L02: The Expectation Operator and Gaussians

* **Def** Expectation Operator

$$E\{x\} = \int_{-\infty}^{+\infty} x\, p(x)\, dx = \mu_x \quad \text{(Weighted average)}$$

- Note: $p(x)$ is a pdf not probability (class conv.)

* Properties of $E\{\}$ — Linear operator

$$E\{A\} = A$$

$$E\{Ax\} = A\,E\{x\}$$

$$E\{A+x\} = A + E\{x\}$$

$$E\{x+y\} = E\{x\} + E\{y\}$$

Proof: $E\{x+y\} = \iint (x+y)\, p(x,y)\, dx\, dy =$

$$= \iint x\, p(x,y)\, dx\, dy + \iint y\, p(x,y)\, dx\, dy =$$

$$= \int x \underbrace{\int p(x,y)\, dy}_{} dx + \int y \int p(x,y)\, dx\, dy =$$

marginalization or total probability (see L01.5 on Canvas)

$$= \int x\, p(x)\, dx + \int y\, p(y)\, dy$$

- Anti - properties

$$E\{x \cdot y\} \neq E\{x\}\, E\{y\} \quad \text{(in general)}$$

If x, y uncorrelated $(\sigma_{xy}^2 = 0) \Rightarrow E\{xy\} = E\{x\}E\{y\}$

If x, y independent $\underset{\not\Leftarrow}{\Rightarrow}$ x, y uncorrelated

- Expectation of a multidimensional r.v.

$$E\left\{\begin{bmatrix} x \\ y \end{bmatrix}\right\} = \begin{bmatrix} E\{x\} \\ E\{y\} \end{bmatrix}$$

- Conditional expectation $\quad E\{x \mid y\} \triangleq \int_{-\infty}^{\infty} x\, p(x \mid y)\, dx$

Covariance   Scalar form

(Cross covariance)

$$\sigma_{xy}^2 = \operatorname{cov}(x, y) = E\{(x - E\{x\})(y - E\{y\})\}$$

$$\sigma_{xx}^2 = \operatorname{cov}(x, x) = E\{(x - E\{x\})^2\}$$

(Autocovariance or variance)

$$= E\{x^2\} - E\{x\}^2 = \operatorname{var}(x)$$

Exercise: Show this

Property: $\sigma_{xx}^2 \geq 0$

Covariance. Vectorial form

(cross covariance)

$$\Sigma_{xy} = cov(x,y) = E\{(x - E\{x\})(y - E\{y\})^T\}$$

$$\Sigma_x = cov(x,x) = cov(x) = E\{(x - E\{x\})(x - E\{x\})^T\}$$

Ex: expand Σ_{xy}

$$E\{xy^T + x\cdot(-E\{y\})^T - E\{x\}y^T + E\{x\}E\{y\}^T\} =$$

$$= E\{xy^T\} - E\{x E\{y\}^T\} - E\{E\{x\}y^T\} + E\{E\{x\}E\{y\}^T\} =$$

$$= E\{xy^T\} - E\{x\}E\{y\}^T - \cancel{E\{x\}E\{y\}^T} + \cancel{E\{x\}E\{y\}^T}$$

*Note: $\Sigma_{xy} = 0 \iff \underbrace{E\{xy^T\} = E\{x\}E\{y\}^T}_{\text{Uncorrelated}}$

Exer: expand Σ_x

Col 1: Σ_x is symetric : $\Sigma_x = \Sigma_x^T$ (not Σ_{xy}!)

$$\Sigma_x \underset{\mu_x = 0}{=} E\{x\cdot x^T\} = E\left\{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}\right\} = E\left\{\begin{bmatrix} x_1x_1 & x_1x_2 & \\ x_2x_1 & x_2x_2 & \\ x_3x_1 & x_3x_2 & x_3x_3 \end{bmatrix}\right\}$$

$x_2x_1 \in \mathbb{R}$, $x_1x_2 = x_2x_1$

Col 2: Σ_x is Positive Semi definite (psd)

$$v^T \cdot \Sigma_x \cdot v \geq 0 \quad , \quad \forall v$$

proof: $v^T E\{(x-\mu)(x-\mu)^T\} v =$

$$= E\{\underbrace{v^T(x-\mu)}_{u \in \mathbb{R}}(x-\mu)^T v\} = E\{u^2\} \geq 0$$

Sample mean and sample covariance.

