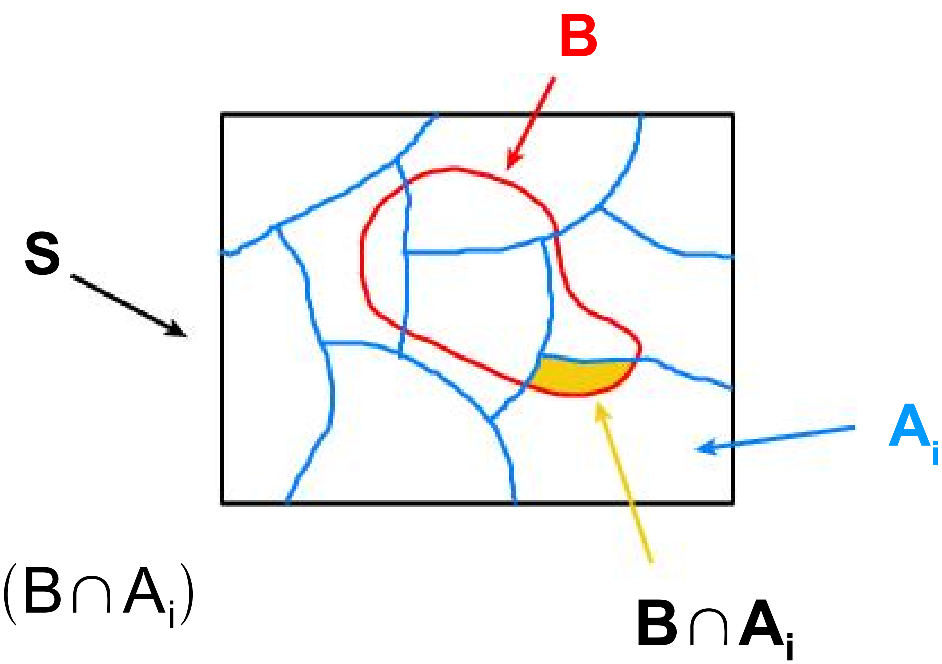
Bayes' Theorem

First published (posthumously) by the
Reverend Thomas Bayes (1702−1761)

An essay towards solving a problem in the doctrine of chances,
Philos. Trans. R. Soc. **53c**(1763) 370; reprinted in Biometrika, **45** (1958) 293.

The law of total probability

Consider a subset *B* of
the sample space *S*,
divided into disjoint subsets A_i
such that $\cup_i A_i = S$:



$$\rightarrow \quad B = B \cap S = B \cap (\cup_i A_i) = \cup_i (B \cap A_i)$$

$$\rightarrow \quad P(B) = P(\cup_i (B \cap A_i)) = \Sigma_i P(B \cap A_i)$$

$$\rightarrow \quad P(B) = \sum_i P(B|A_i)P(A_i)$$

Law of total probability

Bayes' theorem
becomes:

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}}$$

Example: rare disease (1)

Suppose the probability (for anyone) to have the disease A is:

$$P(A) = 0.001$$
$$P(\text{no } A) = 0.999$$

Consider a test for that disease. The result can be 'pos' or 'neg' :

$$P(\text{pos}|A) = 0.98$$
$$P(\text{neg}|A) = 0.02$$

← probabilities to (in)correctly Identify an infected person

$$P(\text{pos}|\text{no } A) = 0.03$$
$$P(\text{neg}|\text{no } A) = 0.97$$

← probabilities to (in)correctly Identify a healthy person

Suppose your result is 'pos'. How worried should you be?

Example: rare disease (2)

The probability to have the disease A, given a 'pos' result is:

$$P(A|pos) =$$

Example: rare disease (2)

The probability to have the disease A, given a 'pos' result is:

$$P(A|\text{pos}) = \frac{P(\text{pos}|A)P(A)}{P(\text{pos}|A)P(A) + P(\text{pos}|\text{no }A)P(\text{no }A)}$$

$$= \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.03 \times 0.999}$$

$$= 0.032$$

i.e. you’re probably OK!

