Politecnico di Milano
Appunti delle lezioni del corso di Statistica (2L) per gli allievi INF e TEL, AA 2008/2009*

Verifica di ipotesi

Ilenia Epifani

17 giugno 2009

Indice

1 Introduzione, notazioni, definizioni

Nelle lezioni precedenti abbiamo affrontato il problema di stimare puntualmente o mediante intervalli di confidenza uno o più parametri incogniti. Molti problemi ingegneristici richiedono invece di verificare ipotesi relative a una distribuzione incognita, sulla base di un campione casuale estratto da essa.

Per chiarire la questione, pensiamo a una ditta produttrice di cinghie di trasmissione per automobili che ha progettato un nuovo congegno che –applicato al sistema di trasmissione– consente di allungare la "vita" di una cinghia di almeno 6 mila chilometri, portando la durata attesa da 50 a 56 mila chilometri. Questa ditta potrebbe essere preoccupata del fatto che il nuovo congegno non migliori le prestazioni e a questo scopo effettua un test su 30 automobili provviste del nuovo congegno, misura la durata delle 30 cinghie e calcola la media campionaria. Quindi la ditta adotta la seguente regola di decisione: se la vita media campionaria osservata è "un pò più grande" di 56 mila chilometri, per esempio se è almeno pari a 57.900 chilometri, allora conclude che la preoccupazione non è concreta". Se invece la vita media campionaria delle cinghie testate non supera la frontiera dei 57.900 chilometri, allora la ditta ritiene verosimilmente concreta la sua preoccupazione.

Quanto appena descritto è un tipico *problema di verifica di ipotesi*, per risolvere il quale useremo la teoria dell'ottimalità dei test di Neyman-Pearson.

Per formalizzare e risolvere un problema di verifica di ipotesi, dobbiamo innanzitutto dotarci di un opportuno vocabolario statistico.

Al solito abbiamo estratto un campione casuale da una popolazione con f.d.r. F completamente o parzialmente incognita.

Definizione 1.1 Un'*ipotesi statistica* è un'asserzione o congettura sulla f.d.r. incognita F. Se F, e quindi la corrispondente funzione di densità f, è nota a meno di un parametro $\theta = (\theta_1, \dots, \theta_m) \in \Theta \subset \mathbb{R}^m$, l'ipotesi statistica è un'asserzione su θ.

Un'ipotesi statistica è *semplice* se specifica completamente la f.d.r. F della popolazione ed è *composta* se non è semplice. Per esempio, l'ipotesi statistica H_A: *"F è la f.d.r. gaussiana standard"* è un'ipotesi semplice mentre l'ipotesi statistica H_B: *"la f.d.r. F appartiene alla famiglia delle f.d.r. gaussiane di media 0"* è un'ipotesi composta, in quanto esistono infinite f.d.r. gaussiane di media 0 compatibili con H_B.

Nell'ultima parte del corso indagheremo la verifica di ipotesi in ambito *non parametrico*, cioè in un ambito di completa ignoranza su F. Per ora abbiamo un campione casuale $X_1, \dots, X_n$ *i.i.d.* $\sim f(x, \theta), \theta \in \Theta \subset \mathbb{R}^m$ e le ipotesi statistiche sono ipotesi su θ. Da questo momento in poi svilupperemo gli ingredienti della verifica di ipotesi in termini parametrici.

Il problema di verifica di ipotesi che noi affrontiamo è il seguente.

Sono state formulate due ipotesi statistiche su θ, chiamiamole H_0 e H_1. Le due ipotesi non sono sullo stesso piano e per distinguerle in letteratura una è detta *nulla* e l'altra è detta *alternativa*. Sia H_0 l'ipotesi nulla e H_1 l'alternativa. Le due ipotesi H_0 e H_1 inducono una partizione dello spazio parametrico Θ in due sottoinsiemi (disgiunti e non vuoti) Θ_0 e Θ_1. Esprimiamo in modo compatto tutto quanto detto sulle due ipotesi nel seguente modo:

$$H_0 : \theta \in \Theta_0, \ H_1 : \theta \in \Theta_1, \text{ con } \Theta_0 \neq \emptyset, \ \Theta_1 \neq \emptyset, \ \Theta_0 \cap \Theta_1 = \emptyset \text{ e } \Theta_0 \cup \Theta_1 = \Theta.$$

In altri termini, il vero valore che il parametro θ ha in natura è compatibile o con H_0 o con H_1 ma non con entrambe; inoltre, in natura non vi sono altri valori possibili per θ se non quelli specificati da H_0 e H_1.

Date le due ipotesi nulla e alternativa, dobbiamo ora costruire una regola decisionale per falsificare (negare) la nulla. Ovviamente dovremo usare i dati per stabilire se H_0 sia falsificabile. Se giungeremo alla conclusione di falsificare H_0 allora opteremo per l'ipotesi alternativa H_1. Attenzione: se H_0 è falsificata, ciò non significa che effettivamente H_0 sia un'ipotesi falsa, così come se H_0 non è falsificata ciò non significa che effettivamente H_0 sia la vera ipotesi. Infatti per decidere sulla falsità di H_0 procediamo in modo induttivo estraendo un campione di dati. Quindi, cambiando la realizzazione campionaria, potrebbe cambiare la decisione presa a favore o contro H_0. Ma non cambia in natura la vera ipotesi. Comunque, una volta presa la decisione, ci comporteremo come se l'ipotesi a favore della quale abbiamo deciso sia quella vera.

Ancora, per capire i diversi ruoli giocati da ipotesi nulla e alternativa, può essere utile leggere una procedura di verifica come un processo all'ipotesi nulla H_0 basato sul principio che l'imputato H_0 è innocente fino a prova contraria. Quindi prima di raccogliere i dati propendiamo per credere che la nulla sia falsa. Una volta raccolti i dati o perseveriamo a credere ciò, o cambiamo opinione e riteniamo verosimile H_0. Comunque, se decidiamo a favore di H_0 ciò non significa che i dati avvalorino fortemente la tesi che H_0 sia vera, ma semplicemente che non vi è sufficiente evidenza sperimentale per falsificarla.

Ma in cosa consiste una regola di decisione? Procediamo a partizionare l'insieme $\mathbb{R}^n$ di tutte le realizzazioni campionarie di $X_1, \ldots, X_n$ in due regioni $\mathcal{G}$ e $\mathcal{G}^c$. Quindi effettuiamo il campionamento, registriamo i risultati $x_1, \ldots, x_n$ e se $(x_1, \ldots, x_n) \in \mathcal{G}$ falsifichiamo H_0, se, invece, $(x_1, \ldots, x_n) \in \mathcal{G}^c$ non falsifichiamo H_0. Riassumendo:

$$\boxed{\text{rifiutiamo } H_0 \text{ se } (x_1, \ldots, x_n) \in \mathcal{G} \text{ e accettiamo } H_0 \text{ se } (x_1, \ldots, x_n) \in \mathcal{G}^c}$$

Definizione 1.2 Se $\mathcal{G}$ è un sottoinsieme di $\mathbb{R}^n$ tale che se $(x_1, \ldots, x_n) \in \mathcal{G}$ rifiutiamo H_0, $\mathcal{G}$ è detta *regione critica o di rifiuto* (di H_0) e $\mathcal{G}^c$ è detta *regione di accettazione* (di H_0).

Ripetiamo: una sola decisione può e deve essere presa: o rifiutare H_0, che equivale ad accettare H_1, o accettare H_0, che equivale a rifiutare H_1.

Definizione 1.3 Un *test di ipotesi* (*hypothesis test*) è una tripletta costituita dal campione $X_1, \ldots, X_n$, dalle ipotesi nulla e alternativa H_0, H_1 e da un sottoinsieme $\mathcal{G}$ di $\mathbb{R}^n$ che ha l'interpretazione di regione critica: $(X_1, \ldots, X_n;\ H_0, H_1;\ \mathcal{G})$.

Può succedere di prendere una decisione sbagliata. L'errore può essere di due tipi:

Errore di I tipo o prima specie: quando rifiutiamo H_0 ma H_0 è vera;

Errore di II tipo o seconda specie: quando accettiamo H_0 ma H_0 è falsa.

Dovremmo evitare questi errori di decisione; d'altro canto, non conoscendo il vero valore di θ non sappiamo con certezza se ne stiamo commettendo uno e quale. Possiamo però calcolare le probabilità di errore.

Siano $\alpha(\theta)$ la probabilità di errore di I tipo al variare di θ in Θ_0 e $\beta(\theta)$ la probabilità di errore di II tipo per θ in Θ_1. Cioè:

$$\alpha(\theta) = P_\theta(\text{“Rifiutare } H_0\text{”}) = P_\theta((x_1, \ldots, x_n) \in \mathcal{G}), \qquad \theta \in \Theta_0 \tag{1}$$

$$\beta(\theta) = P_\theta(\text{“Accettare } H_0\text{”}) = P_\theta((x_1, \ldots, x_n) \in \mathcal{G}^c), \qquad \theta \in \Theta_1 \tag{2}$$

La probabilità di errore di I tipo non può superare il valore

$$\alpha := \sup_{\theta \in \Theta_0} \alpha(\theta) = \sup_{\theta \in \Theta_0} P_\theta((x_1, \dots, x_n) \in \mathcal{G}) \tag{3}$$

Definizione 1.4 L'estremo superiore della probabilità di errore di I tipo $\alpha := \sup_{\theta \in \Theta_0} \alpha(\theta)$ è detto *livello di significatività* del test (*level of significance*) o *ampiezza* o *dimensione* (*size*) della regione critica.

