Statistics for Applications

Chapter 4: Parametric hypothesis testing

Heuristics

Main idea: Decide whether to validate or reject hypotheses

- After tossing a coin many times, how to decide whether it is a fair one ?
- After drawing blood from a patient and measure the concentration of antibody in the sample, how to decide whether the patient is contaminated with a virus ?

Formally, this translates into:

- Decide whether $p = 1/2$ from a sample of n i.i.d. Bernoulli random variables;
- Decide whether $c > c_0$ from a sample of n i.i.d. random variables with $\mathcal{N}(c, \sigma^2)$ distribution.

Heuristics (2)

Example 1: A coin is tossed 80 times, and Heads are obtained 54 times. Can we conclude that the coin is significantly unfair ?

- $n = 80$, $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Ber}(p)$;
- $\bar{X}_n = 54/80 = .68$
- If it was true that $p = .5$: By CLT+Slutsky's theorem,

$$\sqrt{n}\frac{\bar{X}_n - .5}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \approx \mathcal{N}(0,1).$$

- $\sqrt{n}\dfrac{\bar{X}_n - .5}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \approx 3.45$
- Conclusion: It **seems quite** reasonable to reject the hypothesis "$p = .5$".

Heuristics (3)

Example 2: A coin is tossed 30 times, and Heads are obtained 13 times. Can we conclude that the coin is significantly unfair ?

- $n = 30, X_1, \ldots, X_n \overset{iid}{\sim} \text{Ber}(p)$;
- $\bar{X}_n = 13/30 \approx .43$
- If it was true that $p = .5$: By CLT+Slutsky's theorem,

$$\sqrt{n}\frac{\bar{X}_n - .5}{\sqrt{\bar{X}_n(1 - \bar{X}_n)}} \approx \mathcal{N}(0,1).$$

- $\sqrt{n}\frac{\bar{X}_n - .5}{\sqrt{\bar{X}_n(1 - \bar{X}_n)}} \approx -.77$
- Conclusion: It **seems** impossible to reject significantly the hypothesis "$p = .5$".

Statistical formulation (1)

- Consider a sample $X_1, \ldots, X_n$ of i.i.d. random variables and a statistical model $(E, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$.
- Let Θ_0 and Θ_1 be disjoint subsets of Θ.
- Consider the two hypotheses: $\begin{cases} H_0: & \theta \in \Theta_0 \\ H_1: & \theta \in \Theta_1 \end{cases}$
- H_0 is the *null hypothesis*, H_1 is the *alternative hypothesis*.
- If we believe that the true θ is either in Θ_0 or in Θ_1, we may want to *test* H_0 *against* H_1.
- We want to decide whether to *reject* H_0 (look for evidence against H_0).

Statistical formulation (2)

- H_0 and H_1 are not symmetric. E.g., H_0: "Patient is sick" ($c > c_0$) vs. H_1: "Patient is healthy" ($c \leq c_0$).
- A *test* is a statistic $\delta \in \{0, 1\}$ such that:
  - If $\delta = 0$, H_0 is not rejected;
  - If $\delta = 1$, H_0 is rejected.
- Coin example: H_0: "$p = 1/2$" vs. H_1: "$p \neq 1/2$".
- $\delta = \mathbb{1}_{\left|\sqrt{n}\frac{\bar{X}_n - .5}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}\right| > C}$, for some threshold $C > 0$.
- How to choose C ?