Your viewpoint: my degree of belief that I have disease A is 3.2%

Your doctor’s viewpoint: 3.2% of people like this will have disease A

Interpretation of probability

1. Interpretation of probability as **RELATIVE FREQUENCY** (frequentist approach):

   A, B, ... are outcomes of a repeatable experiment:

$$P(A) = \lim_{n\to\infty} \frac{\text{times outcome is } A}{n}$$

   See quantum mechanics, particle scattering, radioactive decays …

2. **SUBJECTIVE PROBABILITY**

   A, B, ... are hypotheses (statements that are true or false)

   P(A) = degree of belief that A is true

In particle physics, frequency interpretation often most useful, but subjective probability can provide a more natural treatment of non-repeatable phenomena

(systematic uncertainties, probability that higgs exists ...)

Frequentist statistics

In frequentist statistics, probabilities are associated only with the data, i.e., outcomes of repeatable observations

Any given experiment can be considered as one of an infinite sequence of possible repetitions of the same experiment, each capable of producing statistically independent results:

Perform experiment N times in identical trials; assume event E occurs k times, then

$$P(E) = \lim_{N \to \infty} k/N$$

BUT:

- Does the limit converge? How large needs N to be?
- What means identical condition? Can 'similar' be sufficient?
- Not applicable for single events

Bayesian probability

In Bayesian statistics, use subjective probability for hypotheses (degree of belief that an hypothesis is true):

Probability of the data assuming hypothesis H (the likelihood)

Prior probability (before seeing the data)

$$P(H|\vec{x}) = \frac{P(\vec{x}|H)\pi(H)}{\int P(\vec{x}|H)\pi(H)dH}$$

Posterior probability (after seeing the data)

Normalization involves sum over all possible hypothesis

Bayes' theorem has an "if-then" character: If your prior probabilities were π(H), then it says how these probabilities should change in the light of the data.

No general prescription for priors (subjective!)

Back to Bayes' theorem

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

Now take: A = theory, B = data:

Likelihood **Prior**

$$P(\text{theory}|\text{data}) = \frac{P(\text{data}|\text{theory})P(\text{theory})}{P(\text{data})}$$

Posterior **Evidence**

Exercises

1. Show that:

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

2. A beam of particles consists of a fraction 10^{-4} electrons and the rest photons. The particles pass through a double-layered detector which gives signals in either zero, one or both layers. The probabilities of these outcomes for electrons (e) and photons (γ) are:

$$P(0|e) = 0.001 \qquad P(0|\gamma) = 0.99899$$
$$P(1|e) = 0.01 \qquad P(1|\gamma) = 0.001$$
$$P(2|e) = 0.989 \qquad P(2|\gamma) = 10^{-5}$$

(a) what is the probability for a particle detected in one layer only to be a photon?

(b) what is the probability for a particle detected in both layers to be an electron?

Exercises

3. Detector for particle identification

   In proton-proton collisions we have: 90% pions, 10% kaons

   1. Kaon identification: 95% efficient
   2. Pion misidentification: 6%

   Question: if the particle identification indicates a kaon, what is the probability that it is a real kaon / a real pion?

Hints

1. Express AUB as the union of three disjoint sets

2.

3. $$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|\text{no}\,A)[1-P(A)]}$$

Random variables

A random variable is a variable whose value results from a measurement on some type of random process.

Formally, it is a function from a probability space, typically to the real numbers, which is measurable.

Intuitively, a random variable is a numerical description of the outcome of an experiment

e.g., the possible results of rolling two dice: (1, 1), (1, 2), etc.