Infine,

Definizione 1.5 La funzione $\pi(\theta)$ definita da $\pi(\theta) = 1 - \beta(\theta) = P_\theta((x_1, \dots, x_n) \in \mathcal{G}), \theta \in \Theta_1$ è la *funzione di potenza* del test (*power function*).

La funzione di potenza di un test ha quindi l'interpretazione di probabilità di prendere la corretta decisione di rifiutare l'ipotesi nulla quando effettettivamente l'ipotesi nulla è falsa.

Esemplifichiamo di seguito le nozioni fin qui introdotte sul familiare modello gaussiano.

Esempio 1.6 Per verificare se una popolazione gaussiana di varianza 1 abbia media non positiva contro l'alternativa che la media sia strettamente positiva, abbiamo estratto un campione di 25 osservazioni e calcolato la media campionaria $\bar{x}$. Quindi abbiamo deciso di rifiutare l'ipotesi nulla che la media teorica sia non positiva se $\bar{x} \geq 0.4$. Procediamo a formalizzare il test di ipotesi e calcoliamone le probabilità di errore di I e II tipo, la significatività del test e la funzione potenza.

Abbiamo il campione casuale $X_1, \dots, X_{25}$ *i.i.d.* $\sim \mathcal{N}(\mu, 1)$ e dobbiamo verificare l'ipotesi $H_0 : \mu \leq 0$ *versus* $H_1 : \mu > 0$ usando la regione critica $\mathcal{G} = \{(x_1, \dots, x_{25}) : \bar{x} \geq 0.4\}$.

1. La probabilità di errore di I tipo è data da

$$\alpha(\mu) = P_\mu(\bar{X} \geq 0.4) = 1 - \Phi\left(\frac{0.4 - \mu}{1/5}\right) = 1 - \Phi(2 - 5\mu), \quad \mu \leq 0$$

dal momento che sotto H_0 $\bar{X} \sim \mathcal{N}(\mu, 1/25)$ con $\mu \leq 0$.

Se per esempio $\bar{x}$ vale 0.3, allora non rifiutiamo H_0. Cioè non rifiutiamo la congettura che μ sia negativa sebbene abbiamo osservato una media campionaria positiva. Ma, non c'è nessun paradosso, in quanto un valore positivo ma "piccolo" della media campionaria è semplicemente dovuto alle fluttuazioni aleatorie della media campionaria.

2. Poiché $\alpha(\mu)$ è funzione strettamente crescente di μ, allora la significatività di questo test è data da

$$\alpha = \sup_{\mu \leq 0} P_\mu(\bar{X} \geq 0.4) = 1 - \Phi(2 - 5 \times 0) = 1 - \Phi(2) \simeq 2.3\%$$

3. Sotto H_1 $\bar{X} \sim \mathcal{N}(\mu, 1/25)$ con $\mu > 0$ e quindi la probabilità di errore di II tipo è

$$\beta(\mu) = P_\mu(\bar{X} < 0.4) = \Phi\left(\frac{0.4 - \mu}{1/5}\right) = \Phi(2 - 5\mu), \quad \mu > 0$$

Infine,

4. la funzione di potenza è

$$\pi(\mu) = P_\mu(\bar{X} \geq 0.4) = 1 - \Phi(2 - 5\mu), \quad \mu > 0$$

π è funzione strettamente crescente di μ e per $\mu \to \infty$ ha un asintoto orizzontale in 1. Se per esempio effettivamente la media è 1, allora la potenza nel vero valore 1 è $\pi(1) = 1 - \Phi(2 - 5 \times 1) = \Phi(3) \simeq 99.9\%$

Notate che il valore di π è strettamente maggiore della significatività del test α.

Un test che ha funzione di potenza $\pi(\theta)$ sempre maggiore della significatività α, per ogni scelta di θ in Θ_1, è detto *test non distorto*. Sui test non distorti non diremo null'altro in questo corso.

2 Lemma di Neyman-Pearson e test del rapporto di verosimiglianza

Un buon test è tale per cui α e β sono trascurabili. In realtà spesso vi è un trade-off fra α e β cosicché è impossibile minimizzare entrambi. Per risolvere il problema e costruire buoni test, una possibile strada è quella di fissare la massima probabilità dell'errore che riteniamo più grave quindi procedere a cercare una regione critica che minimizzi la probabilità dell'altro. Per quanto discusso nella Sezione 1 l'errore più grave fra i due è l'errore di I tipo cioè del rifiuto di H_0 quando H_0 è vera. La questione che si pone è "fissato il livello del test α, esiste una regione critica $\mathcal{G}$ tale che il test con questa regione critica ha probabilità di errore di II tipo β minimo rispetto a ogni altra regione critica di livello minore o uguale a α, uniformemente su ogni $\theta \in \Theta_1$?". Un test che soddisfa queste richieste è detto *test uniformemente più potente di livello α*, dove il nome è giustificato dal fatto che minimizzare β equivale a massimizzare la funzione di potenza π.

Noi costruiremo un test uniformemente più potente nel caso di ipotesi H_0, H_1 entrambe semplici. In particolare il risultato è contenuto nel *Lemma di Neyman-Pearson*.

Lemma 2.1 (Lemma di Neyman-Pearson) *Sia $X_1, \dots, X_n$ un campione casuale con verosimiglianza $L_\theta(x_1, \dots, x_n)$ e supponiamo di voler verificare $H_0 : \theta = \theta_0$ contro $H_1 : \theta = \theta_1$. Sia $\mathcal{G}$ la regione critica definita da*

$$\mathcal{G} = \left\{ (x_1, \dots, x_n) : \ \frac{L_{\theta_0}(x_1, \dots, x_n)}{L_{\theta_1}(x_1, \dots, x_n)} \le \delta \right\} \ .$$

Allora $\mathcal{G}$ è la regione critica che genera massima potenza fra tutte le regioni critiche di ampiezza minore o uguale all'ampiezza di $\mathcal{G}$.

Dimostrazione Dobbiamo dimostrare che per una qualunque fissata regione critica $\mathcal{F}$ di ampiezza al più pari a quella di $\mathcal{G}$, cioè tale che $P_{\theta_0}(\mathcal{F}) \le P_{\theta_0}(\mathcal{G})$, vale $P_{\theta_1}(\mathcal{F}) \le P_{\theta_1}(\mathcal{G})$.

Innanzitutto osserviamo che potendo rappresentare $\mathcal{F}, \mathcal{G}$ come

$$\mathcal{F} = (\mathcal{F} \cap \mathcal{G}) \cup (\mathcal{F} \cap \mathcal{G}^c), \qquad \mathcal{G} = (\mathcal{F} \cap \mathcal{G}) \cup (\mathcal{F}^c \cap \mathcal{G})$$

allora

$$(4) \qquad P_{\theta_i}(\mathcal{F}) \le P_{\theta_i}(\mathcal{G}) \quad \text{se e solo se} \quad P_{\theta_i}(\mathcal{F} \cap \mathcal{G}^c) \le P_{\theta_i}(\mathcal{F}^c \cap \mathcal{G}), \quad \text{per } i = 0, 1 \ .$$

Inoltre,

$$\begin{aligned} L_{\theta_0}(x_1, \dots, x_n) &\le \delta L_{\theta_1}(x_1, \dots, x_n), \quad \forall (x_1, \dots, x_n) \in \mathcal{G} \\ L_{\theta_1}(x_1, \dots, x_n) &< L_{\theta_0}(x_1, \dots, x_n)/\delta, \quad \forall (x_1, \dots, x_n) \in \mathcal{G}^c \end{aligned}$$

da cui otteniamo:

$$P_{\theta_0}(A) \leq \delta P_{\theta_1}(A) \quad \forall A \subset \mathcal{G} \tag{5}$$

$$P_{\theta_1}(B) \leq P_{\theta_0}(B)/\delta \quad \forall B \subset \mathcal{G}^c \tag{6}$$

Infatti, se per esempio le osservazioni $X_1, \ldots, X_n$ sono continue, allora per ogni $B \subset \mathcal{G}^c$ abbiamo:

$$P_{\theta_1}(B) = \int_B L_{\theta_1}(x_1, \ldots, x_n)\ dx_1 \cdots dx_n = \frac{1}{\delta}\int_B L_{\theta_0}(x_1, \ldots, x_n)\ dx_1 \cdots dx_n = \frac{P_{\theta_0}(B)}{\delta}\ .$$

