Review of statistics

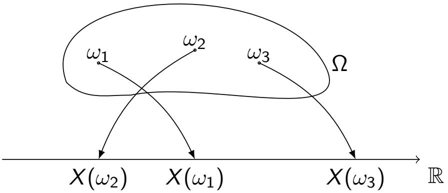
Manuel A. Vázquez, Marcelino Lázaro

January 29, 2024

(Real) Random variable

Function that assigns a real number to the result (outcome) of a **random** experiment.

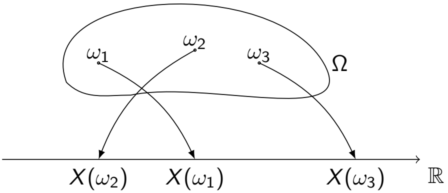
$$X : \Omega \rightarrow \mathbb{R}$$
$$\omega \rightarrow X(\omega)$$



(Real) Random variable

Function that assigns a real number to the result (outcome) of a **random** experiment.

$$X : \Omega \to \mathbb{R}$$
$$\omega \to X(\omega)$$

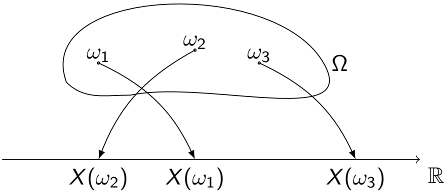


- Range of X: $\text{Range}_X = \{x \in \mathbb{R} : \exists \omega \in \Omega,\ X(\omega) = x\}$
  - Discrete r.v.: the range is a discrete set of values
    - tossing of a coin (X can only take two values)
    - winning number in the lottery (X can only take $100,000$ values)

(Real) Random variable

Function that assigns a real number to the result (outcome) of a **random** experiment.

$$X : \Omega \to \mathbb{R}$$
$$\omega \to X(\omega)$$



- Range of X: $\text{Range}_X = \{x \in \mathbb{R} : \exists \omega \in \Omega,\ X(\omega) = x\}$
  - Discrete r.v.: the range is a discrete set of values
    - tossing of a coin (X can only take two values)
    - winning number in the lottery (X can only take $100,000$ values)
  - Continuous r.v.: continuous range of values
    - the temperature in this room (a *continuum* of values is possible)
    - the price of 1 kg of oranges (*idem*)

Characterizing a random variable

- Distribution function, defined as

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

and depending on whether the variable is continuous or discrete...

Characterizing a random variable

- Distribution function, defined as

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

  and depending on whether the variable is continuous or discrete...

  - Probability density function, defined as

    continuous r.v.

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Characterizing a random variable

- Distribution function, defined as

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

and depending on whether the variable is continuous or discrete...

- Probability density function, defined as $\boxed{\text{continuous r.v.}}$

$$f_X(x) = \frac{d}{dx} F_X(x)$$

- Probability mass function, $\boxed{\text{discrete r.v.}}$

$$p_X(x_i) = P(X = x_i)$$

(there is a **finite** number of x_is)

Properties of $f_X(x)$ (*continuous* r.v.'s)

- $f_X(x) \geq 0$
  (the pdf is always positive)

Properties of $f_X(x)$ (*continuous* r.v.'s)

- $f_X(x) \geq 0$
  (the pdf is always positive)
- $\displaystyle\int_{-\infty}^{\infty} f_X(x) \cdot dx = 1$
  (the area under the pdf)

Properties of $f_X(x)$ (*continuous* r.v.'s)

- $f_X(x) \geq 0$
  (the pdf is always positive)
- $\displaystyle\int_{-\infty}^{\infty} f_X(x) \cdot dx = 1$
  (the area under the pdf)
- $\displaystyle\int_{a}^{b} f_X(x) \cdot dx = P(a < X \leq b)$
  (integral over an interval amounts to probability)

Properties of $f_X(x)$ (*continuous* r.v.'s)

- $f_X(x) \geq 0$
  (the pdf is always positive)
- $\displaystyle\int_{-\infty}^{\infty} f_X(x) \cdot dx = 1$
  (the area under the pdf)
- $\displaystyle\int_a^b f_X(x) \cdot dx = P(a < X \leq b)$
  (integral over an interval amounts to probability)
- In general, $P(X \in A) = \displaystyle\int_A f_X(x) \cdot dx$
  (integral over a set)

Properties of $f_X(x)$ (*continuous* r.v.'s)

