Test d'ipotesi

Struttura di un test d'ipotesi lancio di una moneta

- Definisco l'insieme delle distribuzioni *compatibili* con il campione $X_1, \ldots, X_n$ $B(p)$, $p \in [0, 1]$
- Definisco l'ipotesi da testare, detta *ipotesi nulla* H_0
  Le ipotesi si possono suddividere in due grandi famiglie:
  - **ipotesi parametriche**: la distribuzione del campione è nota a meno di un parametro θ, scalare o vettoriale.

$$H_0 : \qquad \theta \in \Theta_0 \subset \Theta \qquad p = 0.5$$

  - **ipotesi non parametriche**: sono ipotesi sul tipo di distribuzione del campione ovvero sulla sua legge

$$H_0 : \qquad F(x) \in \mathcal{F}_0 \subset \mathcal{F}$$

Ipotesi *semplice* se Θ_0 o $\mathcal{F}_0$ è costituito da un solo elemento, *composta* altrimenti

- Definisco l'ipotesi alternativa H_A che è da considerarsi valida quando si rifiuta H_0.

$$H_A: \quad \theta \in \Theta_1, \quad \Theta_1 := \Theta \setminus \Theta_0 \quad p \neq 0.5$$
$$H_A: \quad F(x) \in \mathcal{F}_1 \quad \mathcal{F}_1 := \mathcal{F} \setminus \mathcal{F}_0$$

- Definisco una statistica $\varphi(X_1, \ldots, X_n)$ con distribuzione nota quando H_0 è vera. $\overline{X}$
- Suddivido lo spazio $\mathcal{G}$ delle possibili osservazioni di φ in due insiemi disgiunti:
  - $\mathcal{A}$, *regione di accettazione di* H_0; $(0.5 - \varepsilon, 0.5 + \varepsilon)$
  - $\mathcal{C} := \mathcal{G} \setminus \mathcal{A}$, *regione critica*. $[0, 0.5 - \varepsilon] \cup [0.5 + \varepsilon, 1]$
- Si formula la regola di decisione:
  - accetto H_0 se $\varphi(x_1, \ldots, x_n) \in \mathcal{A}$; $\overline{x} \in (0.5 - \varepsilon, 0.5 + \varepsilon)$
  - rifiuto H_0 se $\varphi(x_1, \ldots, x_n) \notin \mathcal{A}$, ovvero se $\varphi(x_1, \ldots, x_n) \in \mathcal{C}$.
    $\overline{x} \notin (0.5 - \varepsilon, 0.5 + \varepsilon)$

Errore di prima specie, significatività del test

Sulla base dei dati rigettiamo H_0 quando essa in realtà è vera.
Definisco *livello di significatività del test* la probabilità che ciò accada:

$$\alpha := \mathbb{P}\left(\varphi(X_1, \ldots, X_n) \in \mathcal{C} | H_0\right).$$

$1 - \alpha$ è detto *livello di fiducia del test*.

Errore di seconda specie, potenza del test

Sulla base dei dati accettiamo H_0 quando essa è falsa.

$$\beta := \mathbb{P}\left(\varphi(X_1, \ldots, X_n) \in \mathcal{A} | H_A\right).$$

$1 - \beta$ è detto *potenza del test*.

Remark

se H_A è un'ipotesi composta, allora β è una funzione $\beta(\theta)$, $\theta \in \Theta_1$.

Test d'ipotesi per il valore atteso di campioni gaussiani a varianza nota

$X_1, \ldots, X_n$ campione gaussiano di valore atteso μ incognito e varianza σ^2 nota.

