

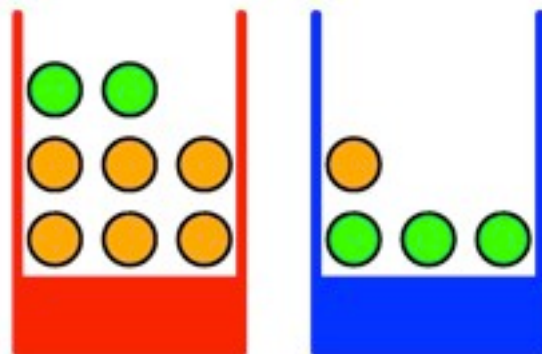
Review

Probability Theory and the Gaussian Distribution

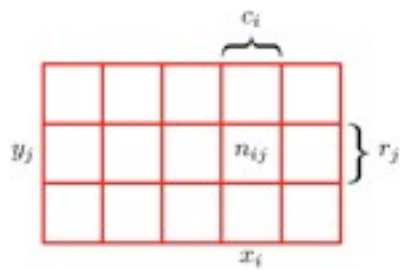
1

Probability Theory

Apples and Oranges



Probability Theory



Marginal Probability

$$p(X = x_i) = \frac{c_i}{N}$$

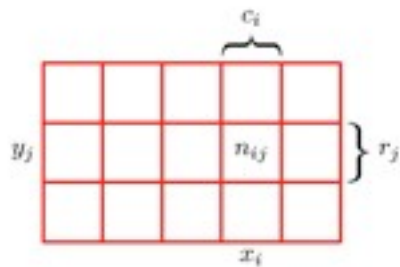
Joint Probability

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

Conditional Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

Probability Theory



Sum Rule

$$\begin{aligned} p(X = x_i) &= \frac{c_i}{N} = \frac{1}{N} \sum_{j=1}^L n_{ij} \\ &= \sum_{j=1}^L p(X = x_i, Y = y_j) \end{aligned}$$

Product Rule

$$\begin{aligned} p(X = x_i, Y = y_j) &= \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} \\ &= p(Y = y_j | X = x_i) p(X = x_i) \end{aligned}$$

The Rules of Probability

Sum Rule

$$p(X) = \sum_Y p(X, Y)$$

Product Rule

$$p(X, Y) = p(Y|X)p(X)$$

Bayes' Theorem

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

$$p(X) = \sum_Y p(X|Y)p(Y)$$

posterior \propto likelihood \times prior

Independent random variables

if $p(X, Y) = p(X)p(Y)$

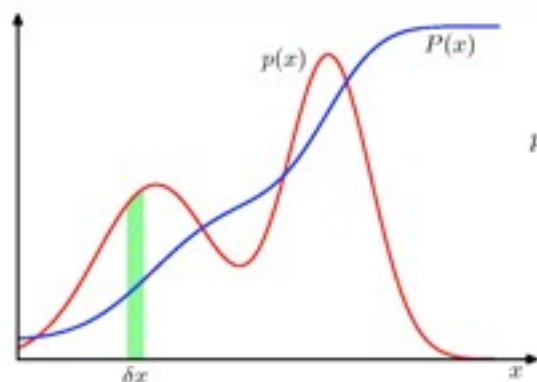
then X and Y are said to be *independent*.

In this case, it holds:

$$p(Y|X) = \frac{p(X, Y)}{p(X)} = p(Y)$$

7

Probability Densities



$$p(x \in (a, b)) = \int_a^b p(x) dx$$

$$p(x) \geq 0$$

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

$$P(z) = \int_{-\infty}^z p(x) dx \quad \text{Cumulative distribution function}$$

The Rules of Probability

for continuous variables

$$\text{Sum Rule} \quad p(x) = \int p(x, y) dy$$

$$\text{Product Rule} \quad p(x, y) = p(y|x)p(x)$$

Expectations

$$E[f] = \sum_x p(x)f(x)$$

$$E[f] = \int p(x)f(x) dx$$

$$E[f] \simeq \frac{1}{N} \sum_{n=1}^N f(x_n)$$

Approximate Expectation
(discrete and continuous)

The approximation becomes exact in the limit

$$N \rightarrow \infty$$

The sample mean is an *unbiased* estimator of the expectation (population mean)