- $f_X(x) \geq 0$
  (the pdf is always positive)
- $\displaystyle\int_{-\infty}^{\infty} f_X(x) \cdot dx = 1$
  (the area under the pdf)
- $\displaystyle\int_{a}^{b} f_X(x) \cdot dx = P(a < X \leq b)$
  (integral over an interval amounts to probability)
- In general, $P(X \in A) = \displaystyle\int_{A} f_X(x) \cdot dx$
  (integral over a set)
- $F_X(x) = \displaystyle\int_{-\infty}^{x} f_X(u) \cdot du$
  (integrating the pdf we get back the distribution function)

Interpretation of the probability density function

Interpretation

The pdf indicates the *relative* likelihood for each value of the random variable

- Regions where $f_X(x)$ is large are associated with values of the random variable that are very likely
- In order to get probabilities, we need to integrate the pdf

Interpretation of the probability density function

Interpretation

The pdf indicates the *relative* likelihood for each value of the random variable

- Regions where $f_X(x)$ is large are associated with values of the random variable that are very likely
- In order to get probabilities, we need to integrate the pdf

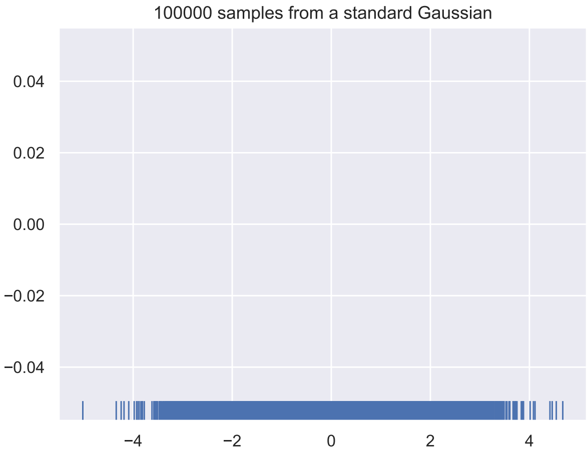
Probability vs density

The probability *density* function at x is **not** the probability of x.

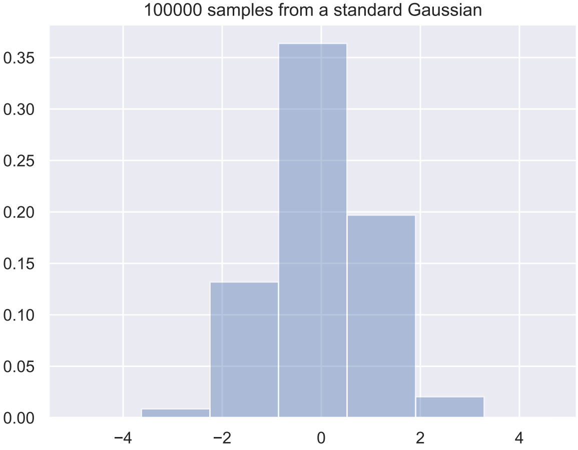
A pdf can be interpreted as a histogram pushed to the limit

- as the width of the *bin* decreases, the histogram resembles more the pdf

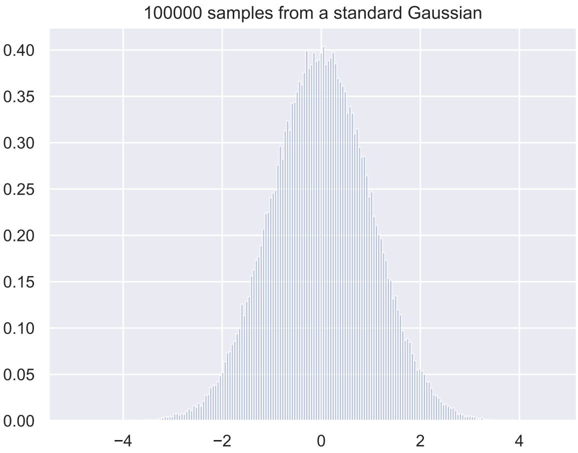
pdf-histogram connection



pdf-histogram connection



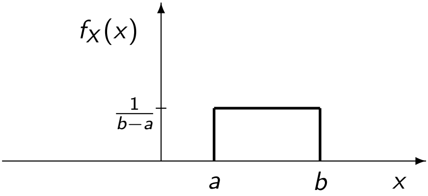
pdf-histogram connection



Uniform distribution

- Continuous distribution[1]
- Parameters: a and b
  - Notation: $\mathcal{U}(a,b)$

