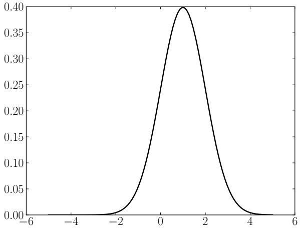
2.4. Normal distribution and maximum likelihood

Prof. Dr. Christoph Lippert

Digital Health & Machine Learning

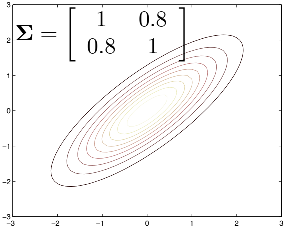
- Normal distribution (Gaussian distribution)

$$p(x;\mu,\sigma^2) = \mathcal{N}\left(x \,\middle|\, \mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$



- Multivariate normal distribution

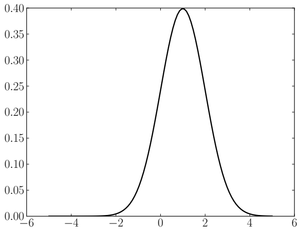
$$p(\mathbf{x}\ ;\ \boldsymbol{\mu},\boldsymbol{\Sigma}) = \mathcal{N}(\mathbf{x}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma})$$
$$= \frac{1}{\sqrt{|2\pi\boldsymbol{\Sigma}|}} \cdot \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

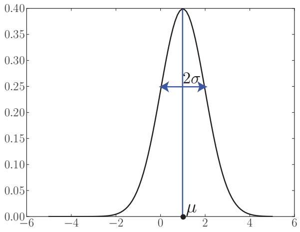


- Gaussian PDF

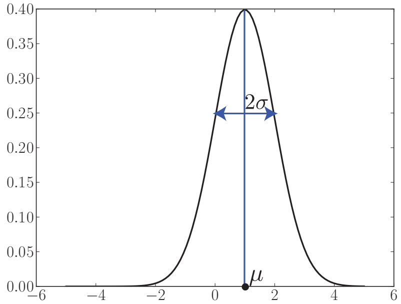
$$\mathcal{N}\left(x \,\middle|\, \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

- Positive: $\mathcal{N}\left(x \,\middle|\, \mu, \sigma^2\right) > 0$
- Normalized: $\int_{-\infty}^{+\infty} \mathcal{N}\left(x \,|\, \mu, \sigma\right) \mathrm{d}x = 1$ (check)
- Expectation: $E(x) = \int_{-\infty}^{+\infty} \mathcal{N}\left(x \,\middle|\, \mu, \sigma^2\right)\, x \mathrm{d}x = \mu$
- Variance: Var$[x] =< x^2 > - < x >^2$
  $= \mu^2 + \sigma^2 - \mu^2 = \sigma^2$





Parameter estimation for the normal distribution



- Data sampled from unknown distribution $p(\mathcal{D}\ ;\ \boldsymbol{\theta}_0)$

$$\mathcal{D} = \{x_1, \dots, x_N\} \sim p(\mathcal{D}\ ;\ \boldsymbol{\theta}_0)$$

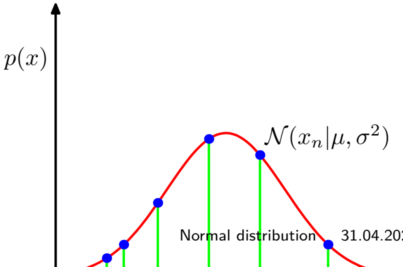
- Model $\mathcal{H}_{\mathsf{Gauss}}$ – normal PDF

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

$$\boldsymbol{\theta} = \{\mu, \sigma^2\}$$

- Likelihood

$$p(\mathcal{D}\ ;\ \boldsymbol{\theta}) = \prod_{n=1}^{N} \mathcal{N}\left(x_n \,\middle|\, \mu, \sigma^2\right)$$



