APPLIED MACHINE LEARNING

Recap: Statistics

Discrete Probabilities
Probability Density Functions

Discrete Probabilities

Consider two random variables x and y taking discrete values over the intervals $[1.....N_x]$ and $[1.....N_y]$ respectively.

$P(x=i)$: the probability that the variable x takes value $i \in [1.....N_x]$.

Two properties:

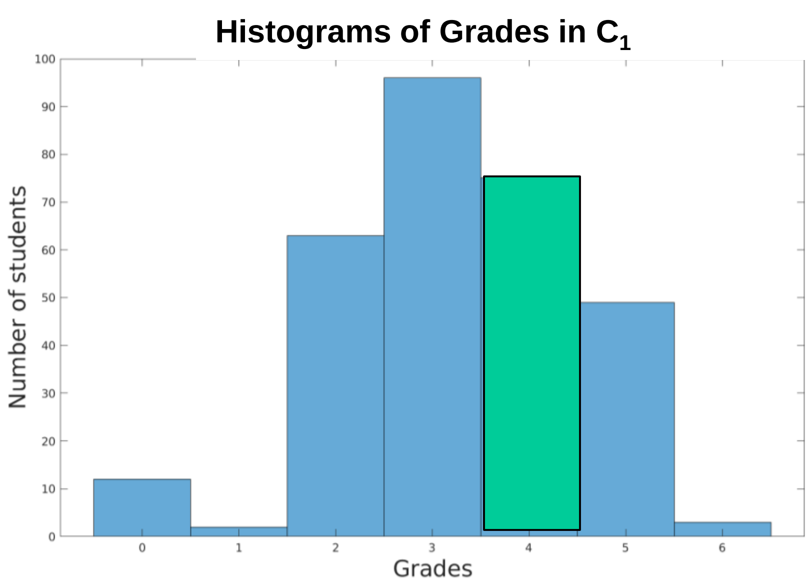
$$0 \le P(x=i) \le 1, \ \forall i = 1,...,N_x,$$

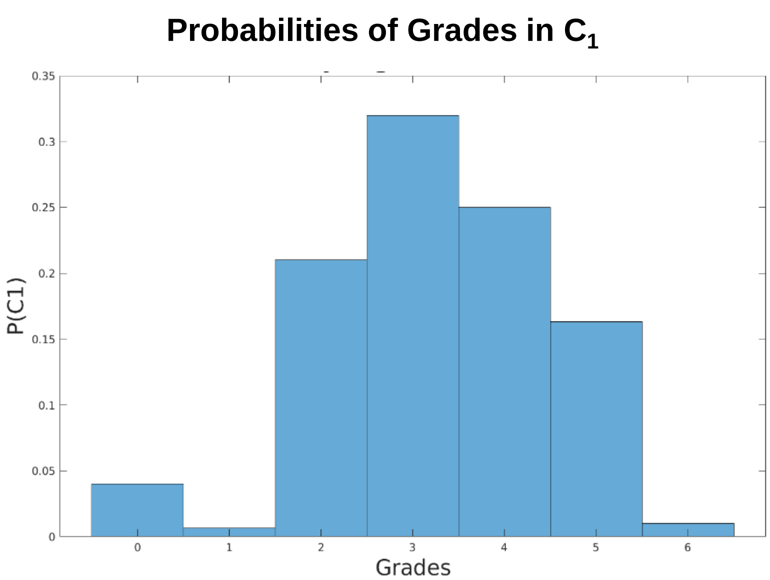
$$\sum_{i=1}^{N_x} P(x=i) = 1.$$

Idem for $P(y=j),\ j=1,...N_y$

Discrete Probabilities

Distribution of grades of 300 students in courses C_1



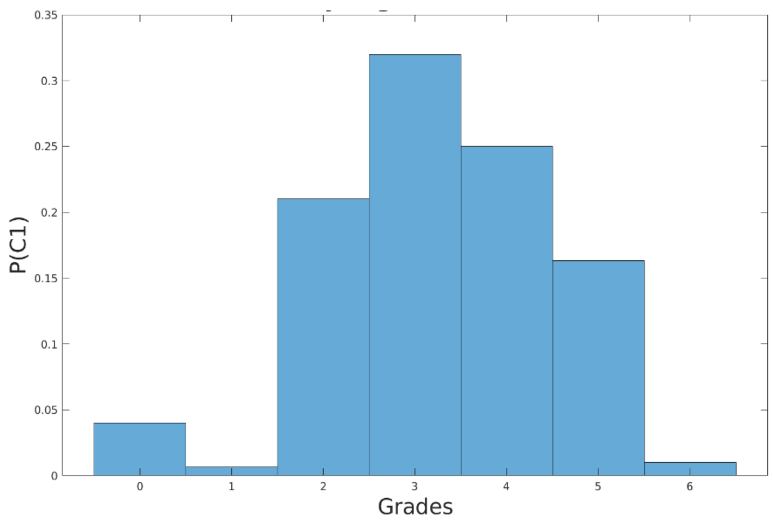


What is the probability that a student receives a grade of 4 in C_1 ?