Statistical formulation (3)

- *Rejection region* of a test δ:

$$R_\delta = \{x \in E^n : \delta(x) = 1\}.$$

- *Type 1 error* of a test δ (rejecting H_0 when it is actually true):

$$\begin{array}{rcccl} \alpha_\delta & : & \Theta_0 & \to & \mathbb{R} \\ & & \theta & \mapsto & \mathbb{P}_\theta[\delta = 1]. \end{array}$$

- *Type 2 error* of a test δ (not rejecting H_0 although H_1 is actually true):

$$\begin{array}{rcccl} \beta_\delta & : & \Theta_1 & \to & \mathbb{R} \\ & & \theta & \mapsto & \mathbb{P}_\theta[\delta = 0]. \end{array}$$

- *Power* of a test δ:

$$\pi_\delta = \inf_{\theta \in \Theta_1} \left(1 - \beta_\delta(\theta)\right).$$

Statistical formulation (4)

- A test δ has *level* α if

$$\alpha_\delta(\theta) \le \alpha, \quad \forall \theta \in \Theta_0.$$

- A test δ has *asymptotic level* α if

$$\lim_{n\to\infty} \alpha_\delta(\theta) \le \alpha, \quad \forall \theta \in \Theta_0.$$

- In general, a test has the form

$$\delta = \mathbb{1}_{T_n > c},$$

for some statistic T_n and threshold $c \in \mathbb{R}$.

- T_n is called the *test statistic*. The rejection region is $R_\delta = \{T_n > c\}$.

Example (1)

- Let $X_1, \dots, X_n \overset{iid}{\sim} \text{Ber}(p)$, for some unknown $p \in (0,1)$.
- We want to test:

$$H_0\text{: "}p = 1/2\text{" vs. } H_1\text{: "}p \neq 1/2\text{"}$$

  with asymptotic level $\alpha \in (0,1)$.

- Let $T_n = \left| \sqrt{n} \ \dfrac{\hat{p}_n - 0.5}{\sqrt{\hat{p}_n(1-\hat{p}_n)}} \right|$, where $\hat{p}_n$ is the MLE.

- If H_0 is true, then by CLT and Slutsky's theorem,

$$\mathbb{P}[T_n > q_{1-\alpha/2}] \xrightarrow[n\to\infty]{} 0.05$$

- Let $\delta_\alpha = \mathbb{1}_{T_n > q_{1-\alpha/2}}$.

Example (2)

Coming back to the two previous coin examples: For $\alpha = 5\%$, $q_{1-\alpha/2} = 1.96$, so:

- In **Example 1**, H_0 is rejected at the asymptotic level 5% by the test $\delta_{5\%}$;
- In **Example 2**, H_0 is not rejected at the asymptotic level 5% by the test $\delta_{5\%}$.

Question: In **Example 1**, for what level α would δ_α not reject H_0 ? And in **Example 2**, at which level α would δ_α reject H_0 ?

p-value

Definition

The (asymptotic) *p-value* of a test δ_α is the smallest (asymptotic) level α at which δ_α rejects H_0. It is random, it depends on the sample.

Golden rule

p-value $\leq \alpha \quad \Leftrightarrow \quad H_0$ is rejected by δ_α, at the (asymptotic) level α.

The smaller the p-value, the more confidently one can reject H_0.

- Example 1: p-value $= \mathbb{P}[|Z| > 3.45] \ll .01$.
- Example 2: p-value $= \mathbb{P}[|Z| > .77] \approx .44$.

Neyman-Pearson's paradigm

Idea: For given hypotheses, among all tests of level/asymptotic level α, is it possible to find one that has maximal power ?

Example: The trivial test $\delta = 0$ that never rejects H_0 has a perfect level ($\alpha = 0$) but poor power ($\pi_\delta = 0$).

Neyman-Pearson's theory provides (the most) powerful tests with given level. In 18.650, we only study several cases.

Reminder about the χ^2 distributions

Definition

For a positive integer d, the χ^2 *distribution with d degrees of freedom* is the law of the random variable $Z_1^2 + Z_2^2 + \ldots + Z_d^2$, where $Z_1, \ldots, Z_d \overset{iid}{\sim} \mathcal{N}(0, 1)$.

