Phy657 – Module 2: Probability Distributions & Inference

Sping 2026

Due: monday 2/16

The solution of pattern recognition problems relies heavily on probability theory. In this module, we will review a few key concepts that will be useful as we explore more complex algorithms for regression and classification.

Learning objectives

1. Review a few key probability distributions. We will concentrate on the Gaussian PDF, but discrete distributions such as the Bernoulli or Binomial will be used in future modules.
2. Understand why Gaussian probability density functions are so prevalent in statistics, connecting with the central limit theorem.
3. Explore the properties of multivariate Gaussian PDFs, with some excursion on the definition of errors (correlated, uncorrelated, and systematic).
4. Characterization of PDFs: we will work with 2D Gaussian PDFs to evaluate the covariance matrix and expectation values of random variates with different statistical properties known.
5. Practice likelihood-based estimation (MLE) and Bayesian inference as core tools for parameter estimation.
6. Connect counting statistics (Poisson) to mean/variance relationships commonly used in experimental physics.

Reading assignment

- Bishop, *Pattern Recognition and Machine Learning*, Chapter 2, with particular emphasis on Sections 2.3–2.6.

Activity 1: Empirical verification of the Central Limit Theorem

Consider N variables $x_1, \ldots, x_N$ each of which has a uniform distribution over the interval $[0, 1]$. Consider the distribution of

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

for $N = 1, 2, 10$. Can you fit the distribution of $\bar{x}$ for $N = 10$ with a Gaussian PDF?

For your laboratory report: review the central limit theorem and its implication in statistics and give examples where the Gaussian PDF is appropriate.

Activity 2: Poisson distribution — mean and variance

Consider a Poisson process with rate parameter λ. The probability of observing k events in a fixed interval is given by the Poisson distribution.

1. Write down the Poisson probability mass function $p(k \mid \lambda)$.
2. Compute the expectation value $\langle k \rangle$.
3. Compute the variance $\mathrm{Var}(k)$.
4. Show explicitly that $\mathrm{Var}(k) = \langle k \rangle$.
5. Briefly discuss the physical meaning of this result for counting experiments.

Activity 3: Practice with multi-dimensional Gaussian distributions

1. Generate 1000 events characterized by coordinates (x_i, y_i) that are random variates distributed according to a 2D Gaussian PDF with mean $[0, 0]$ and covariance matrix
$$\Sigma = \begin{pmatrix} 4 & 2 \\ 2 & 16 \end{pmatrix}.$$
2. Produce a scatter plot of this distribution and draw the contours of constant probability density. Can you give an example of a pair of observables that exhibit such a correlation?

Activity 4: Maximum Likelihood Estimation (by hand) for a Gaussian

Assume a dataset $\{x_n\}_{n=1}^{N}$ is drawn independently from a Gaussian distribution with unknown mean μ and variance σ^2.

1. Write down the likelihood function $p(\{x_n\} \mid \mu, \sigma^2)$.
2. Compute the log-likelihood.
3. Derive the maximum likelihood estimator for μ.
4. Derive the maximum likelihood estimator for σ^2.
5. Compare your expression for σ^2 with the unbiased sample variance estimator. Briefly explain the difference, and discuss when this difference is relevant.

Activity 5: Maximum likelihood estimation for μ and Σ (multivariate Gaussian)

Generate a data set $\vec{X} = (x_1 \cdots x_N)$ where you assume that the observations $\{x_n\}$ are drawn independently from a multivariate Gaussian distribution. Evaluate the maximum likelihood expectations for μ and Σ.

Implement a sequential estimator as described in Section 2.3.5.

For your laboratory report: connect this activity with the concepts of bias in an estimator and remedies to this problem.

Activity 6: Likelihood shape and parameter uncertainty

In this exercise you will visualize how the likelihood function changes with dataset size.

1. Generate $N = 20$ samples from a Gaussian distribution with known parameters.
2. Compute the log-likelihood as a function of μ over a grid of trial values.
3. Plot the log-likelihood versus μ and identify the maximum likelihood estimate.
4. Repeat this procedure for $N = 100$ and $N = 1000$ samples.
5. Compare the width of the likelihood peak in the two cases and discuss how parameter uncertainty scales with N.

Activity 7: Bayesian inference for Gaussian parameters

Generate a data set with a single Gaussian random variable x characterized by a mean μ and variance σ^2.

- Assume it models a data set for which we know σ^2, but we must infer μ from N observations $\{x_n\}$. Develop an algorithm to implement a Bayesian inference of the mean μ for different N. Compare μ_N with μ_{ML}.
- Now analyze the same data set with the assumption that μ is known and we need to infer σ^2. Follow the formalism described by Bishop on pages 97–99.

For your laboratory report: explain the idea of a conjugate prior and how you can develop algorithms to implement Bayesian inference with an iterative process.

Activity 8: Prior sensitivity in Bayesian mean inference

Using your Bayesian mean inference code from Activity 7:

1. Repeat the inference using a narrow prior and a broad prior.
2. Compare the resulting posterior means and variances.
3. Repeat the comparison for a larger dataset size.
4. Discuss how the influence of the prior changes as N increases.