Trace trick

- Trace:

$$\mathrm{Trace}(B) = \mathrm{Tr}(B) = \sum_{i=1}^{n} B_{ii} = \text{“sum of diagonal elements”}$$

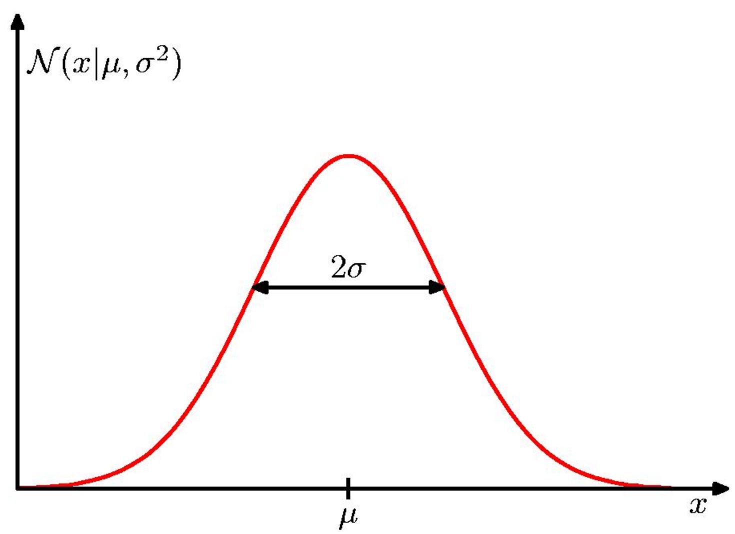
- Cyclic permutations are allowed inside trace

$$\mathrm{Tr}(BCD) = \mathrm{Tr}(DBC) = \mathrm{Tr}(CDB)$$

- Trace of a scalar is a scalar

$$\mathbf{x}^\top B\mathbf{x} = \mathrm{Tr}(\mathbf{x}^\top B\mathbf{x}) = \mathrm{Tr}(\mathbf{x}\mathbf{x}^\top B)$$