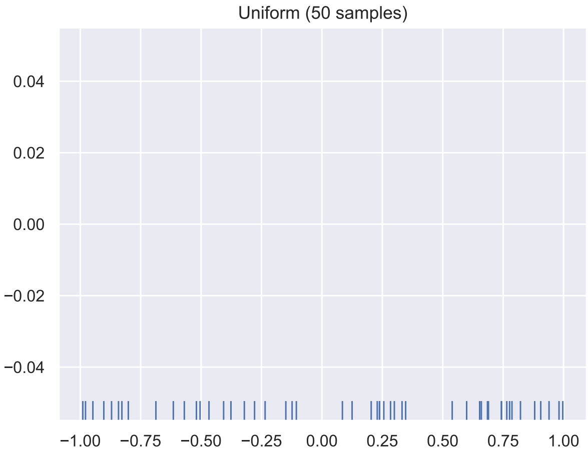
$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$$



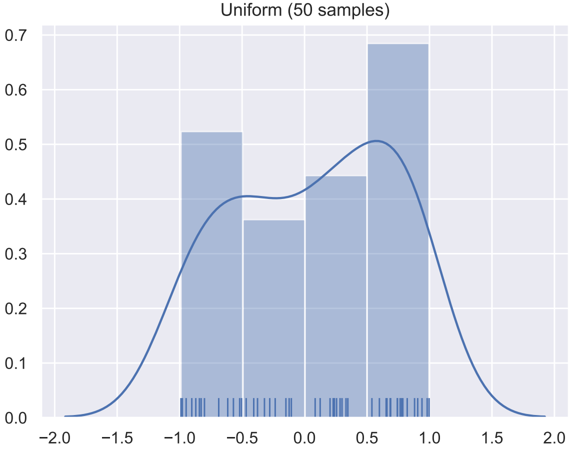
- Example in communications
  - Random phase of a sinusoid: uniform r.v. between 0 and 2π

[1] i.e. distribution for a continuous random variable

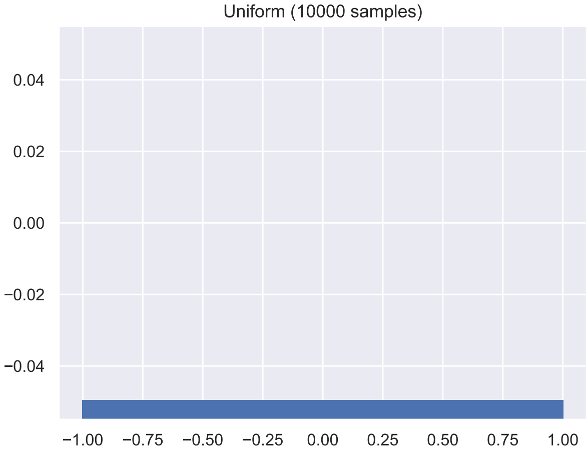
50 samples from a uniform distribution between -1 and 1



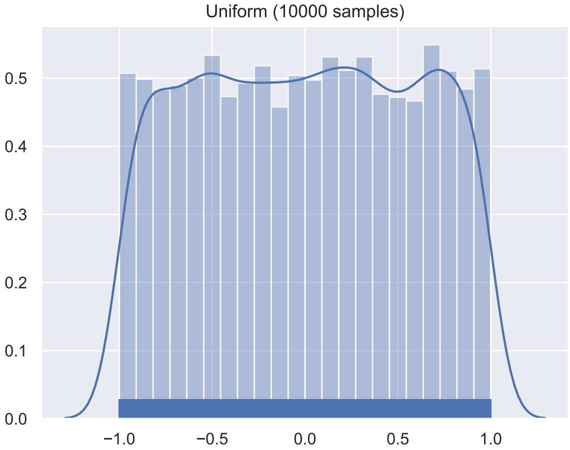
50 samples from a uniform distribution between -1 and 1



10,000 samples from a uniform distribution between -1 and 1



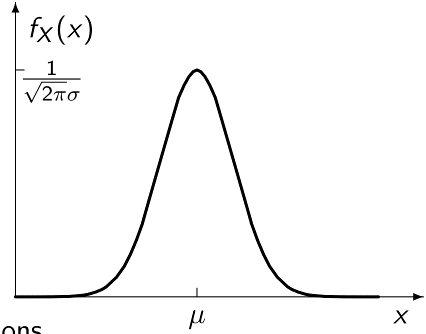
10,000 samples from a uniform distribution between -1 and 1