$x_i \sim p(x)$ iid (independent and identically distributed)

Sample mean: $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$

Sample covariance $\bar{\Sigma}_x = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T$

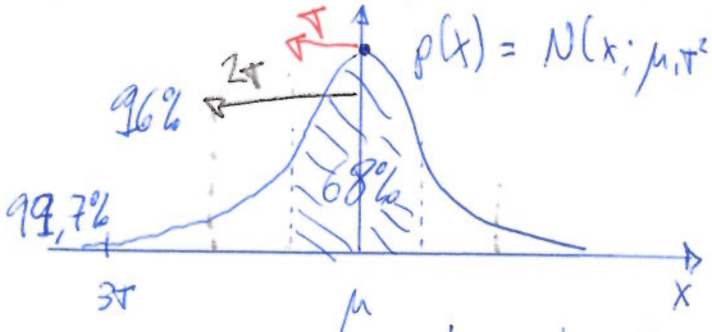
Gaussian distribution (univariate)

$$p(x) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{\text{normalization factor}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} = N(x; \mu, \sigma^2)$$

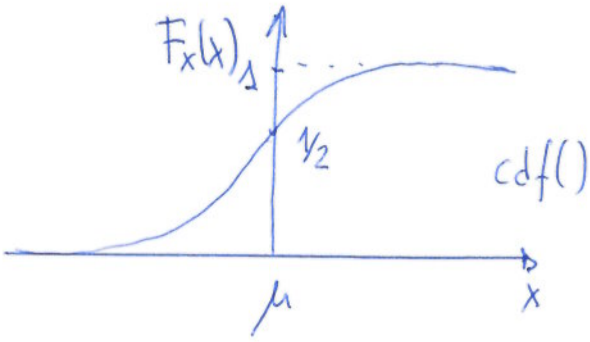
normalization factor $\Rightarrow \int N(x; \mu, \sigma^2) dx = 1$

* Probability density function (pdf)

Cumulative distribution f



mean, median and mode.



$$\mu = E\{x\} = \int_{-\infty}^{\infty} x\, p(x)\, dx$$

$$\sigma^2 = cov(x) = E\{(x-\mu)^2\}$$

All required parameters to describe a Gaussian.

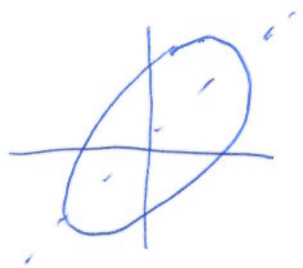
$x \sim N(\mu, \sigma^2)$ draw a sample. Most functions for $\mu = 0, \sigma = 1$

* Multivariate Gaussian.

$$p(x) = \frac{1}{(2\pi)^{K/2} |\Sigma_x|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma_x^{-1}(x-\mu)} = N(x; \mu, \Sigma_x)$$

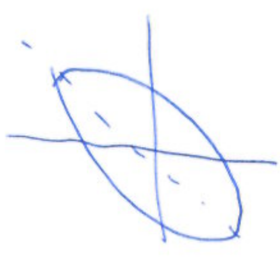
<u>Ex</u>: Intuition on a 2D Gaussian.

$$\Sigma_x = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$



positive correlation.

$$\Sigma_x = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$



negative correlation.

* Covariance Projection

r.v $x \sim N(\mu_x, \Sigma_x)$, r.v $y = f(x)$

+ Affine transformation f: $y = Ax + b$

· $\mu_y = E\{y\} = E\{Ax + b\} = A\,E\{x\} + b = A\mu_x + b$

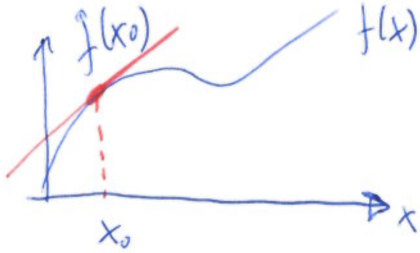
$$\Sigma_y = E\{(y-\mu_y)(y-\mu_y)^T\} = \qquad (AB)^T = B^T A^T$$

$$= E\{(Ax + \cancel{b} - A\mu_x - \cancel{b})(Ax + \cancel{b} - A\mu_x - \cancel{b})^T\} \overset{\downarrow}{=}$$

$$= E\{A\underbrace{(x-\mu_x)(x-\mu_x)^T}_{\Sigma_x}A^T\} = A \cdot \Sigma_x A^T$$

* Non-linear covariance projection

$y = f(x)$



- 1D

$$y = f(x_0) + \underbrace{\left.\frac{df(x)}{dx}\right|_{x_0}}\cdot(x - x_0) + O((x-x_0)^2)$$

computed analytically or numerically.

Taylor expansion

- N-Dimentions

$$y = f(x_0) + \underbrace{\sum_{i=1}^{d} \frac{\partial f(x)}{\partial x^i}\Big|_{x_0}}\cdot(x^i - x_0^i) + O(\|x-x_0\|^2) =$$

Jacobian : $\left[\frac{\partial f}{\partial x^1}, \frac{\partial f}{\partial x^2}, \dots, \frac{\partial f}{\partial x^d}\right] = J$

$$= f(x_0) + J\cdot(x - x_0) + O(\|x-x_0\|^2) \simeq$$

$$\simeq f(x_0) + Jx - Jx_0 = \underbrace{J}_{A}\cdot x + \underbrace{f(x_0) - J\cdot x_0}_{b}$$

$$\Rightarrow y \sim N(f(\mu_x), A\Sigma_x A^T)$$

Linearization is not exempt of problems

Error assumed $O(\Delta x^2)$

Uncented transformation allevates this (L07).

