



Why

What of the generalization to a multivariate normal.

Result

Prop. 1. *Let (x^1, \dots, x^n) be a dataset in \mathbb{R}^d . Let f be a multivariate normal density with mean*

$$\frac{1}{n} \sum_{k=1}^d x^k$$

and covariance

$$\frac{1}{n} \sum_{k=1}^n \left(x^k - \frac{1}{n} \sum_{k=1}^n x^k \right) \left(x^k - \frac{1}{n} \sum_{k=1}^n x^k \right)^\top.$$

Then f is a maximum likelihood multivariate normal density.

Proof. We express the log likelihood

$$\sum_{k=1}^n -\frac{1}{2} (x^k - \mu)^\top \Sigma^{-1} (x^k - \mu) - \frac{1}{2} \log(2\pi)^d - \frac{1}{2} \log \det \Sigma$$

Let $P = \Sigma^{-1}$. The $\log \det \Sigma$ is $\log \det P^{-1}$ is $\log (\det P)^{-1}$ is $-\log \det P$. Use matrix calculus to get

$$\frac{\partial \ell}{\partial P} = \sum_{k=1}^n (x^k - \mu)(x^k - \mu)^\top - P^{-1}.$$

□

We call these two objects the *maximum likelihood mean* or *empirical mean* and *maximum likelihood covariance* or *empirical covariance* of the dataset. We call the normal density with the empirical mean and empirical covariance the *empirical normal* of the dataset.