Gaussian (normal) distribution

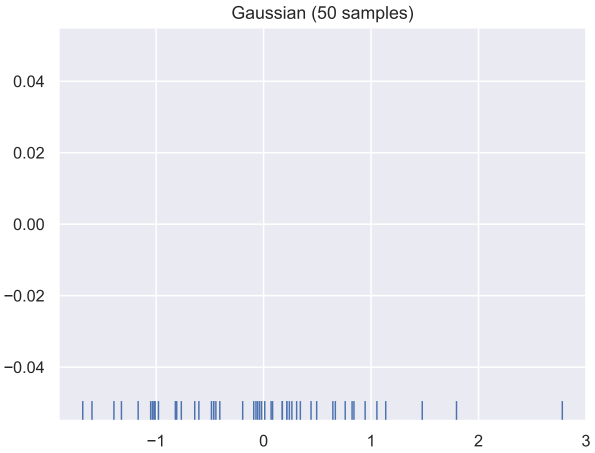
- Continuous distribution
- Parameters: mean (μ), and variance (σ^2)
  - Notation: $\mathcal{N}(\mu, \sigma^2)$

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

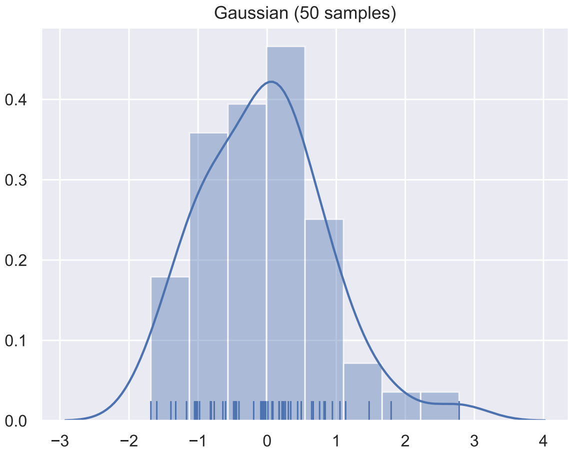


- Example in communications
  - Thermal noise modeling

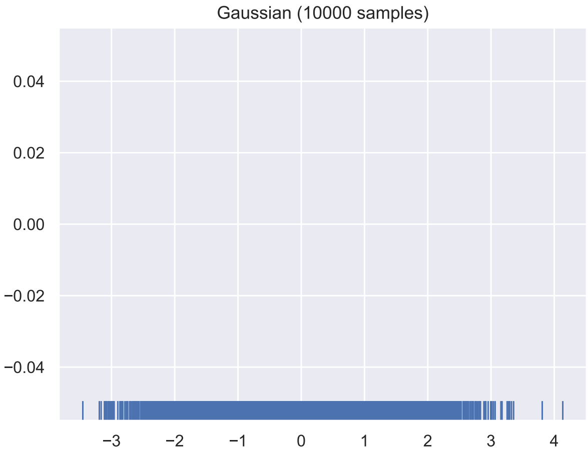
50 samples from a standard Gaussian distribution



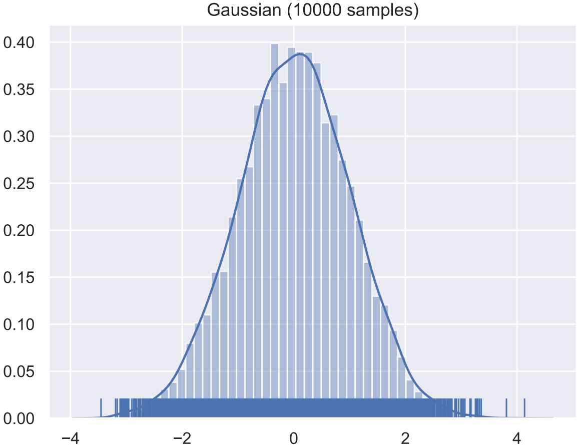
50 samples from a standard Gaussian distribution



10,000 samples from a standard Gaussian distribution



10,000 samples from a standard Gaussian distribution



Q-function

Definition: Q-function

Probability of a Gaussian random variable with zero mean and unit variance (i.e., *standard*) taking on values that are greater than its argument

$$X \sim \mathcal{N}(0,1) \Rightarrow f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