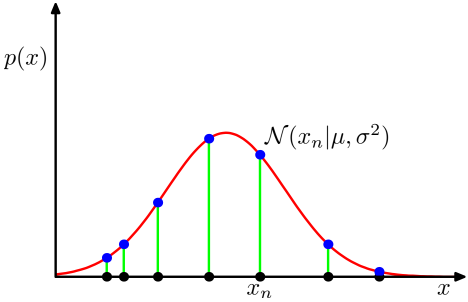
Parameter estimation for the normal distribution

- Likelihood

$$p(\mathcal{D}\ ;\ \boldsymbol{\theta}) = \prod_{n=1}^{N} \mathcal{N}\left(x_n \,\middle|\, \mu, \sigma^2\right)$$

- Maximum likelihood
- Chose parameters $\hat{\mu}$ and $\hat{\sigma}^2$ that *maximize* the likelihood of $\mathcal{D}$

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\mathcal{D}\ ;\ \boldsymbol{\theta})$$



(C.M. Bishop, Pattern Recognition and Machine Learning)

Maximum likelihood estimation in the normal distribution

- Data sample $\mathcal{D}$ of size N modeled by a univariate normal distribution
- Likelihood of the data under the model $p(\mathcal{D}\ ;\ \mu, \sigma^2)$:

$$\prod_{n=1}^{N} \mathcal{N}\left(x_n \,\middle|\, \mu, \sigma^2\right)$$

$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2}(x_n-\mu)^2}$$

- Chose parameters $\hat{\mu}$ and $\hat{\sigma}^2$ that *maximize* the likelihood of $\mathcal{D}$

- Equivalently maximize the *log-Likelihood* $\log p(\mathcal{D}\ ;\ \mu, \sigma^2)$

$$\log p(\mathcal{D}; \mu, \sigma^2) = \sum_{n=1}^{N} \log \mathcal{N}\left(x_n \,\middle|\, \mu, \sigma^2\right)$$

$$= \sum_{n=1}^{N} -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}(x_n - \mu)^2$$

- Chose parameters $\hat{\mu}$ and $\hat{\sigma}^2$ that *maximize* the likelihood of $\mathcal{D}$

$$\log p(\mathcal{D}\ ;\ \mu, \sigma^2) = \sum_{n=1}^{N} -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}(x_n - \mu)^2$$

- Take the derivative of $\log p(\mathcal{D}\ ;\ \mu, \sigma^2)$ with respect to μ:

$$\frac{\partial \log p(\mathcal{D}\ ;\ \mu, \sigma^2)}{\partial \mu} = -\frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n - \mu)$$

- set to zero and solve for $\hat{\mu}$:

$$-\frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n - \hat{\mu}) = 0$$

$$-\frac{1}{\sigma^2}(\sum_{n=1}^{N} x_n) + \frac{N}{\sigma^2}\hat{\mu} = 0$$

$$\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} x_n \quad \text{sample mean}$$

- Take the derivative of $\log p(\mathcal{D}\ ;\ \mu, \sigma^2)$ with respect to σ^2:

$$\frac{\partial \log p(\mathcal{D}\ ;\ \mu, \sigma^2)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \sum_{n=1}^{N}\frac{1}{2\sigma^4}(x_n - \mu)^2$$

- set to zero and solve for $\hat{\sigma}^2$:

$$-\frac{N}{2\hat{\sigma}^2} + \sum_{n=1}^{N}\frac{1}{2\hat{\sigma}^4}(x_n - \hat{\mu})^2 = 0$$

$$\frac{N\hat{\sigma}^2}{2} = \sum_{n=1}^{N}\frac{1}{2}(x_n - \hat{\mu})^2$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - \hat{\mu})^2 \quad \text{sample variance}$$

Parameter estimation for the Gaussian

- Maximum likelihood solutions

$$\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - \hat{\mu})^2$$

Equivalent to common mean and variance estimators (almost).

Summary

- The Normal Distribution
  - Maximum Likelihood Estimation
  - sample mean
  - sample variance