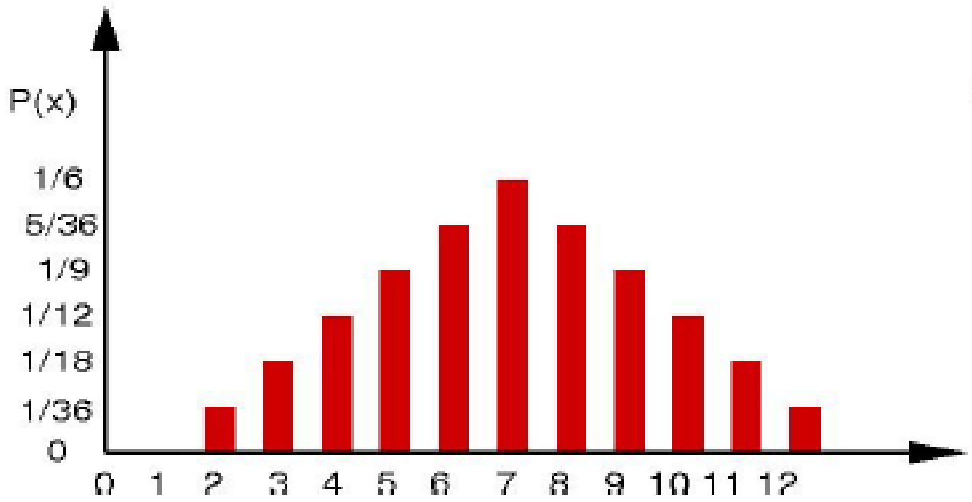
Random variables can be classified as:

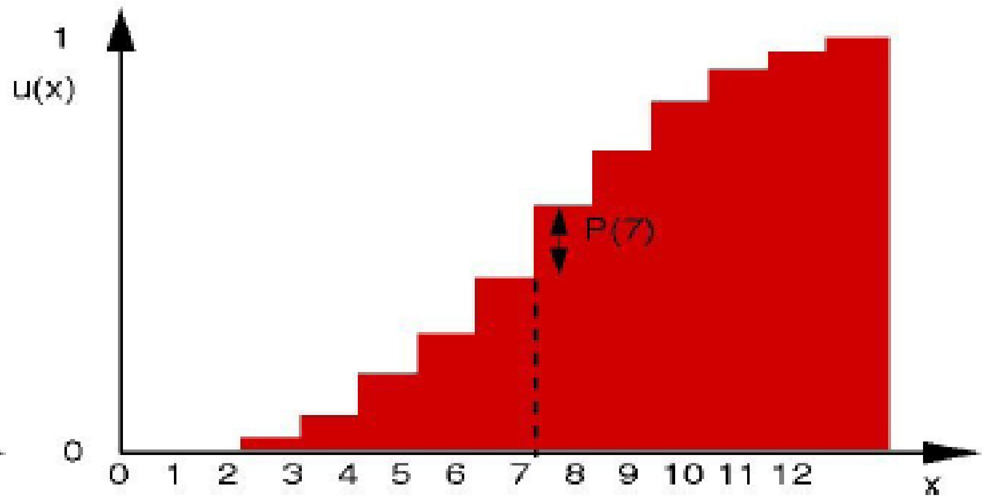
- **discrete** (a random variable that may assume a finite number of values)
- **continuous** (a variable that may assume any numerical value in an interval or collection of intervals).

Random variables

Typical example: throwing two dices

- Result of each "experiment":
  {11, 12, 13, 14, 15, 16, 21, 22, … , 63, 64, 65, 66}
- Random variable x = sum of dices
  → possible values (discrete!): x_i = 2, 3, 4, … , 11, 12
- Probability for each value x_i





$$\sum_i P(x_i) = 1$$

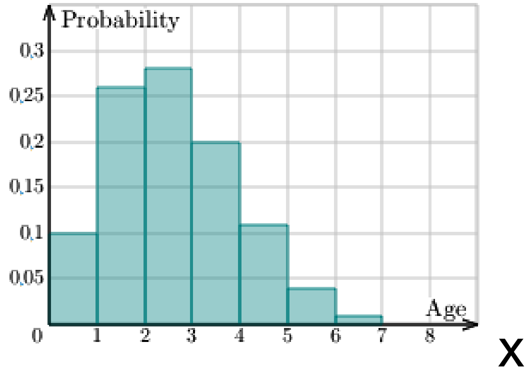
Cumulative distribution u(x):
probability to observe x or smaller value