Examples:

- If $Z \sim \mathcal{N}_d(\mathbf{0}, I_d)$, then $\|Z\|_2^2 \sim \chi_d^2$.
- Cochran's theorem states that for $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, if S_n is the sample variance, then

$$\frac{nS_n}{\sigma^2} \sim \chi_{n-1}^2.$$

- $\chi_2^2 = \mathcal{E}xp(1/2)$.

Reminder about the Student's distributions

Definition

For a positive integer d, the *Student's distribution with* d *degrees of freedom* (denoted by t_d) is the law of the random variable $\frac{U}{\sqrt{V/d}}$, where $U \sim \mathcal{N}(0,1)$, $V \sim \chi^2_d$ and $U \perp\!\!\!\perp V$.

Example:

- Cochran's theorem states that for $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, if S_n is the sample variance, then

$$\sqrt{n-1}\ \frac{\bar{X}_n - \mu}{\sqrt{S_n}} \sim t_{n-1}.$$

Wald's test (1)

- Consider an i.i.d. sample $X_1, \ldots, X_n$ with statistical model $(E, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$, where $\Theta \subseteq \mathbb{R}^d$ ($d \geq 1$) and let $\theta_0 \in \Theta$ be fixed and given.
- Consider the following hypotheses:

$$\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta \neq \theta_0. \end{cases}$$

- Let $\hat{\theta}^{MLE}$ be the MLE. Assume the MLE technical conditions are satisfied.

- If H_0 is true, then

$$\sqrt{n}\ I(\hat{\theta}^{MLE})^{1/2} \left(\hat{\theta}_n^{MLE} - \theta_0\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}_d\left(0, I_d\right) \quad \text{w.r.t. } \mathbb{P}_{\theta_0}.$$

Wald's test (2)

- Hence,

$$\underbrace{n\left(\hat{\theta}_n^{MLE}-\theta_0\right)' I(\hat{\theta}^{MLE})\left(\hat{\theta}_n^{MLE}-\theta_0\right)}_{T_n} \xrightarrow[n\to\infty]{(d)} \chi^2_d \quad \text{w.r.t. } \mathbb{P}_{\theta_0}.$$

- Wald's test with asymptotic level $\alpha \in (0,1)$:

$$\delta = \mathbb{1}_{T_n > q_{1-\alpha}},$$

where q_α is the $(1-\alpha)$-quantile of χ^2_d (see tables).

- Remark: Wald's test is also valid if H_1 has the form "$\theta > \theta_0$" or "$\theta < \theta_0$" or "$\theta = \theta_1$"...

Likelihood ratio test (1)

- Consider an i.i.d. sample $X_1, \ldots, X_n$ with statistical model $(E, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$, where $\Theta \subseteq \mathbb{R}^d$ ($d \geq 1$).

- Suppose the null hypothesis has the form

$$H_0 : (\theta_{r+1}, \ldots, \theta_d) = (\theta_{r+1}^{(0)}, \ldots, \theta_d^{(0)}),$$

  for some fixed and given numbers $\theta_{r+1}^{(0)}, \ldots, \theta_d^{(0)}$.

- Let

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}} \ \ell_n(\theta) \quad \text{(MLE)}$$

  and

$$\hat{\theta}_n^c = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \ \ell_n(\theta) \quad \text{("constrained MLE")}$$

Likelihood ratio test (2)

- Test statistic:

$$T_n = 2\left(\ell_n(\hat{\theta}_n) - \ell_n(\hat{\theta}_n^c)\right).$$

- Theorem

  Assume H_0 is true and the MLE technical conditions are satisfied. Then,

$$T_n \xrightarrow[n\to\infty]{(d)} \chi^2_{d-r} \quad \text{w.r.t. } \mathbb{P}_\theta.$$

- Likelihood ratio test with asymptotic level $\alpha \in (0,1)$:

$$\delta = \mathbb{1}_{T_n > q_{1-\alpha}},$$

  where $q_{1-\alpha}$ is the $(1-\alpha)$-quantile of χ^2_{d-r} (see tables).

