Expectation and Variance in the General Setting

$$L^k(\Omega, \mathcal{A}, P) := \left\{ X : \Omega \to \mathbb{R} \;\middle|\; X \text{ measurable and } \int_\Omega |X|^k \, dP < \infty \right\}$$

$(\Omega, \mathcal{A}, P)$ probability space, $X : \Omega \to \mathbb{R}$ with distribution $P_X = X(P)$, $x \in L^1(\Omega, \mathcal{A}, P)$. The expectation of X is defined as

$$E(X) := \int_\Omega X \, dP = \int_{\mathbb{R}} x \, dP_X(x)$$

$$\left(\text{case of density: } \int_{\mathbb{R}} x \, f(x) \, dx \right)$$

If $X^k \in L^1(\Omega, \mathcal{A}, P)$ then

$E(X^k) = \int x^k \, dP$ is called the k-th moment of X.

If $x \in L^2(\Omega, \mathcal{A}, P)$ we define

$$\mathrm{Var}(X) = E\left((X - E(X))^2 \right)$$

$$\mathrm{Cov}(X, Y) = E\left((X - E(X)) \cdot (Y - E(Y)) \right)$$

Markov and Chebyshev Inequalities

Cauchy - Schwatz - inequality

$X, Y \in L^2(\Omega, \mathcal{A}, P)$ then:

$$E(X \cdot Y)^2 \leq E(X^2) \cdot E(Y^2)$$

Markov inequality: $\varepsilon > 0$, $f: [0,\infty) \to [0,\infty)$, f monotonically increasing. Then

$$P(|Y| > \varepsilon) \leq \frac{E(f(|Y|))}{f(\varepsilon)}$$

In particular,

$$P(|Y| > \varepsilon) \leq \frac{E(|Y|)}{\varepsilon}$$

Chebyshev inequality: $\varepsilon > 0$, $X \in L^2(\Omega, \mathcal{A}, P)$.

Then:

$$\underbrace{P(|X - E(X)| > \varepsilon)}_{\text{important quantity in learning theory}} \leq \frac{Var(X)}{\varepsilon^2}$$

Examples of Probability Distributions

Discrete distributions:

- Uniform distribution on $\{1, \ldots, n\}$

$$P(\{i\}) = \frac{1}{n}$$

- Binomial distribution on $\{0, \ldots, n\}$

Toss a coin n times, independently, each time with probability p of observing head.

Denote head $= 1$, tail $= 0$, $X :=$ # heads.

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

- Poisson distribution on $\mathbb{N}$

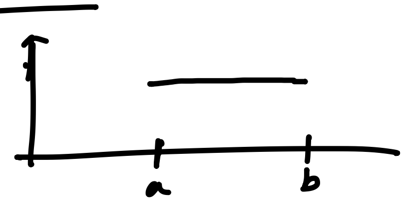
Parameter $\lambda > 0$

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

Intuition: number of jobs submitted to a cloud service.

Continuous distributions:

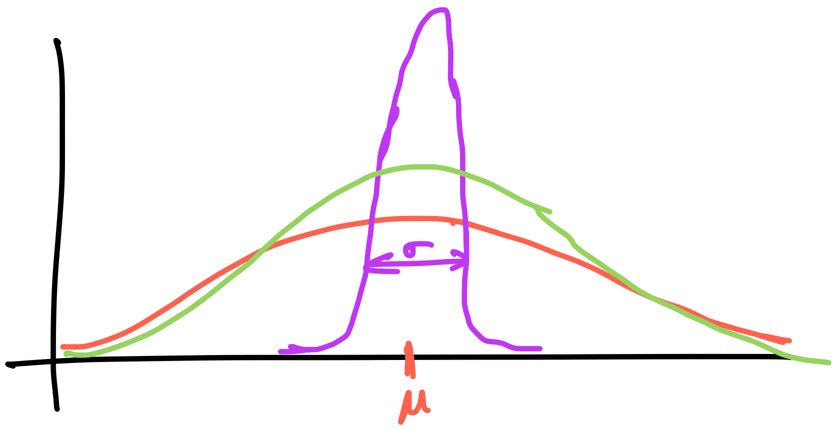
Uniform distribution on $[a, b]$: constant density



Normal distribution on $\mathbb{R}$

Density: parameter μ (mean), σ (std. deviation)

$$f_{\mu,\sigma}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



Notation: $\mathcal{N}(\mu, \sigma^2)$

Some properties:

- $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

  X, Y are independent

  then $X+Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_1^2)$

Normal distribution in higher dimensions

$$X : \Omega \to \mathbb{R}^n, \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mu_i \in E(X_i), \ \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$$

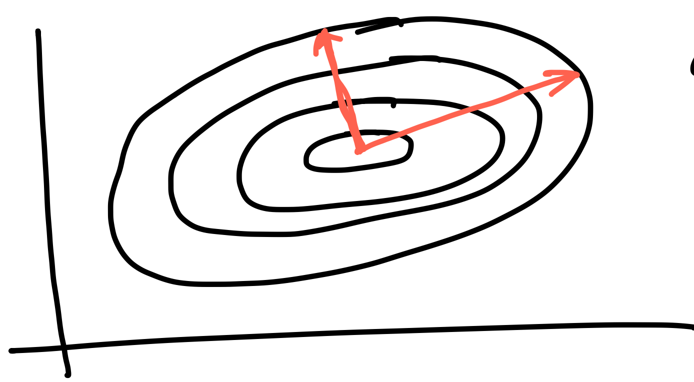
$\Sigma \in \mathbb{R}^{n \times n}$ with $\Sigma_{ij} = \operatorname{cov}(X_i, X_j)$, called the covariance matrix.

$$f_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{n/2} |\det \Sigma|^{1/2}} \exp\left(-\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right)$$

Notation: $\mathcal{N}(\mu, \Sigma)$

Prop: · Σ is positive semi-definite and symmetric.

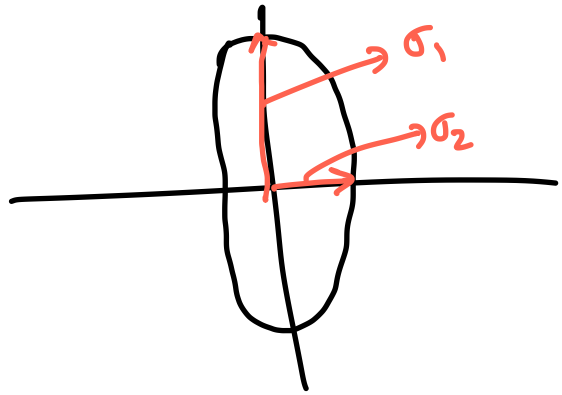
Consequence: Σ has real-valued, non-negative eigs.



contour lines of $f_{\mu,\Sigma}$

directions of eigenvectors.

· $X_1, X_2, \ldots X_n$ are independent $\iff \Sigma = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots \ \sigma_n^2 \end{pmatrix}$



- $X \sim \mathcal{N}(\mu_1, \Sigma_1)$, $Y \sim \mathcal{N}(\mu_2, \Sigma_2)$, independent

  then $\quad X+Y \sim \mathcal{N}(\mu_1+\mu_2, \Sigma_1+\Sigma_2)$

Mixture of Gaussians

Consider $\pi_1, \pi_2, \dots \pi_k$ with $0 \leq \pi_i \leq 1$ & $\sum \pi_i = 1$

Consider the following density:

$$f(x) = \sum_{i=1}^{k} \pi_i \, f_{\mu_i, \Sigma_i}(x)$$