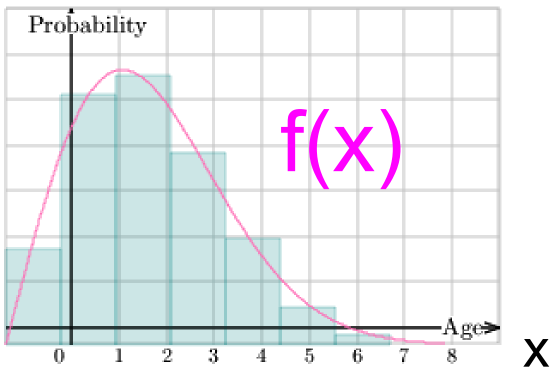
Discrete and continuous

One more example: age of a rented car (x)

Age (Years)	0-1	1-2	2-3	3-4	4-5	5-6	6-7
Probability	.10	.26	.28	.20	.11	.04	.01

The histogram of probabilities can be described by a function f





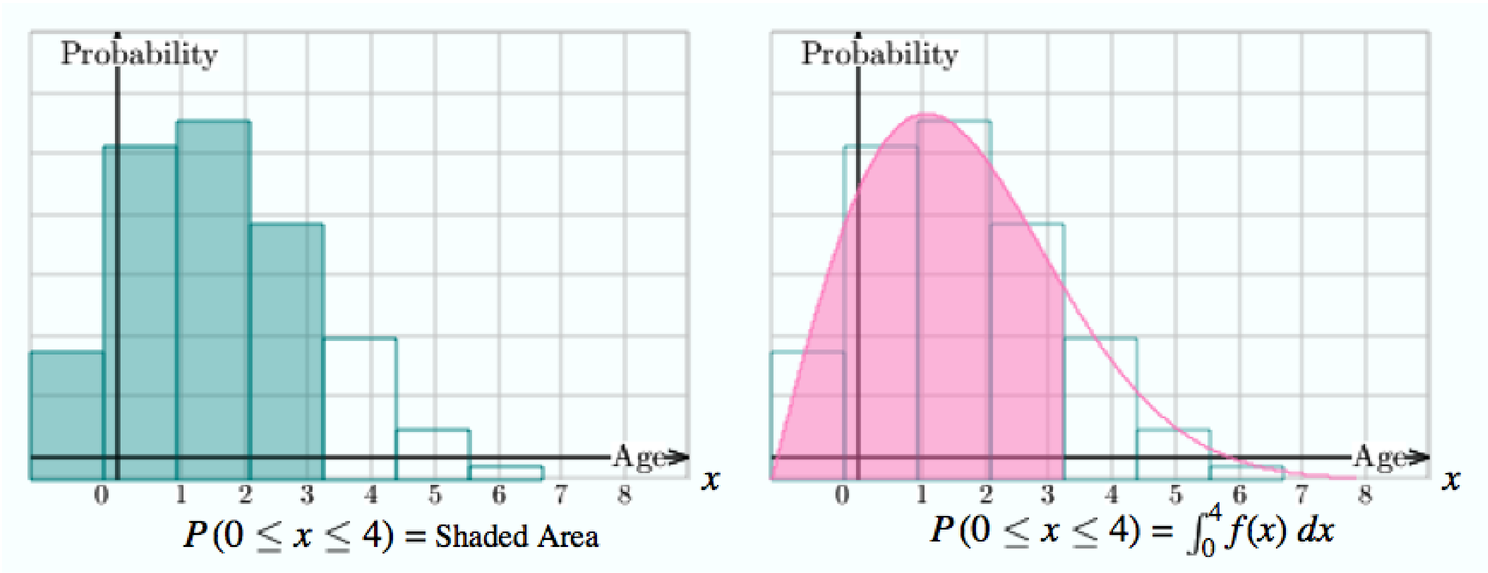
f(x) is called **probability density function**.

The domain of f is the whole range of values which x can take.

