1 MAP and MLE

Note 1 (MLE). Assume that we want to estimate an unobserved population parameter \( \theta \) on the basis of observations \( y \). Let \( f \) be the sampling distribution of \( y \), so that \( f(y|\theta) \) is the probability of \( y \) when the underlying population parameter is \( \theta \). Then the function:

\[ \theta \mapsto f(y|\theta) \]

is known as the likelihood function and the estimate:

\[ \text{MLE}[\theta|y] = \arg \max_{\theta} f(y|\theta) \]

is the maximum likelihood estimate of \( \theta \) (given observations \( y \)).

Note 2 (MAP). Now assume that a prior distribution \( g \) over \( \theta \) exists, with domain \( \Theta \). This allows us to apply Bayes’ rule. Then the posterior distribution of \( \theta \) is as follows:

\[ \theta \mapsto f(\theta|y) = \frac{f(y|\theta)g(\theta)}{\int_{u\in\Theta} f(y|u)g(u)du} \]

The method of maximum a posterior estimation then estimates \( \theta \) as the most frequent values w.r.t. the posterior distribution of this random variable:

\[ \text{MAP}[\theta|y] = \arg \max_{\theta} f(\theta|y) = \arg \max_{\theta} f(y|\theta)g(\theta) \]

The denominator of the posterior distribution does not depend on \( \theta \) and therefore plays no role in the optimization. Observe that the MAP estimate of \( \theta \) coincides with the MLE when the prior \( g \) is uniform (that is, a constant function).

Problem 1. For \( i \in \{0, 1\} \), when \( X = i \), \( Y \sim \text{Poisson}(\mu_i) \). That is,

\[ \mathbb{P}[Y = n|X = i] = \frac{\mu_i^n}{n!}e^{-\mu_i}, n \geq 0 \]

The numbers \( \mu_i \) are known and such that \( \mu_0 < \mu_1 \). Assume that \( \mathbb{P}(X = i) = p_i \), for \( i = 0, 1 \) where the numbers \( p_i \in (0, 1) \) are known and add up to one.

(a) Find MLE\([X|Y = n]\).
(b) Find MAP\([X|Y = n]\).

Problem 2. Let \( X \) be uniformly distributed in \([0, 1]\). Assume that, given \( X = x \), the random variable \( Y \) is exponentially distributed with rate \( x + 1 \).

(a) Calculate \( \mathbb{E}[Y] \).
(b) Find MLE\([X|Y = n]\).
(c) Find MAP\([X|Y = n]\).
2 Hypothesis testing

Type I and Type II errors

Two types of errors can be made: a Type I Error happens when the null hypothesis $H_0$ was rejected though it should have been accepted, and a Type II Error occurs when the alternative hypothesis $H_1$ was rejected though it should have been accepted, or, equivalently, when the null hypothesis $H_0$ was accepted though it should have been rejected.

An example can be seen in the case of a criminal trial, in a democratic state where the rule is “innocent until proven guilty”. The null hypothesis in a democratic State is “he is innocent”, the alternative one is “he is guilty”; a type I error consists of condemning an innocent person, and a type II error consists of letting free a guilty.

A test is a procedure which divides the space of observations into 2 regions, $R$ and $A$. The two important characteristics of a test are called significance and power, which refer to errors of type I and II respectively, i.e. $\alpha$ ($\beta$) can be seen as the probability of making a Type I (Type II) Error:

Significance = $1 - \alpha = 1 - \mathbb{P}(x \in R|H_0) = 1 - \int_R \mathbb{P}(x|H_0)dx = \int_A \mathbb{P}(x|H_0)dx$

Power = $1 - \beta = 1 - \mathbb{P}(x \in A|H_1) = 1 - \int_A \mathbb{P}(x|H_1)dx = \int_R \mathbb{P}(x|H_1)dx$

The determination of a test is usually a trade-off between $\alpha$ and $\beta$. One commonly encountered procedure is to set a priori the significance to a fixed value ($\alpha = 0.01, 0.05, \cdots$) and find the most powerful test. To make $\beta$ as small as possible for a given $\alpha$, the integral over the chosen rejection region $\int_R(x|H_1)dx = 1 - \beta$ must be as large as possible, for a given $\int_R \mathbb{P}(x|H_0)dx = \alpha$.

In the case where data consist of one measurement, say $x$, the choice of $\alpha$ sets $\beta$, through the test $x < x_c$. In other cases, different tests correspond to the same given $\alpha$. It should be noted that in the following nothing is known about the a priori probability, if such a thing exists, of the hypothesis $H_0$ with respect to that of $H_1$. For example, if we are dealing with the one-by-one identification of two types of cell in a test-tube, the formalism makes no use of their relative concentration.

Back to our example of a trial, the procedure is: given a priori a (low) risk of condemning an innocent, what is the most powerful method to convict guilty people? However, in a non-democratic State, where the null hypothesis is “he is guilty”, the procedure would be “given a priori a low risk a releasing a real guilty, what is the most powerful method to prove ones innocence?”. This shows how asymmetric $H_0$ and $H_1$ are.

The Neyman Pearson Test

The Neyman Pearson test applies to the case of a null hypothesis against a alternative hypothesis. The rejection region is determined by the following theorem:

For a given $\alpha$, the most powerful test rejects $H_0$ in a region such as $\frac{\mathbb{P}(x|H_1)}{\mathbb{P}(x|H_0)} > k$

Problem 3. Let be $n$ observations from a Gaussian distribution of unknown mean $\mu$, but of known variance $\sigma^2$. $\mu$ can be either $\mu_0$ (null hypothesis), or $\mu_1$ (alternative hypothesis). Let’s assume $\mu_1 > \mu_0$.

(a) Find the ratio of the likelihood.
(b) What does the Neyman Pearson test tell us about the rejection of the hypothesis $\mu = \mu_0$?
(c) We know that $\mathbb{P}(x > 1.645) = 0.05$ when $x \sim \mathcal{N}(0, 1)$. For $\alpha = 0.05$, when is $H_0$ rejected?