Probabilistic Modeling and Bayesian NEtworks

Probability Distributions

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

INFORMATIKA FAKULTATEA
FACULTAD DE INFORMÁTICA

Konputazio Ingeniaritza eta Sistema Adimentsuak Unibertsitate Masterra

Máster Universitario en Ingeniería Computacional y Sistemas Inteligentes

There are many distributions available; for each application we have to select the most appropriate

DOMAIN

- Categorical, Ordinal or Continuous
- Bounded or Unbounded

SHAPE

- Parameter-dependant
- Represented by the **moments**
- Best known: mean and variance

Moments Around 0

$$\mu'_k = E[X^k] = \int_{-\infty}^{\infty} x^k f(x)dx$$

Moments Around Mean

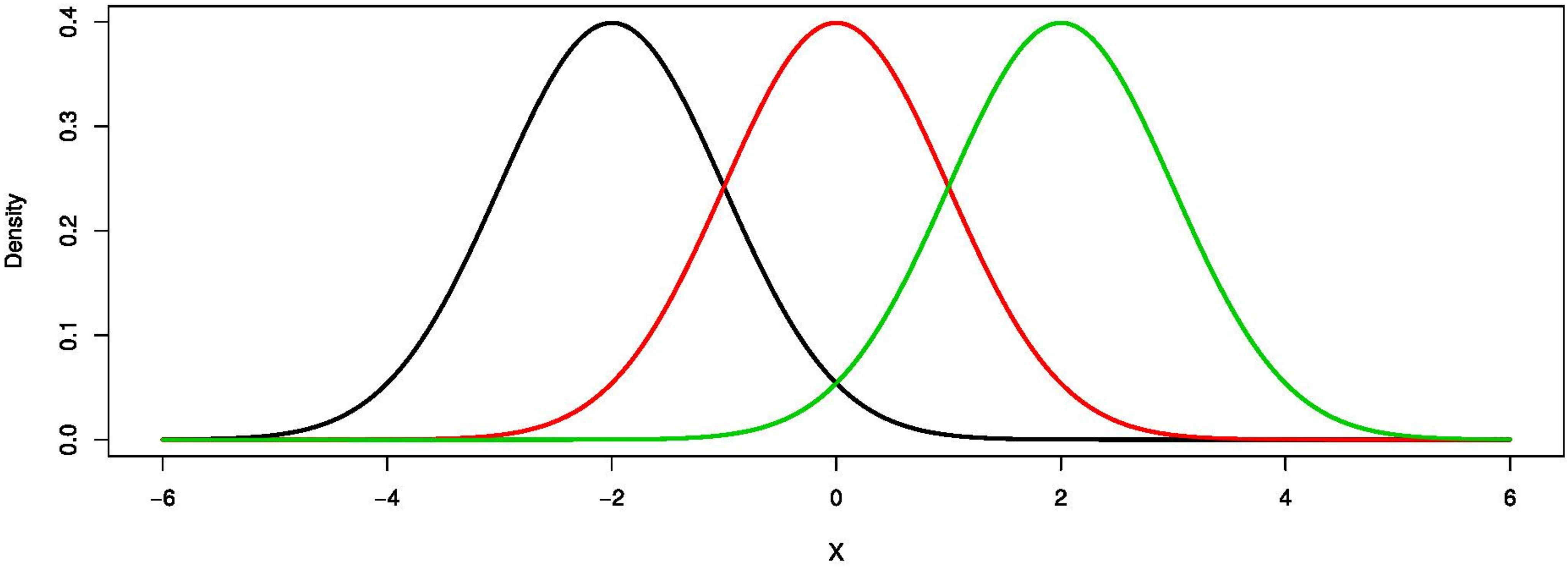
$$\mu_k = E[(X - \mu'_1)^k] = \int_{-\infty}^{\infty} (x - \mu'_1)^k f(x)dx$$

Sample Moments

$$m'_k = \frac{1}{n} \sum_{i=1}^{n} x_i^k \qquad m_k = \frac{1}{n} \sum_{i=1}^{n} (x_i - m'_i)^k$$