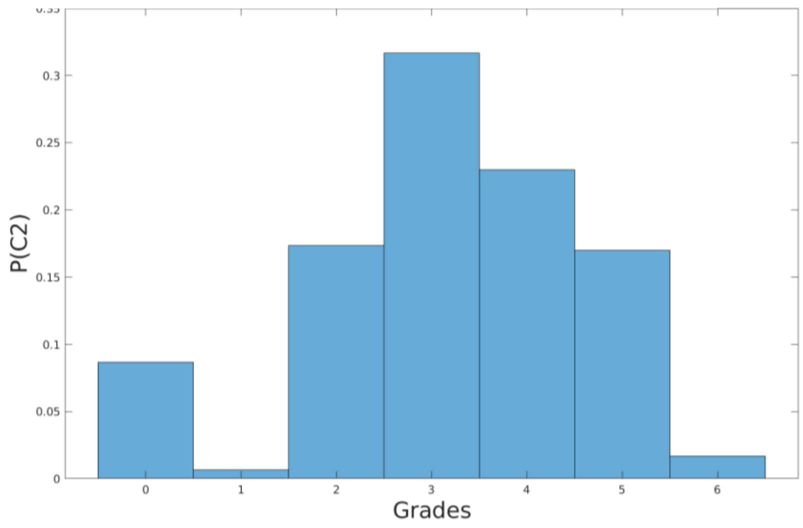
$$P\left(x=4\right)=\frac{\text{Nb of instances of } x=4}{\text{Total nb of measurements } x}$$

Joint Probabilities

Probabilities of Grades in C_1

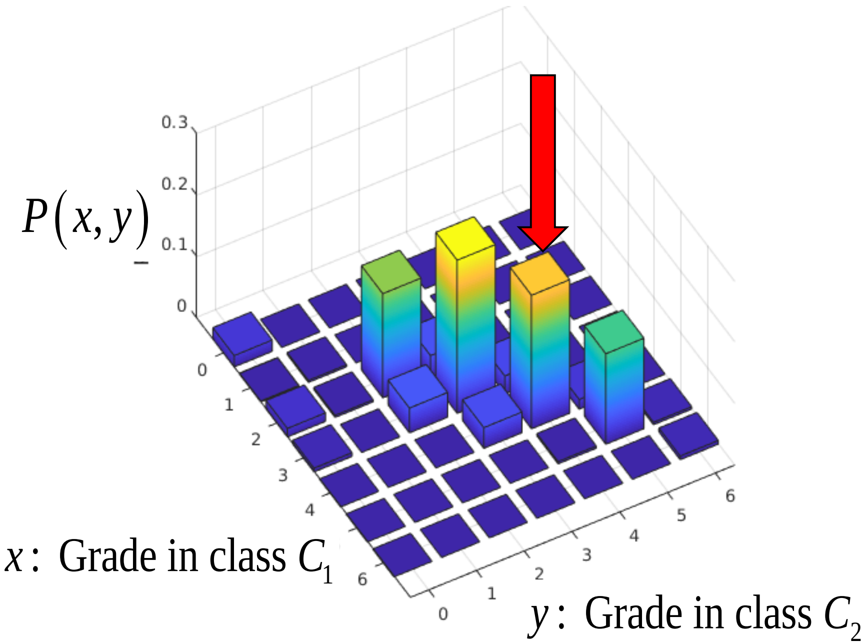


Probabilities of grades in C_2



What is the probability that a student receives a grade of 4 in both C_1 and C_2?

Joint Probabilities



What is the probability that a student receives a grade of 4 in both C_1 and C_2?

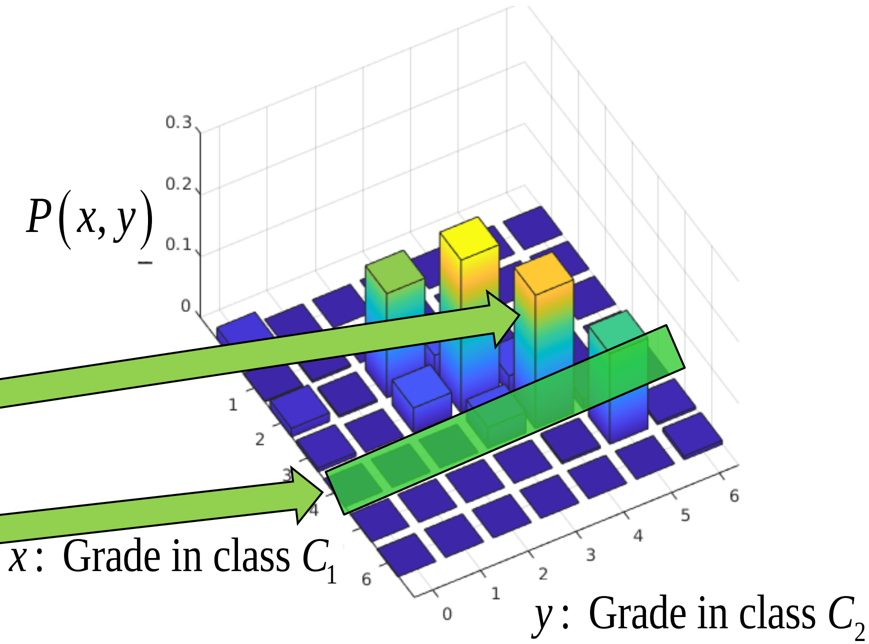
Joint distribution: $P(x=4, y=4) = \dfrac{\text{Nb of instances of } (x=4, y=4)}{\text{Total nb of joint measurements } (x,y)}$