Testing implicit hypotheses (1)

- Let $X_1, \ldots, X_n$ be i.i.d. random variables and let $\theta \in \mathbb{R}^d$ be a parameter associated with the distribution of X_1 (e.g. a moment, the parameter of a statistical model, etc...)
- Let $g : \mathbb{R}^d \to \mathbb{R}^k$ be continuously differentiable (with $k < d$).
- Consider the following hypotheses:

$$\begin{cases} H_0 : & g(\theta) = 0 \\ H_1 : & g(\theta) \neq 0. \end{cases}$$

- E.g. $g(\theta) = (\theta_1, \theta_2)$ ($k = 2$), or $g(\theta) = \theta_1 - \theta_2$ ($k = 1$), or...

Testing implicit hypotheses (2)

- Suppose an asymptotically normal estimator $\hat{\theta}_n$ is available:

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}_d(0, \Sigma(\theta)).$$

- Delta method:

$$\sqrt{n}\left(g(\hat{\theta}_n) - g(\theta)\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}_k\left(0, \Gamma(\theta)\right),$$

where $\Gamma(\theta) = \nabla g(\theta)' \Sigma(\theta) \nabla g(\theta) \in \mathbb{R}^{k\times k}$.

- Assume $\Sigma(\theta)$ is invertible and $\nabla g(\theta)$ has rank k. So, $\Gamma(\theta)$ is invertible and

$$\sqrt{n}\,\Gamma(\theta)^{-1/2}\left(g(\hat{\theta}_n) - g(\theta)\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}_k\left(0, I_k\right).$$

Testing implicit hypotheses (3)

- Then, by Slutsky's theorem, if $\Gamma(\theta)$ is continuous in θ,

$$\sqrt{n}\,\Gamma(\hat{\theta}_n)^{-1/2}\left(g(\hat{\theta}_n) - g(\theta)\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}_k\left(0, I_k\right).$$

- Hence, if H_0 is true, i.e., $g(\theta) = 0$,

$$\underbrace{ng(\hat{\theta}_n)'\Gamma^{-1}(\hat{\theta}_n)g(\hat{\theta}_n)}_{T_n} \xrightarrow[n\to\infty]{(d)} \chi^2_k.$$

- Test with asymptotic level α:

$$\delta = \mathbb{1}_{T_n > q_{1-\alpha}},$$

where $q_{1-\alpha}$ is the $(1-\alpha)$-quantile of χ^2_k (see tables).

The multinomial case: χ^2 test (1)

Let $E = \{a_1, \ldots, a_K\}$ be a finite space and $(\mathbb{P}_{\mathbf{p}})_{\mathbf{p} \in \Delta_K}$ be the family of all probability distributions on E:

- $\Delta_K = \left\{ \mathbf{p} = (p_1, \ldots, p_K) \in (0,1)^K : \sum_{j=1}^{K} p_j = 1 \right\}.$
- For $\mathbf{p} \in \Delta_K$ and $X \sim \mathbb{P}_{\mathbf{p}}$,

$$\mathbb{P}_{\mathbf{p}}[X = a_j] = p_j, \quad j = 1, \ldots, K.$$

The multinomial case: χ^2 test (2)

- Let $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathbb{P}_{\mathbf{p}}$, for some unknown $\mathbf{p} \in \Delta_K$, and let $\mathbf{p}^0 \in \Delta_K$ be fixed.

- We want to test:
$$H_0\text{: " } \mathbf{p} = \mathbf{p}^0 \text{ " vs. } H_1\text{: " } \mathbf{p} \neq \mathbf{p}^0 \text{ "}$$
with asymptotic level $\alpha \in (0, 1)$.

- Example: If $\mathbf{p}^0 = (1/K, 1/K, \ldots, 1/K)$, we are testing whether $\mathbb{P}_{\mathbf{p}}$ is the uniform distribution on E.