Pertanto,

$$\begin{aligned} P_{\theta_1}(\mathcal{F} \cap \mathcal{G}^c) &\leq P_{\theta_0}(\mathcal{F} \cap \mathcal{G}^c)/\delta \quad \text{[per la (6)]} \\ &\leq P_{\theta_0}(\mathcal{F}^c \cap \mathcal{G})/\delta \quad \text{[per la (4) con } i = 0\text{]} \\ &\leq \delta P_{\theta_1}(\mathcal{F}^c \cap \mathcal{G})/\delta \quad \text{[per la (5)]} \\ &= P_{\theta_1}(\mathcal{F}^c \cap \mathcal{G}). \end{aligned}$$

D'altro canto, sappiamo dalla (4) che $P_{\theta_1}(\mathcal{F} \cap \mathcal{G}^c) \leq P_{\theta_1}(\mathcal{F}^c \cap \mathcal{G})$ è equivalente a $P_{\theta_1}(\mathcal{F}) \leq P_{\theta_1}(\mathcal{G})$ e la tesi è dimostrata. ∎

Osservazione 2.2 (Ratio del Test di Neyman-Pearson) Il Lemma ci dice che per ottenere il test più potente dobbiamo usare tutte le informazioni che il modello statistico fornisce e queste informazioni sono riassunte nella funzione di verosimiglianza. Se in corrispondenza della realizzazione campionaria $(x_1, \ldots, x_n)$ L_{θ_0} è più piccola di L_{θ_1}, allora è più verosimile che il campione provenga da una popolazione di densità $f(x, \theta_1)$ che $f(x, \theta_0)$ e, coerentemente con queste considerazioni, il Lemma di Neyman-Pearson porta a rifiutare H_0. La quantificazione di quanto L_{θ_0} debba essere più piccola di L_{θ_1} per rifiutare H_0 è fissata dal livello del test α.

Alla luce di quanto espresso nell'Osservazione 2.2, estendiamo l'uso della funzione di verosimiglianza per costruire regioni critiche per verificare ipotesi non necessariamente entrambe semplici. I test risultanti prendono il nome di *test del rapporto di verosimiglianza (generalizzato)*.

2.1 Test del rapporto di verosimiglianza

Sia $X_1, \ldots, X_n$ un campione casuale con verosimiglianza $L_\theta(x_1, \ldots, x_n)$. Consideriamo il problema di verifica dell'ipotesi nulla $H_0 : \theta \in \Theta_0$ contro $H_1 : \theta \in \Theta_1$.

Definizione 2.3 La quantità

$$\Lambda(x_1, \ldots, x_n) = \frac{\sup_{\theta \in \Theta_0} L_\theta(x_1, \ldots, x_n)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} L_\theta(x_1, \ldots, x_n)}$$

definita per tutte le realizzazioni campionarie $(x_1, \ldots, x_n)$ tali che $\sup_{\theta \in \Theta_0 \cup \Theta_1} L_\theta(x_1, \ldots, x_n) > 0$ è detta *rapporto di verosimiglianza*.

Definizione 2.4 Il test di ipotesi

$$(X_1, \ldots, X_n; H_0 : \theta \in \Theta_0,\ H_1 : \theta \in \Theta_1; \mathcal{G} = \{(x_1, \ldots, x_n) :\ \Lambda(x_1, \ldots, x_n) \leq \delta\})$$

è detto *test del rapporto di verosimiglianza* e ha livello $\alpha = \sup_{\theta \in \Theta_0} P_\theta\left(\Lambda(x_1, \ldots, x_n) \leq \delta\right)$.

La *ratio* del test del rapporto di verosimiglianza è la stessa del test di Neyman-Pearson: rifiuta H_0 se la migliore spiegazione dei dati fornita dalla verosimiglianza non è in H_0 ma in H_1, nel qual caso il denominatore di Λ risulta "molto più grande del numeratore". A cosa corrisponde numericamente "molto più grande" è fissato dal livello α del test.

Osservazione 2.5 (Legame con gli stimatori di massima verosimiglianza) In alcuni problemi di verifica di ipotesi, nella costruzione del test del rapporto di verosimiglianza usiamo gli stimatori di massima verosimiglianza. Per esempio, consideriamo un campione casuale $X_1, \ldots, X_n$ *i.i.d.* $\sim f(x, \theta)$ con $\theta \in \mathbb{R}$ e sia $\hat{\theta}_{ML}$ lo stimatore di massima verosimiglianza di θ. Supponiamo poi di voler verificare $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$. In questo caso il rapporto di verosimiglianza si riduce a

$$\Lambda(x_1, \ldots, x_n) = \frac{L_{\theta_0}(x_1, \ldots, x_n)}{\sup_{\theta \in \mathbb{R}} L_\theta(x_1, \ldots, x_n)} = \frac{L_{\theta_0}(x_1, \ldots, x_n)}{L_{\hat{\theta}_{ML}}(x_1, \ldots, x_n)}$$

3 Test sulla media di una popolazione gaussiana

In questa sezione usiamo i test di Neyman-Pearson e del rapporto di verosimiglianza per fare verifica di ipotesi sulla media di una popolazione gaussiana. Siano $X_1, \ldots, X_n$ *i.i.d.* $\sim \mathcal{N}(\mu, \sigma^2)$. I problemi di verifica di ipotesi da risolvere sono:

1. $H_0 : \mu = \mu_0$ *versus* $H_1 : \mu = \mu_1$ con $\mu_0 \neq \mu_1$
2. $H_0 : \mu = \mu_0$ *versus* $H_1 : \mu < \mu_0$
3. $H_0 : \mu \geq \mu_0$ *versus* $H_1 : \mu < \mu_0$
4. $H_0 : \mu = \mu_0$ *versus* $H_1 : \mu > \mu_0$
5. $H_0 : \mu \leq \mu_0$ *versus* $H_1 : \mu > \mu_0$
6. $H_0 : \mu = \mu_0$ *versus* $H_1 : \mu \neq \mu_0$

La funzione di verosimiglianza di un campione gaussiano $\mathcal{N}(\mu, \sigma^2)$ è data da

$$L_{\mu,\sigma^2}(x_1, \ldots, x_n) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2} - \frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right\} \tag{7}$$

dove $s^2 = \sum_{j=1}^n (x_j - \bar{x})^2/(n-1)$[1].
Ora procediamo trattando separatamente i due casi di varianza nota e varianza incognita.

3.1 Caso di varianza nota: z-test

Sia σ_0^2 il valore noto della varianza. Partiamo dall'affrontare il problema di ipotesi 1. in cui entrambe le ipotesi sono semplici.

[1] cfr. l'Equazione (7), Esempio 5.2, pagina 11 degli appunti sulla Stima puntuale, AA 06/07

Problema 1: $H_0 : \mu = \mu_0$ ***versus*** $H_1 : \mu = \mu_1$ **con** $\mu_0 \neq \mu_1$**.** Applichiamo il Lemma di Neyman-Pearson per costruire il test più potente di livello α. Deriva dall'Equazione (7) che il rapporto delle verosimiglianze L_{μ_0}/L_{μ_1} è

$$\frac{L_{\mu_0}(x_1,\dots,x_n)}{L_{\mu_1}(x_1,\dots,x_n)} = \exp\left\{\frac{n}{\sigma_0^2}\bar{x}(\mu_0-\mu_1) - \frac{n}{2\sigma_0^2}(\mu_0^2-\mu_1^2)\right\}$$

cosicché la regione critica di Neyman Pearson ha forma $\left\{\mathrm{e}^{\frac{n}{\sigma_0^2}\bar{x}(\mu_0-\mu_1)-\frac{n}{2\sigma_0^2}(\mu_0^2-\mu_1^2)} \leq \delta\right\}$.

Se $\boxed{\mu_1 < \mu_0}$, la precedente regione critica è equivalente a $\mathcal{G} = \{\bar{x} \leq k\}$ e il valore di k è determinato imponendo che il test abbia livello α, cioè k deve essere tale che $P_{\mu_0}(\bar{X} \leq k) = \alpha$. Siccome sotto H_0 –cioè se $\mu = \mu_0$– $\bar{X} \sim \mathcal{N}(\mu_0, \sigma_0^2/n)$, allora $k = \mu_0 + z_\alpha \sigma_0/\sqrt{n} = \mu_0 - z_{1-\alpha}\sigma_0/\sqrt{n}$, dove z_a indica il quantile di ordine a della f.d.r. gaussiana standard Φ.

In definitiva, il test più potente di livello α per il problema di ipotesi $H_0 : \mu = \mu_0$ *versus* $H_1 : \mu = \mu_1$ con $\mu_1 < \mu_0$ ha regione critica

$$\mathcal{G} = \left\{(x_1,\dots,x_n) : \bar{x} \leq \mu_0 - z_{1-\alpha}\frac{\sigma_0}{\sqrt{n}}\right\} \tag{8}$$

Osserviamo che la regione critica dipende soltanto da α, da μ_0 e dal fatto che μ_1 sia minore di μ_0. In altri termini, qualunque sia il valore specificato dall'ipotesi alternativa per μ_1, ma tale che $\mu_1 < \mu_0$, la regione critica di massima potenza è la stessa, cioè quella in (8). Questo implica che *la regione critica* (8) *è quella che genera massima potenza uniformemente sull'insieme* $\{\mu < \mu_0\}$ *anche per il problema di ipotesi* 2.*:* $H_0 : \mu = \mu_0$ *versus* $H_1 : \mu < \mu_0$.