Conditional Probabilities

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

$P(x=4, y=4)$

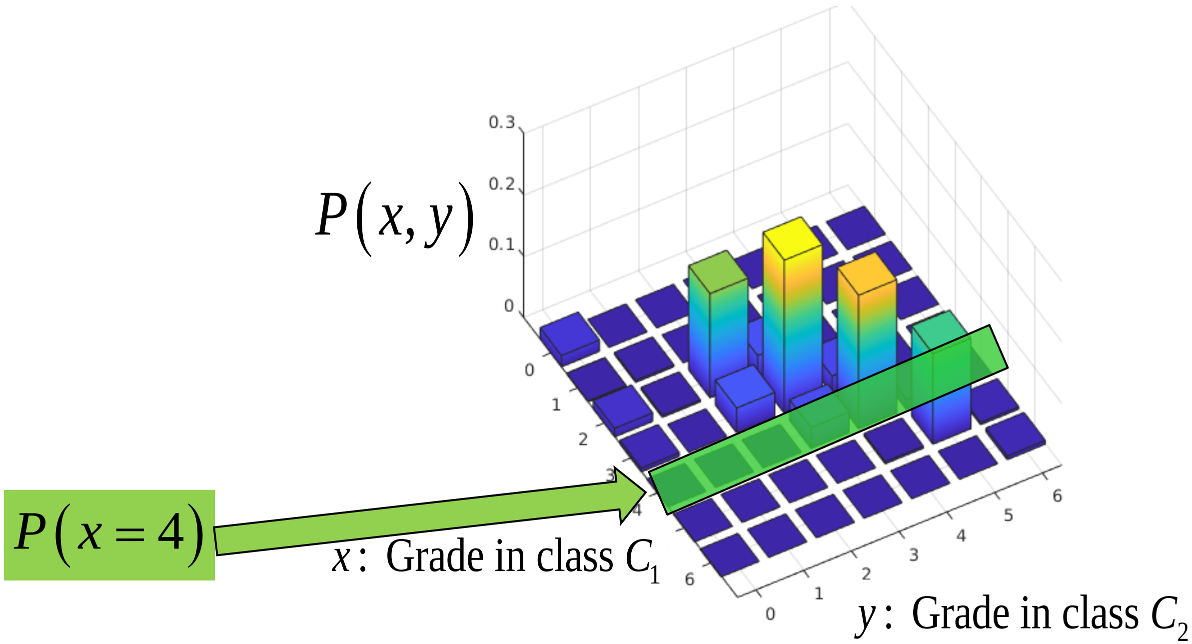
$P(x=4)$



What is the probability that a student receives a grade of 4 in C_2 if s/he has received a grade of 4 in C_1?

Conditional probability: $P(y=4 \mid x=4)$

Marginal Probabilities



The *marginal probability* that variable x takes value x_i is given by:

$$P(x=i) := \sum_{j=1}^{N_y} P(x=i, y=j)$$

Conditional probabilities and statistical independence

$$P(x,y)=P(y\,|\,x)P(x)$$

If x and y are independent

$\Rightarrow P(y\,|\,x)=P(y)$ and $P(x\,|\,y)=P(x)$

$$\Rightarrow P(x,y)=P(x)P(y)$$

How can we tell if x and y are independent?

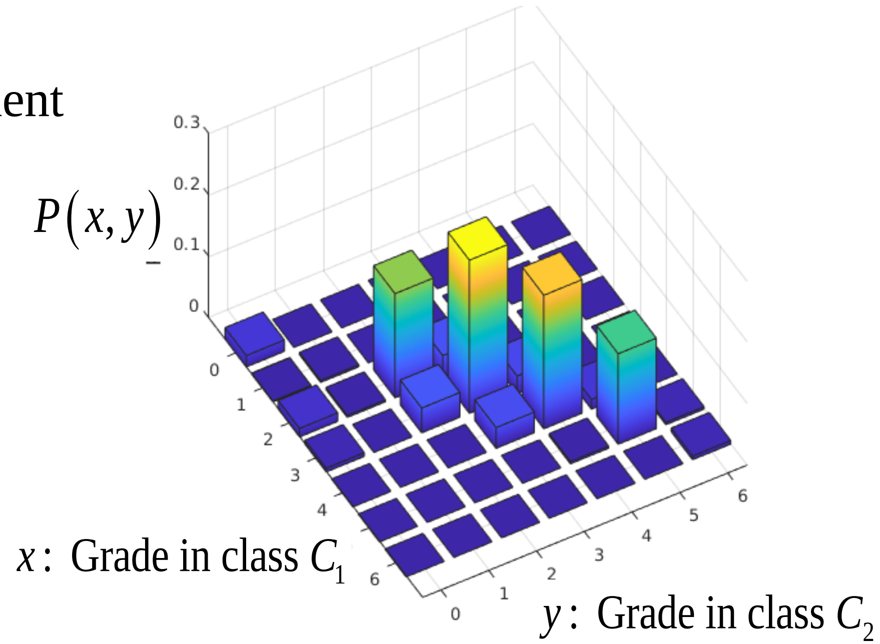
Conditional probabilities and statistical independence

If x and y are independent

$P(y \mid x) = P(y)$

and

$P(x \mid y) = P(x)$



How can we tell if x and y are independent?

