PH7363 Fall 2025

Prof. Saptaparna Bhattacharya

Particle detection and detector physics

The Definition of Likelihood

The likelihood function is a fundamental concept in statistics, particularly in parameter estimation.

- **Core Definition:** The likelihood function, denoted $L(\theta|D)$ or $\mathcal{L}(\theta|D)$, measures the plausibility of a set of parameter values θ given observed data D.
- **Likelihood vs. Probability:** It is defined as being proportional to the probability of observing the data given the parameters, $P(D|\theta)$.

  $$L(\theta|D) = P(D|\theta)$$

  The key difference is in the interpretation:

  - For $P(D|\theta)$, we treat the parameters θ as fixed and consider the probability of different data outcomes D.
  - For $L(\theta|D)$, we treat the observed data D as fixed and consider the function over the space of possible parameters θ.
- **IID Data:** If the data D consists of N independent and identically distributed (IID) samples, $D = \{x_1, x_2, \ldots, x_N\}$, the total likelihood is the product of the individual likelihoods:

  $$L(\theta|D) = P(x_1, \ldots, x_N|\theta) = \prod_{i=1}^{N} P(x_i|\theta)$$

- **Log-Likelihood:** Because multiplying many small probabilities can lead to numerical underflow, it is almost always easier to work with the **log-likelihood**, $l(\theta) = \log L(\theta)$. The log-likelihood turns the product into a sum:

  $$l(\theta|D) = \log\left(\prod_{i=1}^{N} P(x_i|\theta)\right) = \sum_{i=1}^{N} \log P(x_i|\theta)$$

- **Maximum Likelihood Estimation (MLE):** A common goal is to find the parameters $\hat{\theta}$ that maximize the likelihood function. Since the logarithm is a monotonically increasing function, maximizing the likelihood is equivalent to maximizing the log-likelihood.

  $$\hat{\theta}_{\text{MLE}} = \underset{\theta}{\operatorname{argmax}}\, L(\theta|D) = \underset{\theta}{\operatorname{argmax}}\, l(\theta|D)$$

Product of Gaussian PDFs

A key property of the Gaussian distribution is that the product of two Gaussian probability density functions (PDFs) is another Gaussian (though it is unnormalized). This property is the foundation of many algorithms, including the Kalman filter and Bayesian inference with Gaussian priors.

0.1 The 1D Case

Let's consider two 1D Gaussian PDFs:

$$\mathcal{N}(x|\mu_1, \sigma_1^2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right)$$

$$\mathcal{N}(x|\mu_2, \sigma_2^2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right)$$

The product $P(x) = \mathcal{N}(x|\mu_1, \sigma_1^2) \times \mathcal{N}(x|\mu_2, \sigma_2^2)$ will be proportional to a new Gaussian, $\mathcal{N}(x|\mu_{\text{new}}, \sigma_{\text{new}}^2)$.

The parameters of the new Gaussian are:

- **New Precision:** The precisions (inverse variance, $\tau = 1/\sigma^2$) add up.

$$\frac{1}{\sigma_{\text{new}}^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

  This means the new variance is:

$$\sigma_{\text{new}}^2 = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

- **New Mean:** The new mean is a precision-weighted average of the old means.

$$\mu_{\text{new}} = \sigma_{\text{new}}^2 \left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}\right)$$

This result shows that combining two Gaussian beliefs (e.g., from two sensors) yields a new Gaussian belief that is more certain (has a smaller variance) than either of the individuals.

The Multivariate Case

The same property holds for multivariate Gaussians. Let two d-dimensional Gaussians be:

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \quad \text{and} \quad \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$$

where $\mathbf{x}$ and $\boldsymbol{\mu}$ are $d \times 1$ vectors and $\boldsymbol{\Sigma}$ is a $d \times d$ covariance matrix.

The product $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \times \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ is proportional to a new Gaussian $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{\text{new}}, \boldsymbol{\Sigma}_{\text{new}})$.

The parameters of the new Gaussian are:

- **New Precision Matrix:** The precision matrices (inverse covariance matrices) add.

$$\mathbf{\Sigma}_{\text{new}}^{-1} = \mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1}$$

  This means the new covariance matrix is:

$$\mathbf{\Sigma}_{\text{new}} = \left(\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1}\right)^{-1}$$

- **New Mean Vector:** The new mean is a precision-weighted average.

$$\boldsymbol{\mu}_{\text{new}} = \mathbf{\Sigma}_{\text{new}} \left(\mathbf{\Sigma}_1^{-1} \boldsymbol{\mu}_1 + \mathbf{\Sigma}_2^{-1} \boldsymbol{\mu}_2\right)$$

This is the core calculation in Bayesian updates where the prior and likelihood are both Gaussian (a conjugate prior relationship).