

Lecture 3

Gidon Rosalki

2025-10-26

Last lecture we discussed the σ function, that gives a weight to each pair of letters, for example:

$$\begin{aligned}\sigma : \Sigma \times \Sigma &\rightarrow \mathbb{R} \sigma(A, T) = -3 \\ \sigma(A, A) &= 1\end{aligned}$$

1 Probabilistic models and Decision

Let us define probability. Probability is a method to define uncertainty about the world. We can put a probability that it will rain tomorrow, but we cannot *know*. How can we make probabilities? Well, we can assign them based off real world data, such as what was the most frequent occurrence, or perhaps based off subjective feelings.

So, we want to describe the uncertain world with probabilities. Let us consider something where we have information. Consider taking blood tests: We have lots of data from lots of blood tests taken over the decades. So, we may consider the probability that a person is healthy

$$\mathbb{P}_H(X)$$

Where X is the results of the blood tests.

1.1 Hypothesis testing

We have some types of hypothesis testing:

1. One hypothesis: Is someone healthy or not
2. Two hypothesis: Consider we know what people sick with flu appear to be. So now its “healthy” or “sick with flu”

Before we start making decisions on our hypotheses, we need to consider what is the meaning of the quality of a decision.

Definition 1.1 (Decision rule).

$$\tau(x) \rightarrow \{H, F\}$$

From this we create a matrix, between the truth, and the predicted result:

$$\begin{bmatrix} & H(-) & F(+) \\ H(-) & TN & FN \\ F(+) & FP & TP \end{bmatrix}$$

Where the columns are the true result, and the rows are the predicted result. Thus we see true negative, false negative, false positive, and true positive. We may now create the probability matrix of results:

$$\begin{bmatrix} & H & F \\ H & \frac{TN}{N} & \frac{FN}{P} \\ F & \frac{FP}{N} & \frac{TP}{P} \end{bmatrix}$$

We may also create the specificity matrix:

$$\begin{bmatrix} & H & F \\ H & \frac{TN}{TN+FN} & \frac{FN}{TN+FN} \\ F & \frac{FP}{FP+TP} & \frac{TP}{FP+TP} \end{bmatrix}$$

From all this, we call $\frac{TP}{FP+TP}$ the **specificity**, and $\frac{TN}{TN+FN}$ the **sensitivity**. On a graph of sensitivity against specificity, we ideally we both to be as high as possible. Any hypothesised rule can be placed as a point at some point on that graph. Given two different points, of equal distance from the origin, we cannot necessarily say that one is better than the other, just that one is more sensitive, and one is more specific. The decision of which is better in this case is down to the user, and what their objective is.

1.2 Neiman Peerson lemma

Theorem 1.

$$\tau_t(x) = \begin{cases} +, & \text{if } \frac{\mathbb{P}_F(x)}{\mathbb{P}_H(x)} \geq t \\ -, & \text{if } \frac{\mathbb{P}_F(x)}{\mathbb{P}_H(x)} < t \end{cases}$$

τ is optimal / maximal **if and only if** $\exists t \forall x : \tau(x) = \tau_t(x)$

We will also need **Bayesian probability**:

$$\begin{aligned} \mathbb{P}(+|X) &= \frac{\mathbb{P}(X|+)\mathbb{P}(+)}{\mathbb{P}(X)} \\ &= \frac{\mathbb{P}(X|+)\mathbb{P}(+)}{\mathbb{P}(X|+)\mathbb{P}(+) + \mathbb{P}(X|-)\mathbb{P}(-)} \\ \mathbb{P}(X|+) &= \mathbb{P}_F(X) \\ \mathbb{P}(X|-) &= \mathbb{P}_H(X) \\ \mathbb{P}(+) &= \\ \mathbb{P}(-) &= \end{aligned}$$

1.2.1 Single hypothesis

For the purpose of demonstration, let us say that we are measuring only one number, for example body temperature.

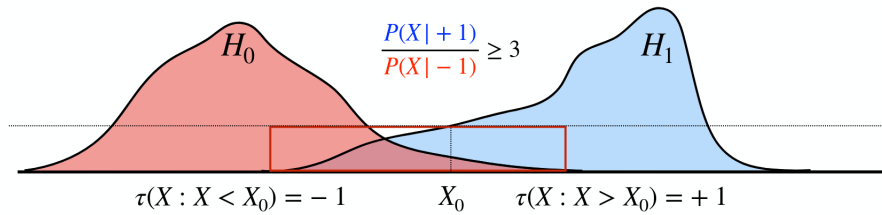


Figure 1: Types of error

A lower temperature generally indicates healthy, and higher generally unhealthy, but there is an amount of crossover in the middle, where we need to make a decision of a cut-off point, where above that indicates unhealthy, and below indicates healthy. For a probability $p_H(x)$, we have H_0 , which indicates healthy, and H_1 which indicates **not** H_0 . Since we are working with continuous probabilities, rather than discrete, we might for example find $\mathbb{P}_H(\text{Temp} > 38.5)$ (the area under the graph). This gives us the p -value.

1.3 Sequences

Let there be 2 sequences s, t . We shall create H_0 , that states that s, t are independent, and H_1 that s, t share some common ancestor. We will write $\mathbb{P}_{H_0} = \mathbb{P}_0 \wedge \mathbb{P}_{H_1} = \mathbb{P}_1$. We will make some assumptions to decide this:

- For H_0 , we will say that positions are i.i.d. i.e. each position is not dependent on what comes before, or after it:

$$\begin{aligned} \mathbb{P}(st) &= \mathbb{P}_0(s) \mathbb{P}_0(t) \\ \mathbb{P}(s) &= \prod_{i=1}^n p_0(s_i) \\ \mathbb{P}(t) &= \prod_{i=1}^n p_0(t_i) \end{aligned}$$

- We will still say that the positions are i.i.d, but we will also say that the sequences are dependent:

$$\mathbb{P}_1(s, t) = \prod_{i=1}^n p_1(s_i, t_i)$$

So, let us consider:

$$\begin{aligned}\frac{\mathbb{P}_1(s, t)}{\mathbb{P}_0(s, t)} &= \prod_{i=1}^n \frac{p_1(s_i, t_i)}{p_0(s_i) p_0(t_i)} \\ \log\left(\frac{\mathbb{P}_1(s, t)}{\mathbb{P}_0(s, t)}\right) &= \sum_{i=1}^n \log\left(\frac{p_1(s_i, t_i)}{p_0(s_i) p_0(t_i)}\right) \\ \textit{Score}(s, t) &= \sum_{i=1}^n \sigma(s_i, t_i) \\ \implies \sigma(s, t) &= \log\left(\frac{\mathbb{P}_1(s, t)}{\mathbb{P}_0(s) \mathbb{P}_0(t)}\right)\end{aligned}$$