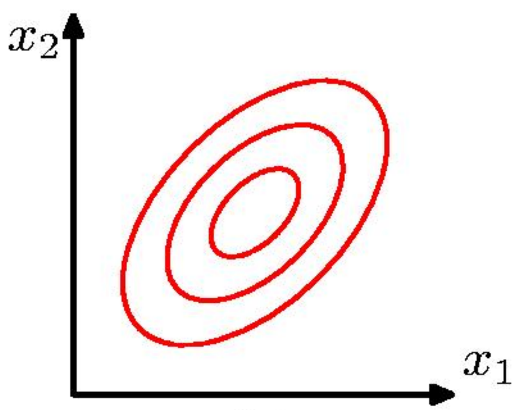
Gaussian

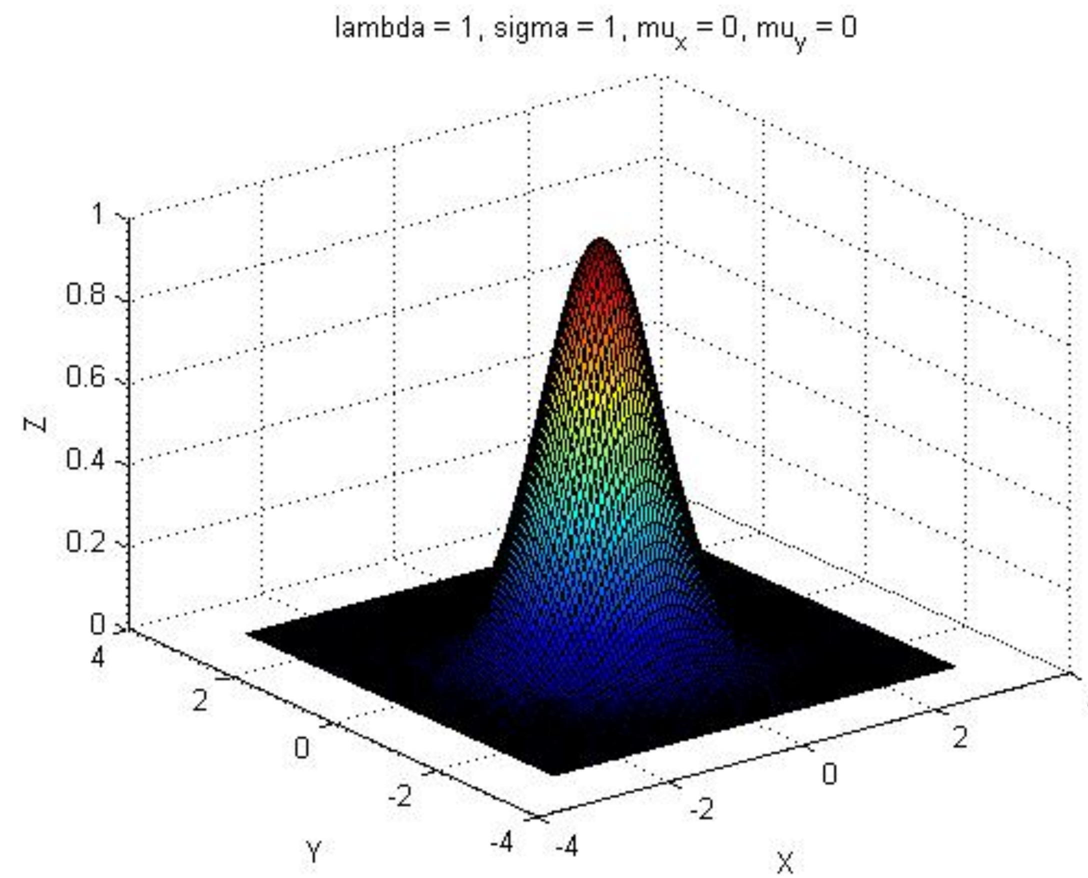


$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

or Normal

Multivariate Gaussian





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

Why a Gaussian?

- Central Limit Theorem
  - Distribution of the sum of N iid random variables becomes increasingly Gaussian as N grows.
  - example

Gaussian

- Expectation

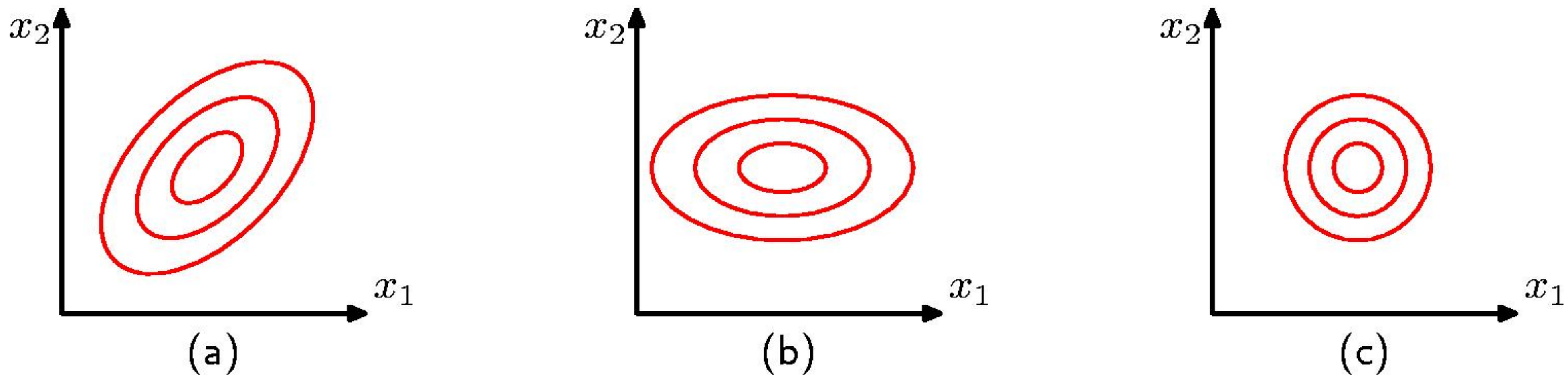
$$\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}$$

- Second moment

$$\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathrm{T}}] = \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} + \boldsymbol{\Sigma}.$$

- Covariance

$$\operatorname{cov}[\mathbf{x}] = \mathbb{E}\left[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^{\mathrm{T}}\right] = \boldsymbol{\Sigma}$$



Gaussian

- Lots of good properties.
  - Linear transform of a gaussian variable is a gaussian

$$\left.\begin{aligned} \mathbb{E}(AX+b) &= A\mathbb{E}(X)+b \\ \mathtt{Cov}(AX+b) &= A\mathtt{Cov}(X)A^T \end{aligned}\right\}\text{General!}$$

so:

$$X \sim \mathcal{N}(\mu, \Sigma) \quad \Rightarrow \quad AX+b \sim \mathcal{N}(A\mu+b, A\Sigma A^T).$$