- $H_0: \quad \mu = \mu_0, \qquad H_A: \quad \mu \neq \mu_0$
- Sappiamo che $\mathbb{P}_{X_i} = N\left(\mu_0, \sigma^2\right) \iff \mathbb{E}\left[\overline{X}\right] = \mu_0$
- Criterio di accettazione: accetto H_0 se e solo se $|\overline{x} - \mu_0| < \varepsilon$.
- Livello di significatività

  $$\alpha = \mathbb{P}\left(\left|\overline{X} - \mu_0\right| \geqslant \varepsilon | \mu = \mu_0\right)$$

  Devo determinare ε in funzione di α
- H_0 vera $\implies \mathbb{P}_{\overline{X}} = N\left(\mu_0, \dfrac{\sigma^2}{n}\right)$
- $Z := \dfrac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \implies \mathbb{P}_Z = N(0, 1).$

$$\alpha = \mathbb{P}\left(\left|\overline{X} - \mu_0\right| \geqslant \varepsilon | \mu = \mu_0\right) = \mathbb{P}\left(\frac{\left|\overline{X} - \mu_0\right|}{\frac{\sigma}{\sqrt{n}}} \geqslant \frac{\varepsilon}{\frac{\sigma}{\sqrt{n}}} | \mu = \mu_0\right)$$

$$= \mathbb{P}\left(|Z| \geqslant \frac{\varepsilon\sqrt{n}}{\sigma}\right)$$

$$= \mathbb{P}\left(Z \geqslant \frac{\varepsilon\sqrt{n}}{\sigma}\right) + \mathbb{P}\left(Z \leqslant \frac{-\varepsilon\sqrt{n}}{\sigma}\right)$$

$$= 1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) + \Phi\left(\frac{-\varepsilon\sqrt{n}}{\sigma}\right)$$

$$= 2\left(1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)\right) \iff$$

$$\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) = 1 - \frac{\alpha}{2} \iff \frac{\varepsilon\sqrt{n}}{\sigma} = z_{1-\frac{\alpha}{2}} \iff \varepsilon = \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$$

Formulazione del criterio

Presi i dati $x_1, \ldots, x_n$, sia $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ la loro media:

accetto H_0 se $|\overline{x} - \mu_0| < \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$ e la rifiuto altrimenti.

Test unilaterale inferiore con H_0 semplice

$X_1, \ldots, X_n$ un campione gaussiano di valore atteso μ incognito e varianza σ^2 nota.

- $H_0: \quad \mu = \mu_0 \qquad H_A: \quad \mu > \mu_0$
- Criterio di accettazione: accetto H_0 se e solo se $\overline{x} < \mu_0 + \varepsilon$.
- Livello di significatività

$$\alpha = \mathbb{P}\left(\overline{X} \geqslant \mu_0 + \varepsilon | \mu = \mu_0\right).$$

Devo determinare ε in funzione di α

- H_0 vera $\implies \mathbb{P}_{\overline{X}} = N\left(\mu_0, \frac{\sigma^2}{n}\right)$
- $Z := \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \implies \mathbb{P}_Z = N(0, 1)$

$$\alpha = \mathbb{P}\left(\overline{X} \geqslant \mu_0 + \varepsilon | \mu = \mu_0\right) = \mathbb{P}\left(\frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geqslant \frac{\varepsilon}{\frac{\sigma}{\sqrt{n}}} | \mu = \mu_0\right)$$

$$= \mathbb{P}\left(Z \geqslant \frac{\varepsilon\sqrt{n}}{\sigma}\right) = 1 - \mathbb{P}\left(Z \leqslant \frac{\varepsilon\sqrt{n}}{\sigma}\right) = 1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) \iff$$

$$\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) = 1 - \alpha \iff \varepsilon = \frac{\sigma}{\sqrt{n}} z_{1-\alpha}.$$

Formulazione del criterio

Presi i dati $x_1, \ldots, x_n$, sia $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ la loro media:

accetto H_0 se $\overline{x} < \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$ e la rifiuto altrimenti.

Test unilaterale inferiore con H_0 composta

$X_1, \ldots, X_n$ un campione gaussiano di valore atteso μ incognito e varianza σ^2 nota.