We use it to calculate the probability of the given variable x to be in an interval [a,b]

Discrete and continuous

Probability for the rental car to have age between 0 and 4 years:



Probability density functions

Suppose outcome of experiment is **continuous** value x:

$$P(x \text{ found in } [x, x+dx]) = f(x)dx$$

→ **f(x) = probability density function (pdf)**

With:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Normalization
(x must be somewhere)

Note:

- $f(x) \geq 0$
- f(x) is NOT a probability ! It has dimension 1/x !

Probability density functions

Otherwise, for a **discrete** outcome x_i, with i=1,2,... :

$$P(x_i)=p_i$$ Probability mass function

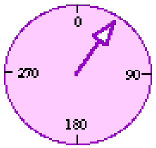
$$\sum_i P(x_i)=1$$ x must take on one of its possible values

Exercises

1. For what constant k is $f(x)=ke^{-x}$ a probability density function on [0,1]?

   If f is any non-negative function with domain some interval (a,b), then the process of choosing a suitable constant k to make $\int_a^b k\, f(x)dx = 1$ is called **normalizing** the function f

2. Suppose that you spin the dial shown in the figure so that it comes to rest at a random position. Model this with a suitable probability density function, and use it to find the probability that the dial will land somewhere between 5° and 300°.



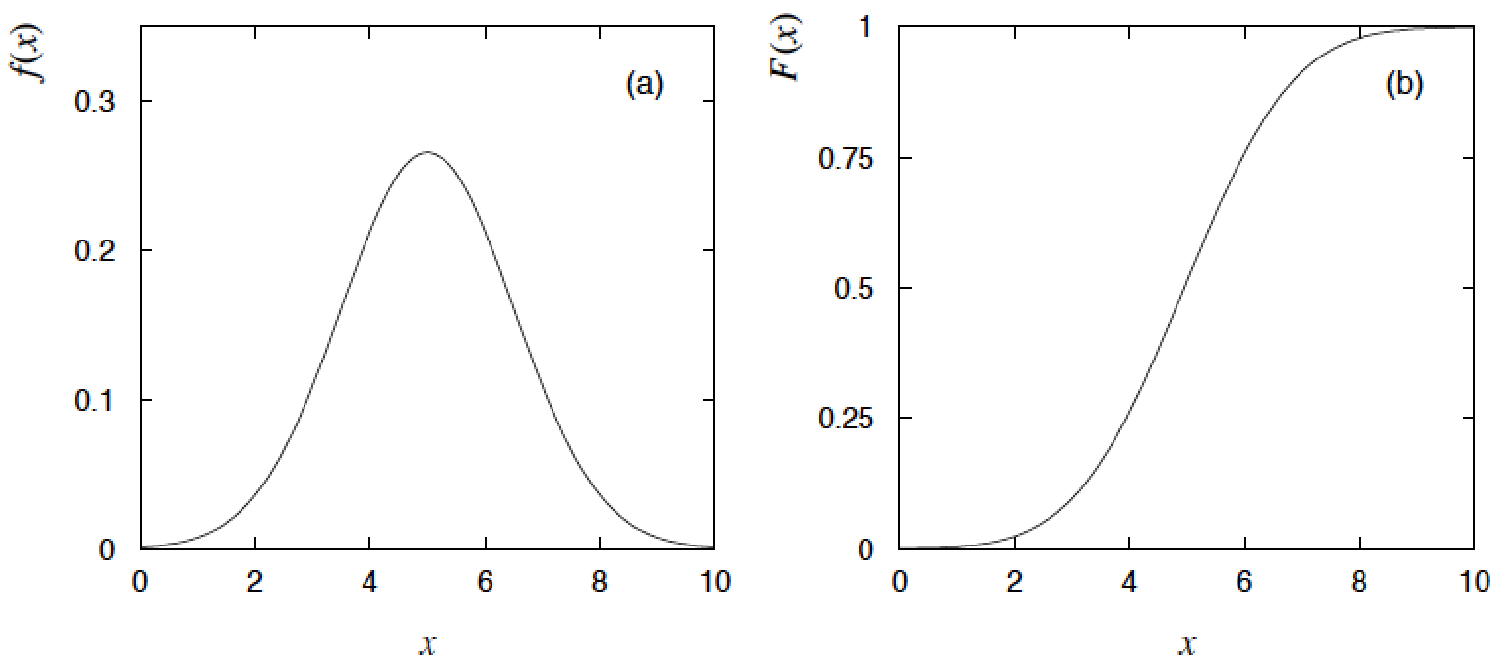
   The **uniform density function** on the interval [a,b] is the constant function defined by $f(x) = \frac{1}{b-a}$