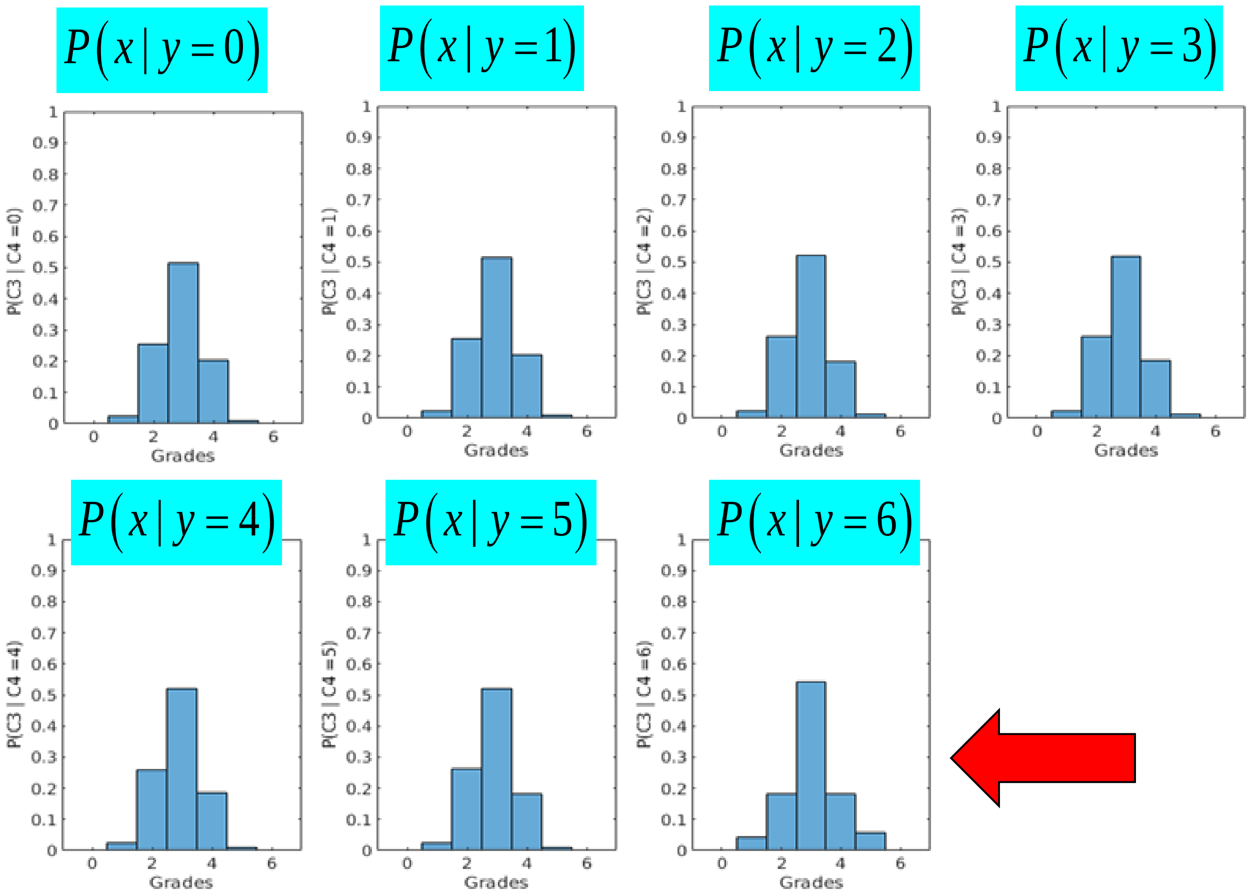
Conditional probabilities and statistical independence

If x and y are independent

$P(y \mid x) = P(y)$

and

$P(x \mid y) = P(x)$



If all the conditional distributions are identical, the variables are independent.

Marginal, Joint and Conditional Probabilities

- To compute the marginal, one needs the joint distribution p(x,y).
- If x is a multidimensional variable → the marginal is a joint distribution!

❖ The marginals of N variables taking K values corresponds to $N(K-1)$ probabilities.

❖ The joint distribution corresponds to $\sim N^K$ probabilities.

❖ The joint distribution is far richer than the marginals.

Pros of computing the joint distribution:
Provides statistical dependencies across all variables and the marginal distributions

Cons:
Computational costs grow exponentially with number of dimensions
(statistical power: 10 samples to estimate each parameter of a model)

→ Compute solely the conditional if you care only about dependencies across variables (see class on non-linear regression).