- $H_0 : \quad \mu \leqslant \mu_0 \qquad H_A : \quad \mu > \mu_0$
- Criterio di accettazione: accetto H_0 se e solo se $\overline{x} < \mu_0 + \varepsilon$.
- Livello di significatività

$$\mathbb{P}\left(\overline{X} \geqslant \mu_0 + \varepsilon | \mu \leqslant \mu_0\right)$$

- Se $\mathbb{P}_{X_i} = N\left(\mu, \sigma^2\right)$

$$\Longrightarrow \mathbb{P}_{\overline{X}} = N\left(\mu, \frac{\sigma^2}{n}\right), \quad Z := \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}, \quad \mathbb{P}_Z = N(0, 1)$$

$\Longrightarrow$

$$\mathbb{P}\left(\overline{X} \geqslant \mu_0 + \varepsilon | \mathbb{E}\left[X_i\right] = \mu\right) = \mathbb{P}\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geqslant \frac{\mu_0 - \mu + \varepsilon}{\frac{\sigma}{\sqrt{n}}} | \mathbb{E}\left[\overline{X}\right] = \mu\right)$$

$$= \mathbb{P}\left(Z \geqslant \frac{(\mu_0 - \mu + \varepsilon)\sqrt{n}}{\sigma}\right) = 1 - \mathbb{P}\left(Z \leqslant \frac{(\mu_0 - \mu + \varepsilon)\sqrt{n}}{\sigma}\right)$$

$$= 1 - \Phi\left(\frac{(\mu_0 - \mu + \varepsilon)\sqrt{n}}{\sigma}\right) \leqslant 1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) \leqslant \alpha$$

Voglio

$$\mathbb{P}\left(\overline{X} > \mu_0 + \varepsilon | \mathbb{E}\left[\overline{X}\right] = \mu\right) \leqslant \alpha \qquad \forall \mu \leqslant \mu_0$$

Scelgo ε tale che modo da avere $1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) = \alpha$ cioè

$$\varepsilon = \frac{\sigma}{\sqrt{n}} z_{1-\alpha}.$$

Criterio di accettazione

Presi i dati $x_1, \ldots, x_n$, sia dunque $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ la loro media.

Accetto H_0 se $\overline{x} < \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$ e la rifiuto altrimenti.

Test unilaterale superiore con H_0 semplice

$X_1, \ldots, X_n$ un campione gaussiano di valore atteso μ incognito e varianza σ^2 nota.

- $H_0 : \quad \mu = \mu_0 \qquad H_A : \mu < \mu_0$
- Criterio di accettazione: accetto H_0 se se $\overline{x} > \mu_0 - \varepsilon$.
- Livello di significatività:

$$\mathbb{P}\left(\overline{X} \leqslant \mu_0 - \varepsilon | \mu = \mu_0\right) = \alpha$$

- $\varepsilon = \dfrac{\sigma}{\sqrt{n}} z_{1-\alpha}$

Criterio di accettazione

Presi i dati $x_1, \ldots, x_n$, sia dunque $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ la loro media.

Accetto H_0 se $\overline{x} > \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$ e la rifiuto altrimenti.

Test unilaterale superiore con H_0 composta

$X_1, \ldots, X_n$ un campione gaussiano di valore atteso μ incognito e varianza σ^2 nota.

- $H_0: \quad \mu \geqslant \mu_0 \qquad H_A: \mu < \mu_0$.
- Criterio di accettazione: accetto H_0 se $\overline{x} > \mu_0 - \varepsilon$.
- Livello di significatività:

$$\alpha = \mathbb{P}\left(\overline{X} \leqslant \mu_0 - \varepsilon \middle| \mathbb{E}\left[\overline{X}\right] \leqslant \mu_0\right).$$

- $\varepsilon = \dfrac{\sigma}{\sqrt{n}} z_{1-\alpha}$

Criterio di accettazione

Presi i dati $x_1, \ldots, x_n$, sia dunque $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ la loro media.

Accetto H_0 se $\overline{x} > \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$ e la rifiuto altrimenti.