L 03 : Gaussians II

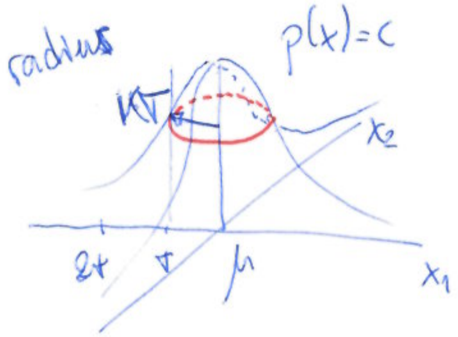
* Visualizing Gaussians (2D)

$$p(x) = \alpha \cdot e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

find contours of constant $p(x)$

$$K^2 = (x-\mu)^T \Sigma^{-1} (x-\mu) \quad , \quad K = 1, 2, \ldots$$

$K=1 \rightarrow 1\text{-}\sigma$



Isocontours are eqs of a quadratic curve.

Ellipse.

* Mahalanobis distance.

$$\|x-y\|_\Sigma = \sqrt{(x-y)^T \Sigma^{-1}(x-y)}$$

* Non-spherical Σ : we use covariance projection and L.A.

(recall:) if $y = Ax + b$ \ $x \sim N(0, \mathbb{I})$

then $\Sigma_y = A \Sigma_x A^T$

Problem: find A such that projects into Σ_y.

given $\Sigma_x = \mathbb{I}$ (standard $N(0, \mathbb{I})$.

$$\Sigma_y = A \cdot \Sigma_x \cdot A^T = A \cdot \mathbb{I} \cdot A^T = AA^T$$

1) SVD decomposition:

$$\Sigma_y = U \cdot D V^T = U \cdot D \cdot U^T = \underbrace{U \cdot D^{1/2}}_{A} \cdot \underbrace{D^{T/2} U^T}$$

symetric.

2) Cholesky decomposition. (More efficient.)

$$\Sigma_y = L \cdot L^T$$

Ex:

$$\Sigma = \begin{bmatrix} 4 & -2 \\ -2 & 10 \end{bmatrix}$$ isocontour for $K=1$? (1-sigma)

$$= \underbrace{\begin{bmatrix} a & 0 \\ b & c \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}}_{L^T} = \begin{bmatrix} 4 & -2 \\ -2 & 10 \end{bmatrix}$$

$$a^2 = 4 \Rightarrow a = \pm 2$$

$$ba = ab = -2 \Rightarrow b = -1$$

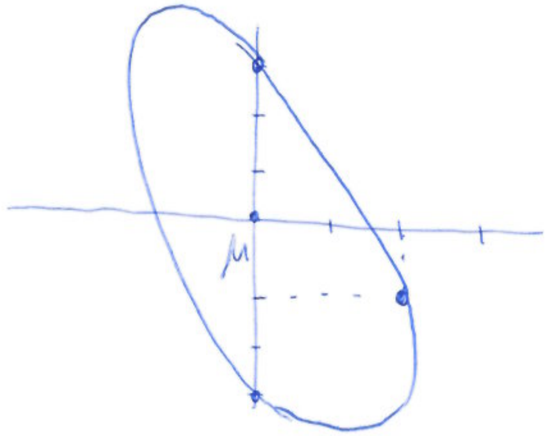
$$b^2 + c^2 = 10 \Rightarrow c = \sqrt{10 - b^2} = \pm 3$$

We choose positive values for diagonal $\Rightarrow \exists ! L$
(unique solution)

Project points from the circumference $r=1$

$$\begin{bmatrix} 2 & 0 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} & L & \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$



note: Solution is centered at μ, so we need to translate.

* Sampling from Gaussians

Most functions sample $x \sim N(0,1)$

Problem: how to sample $y \sim N(\mu_y, \Sigma_y)$?

Solution: use covariance projection. Affine $y = Ax + b$

1) Select $x \sim N(0, I_D)$. $\quad x = \begin{bmatrix} N(0,1) \\ N(0,1) \\ \vdots \\ N(0,1) \end{bmatrix}$ iid

2) $y \sim N(A\mu_x + b, A\Sigma_x A^T)$

$\Rightarrow \boxed{b = \mu_y}$

3) Find A such as $\boxed{A \cdot A^T = \Sigma_y}$ (Cholesky)

4) $x \sim N(0, I)$. then $y = A \cdot x + b$

equiv. $y \sim N(\mu_y, \Sigma_y)$.