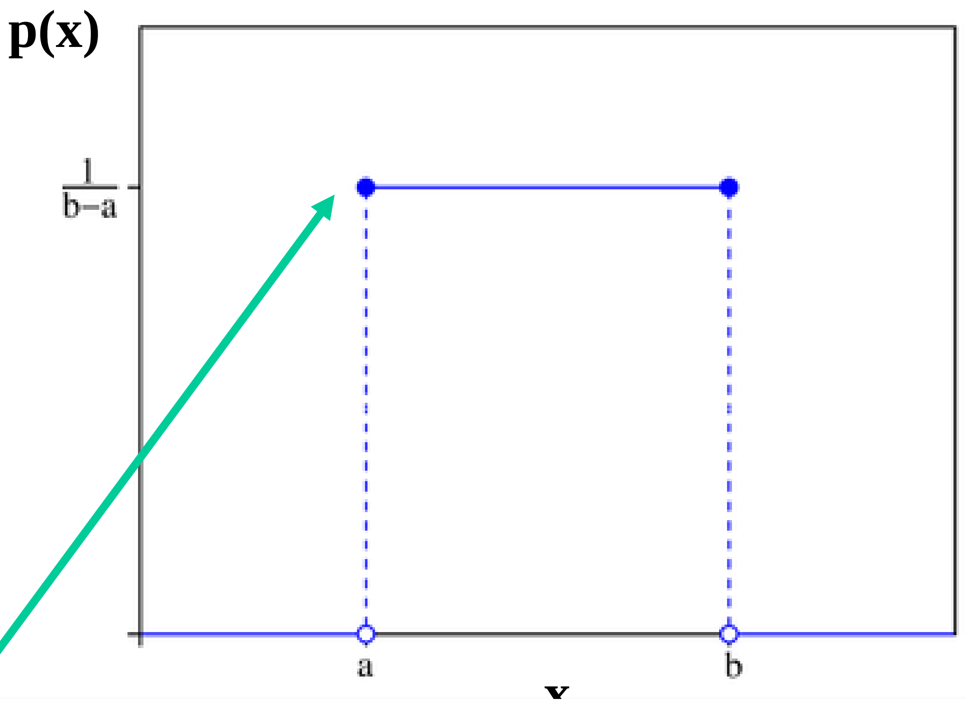
Probability Distributions, Density Functions

p(x) a ***continuous function*** is the *probability density function or probability distribution function (PDF) (sometimes also called probability distribution* or simply *density)* of variable *x*.

$$p(x) \geq 0, \quad \forall x \in \mathbb{R}$$
$$\int_{-\infty}^{\infty} p(x)dx = 1$$

Probability Distributions, Density Functions

- The pdf is not bounded by 1.
- It can grow unbounded, depending on the value taken by x.



p(X=x) does not give the probability. The probability must be computed from the cumulative density function.

PDF equivalency with Discrete Probability

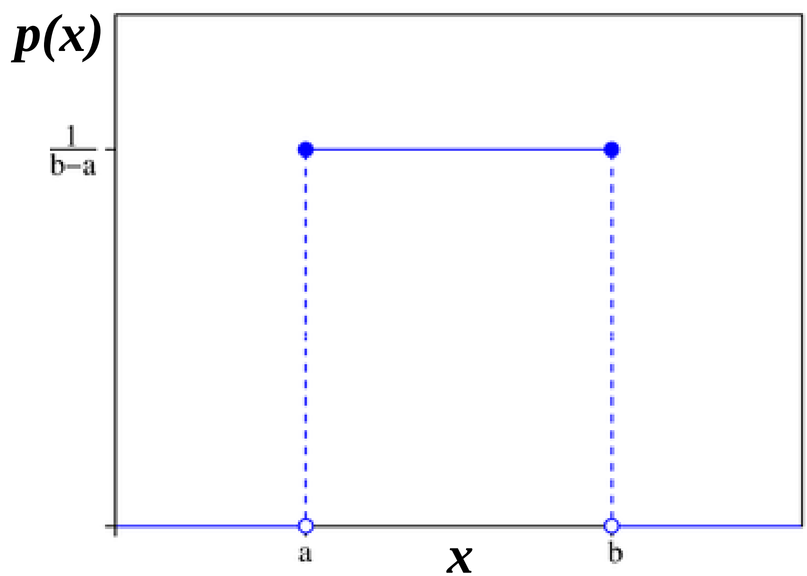
The cumulative distribution function (or simply *distribution function)* of X is:

$$D_x\left(x^*\right) = P\left(x \le x^*\right)$$

$$D_x\left(x^*\right) = \int_{-\infty}^{x^*} p(x)dx, \qquad x \in \mathbb{R}$$

$p(x)$ dx ~ probability of x to fall within an infinitesimal interval [x, x + dx]

PDF equivalency with Discrete Probability

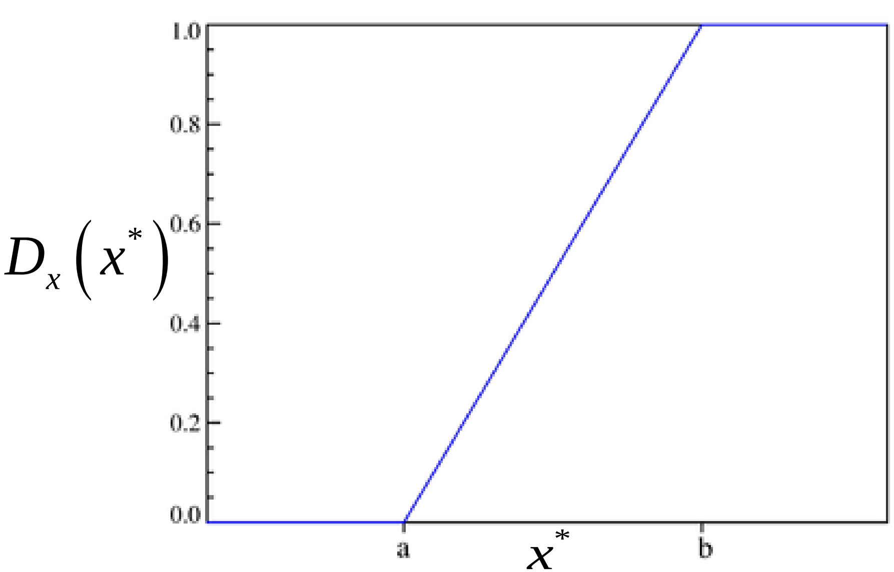


Probability that x takes a value in the subinterval [a,b] is given by:

$$P(x \le b) := D_x(x \le b) = \int_{-\infty}^{b} p(x)dx$$

$$P(a \le x \le b) = D_x(x \le b) - D_x(x \le a)$$

$$P(a \le x \le b) = \int_{a}^{b} p(x)dx = 1$$



Expectation

The *expectation* of the random variable *x* with probability P(x) (in the discrete case) and pdf p(x) (in the continuous case), also called the *expected value* or *mean*, is the mean of the observed value of *x* weighted by *p(x)*.
If *X* is the set of observations of *x*, then:

When x takes discrete values: $$\mu = E\{x\} = \sum_{x \in X} xP(x)$$

For continuous distributions: $$\mu = E\{x\} = \int_X x \cdot p(x) \cdot dx$$

Statistical Independence and and uncorrelatedness

x_1 and x_2 are statistically independent if:

$$p(x_1 \mid x_2) = p(x_1) \quad \text{and} \quad p(x_2 \mid x_1) = p(x_2)$$

$$\Rightarrow p(x_1, x_2) = p(x_1)\, p(x_2)$$

x_1 and x_2 are uncorrelated if $\operatorname{cov}(x_1, x_2) = 0.$

$$\operatorname{cov}(x_1, x_2) = E\{x_1, x_2\} - E\{x_1\} E\{x_2\}$$

$$\Rightarrow E\{x_1, x_2\} = E\{x_1\} E\{x_2\}$$

Statistical Independence and and uncorrelatedness

Independent $\underset{\not\longleftarrow}{\longrightarrow}$ **Uncorrelated**

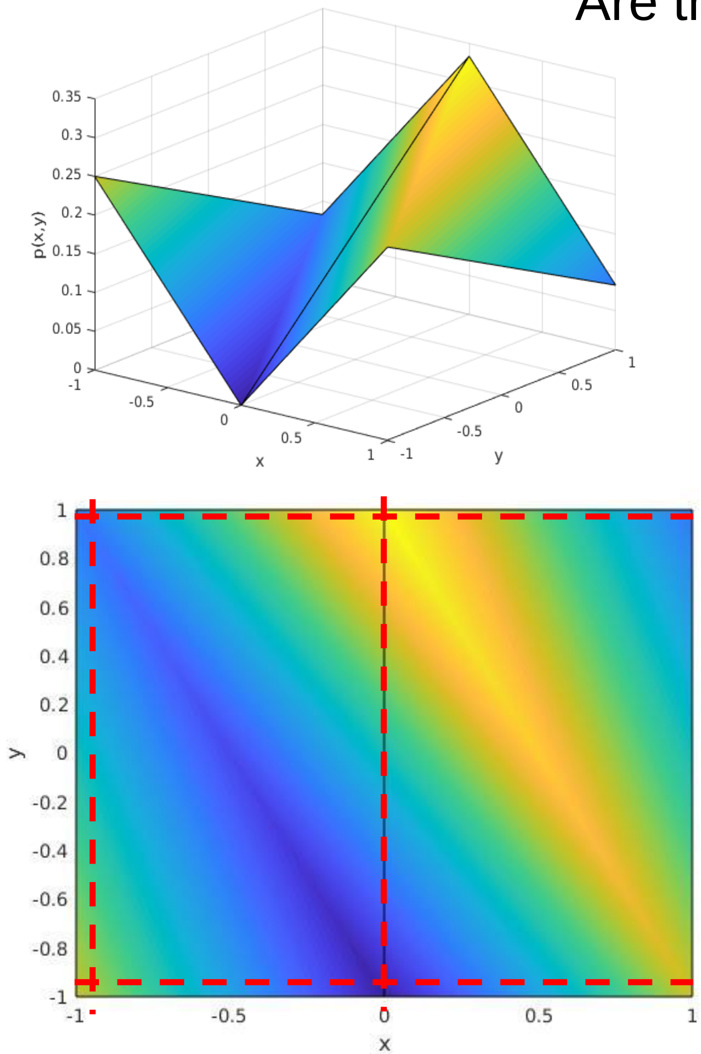
$$p\left(x_1, x_2\right) = p\left(x_1\right) p\left(x_2\right) \quad \Rightarrow \quad E\left\{x_1, x_2\right\} = E\left\{x_1\right\} E\left\{x_2\right\}$$

