Applications - Lecture 7

The Formalism of Statistical Hypothesis Testing

Ingénierie Mathématique & Informatique

Section	Notion	Used in
7.1.2	Type I/II risk, level α, power function	App. 2
7.1.3	Construction via $g(\theta)$, Z_n, quantiles φ_r	App. 2
7.2.2	Family-Wise Error Rate, Bonferroni correction	App. 1
7.2 (Ex. 7.A.3)	Benjamini-Hochberg procedure, False Discovery Rate	App. 1

1. **Correct** for thousands of simultaneous tests: gene expression screening.
2. **Plan** the sample size n so that a test reaches a target power: clinical and industrial experiments.

App. 1 - Multiple Testing and False Discovery Rate in Genomics

Correcting for m simultaneous comparisons

Problem. A bioinformatician tests whether each of $m = 1\,000$ genes is differentially expressed between a diseased and a healthy group: this gives m pairs (H_0^k, H_1^k), each with its own p-value p_k and rejection region W_n^k. Left uncorrected, the look-elsewhere effect makes
$\mathbb{P}_{\theta_0}(\text{at least one false positive}) = 1 - (1-\alpha)^m \to 1$.

Bonferroni. Reject H_0^k as soon as $p_k \le \alpha/m$:

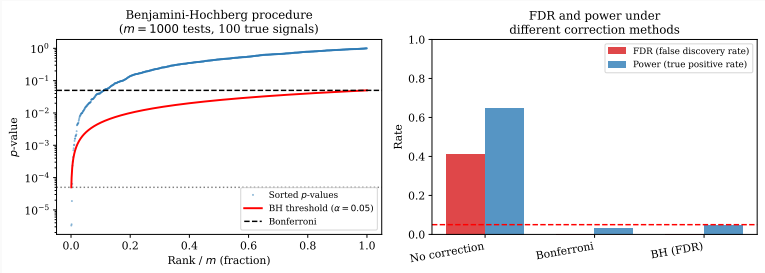
$$\mathrm{FWER} = \sup_{\theta \in H_0} \mathbb{P}_\theta\big(\exists k : X_n \in W_n^k\big) \;\le\; \sum_{k=1}^{m} \alpha_k \;\le\; \alpha \quad (\alpha_k = \alpha/m).$$

Benjamini-Hochberg (Exercise 7.A.3). With $p_{(1)} \le \cdots \le p_{(m)}$,

$$J = \max\Big\{ j \in \{1,\dots,m\} : p_{(j)} \le \tfrac{j}{m}\alpha \Big\}, \qquad \text{reject } H_0^k \text{ for all } k \text{ with } p_k \le p_{(J)}.$$

Under independence, the **False Discovery Rate** $\mathrm{FDR}(\theta) = \mathbb{E}_\theta[V/R]$ satisfies $\mathrm{FDR}(\theta) = \frac{m_0}{m}\alpha \le \alpha$.

Reading the BH plot and take-aways



Left: sorted p-values $p_{(k)}$ against k/m (semi-log scale), the BH threshold line $\frac{k}{m}\alpha$ (red), and the Bonferroni threshold α/m (dashed). Right: empirical FDR and power for no correction, Bonferroni, and BH at $\alpha = 0.05$.

- Without correction, FDR is $\approx 60\%$; Bonferroni controls FWER but sacrifices most of the power.
- BH controls the weaker FDR instead of the FWER, and thereby retains substantial power: the operating point genomics needs.

App. 2 - Power Analysis and Sample Size Planning

Level, power, and the choice of n

Problem. Before running a clinical trial or an industrial experiment, choose the sample size n per group so that a real effect is detected with high probability.

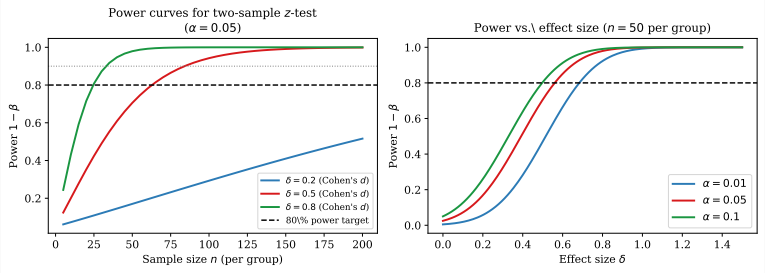
Model. $H_0 = \{g(\theta) = 0\}$, $H_1 = \{g(\theta) \neq 0\}$ with $g(\theta) = \mu_1 - \mu_2$, estimator $Z_n = \bar{X}_{1,n} - \bar{X}_{2,n}$. The symmetric two-sided level-α rejection region is

$$W_n = \Big\{ |Z_n| \geq \sigma \sqrt{2/n}\, \varphi_{1-\alpha/2} \Big\},$$

with power at effect size $\delta = g(\theta)$

$$1 - \beta(n) = \Phi\Big(\tfrac{\delta}{\sigma\sqrt{2/n}} - \varphi_{1-\alpha/2} \Big) + \Phi\Big(-\tfrac{\delta}{\sigma\sqrt{2/n}} - \varphi_{1-\alpha/2} \Big).$$

Power curves, sample size, and take-aways



Left: power $1-\beta$ vs. n per group for $\delta \in \{0.2, 0.5, 0.8\}$, $\alpha = 0.05$, $\sigma = 1$; dashed line at 80% power. Right: power vs. δ for $n = 50$, at $\alpha \in \{0.01, 0.05, 0.10\}$.

- Fixing a power level ρ and threshold δ and inverting the power formula gives $n = 2\sigma^2(\varphi_{1-\alpha/2} + \varphi_\rho)^2/\delta^2$: halving δ quadruples n ($n \approx 26$ for $\delta = 0.8$, $n \approx 395$ for $\delta = 0.2$, at $\rho = 80\%$).
- Consistency gives power $\to 1$ as $n \to \infty$; lowering α raises $\varphi_{1-\alpha/2}$ and lowers power at fixed n