Cumulative distribution function (cdf)

Given a pdf f(x'), probability to have outcome less then or equal to x, is:

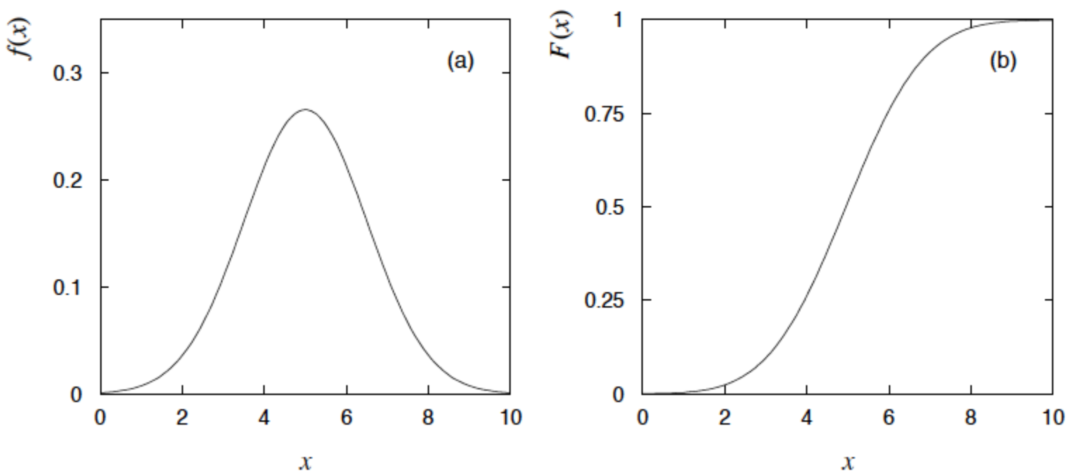
$$\int_{-\infty}^{x} f(x')dx' = F(x)$$

Cumulative distribution function



Cumulative distribution function (cdf)

$$\int_{-\infty}^{x} f(x')dx' = F(x)$$ **Cumulative distribution function**



- F(x) is a continuously non-decreasing function
- $F(-\infty)= 0, F(\infty)=1$
- For well behaved distributions:

$$\text{pdf}: \quad f(x) = \frac{\partial F(x)}{\partial x}$$

Exercise

1. Given the probability density function:

$$f(x) = \begin{cases} |1-x| & \text{for } x \text{ in } [0,2] \\ 0 & \text{elsewhere} \end{cases}$$

- compute the cdf F(x)
- what is the probability to find x > 1.5 ?
- what is the probability to find x in [0.5,1] ?