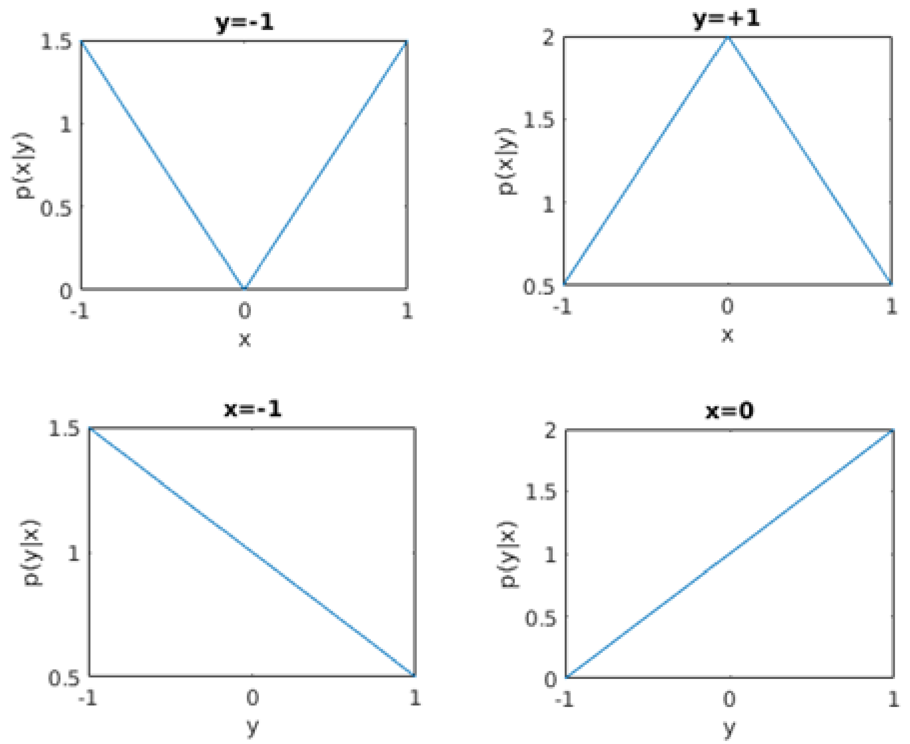
$$p\left(x_1, x_2\right) = p\left(x_1\right) p\left(x_2\right) \quad \not\Leftarrow \quad E\left\{x_1, x_2\right\} = E\left\{x_1\right\} E\left\{x_2\right\}$$

Statistical independence ensures uncorrelatedness.
The converse is not true

Statistical Independence and and uncorrelatedness

Are x,y correlated?
Are they dependent?





Variance

σ^2, the variance of a distribution measures the amount of spread of the distribution around its mean:

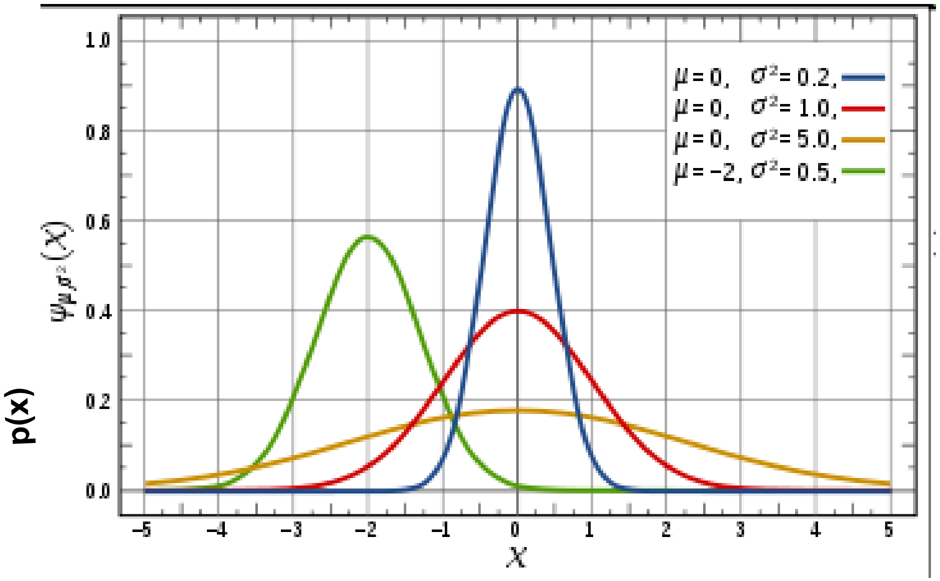
$$\sigma^2 = Var(x) = E\left\{\left(x-\mu\right)^2\right\} = E\left\{x^2\right\} - \left[E\left\{x\right\}\right]^2$$

σ is the *standard deviation* of x.

Parametric PDF

The uni-dimensional Gaussian or Normal distribution is a distribution with pdf given by:

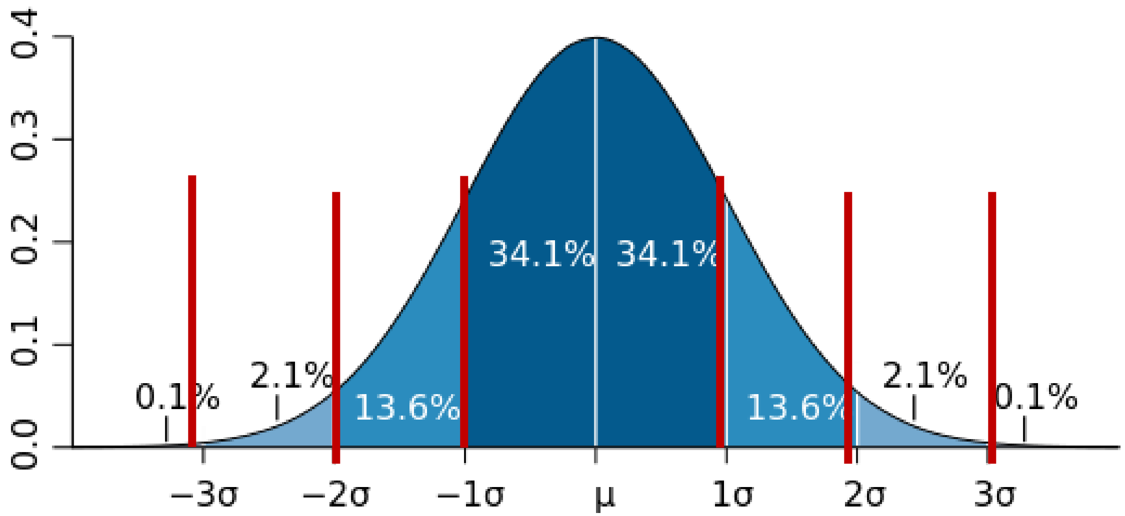
$$p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)},\quad \mu\text{:mean},\quad \sigma^2\text{:variance}$$