Useremo ancora la regione critica (8) anche per il problema di ipotesi 3. dato da $H_0 : \mu \geq \mu_0$ *versus* $H_1 : \mu < \mu_0$. Infatti la regione critica (8) è quella cui si arriva con un test del rapporto di verosimiglianza. Ma noi non faremo questo conto. Si può dimostrare che la regione critica (8) è quella che genera la massima potenza per il problema di ipotesi 3. nella classe dei test non distorti di livello α, (che è ovviamente una classe più piccola rispetto a quella dei test semplicemente di livello α). Ma noi non svilupperemo nemmeno questo punto.

Se invece $\boxed{\mu_1 > \mu_0}$, la regione critica che genera massima potenza ha forma $\mathcal{G} = \{\bar{x} \geq k\}$ ed è di livello α se $k = \mu_0 + z_{1-\alpha}\sigma_0/\sqrt{n}$. Le osservazioni su svolte per il caso $\mu_1 < \mu_0$ possono analogamente essere ripetute (con gli ovvi distinguo) portando alla seguente regola: usiamo la regione critica

$$\mathcal{G} = \left\{(x_1,\dots,x_n) : \bar{x} \geq \mu_0 + z_{1-\alpha}\frac{\sigma_0}{\sqrt{n}}\right\} \tag{9}$$

per decidere a livello α in ciascuno dei seguenti tre problemi di verifica di ipotesi:

1. $H_0 : \mu = \mu_0$ *versus* $H_1 : \mu = \mu_1$ con $\mu_1 > \mu_0$
4. $H_0 : \mu = \mu_0$ *versus* $H_1 : \mu > \mu_0$
5. $H_0 : \mu \leq \mu_0$ *versus* $H_1 : \mu > \mu_0$.

Rimane da costruire la regione critica per il Problema 6.: $H_0 : \mu = \mu_0$ *versus* $H_1 : \mu \neq \mu_0$. Qui usiamo il test del rapporto di verosimiglianza.

a) Lo stimatore ML di μ è $\bar{X}$. Per quanto detto nell'Osservazione 2.5 e per l'Equazione (7) il rapporto di verosimiglianza Λ è dato da

$$\frac{L_{\mu_0,\sigma_0^2}(x_1,\dots,x_n)}{L_{\bar{x},\sigma_0^2}(x_1,\dots,x_n)} = \frac{\left(\frac{1}{2\pi\sigma_0^2}\right)^{\frac{n}{2}}\exp\left\{-\frac{(n-1)s^2}{2\sigma_0^2} - \frac{n(\bar{x}-\mu_0)^2}{2\sigma_0^2}\right\}}{\left(\frac{1}{2\pi\sigma_0^2}\right)^{\frac{n}{2}}\exp\left\{-\frac{(n-1)s^2}{2\sigma_0^2}\right\}} = \exp\left\{-\frac{n(\bar{x}-\mu_0)^2}{2\sigma_0^2}\right\}$$

b) $\Lambda \le \lambda$ se e solo se $\dfrac{\sqrt{n}|\bar{x}-\mu_0|}{\sigma_0} \ge k$ con k^2 tale che $P_{\mu_0}\left(\dfrac{\sqrt{n}|\bar{X}-\mu_0|}{\sigma_0} \ge k\right) = \alpha$. Siccome sotto H_0 $\bar{X} \sim \mathcal{N}(\mu_0, \sigma_0^2/n)$, segue che $k = z_{1-\alpha/2}$. In definitiva,

$$\left\{(x_1,\dots,x_n):\ \frac{|\bar{x}-\mu_0|}{\sigma_0/\sqrt{n}} \ge z_{1-\frac{\alpha}{2}}\right\} \tag{10}$$

è la regione critica di livello α per verificare $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$ nel caso di campione gaussiano con varianza nota σ_0^2.

Osservazione 3.1 (Nozione di statistica test) Notiamo che in tutti i problemi di ipotesi sulla media di popolazione gaussiana, non abbiamo mai esplicitamente calcolato né la frontiera δ né il rapporto di verosimiglianza $L_{\mu_0}(x_1,\dots,x_n)/L_{\mu_1}(x_1,\dots,x_n)$ (nel caso di ipotesi semplici) o $\Lambda(x_1,\dots,x_n)$ (nel caso di ipotesi composte). Questo perché siamo sempre riusciti a ridurre le diseguaglianze su L_{μ_0}/L_{μ_1} e Λ in diseguaglianze sulla statistica $\bar{X}$. In altri termini, è la realizzazione campionaria di $\bar{X}$ che è decisiva per stabilire quale ipotesi riteniamo vera. Per questo motivo la statistica $\bar{X}$ è detta *statistica test*.

In generale, ogni qualvolta esista una statistica T tale che una regione critica $\mathcal{G}$ ha forma $\mathcal{G} = \{T \in A\}$, per qualche A sottoinsieme di $\mathbb{R}$, allora la statistica T è detta *statistica test* (*test statistics*).

Infine, i test sulla media con varianza nota sono detti *z-test*.

3.2 Caso di varianza incognita: t-test

Nel caso di varianza incognita, in tutti i problemi di verifica di ipotesi 1.–6. le ipotesi sono composte. Si può così usare il test del rapporto di verosimiglianza per costruire le regioni critiche di livello α. Calcolando il rapporto di verosimiglianza e poi riducendo *il più possibile* la diseguaglianza $\Lambda(x_1,\dots,x_n) \le \delta$, otteniamo che la decisione dipende dalla statistica test $\sqrt{n}(\bar{X}-\mu_0)/\sqrt{S^2}$. In particolare, i test di ipotesi cui si giunge sono i seguenti:

$\mathbf{H_0}$	$\mathbf{H_1}$	**Si rifiuta $\mathbf{H_0}$ se**
$\mu=\mu_0$ $\mu=\mu_0$ $\mu\le\mu_0$	$\mu=\mu_1$ *con* $\mu_0<\mu_1$ $\mu>\mu_0$ $\mu>\mu_0$	$\bar{x} \ge \mu_0 + t_{n-1}(1-\alpha)\dfrac{s}{\sqrt{n}}$
$\mu=\mu_0$ $\mu=\mu_0$ $\mu\ge\mu_0$	$\mu=\mu_1$ *con* $\mu_0>\mu_1$ $\mu<\mu_0$ $\mu<\mu_0$	$\bar{x} \le \mu_0 - t_{n-1}(1-\alpha)\dfrac{s}{\sqrt{n}}$
$\mu=\mu_0$	$\mu\neq\mu_0$	$\dfrac{|\bar{x}-\mu_0|}{s/\sqrt{n}} \ge t_{n-1}(1-\dfrac{\alpha}{2})$

(Per una ricostruzione di queste regioni critiche potete anche consultare il Theorem $III.2.3$ e la Remark 1 pagina 156 in Pestman 1998).

Notate le uniche differenze rispetto al caso di varianza nota: dove nel caso di varianza nota trovavamo σ_0^2 ora, con varianza incognita, abbiamo la realizzazione della varianza campionaria, s^2, dove invece c'era il quantile z_a ora abbiamo il quantile $t_{n-1}(a)$ della f.d.r. t di student con $n-1$ gradi di libertà.

I test sulla media con varianza incognita sono detti *t-test*.

[2] $k = \sqrt{(1/\delta - 1)(n-1)}$

4 Nozione di p-value

La procedura di verifica di ipotesi presentata è conservativa sull'ipotesi nulla: si rifiuta H_0 solo se vi è una forte evidenza dei dati contro di essa. D'altro canto, la decisione è fondata sul parametro di controllo dato dal livello del test α. Avendo a disposizione i dati, possiamo discriminare fra gli α per cui rifiutiamo H_0 e quelli per cui accettiamo H_0? In altri termini, guardando i dati, possiamo arrivare a una decisione di rifiuto di H_0 forte, robusta rispetto alla scelta di α? Si, usando la nozione di p-value o *livello descrittivo* del test.

Per descrivere il p-value, partiamo dall'unica procedura test generale che abbiamo elaborato, cioè il test del rapporto di verosimiglianza. Supponiamo di voler verificare l'ipotesi nulla $H_0 : \theta \in \Theta_0$ e sia λ il valore assunto dal rapporto di verosimiglianza in corrispondenza della realizzazione campionaria $(x_1, \ldots, x_n)$: $\lambda = \Lambda((x_1, \ldots, x_n)$. Sia p il livello del test del rapporto di verosimiglianza con frontiera $\delta = \lambda$, cioè $p = \sup_{\theta \in \Theta_0} P_\theta(\Lambda \leq \lambda)$ e sia $\{\Lambda \leq \delta(\alpha)\}$ la regione critica di livello α. Ho scritto $\delta(\alpha)$, perché il valore di frontiera δ cresce con α. Quindi la diseguaglianza $\lambda \leq \delta(\alpha)$ è soddisfatta se e solo se $p \leq \alpha$. Ma se $\lambda \leq \delta(\alpha)$ noi rifiutiamo H_0 a livello α. Concludiamo che

$$\boxed{\text{per ogni } \alpha \geq p \text{ rifiutiamo } H_0 \text{ e per ogni } \alpha < p \text{ accettiamo } H_0}.$$

Un valore $p \in (0, 1)$ per cui avviene ciò è detto p-value del test.

