Statistical Machine Learning
Multiple Testing

Horacio Gómez-Acevedo
Department of Biomedical Informatics
University of Arkansas for Medical Sciences

March 17, 2022

Hypothesis test

- A *statistical hypothesis* H ia a conjecture about the probability distribution of a population.
- A hypothesis H is said to be *simple* if the distribution of the population is completely specified by H. If not, then H is called a *composite* hypothesis.
- H_0 is called the *null hypothesis* and H_1 the *alternative hypothesis*.
- We say that we commit a *type I error* if we decide to accept H_1, whereas in reality H_0 is true.
- The probability of committing a type I error will be denoted by α.
- Acceptance of H_0 whereas H_1 is true is called a *type II error*.
- The probability of committing a type II error will be denoted by β.

Example

A factory has packets of coffee with and adjusted weight of 500 grams. We assume that the weight of the packages is $N(500, 50)$-distributed. A container with coffee packages contains packages wrongly processed with weight $N(490, 50)$-distributed. To determine the container that has incorrectly weight the packages, we draw a sample (package) of each of two containers X_1, X_2 . Based on the outcome of the 2-vector (X_1, X_2) we will make a conjecture about the distribution of the population. Namely

$$H_0\text{: the population is } N(500, 50) - \text{distributed}$$
$$H_1\text{: the population is } N(490, 50) - \text{distributed}$$

Let's suppose our decision as
If both X_1 and $X_2 \leq 496$ then we accept H_1, otherwise we accept H_0.

Example (cont)

Let's define our *critical region* G as

$$G = \{(x_1, x_2) \in \mathbb{R}^2 : x_1, x_2 \leq 496\}$$

Then, we can formulate our decision as

$$= \begin{cases} \text{if } (X_1, X_2) \in G & \text{then we choose } H_1 \text{ as our conjecture.} \\ \text{if } (X_1, X_2) \notin G & \text{then we choose } H_0 \text{ as our conjecture.} \end{cases}$$

We can calculate α and β as follows

$$\begin{aligned} \alpha &= P(\text{acceptance of } H_1 | H_0 \text{ is true}) \\ &= P((X_1, X_2) \in G | \mu = 500, \sigma^2 = 50) \\ &= P(X_1 \leq 496 | \mu = 500, \sigma^2 = 50) \cdot P(X_2 \leq 496 | \mu = 500, \sigma^2 = 50) \\ &= 0.081 \end{aligned}$$

Example (cont)

$$
\begin{aligned}
\beta &= P(\text{acceptance of } H_0 | H_1 \text{ is true}) \\
&= P((X_1, X_2) \notin G | \mu = 490, \sigma^2 = 50) \\
&= 1 - P((X_1, X_2) \in G | \mu = 490, \sigma^2 = 50) \\
&= 1 - P(X_1 \leq 496 | \mu = 490, \sigma^2 = 50) \cdot P(X_2 \leq 496 | \mu = 490, \sigma^2 = 50) \\
&= 0.356
\end{aligned}
$$

Hypothesis test

A *Hypothesis test* is a collection

$$(X_1, \ldots, X_n; H_0; H_1 : G)$$

where $X_1, \ldots, X_n$ is a sample, H_0 and H_1 hypotheses concerning the probability distribution of the population and $G \subset \mathbb{R}^n$ a Borel set (meaning a collection of open sets)
If H_0 is a simple statistical hypothesis. The *level of significance* of the hypothesis test $(X_1, \ldots, X_n; H_0; H_1; G)$ is understood to be the number

$$\alpha = P^{H_0}_{X_1,\ldots,X_n}(G)$$

Thus, we say that α represents the probability of committing a type I error.

The power function

With our previous setup the β could not be used for composite hypothesis. Thus, we have to define something more general.

- Let $f(\cdot, \theta))_{\theta \in \Theta}$ be a family of probability densities
- Let's assume that the population $X_1, \ldots, X_n$ has a probability density $f(\cdot, \theta)$ where $\theta \in \Theta$.
- Let's assume that H_0 and H_1 are statements of the type

$$H_0 \colon \theta \in \Theta_0 \quad \text{and} \quad H_1 \colon \theta \in \Theta_1$$

where $\Theta_0 \cup \Theta_1 = \Theta$ and $\Theta_1 \cap \Theta_1 = \emptyset$.

For a fixed $\theta \in \Theta$, the probability distribution of the population is completely specified. Then, we define for every $\theta \in \Theta_1$

$$\beta(\theta) = P^{\theta}_{X_1, \ldots, X_n}(G^c)$$

The expression $1 - \beta(\theta)$ is called the *power function* for $\theta \in \Theta_1$.

Example (cont)

From our previous example, given is a $N(\mu, 50)$-distributed population, where $\mu \leq 500$. The family of probability densities $f(:\ , \mu)$ where $\mu \leq 500$ and

$$f(x, \mu) = \frac{1}{\sqrt{100\pi}} \exp(\frac{-(x-\mu)^2}{100})$$

The parameter space is defined by $\Theta = (-\infty, 500]$.
If we draw a sample X_1, X_2 of size 2 from this population. We can define $\Theta_0 = \{500\}$ and $\Theta_1 = (-\infty, 500)$. This corresponds to the following hypotheses

$$H_0 : \mu = 500 \quad \text{against} \quad H_1 : \mu < 500$$

If we choose the following critical region

$$G = \{(x_1, x_2) \in \mathbb{R}^2 : x_1, x_2 \leq 494.63\}$$

Example (cont)

Then, our 5-tuple $(X_1, X_2; H_0; H_1; G)$ constitutes a hypothesis test. The size α of the critical region G is $\alpha = 0.05$, and the power function

$$
\begin{aligned}
1 - \beta(\mu) &= 1 - P^{\mu}_{X_1,X_2}(G^c) = P^{\mu}_{X_1,X_2}(G) = P((X_1, X_2) \in G) \\
&= P(X_1 \leq 494.63 \text{ and } X_2 \leq 494.63 | \mu = \mu, \sigma^2 = 50) \\
&= P(X_1 \leq 494.63 | \mu = \mu, \sigma^2 = 50) \\
&\quad \cdot P(X_2 \leq 494.63 | \mu = \mu, \sigma^2 = 50)
\end{aligned}
$$

References

Materials and some of the pictures are from (Calin, 2019).

Calin, O. (2019). *Deep Learning Architectures*. Springer Series in the Data Sciences. Springer. ISBN: 978-3-030-36723-7.

I have used some of the graphs by hacking TiKz code from StakExchange, Inkscape for more aesthetic plots and other old tricks of TEX