The Gaussian function is entirely determined by its mean and variance.
For this reason, it is referred to as a parametric distribution.

Illustrations from Wikipedia

Mean and Variance in PDF



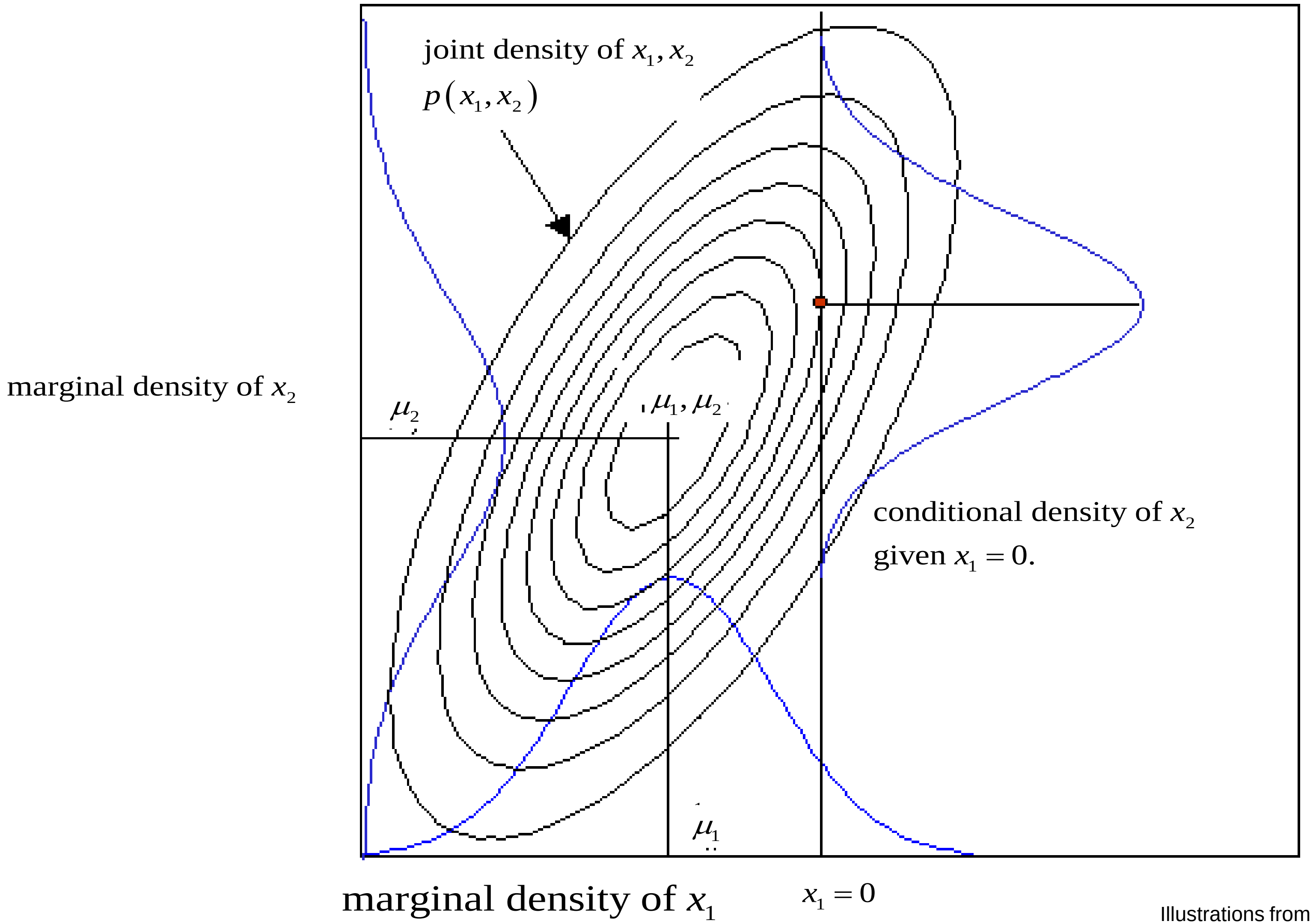
~68% of the data are comprised between +/- 1σ
~96% of the data are comprised between +/- 2σ
~99% of the data are comprised between +/- 3σ

This is no longer true for arbitrary pdf-s!

Illustrations from Wikipedia

Marginal, Conditional Pdf of Gauss Functions

The conditional and marginal pdf of a multi-dimensional Gauss function are all Gauss functions!



Illustrations from Wikipedia