Come calcoliamo praticamente i valori del p-value? Se esiste una statistica test, calcoliamo prima la statistica test usando i dati, poi, ...è più semplice descrivere con un esempio. Vediamolo su un esempio gaussiano.

Esempio 4.1 Per verificare $H_0 : \mu = \mu_0$ contro $H_1 : \mu = \mu_1$, con $\mu_1 < \mu_0$, a livello α, se la varianza σ_0^2 è nota e abbiamo n osservazioni usiamo la regione critica $\mathcal{G} = \{\sqrt{n}(\bar{x} - \mu_0)/\sigma_0 \leq z_\alpha\}$, da cui ricaviamo che, fissato il valore di $\bar{x}$, rifiutiamo per tutti gli α tali che $\alpha \geq \Phi(\frac{\bar{x}-\mu_0}{\sigma_0/\sqrt{n}})$. Quindi il p-value è dato da $\Phi(\frac{\bar{x}-\mu_0}{\sigma_0/\sqrt{n}})$.

Per esempio, se $\mu_0 = 2, \mu_1 = 1, n = 25, \sigma_0^2 = 4$ e $\bar{x} = 1.4$, allora il p-value vale $\Phi(-0.75) \simeq 1 - 0.7734 = 22.66\%$.

Come interpretiamo praticamente i valori del p-value? L'idea è la seguente: un p-value prossimo a zero indica forte evidenza empirica contro H_0, mentre, un p-value prossimo a uno indica mancanza di evidenza empirica contro H_0. Ma attenzione a non assegnare al p-value più del suo significato di discrimine fra gli α basato sui dati. Infatti, al variare dei dati il p-value cambia, quindi anche il p-value è una variabile aleatoria e si può dimostrare che se questi dati provengono da un modello continuo, la variabile aleatoria p-value sotto H_0 ha distribuzione uniforme su $(0, 1)$. Ciò significa che possiamo trovarci di fronte alla seguente situazione: l'ipotesi nulla H_0 è vera ma osserviamo dei valori del p-value prossimi a 0 e quindi rifiutiamo H_0.

5 Dualità fra verifica di ipotesi e stima intervallare

Esiste una stretta relazione fra la teoria della verifica di ipotesi e degli intervalli di confidenza. Diamone qui un'idea facendo vedere che

i) un intervallo di confidenza bilatero di livello γ per un parametro unidimensionale θ può essere usato per costruire un test di ipotesi di livello $\alpha = 1 - \gamma$ per verificare $H_0 = \theta_0$ contro $H_1 : \theta \neq \theta_0$ e, viceversa, che *ii*) una famiglia di test: $\{(X_1, \ldots, X_n; H_0 = \theta_0, H_1 : \theta \neq \theta_0; G_{\theta_0})\}_{\theta_0 \in \Theta}$ può essere usata per costruire una “regione di confidenza” per θ.

5.1 Dall'intervallo di confidenza al test di ipotesi

Sia $(x_1, \ldots, x_n)$ la realizzazione di un campione casuale $X_1, \ldots, X_n$ estratto da $f(x, \theta)$ e sia $IC(x_1, \ldots, x_n)$ un intervallo di confidenza per θ di livello γ, cioè tale che $P_\theta(\theta \in IC(X_1, \ldots, X_n)) = \gamma$. Siamo interessati a verificare: $H_0 = \theta_0$ contro $H_1 : \theta \neq \theta_0$. Consideriamo il sottoinsieme di $\mathbb{R}^n$

$$A_0 = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \theta_0 \in IC(x_1, \ldots, x_n)\} \tag{11}$$

Allora $(x_1, \ldots, x_n) \in A_0$ se e solo se $\theta_0 \in IC(x_1, \ldots, x_n)$ e quindi

$$P_{\theta_0}(A_0) = P_{\theta_0}(\theta_0 \in IC(X_1, \ldots, X_n)) = \gamma$$

Segue che $G := A_0^c$ è una regione critica per $H_0 = \theta_0$ contro $H_1 : \theta \neq \theta_0$ di ampiezza $\alpha = 1 - \gamma$.

Praticamente, *per verificare $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$ a un livello di significatività α, determiniamo un intervallo di confidenza bilatero $(1 - \alpha)100\%$ per θ. Se l'intervallo contiene θ_0 accettiamo H_0, altrimenti la rifiutiamo.*

Esempio 5.1 Sia $X_1, \ldots, X_n$ *i.i.d.* $\sim N(\mu, \sigma^2)$. Se la varianza è nota, un intervallo di confidenza bilatero $\gamma 100\%$ per la media μ è

$$IC(x_1, \ldots, x_n) = \left(\bar{x} - \sqrt{\frac{\sigma^2}{n}} z_{\frac{1+\gamma}{2}}, \bar{x} + \sqrt{\frac{\sigma^2}{n}} z_{\frac{1+\gamma}{2}}\right)$$

Procedendo come in (11), definiamo

$$\begin{aligned} G &= \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n : \ \mu_0 \notin \left(\bar{x} - \sqrt{\frac{\sigma^2}{n}} z_{\frac{1+\gamma}{2}}, \bar{x} + \sqrt{\frac{\sigma^2}{n}} z_{\frac{1+\gamma}{2}}\right)\right\} \\ &= \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n : \ \frac{|\bar{x} - \mu_0|}{\sqrt{\frac{\sigma^2}{n}}} \geq z_{\frac{1+\gamma}{2}}\right\} \end{aligned}$$

Ritroviamo la regione critica del test del rapporto di verosimiglianza di livello $\alpha = 1 - \gamma$ per $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$, per popolazione gaussiana con varianza nota.

Esempio 5.2 Sia $X_1, \ldots, X_n$ *i.i.d.* $\sim N(\mu, \sigma^2)$. Se σ^2 è incognita, un IC bilatero $\gamma 100\%$ per la media μ è

$$IC(x_1, \ldots, x_n) = [\bar{x} - \sqrt{\frac{s^2}{n}} t_{\frac{1+\gamma}{2}}(n-1), \bar{x} + \sqrt{\frac{s^2}{n}} t_{\frac{1+\gamma}{2}}(n-1)]$$

Definiamo ora

$$\begin{aligned} G &= \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n : \ \bar{x} - \sqrt{\frac{s^2}{n}} t_{\frac{1+\gamma}{2}}(n-1) < \mu_0 < \bar{x} + \sqrt{\frac{s^2}{n}} t_{\frac{1+\gamma}{2}}(n-1)\right\}^c \\ &= \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n : \ \frac{|\bar{x} - \mu_0|}{\sqrt{\frac{s^2}{n}}} \geq t_{\frac{1+\gamma}{2}}(n-1)\right\} \end{aligned}$$

Scopriamo che G è la regione critica del test del rapporto di verosimiglianza di livello $\alpha = 1 - \gamma$ per $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$, per popolazione gaussiana con varianza incognita.

5.2 Dalla teoria delle ipotesi alla regione di confidenza

Viceversa, consideriamo una famiglia di test di ipotesi tutti di livello α: $\{(X_1, \ldots, X_n; H_0 = \theta_0, H_1 : \theta \neq \theta_0; G_{\theta_0})\}_{\theta_0 \in \Theta}$, cioè abbiamo un test di ipotesi per ogni possibile specificazione di θ_0 in Θ, Allora

$$SC(x_1, \ldots, x_n) := \{\theta_0 \in \Theta : \ (x_1, \ldots, x_n) \notin G_{\theta_0}\} \tag{12}$$

è un sottoinsieme di Θ aleatorio (perché varia al variare delle realizzazioni campionarie) tale che

$$P_{\theta_0}(\theta_0 \in SC(X_1, \ldots, X_n))) = P_{\theta_0}((x_1, \ldots, x_n) \notin G_{\theta_0}) = 1 - P_{\theta_0}(G_{\theta_0}) = 1 - \alpha.$$

Quindi, se $x_1, \ldots, x_n$ è una realizzazione del campione casuale, allora $SC(x_1, \ldots, x_n)$ è una regione di confidenza $(1-\alpha)100\%$ per θ.

Esempio 5.3 Sia $X_1, \ldots, X_n$ *i.i.d.* $\sim N(\mu, \sigma^2)$ con σ^2 noto. Allora

$$A(\mu_0) = \left\{ (x_1 \ldots, x_n) \in \mathbb{R}^n : \ \frac{|\bar{x} - \mu_0|}{\sqrt{\frac{\sigma^2}{n}}} < z_{1-\frac{\alpha}{2}} \right\}$$

è la regione di accettazione di ampiezza α del test del rapporto di verosimiglianza per $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$. Procedendo come in (12) possiamo definire:

$$IC(x_1, \ldots, x_n) := \left\{ \mu_0 \in \mathbb{R} : \ \frac{|\bar{x} - \mu_0|}{\sqrt{\frac{\sigma^2}{n}}} < z_{1-\frac{\alpha}{2}} \right\} = \left(\bar{x} - \sqrt{\frac{\sigma^2}{n}} z_{1-\frac{\alpha}{2}}, \bar{x} + \sqrt{\frac{\sigma^2}{n}} z_{1-\frac{\alpha}{2}} \right)$$

riottenendo così l'intervallo di confidenza per la media di una popolazione gaussiana di varianza nota, di livello $\gamma = 1 - \alpha$ e di minima lunghezza.