The multinomial case: χ^2 test (3)

- Likelihood of the model:

$$L_n(X_1, \dots, X_n, \mathbf{p}) = p_1^{N_1} p_2^{N_2} \dots p_K^{N_K},$$

where $N_j = \#\{i = 1, \dots, n : X_i = a_j\}$.

- Let $\hat{\mathbf{p}}$ be the MLE:

$$\hat{\mathbf{p}}_j = \frac{N_j}{n}, \qquad j = 1, \dots, K.$$

⚠ $\hat{\mathbf{p}}$ maximizes $\ln L_n(X_1, \dots, X_n, \mathbf{p})$ **under the constraint**

$$\sum_{j=1}^{K} p_j = 1.$$

The multinomial case: χ^2 test (4)

- If H_0 is true, then $\sqrt{n}(\hat{\mathbf{p}} - \mathbf{p}^0)$ is asymptotically normal, and the following holds.

Theorem

$$\underbrace{n \sum_{j=1}^{k} \frac{\left(\hat{\mathbf{p}}_j - \mathbf{p}_j^0\right)^2}{\mathbf{p}_j^0}}_{T_n} \xrightarrow[n \to \infty]{(d)} \chi^2_{K-1}.$$

- χ^2 test with asymptotic level α: $\quad \delta_\alpha = \mathbb{1}_{T_n > q_{1-\alpha}}$, where $q_{1-\alpha}$ is the $(1-\alpha)$-quantile of χ^2_{K-1}.
- Asymptotic p-value of this test: $p - \text{value} = \mathbb{P}\left[Z > T_n | T_n\right]$, where $Z \sim \chi^2_{K-1}$ and $Z \perp\!\!\!\perp T_n$.

The Gaussian case: Student's test (1)

- Let $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, for some unknown $\mu \in \mathbb{R}, \sigma^2 > 0$ and let $\mu_0 \in \mathbb{R}$ be fixed, given.

- We want to test:

$$H_0\text{: "}\mu = \mu_0\text{" vs. } H_1\text{: "}\mu \neq \mu_0\text{"}$$

  with asymptotic level $\alpha \in (0, 1)$.

- **If σ^2 is known:** Let $T_n = \sqrt{n}\ \dfrac{\bar{X}_n - \mu_0}{\sigma}$. Then, $T_n \sim \mathcal{N}(0, 1)$ and

$$\delta_\alpha = \mathbb{1}_{|T_n| > q_{1-\alpha/2}}$$

  is a test with (non asymptotic) level α.

The Gaussian case: Student's test (2)

If σ^2 is unknown:

- Let $\widetilde{T_n} = \sqrt{n-1}\ \dfrac{\bar{X}_n - \mu_0}{\sqrt{S_n}}$, where S_n is the sample variance.
- Cochran's theorem:
  - $\bar{X}_n \perp\!\!\!\perp S_n$;
  - $\dfrac{nS_n}{\sigma^2} \sim \chi^2_{n-1}$.
- Hence, $\widetilde{T_n} \sim t_{n-1}$: Student's distribution with $n-1$ degrees of freedom.

The Gaussian case: Student's test (3)

- Student's test with (non asymptotic) level $\alpha \in (0,1)$:

$$\delta_\alpha = \mathbb{1}_{|\widetilde{T_n}| > q_{1-\alpha/2}},$$

where $q_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of t_{n-1}.

- If H_1 is "$\mu > \mu_0$", Student's test with level $\alpha \in (0,1)$ is:

$$\delta'_\alpha = \mathbb{1}_{\widetilde{T_n} > q_{1-\alpha}},$$

where $q_{1-\alpha}$ is the $(1-\alpha)$-quantile of t_{n-1}.

- Advantage of Student's test:
  - Non asymptotic
  - Can be run on small samples

- Drawback of Student's test: It relies on the assumption that the sample is Gaussian.