Histograms

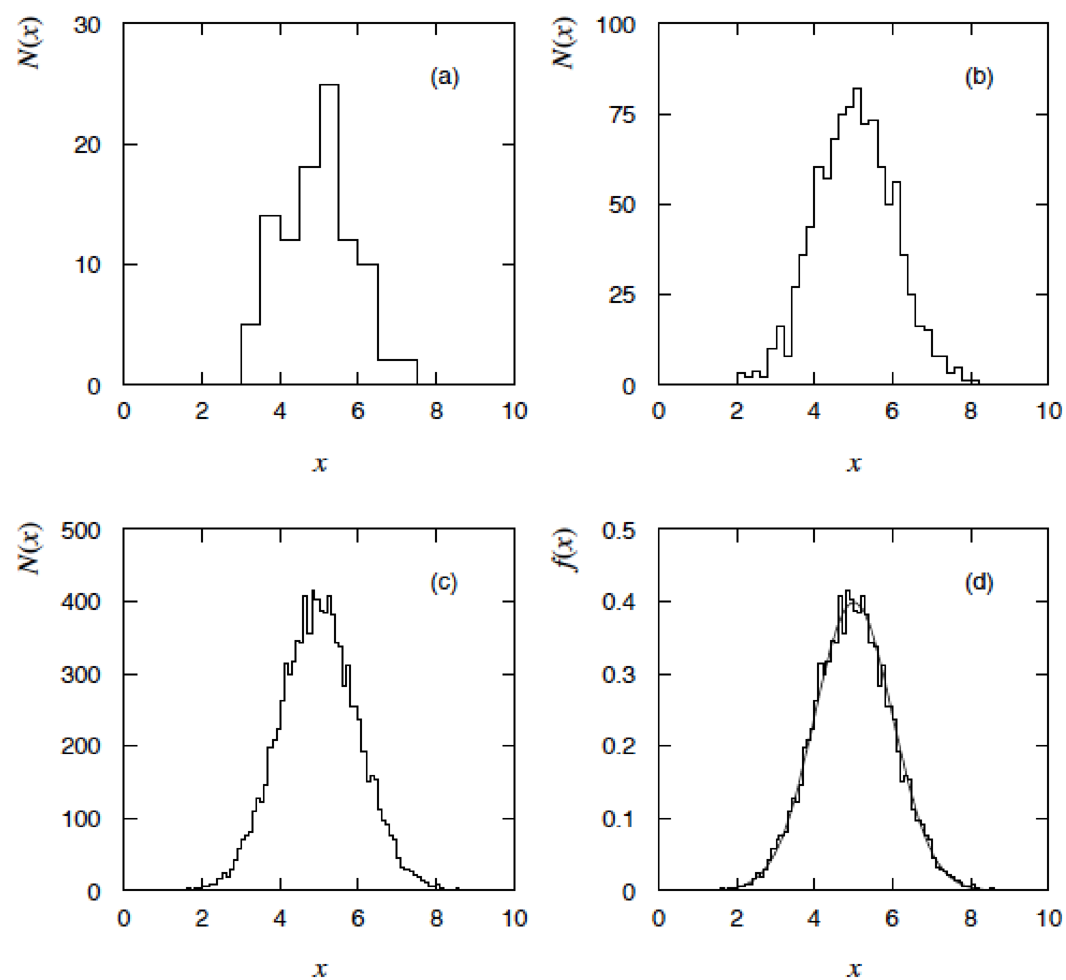
A probability density function can be obtained from a histogram at the limit of:

- Infinite data sample
- Zero bin width
- Normalized to unit area

$$f(x) = \frac{N(x)}{n \Delta x}$$

N = number of entries

Δx = bin width



Exercises

- **T-Shirts** The age (in years) of randomly chosen T-shirts in your wardrobe from last summer is distributed according to the density function $f(x)=10/9x^2$ with $1 \le x \le 10$. Find the probability that a randomly chosen T-shirt is between 2 and 8 years old.

- **The Doomsday Meteor** The probability that a "doomsday meteor" will hit the earth in any given year and release a billion megatons or more of energy is on the order of 0.000 000 01. If X is the year in which a doomsday meteor hits the earth, then it may be modeled with an associated probability density function given by $f(x)=ae^{-ax}$ with a=0000 000 01.

  (a) What is the probability that the earth will be hit by a doomsday meteor at least once during the next 100 years? (Give the answer correct to 2 significant digits.)

  (b) What is the probability that the earth has been hit by a doomsday meteor at least once since the appearance of life (about 4 billion years ago)?

Wrapping up lecture 2, next time

This lecture:
- Abstract properties of probability
- Axioms
- Interpretations of probability
- Bayes' theorem
- Random variables
- Probability density function
- Cumulative distribution function

Next time:
- Expectation values
- Error propagation
- Catalog of pdfs