Variations and Covariances

$$\text{var}[f] = \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

$$\text{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2$$

The variance measures the variability of $f(x)$ around its mean value

Variations and Covariances

$$\begin{aligned}\text{cov}[x, y] &= \mathbb{E}_{x,y}[\{x - \mathbb{E}[x]\} \{y - \mathbb{E}[y]\}] \\ &= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y]\end{aligned}$$

$$\begin{aligned}\text{cov}[\mathbf{x}, \mathbf{y}] &= \mathbb{E}_{\mathbf{x}, \mathbf{y}}[\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\} \{\mathbf{y}^T - \mathbb{E}[\mathbf{y}^T]\}] \\ &= \mathbb{E}_{\mathbf{x}, \mathbf{y}}[\mathbf{x}\mathbf{y}^T] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}^T]\end{aligned}$$

The covariance of two random variables measures the extent to which they vary together.

If x and y are independent, then $\text{cov}[x, y] = 0$

Variances and Covariances

covariance of \mathbf{x} :

$$\text{cov}[\mathbf{x}, \mathbf{x}] = \text{cov}[\mathbf{x}] = E_{\mathbf{x}}[(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])^T]$$

Given a random sample $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ of \mathbf{x} ,

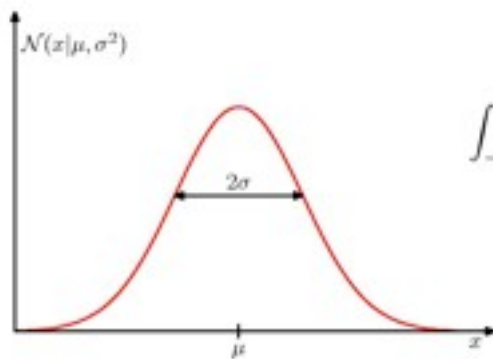
the *sample covariance matrix* is $Q = [q_{ij}]$

$$q_{ij} = \frac{1}{N-1} \sum_{k=1}^N (x_{ik} - \mu_i)(x_{jk} - \mu_j)$$

The sample covariance matrix is an unbiased estimate of the covariance matrix.

The Gaussian Distribution

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}$$



$$\mathcal{N}(x|\mu, \sigma^2) > 0$$

$$\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) dx = 1$$

Gaussian Mean and Variance

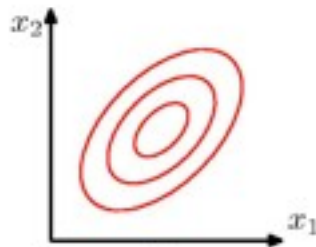
$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) x dx = \mu$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) x^2 dx = \mu^2 + \sigma^2$$

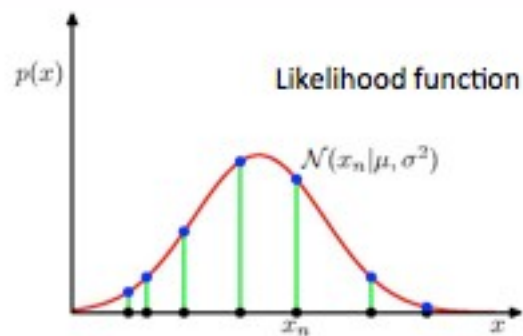
$$\text{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$$

The Multivariate Gaussian

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}$$



Gaussian Parameter Estimation



$$p(\mathbf{x}|\mu, \sigma^2) = \prod_{n=1}^N \mathcal{N}(x_n|\mu, \sigma^2)$$

Maximum (Log) Likelihood

$$\ln p(\mathbf{x}|\mu, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{n=1}^N (x_n - \mu)^2 - \frac{N}{2} \ln \sigma^2 - \frac{N}{2} \ln(2\pi)$$

$$\mu_{\text{ML}} = \frac{1}{N} \sum_{n=1}^N x_n \quad \sigma_{\text{ML}}^2 = \frac{1}{N} \sum_{n=1}^N (x_n - \mu_{\text{ML}})^2$$

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}$$

Properties of μ_{ML} and σ_{ML}^2

$$E[\mu_{\text{ML}}] = \mu$$

$$E[\sigma_{\text{ML}}^2] = \left(\frac{N-1}{N}\right) \sigma^2$$

$$\begin{aligned} \tilde{\sigma}^2 &= \frac{N}{N-1} \sigma_{\text{ML}}^2 \\ &= \frac{1}{N-1} \sum_{n=1}^N (x_n - \mu_{\text{ML}})^2 \end{aligned}$$