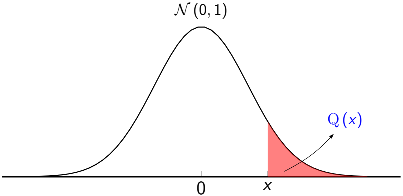
then

$$Q(x) = P(X > x) = \int_x^{+\infty} f_X(z)\, dz = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz$$

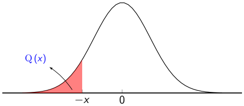
- allows computing the probability over any interval of **any** Gaussian r.v.
- numerically computed by MATLAB/Python/R...
- historically tabulated **only for** $x > 0$
- From the definition, it's clear that
  - $Q(0) = \frac{1}{2}$
  - $Q(\infty) = 0$

Q-function: interpretation

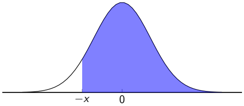
Graphical interpretation:



Due to the symmetry in Gaussian pdf we have



and hence $\mathrm{Q}(-x) = 1 - \mathrm{Q}(x) \equiv$



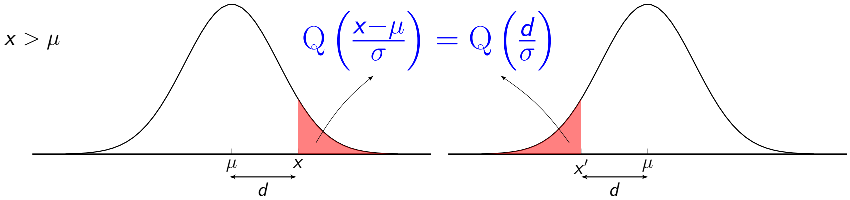
Here, we are assuming $\boxed{x > 0}$

Integral of a $\mathcal{N}\left(\mu, \sigma^2\right)$ (Gaussian) pdf

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right) \Rightarrow P(X > x) = \mathrm{Q}\left(\frac{x-\mu}{\sigma}\right)$$

Integral of a $\mathcal{N}\left(\mu, \sigma^2\right)$ (Gaussian) pdf

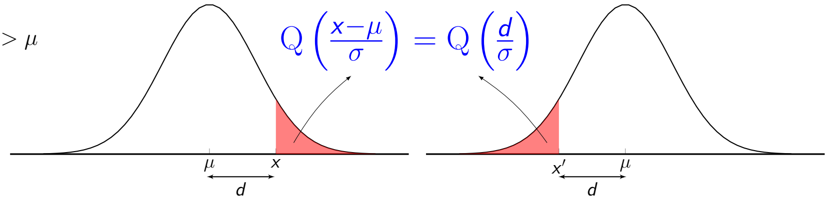
$$X \sim \mathcal{N}\left(\mu, \sigma^2\right) \Rightarrow P(X > x) = \mathrm{Q}\left(\frac{x-\mu}{\sigma}\right)$$



Integral of a $\mathcal{N}\left(\mu, \sigma^2\right)$ (Gaussian) pdf

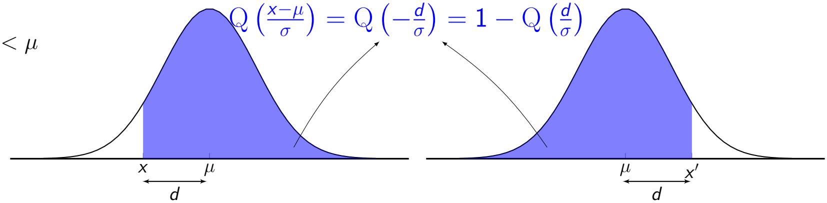
$$X \sim \mathcal{N}\left(\mu, \sigma^2\right) \Rightarrow P(X > x) = \mathrm{Q}\left(\frac{x-\mu}{\sigma}\right)$$

$x > \mu$



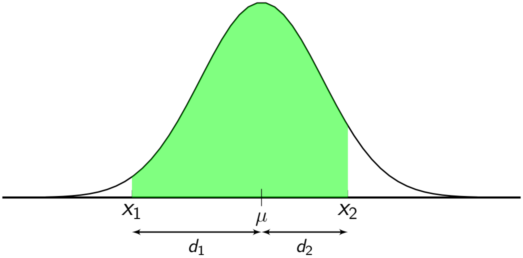
and conversely

$x < \mu$

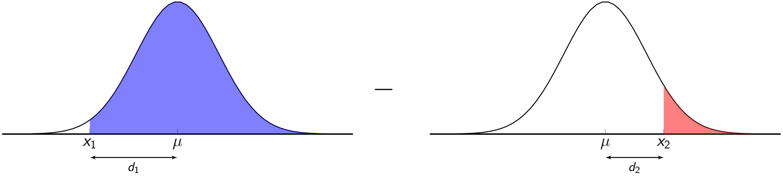


Integrals of $\mathcal{N}\left(\mu, \sigma^2\right)$ over intervals

Any interval, e.g.,



can be computed by adding/subtracting Q-function's



Statistical moments

- Expectation (mean)

