Statistics

Jihang Li

November 21, 2019

Contents

Part I

Distributions

Chapter 1

Normal Distribution

1.1 Basic Form

- X: random variable.
- μ: mean or expectation.
- σ: standard deviation.
- σ^2: variance.
- $X \sim \mathcal{N}(\mu, \sigma^2)$: X is distributed normally with μ and σ^2,

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}). \tag{1.1}$$

1.2 Multivariate Form

- $\mathbf{X}$: random vector, $\mathbf{X} = (X_1, \ldots, X_k)$.
- $\boldsymbol{\mu}$: mean or expectation vector, $\boldsymbol{\mu} = \mathrm{E}[\mathbf{X}] = (\mathrm{E}[\mathbf{X_1}], \mathrm{E}[\mathbf{X_2}], \ldots, \mathrm{E}[\mathbf{X_k}])$.
- $\mathbf{\Sigma}$: $k \times k$ convariance matrix, $\Sigma_{i,j} = \mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)] = \mathrm{Cov}[X_i, X_j]$, where $1 \leq i, j \leq k$.
- Q: precision matric $\mathbf{\Sigma}^{-1}$
- $X \sim \mathcal{N}_{\|}(\boldsymbol{\mu}, \mathbf{\Sigma})$: X is distributed normally with $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ and k indicates k-dimension.

For **non-degenerate case:**

$$f_{\mathbf{X}}(x_1, \ldots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\mathbf{\Sigma}|}} \exp(-\frac{(\mathbf{X}-\boldsymbol{\mu})^{\mathrm{T}} \mathbf{\Sigma}^{-1} (\mathbf{X}-\boldsymbol{\mu})}{2}) \tag{1.2}$$

The descriptive statistic $\sqrt{(\mathbf{X}-\boldsymbol{\mu})^{\mathrm{T}} \mathbf{\Sigma}^{-1} (\mathbf{X}-\boldsymbol{\mu})}$ is known as the Mahalanobis distance.

Chapter 2

Poisson Distribution

$$f(k;\lambda) = \mathrm{P}(X = k) = \frac{\lambda^k e^- - \lambda}{k!} \tag{2.1}$$

where the positive real number $\lambda = \mathrm{E}(X) = \mathrm{Var} X$ is the average number of events per intervel, and $k = 0, 1, 2, \ldots$.

Part II

Pattern Recognition

Chapter 3

Likelihood

[References: Likelihood Function, Bayes' Theorem]

The likelihood expresses how likely particular values of statistical model parameters are for a given sample of data. It is **equal** to the joint probability distribution of a random sample, but with the random variable fixed at the given observations.

The likelihood describes a hypersurface whose peak, if it exists, represents the combination of model parameter values that **maximize** the probability of drawing the sample obtained.

$$\text{Posterior Probability} = \frac{\text{Likelihood} \cdot \text{Prior Probability}}{\text{Evidence}}$$

3.1 Discrete Probability Distribution

Let X be a discrete random variable with probability mass function p depending on a parameter θ, then the likelihood function is

$$\mathcal{L}(\theta|x) = p_\theta(x) = P_\theta(X = x).$$

Example: Consider a statistical model of a coin flip. $p_H \in [0.0, 1.0]$ expresses the "fairness" of the coin, which is the probability that a coin lands heads up ("H") when tossed. For a perfectly fair coin, $p_H = 0.5$. In a case that two heads observed in two tosses ("HH") and assuming each successive coin flip is i.i.d., then

$$P(\text{HH}|p_H = 0.5) = 0.5^2 = 0.25.$$

Hence, given this observed data HH, the likelihood that $p_H = 0.5$ is 0.25:

$$\mathcal{L}(p_H = 0.5|\text{HH}) = 0.25.$$

This is **NOT** the same as saying that the probability that $p_H = 0.5$ is 0.25 given the observation HH.

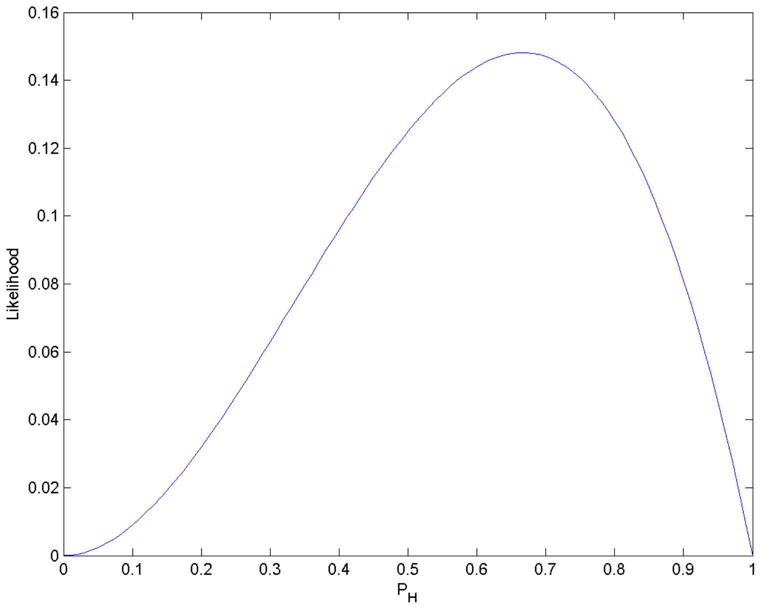

Figure 3.1: The likelihood function $(p_H^2(1 - p_H))$ for the probability of a coin landing heads-up (without prior knowledge of the coin's fairness), given that we have observed HHT.