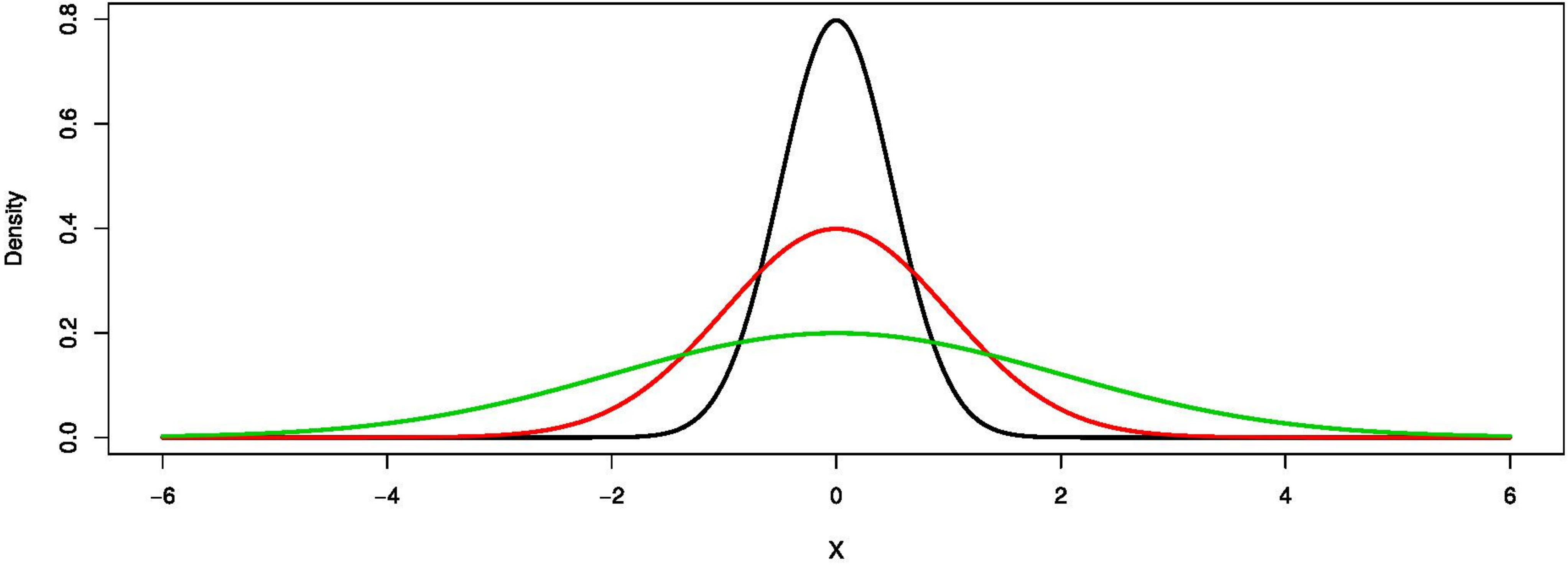
1st Moment: Mean

Information about the **LOCATION**



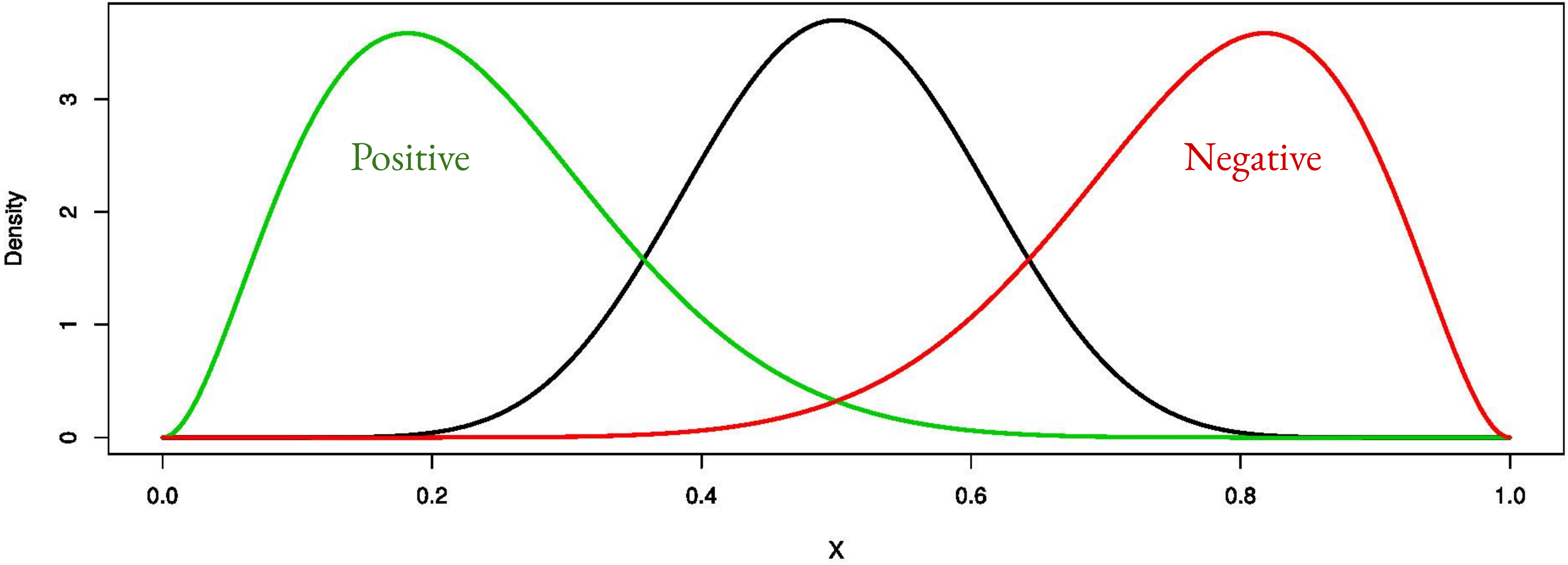
2^{ND} MOMENT: VARIANCE

Information about the **SPREADNESS**



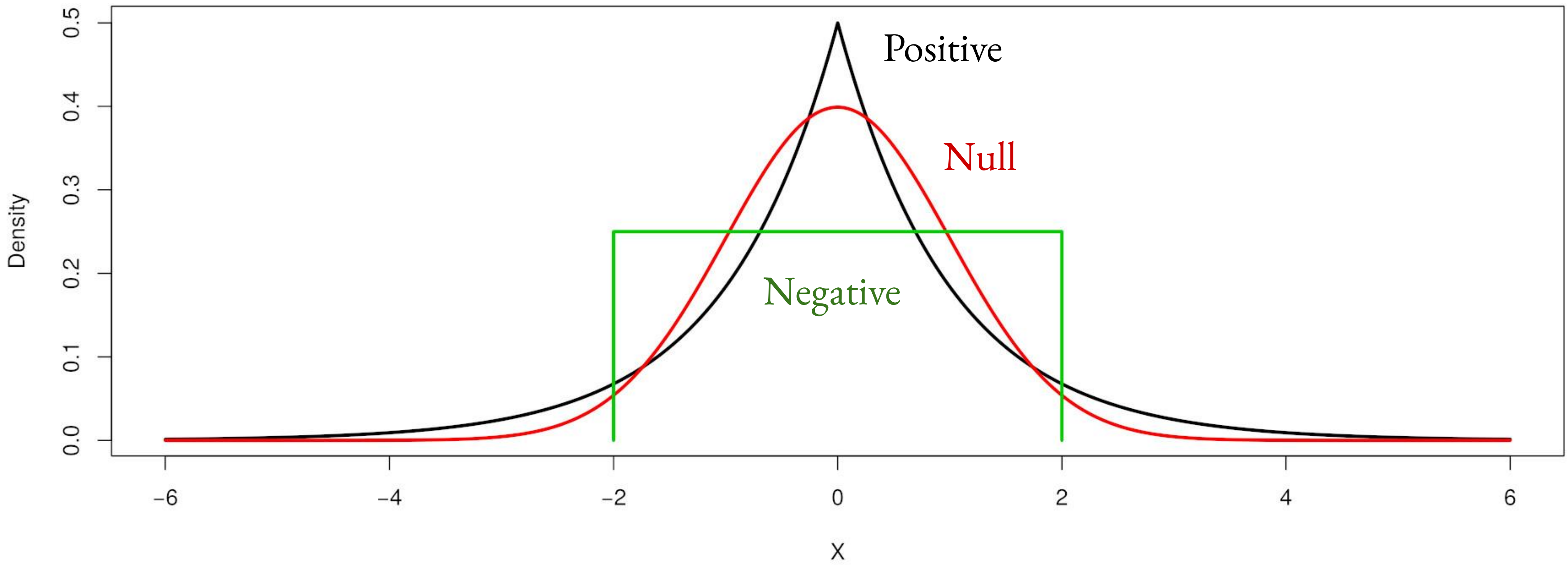
3^{RD} Moment: Skewness

Information about the **SYMMETRY**



4^{TH} MOMENT: KURTOSIS

Information about the tendency to produce **OUTLIERS**



Some Useful Properties

Expectation

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

Variance

$$VAR(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

$$VAR(c) = 0$$

Probability Distributions

Discrete Variables

Multinomial Distribution

Probability distribution associated to categorical variables (discrete values with no particular order)

Domain $X \in \{1, \ldots, r\}$

Parameters $0 \leq p_1, \ldots, p_r \leq 1; \sum_{i=1}^{r} p_i = 1$

Function $P(X = x_i) = p_i$

Special Cases Bernouilli distribution (r=2)
Uniform distribution

Binomial Distribution

In a sample of size n drawn from a Bernouilli distribution, the binomial counts the number of outcomes of a certain type