$$\mu_X = \mathbb{E}\left[X\right] = \int_{-\infty}^{\infty} x\, f_X(x)dx$$

- Expectation of a function, g, of X

$$\mathbb{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)\, f_X(x)dx$$

- Variance

$$\sigma_X^2 = \mathbb{E}\left[(x-\mu_X)^2\right] = \int_{-\infty}^{\infty} (x-\mu_X)^2\; f_X(x)$$

Variance and the expectation of the square

$$\sigma_X^2 = \mathbb{E}\left[(X-\mu_X)^2\right] = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}\left[X\right]\right)^2 = \mathbb{E}\left[X^2\right] - \mu_X^2$$

Properties of moments

- $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] = \mu_X + \mu_Y$
  (expectation is a linear operator)
- $\mathbb{E}[c] = c$ for any constant c
- $\mathbb{E}[cX] = c\mathbb{E}[X]$
  (still a linear operator)
- $\mathbb{E}[X+c] = \mathbb{E}[X] + c$
  (from the first two)
- $\mathrm{Var}(c) = 0$
  (a constant is not random)
- $\mathrm{Var}(c \cdot X) = c^2 \cdot \mathrm{Var}(X)$
  (variance is a squared magnitude)
- $\mathrm{Var}(X+c) = \mathrm{Var}(X)$
  (adding a constant doesn't increase the variance of an r.v.)

Multidimensional random variables

X, Y two r.v.'s defined over the same sample space Ω

We characterize the random experiment using *joint* probabilistic modeling

- *Joint* distribution function

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$$

- *Joint* probability density function

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

Properties of $f_{X,Y}(x,y)$

- *Marginal* pdf for X

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

- *Marginal* pdf for Y

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

- The whole area under the pdf

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$$

- Probability of a set

$$P((X,Y)\in A) = \int\int_{(x,y)\in A} f_{X,Y}(x,y)dxdy$$

Conditional probability density function

- The knowledge of one variable modifies the probabilities of the other.

$$f_{Y|X}(y|x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)}, & f_X(x) \neq 0 \\ \text{undefined}, & f_X(x) = 0 \end{cases}$$

- Definition of statistical independence:

$$f_{Y|X}(y|x) = f_Y(y)$$

$$f_{X|Y}(x|y) = f_X(x)$$

  - Consequence: for independent random variables

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

Statistical moments

In general, the expectation of a function, g, of X and Y is

$$\mathbb{E}\left[g(X,Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y)dxdy,$$

and two well-known moments result from the choice of g

$$g(X,Y) = XY \rightarrow \text{correlation}$$

$$g(X,Y) = (X-\mu_X)(Y-\mu_Y) \rightarrow \text{covariance} \equiv \mathrm{Cov}\left(X,Y\right)$$

Definition: Uncorrelatedness

Random variables X and Y are **uncorrelated** if $\mathrm{Cov}\left(X,Y\right) = 0$

- Independence $\Rightarrow$ uncorrelatedness
- Uncorrelatedness $\not\Rightarrow$ independence...with one exception

Only for Gaussian r.v.'s...

Uncorrelatedness $\Rightarrow$ independence

Sum of independent random variables

Theorem: Central Limit Theorem (CLT)

If $(X_1, X_2, \cdots, X_n)$ are independent and identically distributed (i.i.d.) r.v.'s with means $\mu_1, \mu_2, \cdots, \mu_n$, and variances $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$, then the distribution of

$$Y = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu_i}{\sigma_i}$$

approaches,as $n \to \infty$, a Gaussian distribution with zero mean and unit variance, i.e., $Y \sim \mathcal{N}(0, 1)$.

- Special case: for i.i.d. r.v's with the *same mean, μ, and variance, σ^2*, the average

$$Y = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

approaches a distribution $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$. This is true even when the original distribution is not Gaussian. It is easily shown via simulation.