6 Test sulla varianza di una popolazione gaussiana

In questa sezione usiamo la dualità fra verifica di ipotesi e stima intervallare per costruire i test di ipotesi sulla varianza di una popolazione gaussiana. Al solito trattiamo distintamente i due casi di media incognita e media nota.

6.1 Caso di media incognita

[Test bilatero] Partiamo dal test bilatero per verificare $H_0 : \sigma^2 = \sigma_0^2$ contro $\sigma^2 \neq \sigma_0^2$ e supponiamo che la media sia incognita.

Sia $X_1, \ldots, X_n$ *i.i.d.* $\sim N(\mu, \sigma^2)$. Se la media μ è incognita, un IC bilatero $(1-\alpha)100\%$ per la varianza σ^2 è

$$IC(x_1, \ldots, x_n) = \left(\frac{s^2(n-1)}{\chi^2_{n-1}\left(1 - \frac{\alpha}{2}\right)}, \frac{s^2(n-1)}{\chi^2_{n-1}\left(\frac{\alpha}{2}\right)} \right)$$

Definiamo ora

$$G = \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n : \ \sigma_0^2 \notin \left(\frac{s^2(n-1)}{\chi^2_{n-1}\left(1-\frac{\alpha}{2}\right)}, \frac{s^2(n-1)}{\chi^2_{n-1}\left(\frac{\alpha}{2}\right)} \right) \right\}$$
$$= \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n : \ \text{or } \frac{s^2(n-1)}{\sigma_0^2} \le \chi^2_{n-1}\left(\frac{\alpha}{2}\right) \ \text{or } \frac{s^2(n-1)}{\sigma_0^2} \ge \chi^2_{n-1}\left(1-\frac{\alpha}{2}\right) \right\}$$

Allora G è una regione critica di livello α per $H_0 : \sigma^2 = \sigma_0^2$ contro $H_1 : \sigma^2 \neq \sigma_0^2$, per popolazione gaussiana con media incognita.

[Test unilateri] Costruiamo ora un test di ipotesi per $H_0 : \sigma^2 \ge \sigma_0^2$ contro $H_1 : \sigma^2 < \sigma_0^2$, quando la media è incognita. Sappiamo che $\left(0, \frac{s^2(n-1)}{\chi^2_{n-1}(1-\gamma)}\right)$ è un intervallo di confidenza $\gamma 100\%$ a una coda inferiore per σ^2 . Quindi abbiamo alta confidenza (γ) che il vero valore di σ^2 sia minore o uguale di $\frac{s^2(n-1)}{\chi^2_{n-1}(1-\gamma)}$. Cioè, stando ai dati, un valore di $\sigma^2 \ge \frac{s^2(n-1)}{\chi^2_{n}(1-\gamma)}$ è altamente implausibile. Pertanto, se $\sigma_0^2 \ge \frac{s^2(n-1)}{\chi^2_{n-1}(1-\gamma)}$, allora ogni σ^2 compatibile con H_0 non è plausibile. Segue che *"Rifiutare $H_0 : \sigma^2 \ge \sigma_0^2$ a favore di $H_1 : \sigma^2 < \sigma_0^2$ se $\sigma_0^2 \ge \frac{s^2(n-1)}{\chi^2_{n-1}(1-\gamma)}$"* è una regola decisionale sensata. Il test per la varianza da popolazione gaussiana con media incognita, costruito sulla base di questa regola, ha regione critica

$$\mathcal{G} = \left\{ (x_1, \ldots_n) : \frac{s^2(n-1)}{\sigma_0^2} \le \chi^2_{n-1}(1-\gamma) \right\} \tag{13}$$

Se la media è incognita, possiamo usare la regione critica (13) come regione di rifiuto dell'ipotesi nulla nei seguenti problemi:

1. H_0: $\sigma^2 \ge \sigma_0^2$ *versus* H_1: $\sigma^2 < \sigma_0^2$
2. H_0: $\sigma^2 = \sigma_0^2$ *versus* H_1: $\sigma^2 < \sigma_0^2$

e anche

3. H_0: $\sigma^2 = \sigma_0^2$ *versus* H_1: $\sigma^2 = \sigma_1^2$ con $\sigma_1^2 < \sigma_0^2$.

Il livello di tutti questi test è $\alpha = 1-\gamma$. Infatti, se $\alpha = 1-\gamma$, allora

$$\alpha = P_{\sigma_0^2}\left(\frac{S^2(n-1)}{\sigma_0^2} \le \chi^2_{n-1}(\alpha) \right) = \sup_{\sigma^2 \ge \sigma_0^2} P_{\sigma^2}\left(\frac{S^2(n-1)}{\sigma_0^2} \le \chi^2_{n-1}(\alpha) \right)$$

Invece, partendo dall'intervallo unilatero della varianza di forma (c, ∞) e confidenza $(1-\alpha)100\%$ dato da $(s^2(n-1)/\chi^2_{n-1}(1-\alpha))$, e ragionando in modo analogo a prima, troviamo che una regione critica di ampiezza α per i problemi di ipotesi

4. H_0: $\sigma^2 \le \sigma_0^2$ *versus* H_1: $\sigma^2 > \sigma_0^2$
5. H_0: $\sigma^2 = \sigma_0^2$ *versus* H_1: $\sigma^2 > \sigma_0^2$
6. H_0: $\sigma^2 = \sigma_0^2$ *versus* H_1: $\sigma^2 = \sigma_1^2$ con $\sigma_1^2 > \sigma_0^2$

è data da

$$\mathcal{G} = \left\{ (x_1, \ldots_n) : \frac{s^2(n-1)}{\sigma_0^2} \ge \chi^2_{n-1}(1-\alpha) \right\} \tag{14}$$

6.2 Caso di media nota

Nel caso di media nota e pari a μ_0 faremo verifica di ipotesi sulla varianza da popolazione gaussiana come nella Sezione 6.1, con la sola differenza che dove con media incognita trovavamo la varianza campionaria S^2 ora, con media nota, abbiamo la statistica $S_0^2 = \sum_{j=1}^n (X_j - \mu_0)^2/n$ e dove c'era il quantile della f.d.r. chiquadrato con $n-1$ gradi di libertà $\chi^2_{n-1}(a)$ ora abbiamo il quantile $\chi^2_n(a)$ della f.d.r. χ^2 con n gradi di libertà.

7 Verifica di ipotesi per una popolazione bernoulliana

In questa sezione affrontiamo il problema di verificare ipotesi sulla probabilità che accada un certo evento, oppure sulla percentuale di unità di una certa popolazione che presentano una particolare caratteristica. Per esempio siamo interessati a verificare ipotesi sulla percentuale di pezzi difettosi prodotti da una certa catena produttiva. Se un evento accade, o un'unità presenta la particolare caratteristica, per convenzione, indichiamo questo risultato come un "successo".

Per la verifica di ipotesi sulla probabilità di un successo θ, estraiamo con reimmissione un campione di n individui da una popolazione (ovvero ripetiamo n volte un esperimento sempre nelle stesse condizioni), registriamo il numero di successi osservati e stimiamo θ con la frequenza relativa campionaria di successo, cioè con il rapporto fra numero di successi e numero di prove n.

Dal punto di vista della modellizzazione, abbiamo estratto un campione casuale dalla popolazione bernoulliana di parametro θ incognito: $X_1, \dots, X_n$ *i.i.d.* $\sim \theta^x(1-\theta)^{1-x}\mathbf{1}_{\{0,1\}}(x)$, $\theta \in (0,1)$. La verosimiglianza del campione è

$$L_\theta(x_1, \dots, x_n) = \theta^{\sum_{j=1}^n x_j}(1-\theta)^{n-\sum_{j=1}^n x_j} = \theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}} \tag{15}$$

con $\hat{\theta} = \sum_{j=1}^n x_j/n$. Deduciamo da (15) che la verifica di ipotesi su θ si baserà sulla statistica test $\sum_{j=1}^n X_j$ che registra il numero assoluto di successi nelle n prove o equivalentemente $\hat{\theta}$ che ne registra il numero relativo. Per quanto concerne la distribuzione del numero totale di successi, vale che $\sum_{j=1}^n X_j$ ha densità binomiale di parametri n, θ: $\sum_{j=1}^n X_j \sim \mathbf{Bi}(n, \theta)$.