Domain $X \in \{0, 1, \ldots, n\}$

Parameters $n \geq 0; 0 \leq p \leq 1$

Function $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$

Moments $\mu_1' = np; \mu_2 = np(1 - p)$

Hypergeometric Distribution

In a set of N elements there are M of a special type. If we draw n elements without replacement, this variable counts the number of elements of the special type

Domain $$max\{0, M - N + n\} \leq X \leq min\{n, N\}$$

Parameters $$N \geq 0; 0 \leq M \leq N; 0 \leq n \leq N$$

Function $$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

Moments $$\mu_1' = \frac{nM}{N}; \mu_2 = \frac{nM}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}$$

Poisson Distribution

This distribution counts the number of events that happen in a certain period of time as long a the event rate is constant and events are independent

Domain $X \in \{0, 1, 2, \ldots\}$

Parameters $\lambda > 0$

Function $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$

Moments $\mu_1' = \lambda; \mu_2 = \lambda$

Probability Distributions

Continuous Variables

Normal Distribution

This distributions represents many natural situations. In particular, it represents the distribution of the sum of random iid samples.

Domain $X \in \mathbb{R}$

Parameters $\mu \in \mathbb{R}; \sigma > 0$

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

Moments $\mu_1' = \mu; \mu_2 = \sigma^2$

Gamma Distribution

Given a Poisson distribution, the time until the a-th event happens is represented by this type of distribution

Domain	$X > 0$
Parameters	$a > 0; b > 0$
Function	$f(X = x) = \frac{1}{\Gamma(a)b^a} e^{-\frac{x}{b}} x^{(a-1)}$
Moments	$\mu_1' = ab; \mu_2 = ab^2$

Beta Distribution

This distribution represents the ratio between two Gamma variables. Its main interest is that it is bounded between 0 and 1

Domain

$$0 \leq X \leq 1$$

Parameters

$$a > 0; b > 0$$

Function

$$f(X = x) = \frac{1}{Beta(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

Moments

$$\mu_1' = \frac{\alpha}{\alpha+\beta}; \mu_2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Chi Square Distribution

This distribution represents the sum of n Gaussian distributed variables with mean 0 and variance 1. It is used to model variance. It is also used in some statistical tests

Domain	$X > 0$
Parameters	$n > 0$
Function	$f(X = x) = \frac{1}{2^{n/2}\Gamma(n/2)} e^{-x/2} x^{n/2-1}$
Moments	$\mu_1' = n; \mu_2 = 2n$

Probability Distributions

Random vectors

Bivariate Normal Distribution

Extension of the Gaussian distribution to two dimensions

Domain $(X, Y) \in \mathbb{R}^2$

Parameters $\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$

Function

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-a}$$

$$a = -\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]$$

The marginals of the bivariate distribution are Gaussian distributions

Joint Moments, bivariate case

First Order

$$E[X] = \int_{\Omega_X} \int_{\Omega_Y} x f(x,y) dx dy \qquad E[Y] = \int_{\Omega_X} \int_{\Omega_Y} y f(x,y) dx dy$$

Second order

$$E[X^2] = \int_{\Omega_X} \int_{\Omega_Y} x^2 f(x,y) dx dy \qquad E[Y^2] = \int_{\Omega_X} \int_{\Omega_Y} y^2 f(x,y) dx dy$$

$$E[XY] = \int_{\Omega_X} \int_{\Omega_Y} xy f(x,y) dx dy$$

Second order around the mean: Covariance

$$E[(X - E[X])(Y - E[Y])] = \int_{\Omega_X} \int_{\Omega_Y} (x - E[X])(y - E[Y]) f(x,y) dx dy$$

Useful properties

Variance of Sum

$$VAR(aX + bY + c) = a^2VAR(X]) + b^2VAR(Y) + 2abCOV(X, Y)$$

Correlation

$$\rho_{XY} = \frac{COV(X,Y)}{\sqrt{VAR(X)VAR(Y)}}$$

Independent variables

$$COV(X, Y) = 0 \rightarrow \rho_{XY} = 0$$

Dirichlet Distribution

Extension of the Beta distribution to two or more dimensions

Domain $(X_1, \ldots, X_n) \in [0,1]^n; \sum_i x_i = 1$

Parameters $(\alpha_1, \ldots, \alpha_n); \alpha_i > 0$

Function $f(x_1, \ldots, x_n) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \prod_i x_i^{\alpha_i - 1}$

The marginals of the Dirichlet distribution are Beta distributions