Gaussian

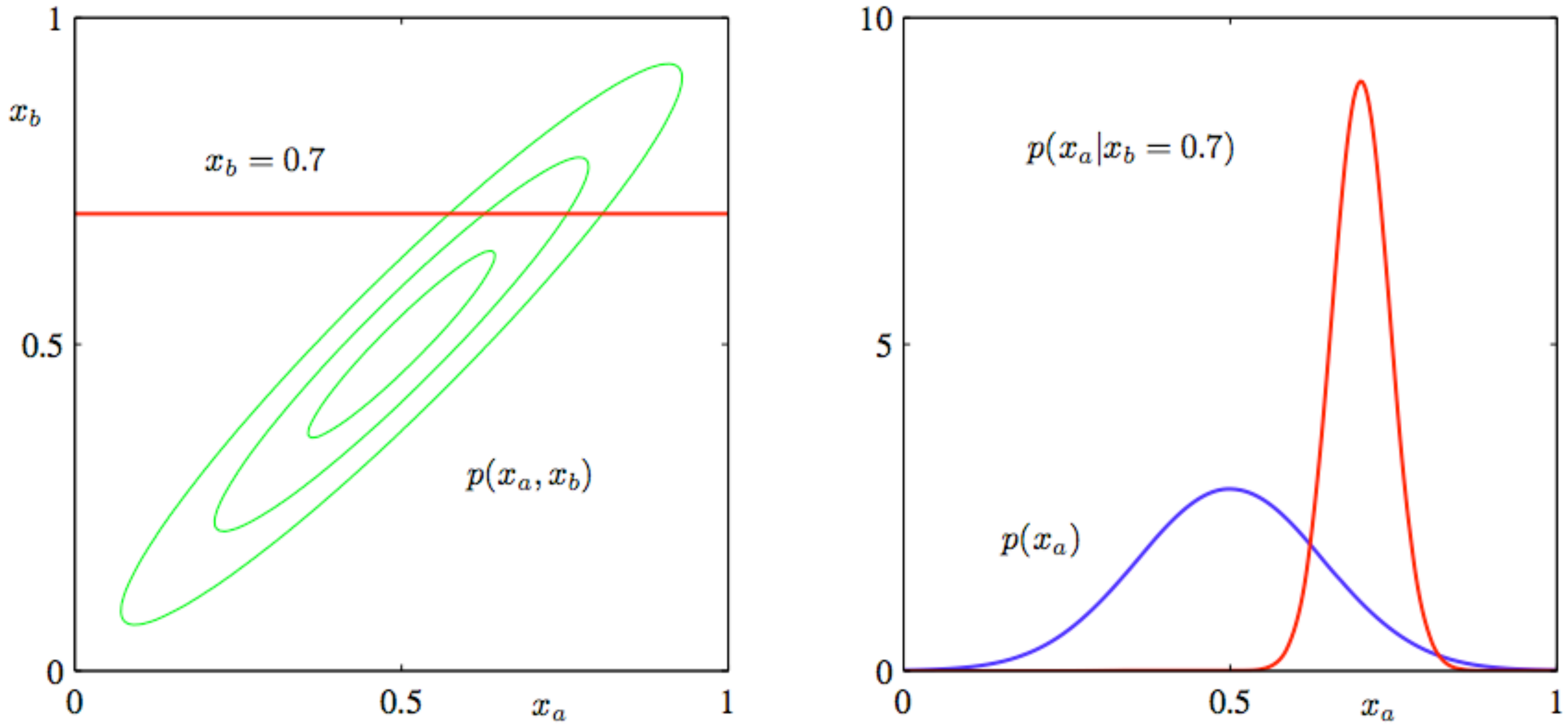
- Let X_1, X_2 ~ Normal, indep

$$Y = X_1 + X_2, X_1 \perp X_2 \quad \Rightarrow \quad \mu_Y = \mu_1 + \mu_2, \quad \Sigma_Y = \Sigma_1 + \Sigma_2$$

  - $p(X_1) \cdot p(X_2)$ ~ Multivariate Normal
  - $X_1 \mid X_2$ ~ Normal

- $p(\mathbf{x}_a, \mathbf{x}_b)$ Normal:
  - marginal is Normal: $p(\mathbf{x}_a) = \int p(\mathbf{x}_a, \mathbf{x}_b) \, \mathrm{d}\mathbf{x}_b$

Conditional



Working with Gaussians

- Goal: End up with a gaussian form
- Expand, collect all terms, complete the square.
- See Bishop, Section 2.3.

Working with Gaussians

- Life saver: Equations 2.113 - 2.117
  - Try to derive them!
  - -> Geometric intuition

Marginal and Conditional Gaussians

Given a marginal Gaussian distribution for $\mathbf{x}$ and a conditional Gaussian distribution for $\mathbf{y}$ given $\mathbf{x}$ in the form

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}) \tag{2.113}$$

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1}) \tag{2.114}$$

the marginal distribution of $\mathbf{y}$ and the conditional distribution of $\mathbf{x}$ given $\mathbf{y}$ are given by

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}}) \tag{2.115}$$

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\Sigma}\{\mathbf{A}^{\mathrm{T}}\mathbf{L}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Lambda}\boldsymbol{\mu}\}, \boldsymbol{\Sigma}) \tag{2.116}$$

where

$$\boldsymbol{\Sigma} = (\boldsymbol{\Lambda} + \mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{A})^{-1}. \tag{2.117}$$

Additional infos

- List of cribsheets and additional material soon on the website.