Applicando il Lemma di Neyman-Pearson, otteniamo che il test più potente per verificare $H_0 : \theta = \theta_0$ contro $H_1 : \theta = \theta_1$ ha regione critica di forma $\mathcal{G} = \{\sum_{j=1}^n x_j \geq k_1\}$ se $\theta_1 > \theta_0$ e regione critica di forma $\mathcal{G} = \{\sum_{j=1}^n x_j \leq k_2\}$ se $\theta_1 < \theta_0$. Determiniamo i valori di k_1, k_2 imponendo che il livello del test sia al più pari ad α. Per esempio, nel secondo caso richiediamo

$$P_{\theta_0}\Big(\sum_{j=1}^n X_j \leq k_2\Big) \leq \alpha \tag{16}$$

e andiamo a calcolare il più grande valore di k_2 per cui questa diseguaglianza è soddisfatta. A sinistra della disequazione non abbiamo altro che la f.d.r. $\mathbf{Bi}(n, \theta_0)$, cioè

$$P_{\theta_0}\Big(\sum_{j=1}^n X_j \leq k_2\Big) = \sum_{k \leq k_2} \binom{n}{k} \theta_0^j (1-\theta_0)^{n-j}$$

La f.d.r. $\mathbf{Bi}(n, \theta)$ è discreta, per cui non è detto che riusciamo a trovare per tutti gli α un k_2 che realizzi l'uguaglianza in (16). Di contro, è semplice calcolare il p-value del test: se

$\sum_{j=1}^{n} x_j = s$, il p-value è $\sum_{k\leq s}\binom{n}{k}\theta_0^k(1-\theta_0)^{n-k}$ per il caso $\theta_1 < \theta_0$ ed è $\sum_{k>s}\binom{n}{k}\theta_0^k(1-\theta_0)^{n-kj}$, per il caso $\theta_1 > \theta_0$.

Se invece abbiamo un campione "numeroso", allora possiamo costruire un test approssimato asintotico per verificare $H_0 : \theta = \theta_0$ contro $H_1 : \theta = \theta_1$ che non richiede i calcoli binomiali. Infatti, in virtù del teorema centrale del limite, se n è grande, approssimativamente il numero totale di successi ha f.d.r. gaussiana di media $n\theta$ e varianza $n\theta(1-\theta)$. Quindi, se n è grande, possiamo trovare un k_2 che approssimativamente realizzi l'uguaglianza in (16) per ogni valore di α fissato; in tal caso, l'uguaglianza dell'Equazione (16) si riduce a

$$\Phi\Big(\frac{k_2 - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}\Big) = \alpha$$

che ha soluzione $k_2 = n\theta_0 - z_{1-\alpha}\sqrt{n\theta_0(1-\theta_0)}$. Esprimendo il tutto in termini della frequenza relativa campionaria di successo $\hat{\theta}$, otteniamo in definitiva che se n è grande, allora:

a) per verificare $H_0 : \theta = \theta_0$ contro $H_1 : \theta = \theta_1$ con $\theta_1 < \theta_0$ usiamo la regione critica di livello approssimativamente α data da

$$\mathcal{G} = \Big\{\hat{\theta} \leq \theta_0 - z_{1-\alpha}\sqrt{\frac{\theta_0(1-\theta_0)}{n}}\Big\} \tag{17}$$

b) per verificare $H_0 : \theta = \theta_0$ contro $H_1 : \theta = \theta_1$ con $\theta_1 > \theta_0$ usiamo la regione critica di livello approssimativamente α data da

$$\mathcal{G} = \Big\{\hat{\theta} \geq \theta_0 + z_{1-\alpha}\sqrt{\frac{\theta_0(1-\theta_0)}{n}}\Big\} \tag{18}$$

c) per i problemi $H_0 : \theta = \theta_0$ versus $H_1 : \theta < \theta_0$ e $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$ useremo la regione critica (17);

d) per i problemi $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$ e $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ useremo la regione critica (18);

e) per il problema $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ useremo la regione critica

$$\mathcal{G} = \Big\{|\hat{\theta} - \theta_0| \geq z_{1-\frac{\alpha}{2}}\sqrt{\frac{\theta_0(1-\theta_0)}{n}}\Big\} \tag{19}$$

Per stabilire se il campione sia sufficientemente numeroso da poter procedere con i precedenti test asintotici, usiamo una delle solite regolette: almeno 30 osservazioni e deve verificarsi che $n\theta_0 > 5$ e $n(1-\theta_0) > 5$.

Inoltre, in ciascuno dei precedenti test asintotici, un'approssimazione gaussiana della funzione di potenza o della probabilità di errore di II tipo in un generico valore $\theta \in \Theta_1$ funziona se $n\theta > 5$ e $n\theta(1-\theta) > 5$.

Infine, ricordiamo che per n grande un intervallo di confidenza per θ asintotico bilatero con confidenza approssimativamente γ ha estremi $\hat{\theta} \pm z_{\frac{1+\gamma}{2}}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$. In questo caso possiamo controllare se sia raggiunta la numerosità necessaria perché l'approssimazione funzioni verificando se $n\hat{\theta} > 5$ e $n(1-\hat{\theta}) > 5$.

8 Verifica di ipotesi per dati gaussiani accoppiati

Supponiamo di essere interessati a confrontare due diverse misurazioni X e Y rilevate su ciascun individuo di una data popolazione. Potremmo essere interessati a confrontarne i valori medi oppure, se per esempio X e Y sono misurazioni di due caratteri diversi, potremmo essere interessati a stabilire se questi caratteri siano o no indipendenti. Gli n dati che raccogliamo sono accoppiati e il campione casuale che otteniamo è il campione casuale di dati bivariati (*paired sample*) $(X_1, Y_1), \dots, (X_n, Y_n)$.

In questa sezione prima confrontiamo le medie di X e Y, poi, descriviamo un test di indipendenza fra X e Y nel caso di dati accoppiati gaussiani. Nel capitolo dell'inferenza non parametrica, affronteremo il problema del confronto di dati accoppiati nell'ipotesi che nulla sappiamo della distribuzione congiunta di (X, Y).

Sia $(X_1, Y_1), \dots, (X_n, Y_n)$ un campione casuale (semplice) bidimensionale estratto dalla densità gaussiana bidimensionale $N\left(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho\right)$, cioè la comune densità è data da

$$(20) \quad f(x, y; \mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho) = \\ = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}, \\ \mu_X \in \mathbb{R},\ \mu_Y \in \mathbb{R},\ \sigma_X^2 > 0,\ \sigma_Y^2 > 0,\ \rho \in (-1, 1)$$

I parametri μ_X e σ_X^2 rappresentano media e varianza di X_1, μ_Y e σ_Y^2 sono media e varianza di Y_1 e ρ è il coefficiente di correlazione lineare

$$\rho = \rho(X_1, Y_1) = \frac{\operatorname{Cov}(X_1, Y_1)}{\sqrt{\operatorname{Var}(X_1)\operatorname{Var}(Y_1)}}$$

8.1 t-test per il confronto delle medie

Supponiamo tutti i parametri siano incogniti e sia $Z = X - Y$. Allora Z ha media $\mu_Z = \mu_X - \mu_Y$ e $Z_1, \dots, Z_n$ è un campione casuale estratto da una popolazione gaussiana con media e varianza entrambe incognite: $Z_1, \dots, Z_n\ i.i.d. \sim N(\mu_Z, \sigma_Z^2)$. I seguenti problemi di verifica di ipotesi:

1. $H_0 : \mu_X \le \mu_Y + \Delta$ contro $H_1 : \mu_X > \mu_Y + \Delta$
2. $H_0 : \mu_X \ge \mu_Y + \Delta$ contro $H_1 : \mu_X < \mu_Y + \Delta$
3. $H_0 : \mu_X = \mu_Y + \Delta$ contro $H_1 : \mu_X \ne \mu_Y + \Delta$

sono rispettivamente equivalenti a

1* $H_0 : \mu_Z \le \Delta$ contro $H_1 : \mu_Z > \Delta$

2* $H_0 : \mu_Z \ge \Delta$ contro $H_1 : \mu_Z < \Delta$

3* $H_0 : \mu_Z = \Delta$ contro $H_1 : \mu_Z \ne \Delta$.

Quindi, un possibile approccio per confrontare μ_X e μ_Y è di basarci sulle realizzazioni del campione delle differenze $Z_1, \dots, Z_n$ e di usare i t-test elaborati per verificare ipotesi sulla media di popolazione gaussiana con varianza incognita. Questi t-test ci forniscono delle regole per risolvere i tre problemi ipotetici (1-3), formulati in termini dei corrispondenti (1*-3*).

A titolo di esempio, una regione critica di ampiezza α per verificare $H_0 : \mu_X = \mu_Y + \Delta$ contro $H_1 : \mu_X \neq \mu_Y + \Delta$ è data da $\mathcal{G} = \{(z_1, \ldots, z_n) : \sqrt{n}|\bar{Z} - \Delta|/\sqrt{S_Z^2} \geq t_{n-1}(1 - \alpha/2\}$, dove S_Z^2 indica la varianza campionaria del campione $Z_1, \ldots, Z_n$.

Esempio 8.1 Un gruppo di sei pazienti si sono sottoposti a una cura dimagrante. All'inizio della cura pesavano chilogrammi 77, 87, 104, 98, 91, 78 . Dopo un mese i dati sono così variati: 75, 88, 97, 99, 86, 70.

1. Si può affermare a un livello $\alpha = 0.05$ che la cura sia stata efficace?
2. Se i depliant sulla cura promettevano un calo di almeno 2 Kg, la cura è stata efficace?
3. Determinate per quali valori di α si rifiuta l'ipotesi di perdere con la cura in media $2Kg$.

Per rispondere alle domande assumete l'ipotesi che il vettore aleatorio (X, Y), con X=peso prima della cura e Y=peso dopo la cura, sia gaussiano.

Soluzione Abbiamo 6 dati accoppiati gaussiani con medie, varianze e coefficiente di correlazione incogniti.

1. Impostiamo il problema $H_0 : \mu_X - \mu_Y \leq 0$ versus $H_1 : \mu_X - \mu_Y > 0$ e usiamo la sintesi del campione $Z_i = X_i - Y_i$, $i = 1, \ldots, 6$, la cui realizzazione campionaria è $2,\ -1,\ 7,\ -1,\ 5,\ 8$. Rifiutiamo H_0 che la cura non sia stata efficace se $\sqrt{6}(\bar{Z} - 0)/S_Z \geq t_{6-1}(1 - 0.05) \simeq 2.015$. Poiché $\sqrt{6}\bar{Z}/S_Z \simeq \sqrt{6} \times 3.33/\sqrt{15.4677} \simeq 2.074 > 2.015$, allora a livello $\alpha = 0.05$ accettiamo H_1 cioè concludiamo che la cura è stata efficace.

2. Per verificare se in media si perdono $2Kg$ al mese, impostiamo la procedura test per $H_0 : \mu_X - \mu_Y \leq 2$ versus $H_1 : \mu_X - \mu_Y > 2$ con regione critica $G = \{(z_1, \ldots, z_6) \in \mathbb{R}^6 : \sqrt{6}(\bar{z} - 2)/S_z > 2.015\}$. Con le nostre realizzazioni campionarie abbiamo $\sqrt{6}(\bar{z} - 2)/s_Z \simeq 0.82837$ che è minore di 2.015, quindi accettiamo H_0 cioè concludiamo che la cura non è stata così efficace come pubblicizzato.

3. Calcoliamo il p-value per l'ultimo problema di verifica di ipotesi: p–value$= 1 - F_{t_5}(0.82837) \simeq 0.2226$. Vi è una forte evidenza dei dati ad accettare H_0: rifiutiamo H_0 del punto 2. per ogni $\alpha \geq 22.26\%$. (Con le tavole distribuitevi, per n "piccolo" siamo in grado di determinare soltanto un intervallo in cui varia il p–value: poiché $0.75 = F_{t_5}(0.727) < F_{t_5}(0.82837) < F_{t_5}(1.476) = 0.90$, allora il p–value è compreso fra il 10% e il 25%.) ∎

8.2 t-test di indipendenza per dati gaussiani accoppiati

Se la coppia (X, Y) è congiuntamente gaussiana allora X, Y sono indipendenti se e solo se il coefficiente di correlazione lineare ρ è nullo. Infatti, la densità congiunta di X, Y è data dalla funzione $f(x, y; \mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ in (20), che fattorizza nel prodotto di due densità gaussiane $\mathcal{N}(\mu_X, \sigma_X^2)$ e $\mathcal{N}(\mu_Y, \sigma_Y^2)$ quando $\rho = 0$.

Inoltre, un valore di ρ positivo indica una concordanza fra i due caratteri X, Y di tipo lineare, mentre un valore negativo indica una discordanza, cioè una tendenza dei due caratteri ad associarsi in modo tale che all'aumentare dell'uno l'altro tende a diminuire. Segue che per verificare se dati accoppiati gaussiani siano indipendenti o concordanti, usiamo dei test di ipotesi sul parametro ρ. Le ipotesi che può essere interessante mettere a confronto sono le seguenti:

$H_0 : \rho = 0$ contro $H_1 : \rho \neq 0$ per escludere l'indipendenza

$H_0 : \rho = 0$ o $H_0 : \rho \leq 0$ contro $H_1 : \rho > 0$ per escludere la dipendenza negativa
$H_0 : \rho = 0$ o $H_0 : \rho \geq 0$ contro $H_1 : \rho < 0$ per escludere la dipendenza positiva.

Nella costruzione delle regioni critiche, qui andiamo speditamente partendo da una statistica test opportuna. Nel caso di un campione accoppiato gaussiano $(X_1, Y_1), \ldots, (X_n, Y_n)$ con parametri tutti incogniti, uno stimatore per ρ è dato dal *coefficiente di correlazione campionario o empirico*:

$$R = \frac{\sum_{j=1}^n (X_j - \bar{X})(Y_j - \bar{Y})}{\sqrt{\sum_{j=1}^n (X_j - \bar{X})^2 \sum_{j=1}^n (Y_j - \bar{Y})^2}}$$

La statistica R è in modulo minore o uguale di 1: $-1 \leq R \leq 1$ e per quanto concerne la sua distribuzione vale il seguente risultato:

Teorema 8.2 *Sia* $(X_1, Y_1), \ldots, (X_n, Y_n)$ *i.i.d.* $\sim \mathcal{N}$ *e* $\rho = 0$. *Allora*

$$\frac{R}{\sqrt{1 - R^2}} \sqrt{n-2} \sim t_{n-2}, \quad n \geq 3.$$

Dimostrazione Si può vedere per esempio la Sezione 7.6 pagine 325-332 in Rohatgi e Saleh (1999). Ma ovviamente non è in programma. ▮

Poiché $\frac{x}{\sqrt{1-x^2}}\sqrt{n-2}$ è funzione crescente di x sull'intervallo $(-1, 1)$, allora usiamo $\dfrac{R\sqrt{n-2}}{\sqrt{1-R^2}}$ come statistica test per i diversi problemi di verifica di ipotesi su ρ sopra descritti. Ovviamente, a seconda delle ipotesi nulla e alternativa specificate, cambierà il quantile con cui confrontare il valore della statistica test. Più precisamente:

$$\mathcal{G}_1 = \left\{ \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \geq t_{n-2}(1-\alpha) \right\} \text{ è una regione critica per } H_0 : \rho = 0 \text{ o } H_0 : \rho \leq 0 \text{ contro } H_1 : \rho > 0$$

$$\mathcal{G}_2 = \left\{ \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} < -t_{n-2}(1-\alpha) \right\} \text{ è una regione critica per } H_0 : \rho = 0 \text{ o } H_0 : \rho \geq 0 \text{ contro } H_1 : \rho < 0$$

$$\mathcal{G}_3 = \left\{ \left| \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \right| > t_{n-2}(1 - \frac{\alpha}{2}) \right\} \text{ è una regione critica per } H_0 : \rho = 0 \text{ contro } H_1 : \rho \neq 0$$

Osservazione 8.3 Osservate che per calcolare R è sufficiente conoscere i valori delle seguenti statistiche: $\bar{x}$, $\bar{y}$, $\sum_{j=1}^n x_j^2$, $\sum_{j=1}^n y_j^2$ e $\sum_{j=1}^n x_j y_j$ dal momento che

$$\sum_{j=1}^n (X_j - \bar{X})(Y_j - \bar{Y}) = \sum_{j=1}^n X_j Y_j - n\bar{X}\bar{Y}, \ \sum_{j=1}^n (X_j - \bar{X})^2 = \sum_{j=1}^n X_j^2 - n\bar{X}^2 \text{ e } \sum_{j=1}^n (Y_j - \bar{Y})^2 = \sum_{j=1}^n Y_j^2 - n\bar{Y}^2.$$

9 Verifica di ipotesi per confrontare dati gaussiani indipendenti

Affrontiamo ora il problema del confronto di dati gaussiani indipendenti. Il confronto di dati indipendenti sotto ipotesi di completa ignoranza circa le distribuzioni sottostanti i due campioni sarà oggetto dei test di omogeneità non parametrici.

Abbiamo estratto due campioni di dati indipendenti $X_1, \ldots, X_m$ *i.i.d.* $\sim \mathcal{N}(\mu_X, \sigma_X^2)$ e $Y_1, \ldots, Y_n$ *i.i.d.* $\sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. I due campioni di dati possono contenere un numero diverso di osservazioni, cioè m può essere diverso da n.

Procediamo prima a confrontare le medie e poi le varianze.

9.1 Confronto delle medie

Caso di varianze note: z-test. Supponiamo le varianze siano entrambe note. La statistica da usare per confrontare le medie è la differenza delle medie campionarie, $\bar{X} - \bar{Y}$ che è un "buon" stimatore di $\mu_X - \mu_Y$. Inoltre, in quanto differenza di va gaussiane indipendenti, $\bar{X} - \bar{Y}$ è anche essa va gaussiana con media la differenza delle medie $\mu_X - \mu_Y$ e varianza la somma delle varianze $\sigma_X^2/m + \sigma_Y^2/n$, per cui

$$\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \sim \mathcal{N}(0, 1)$$

Segue che

$$\mathcal{G}_1 = \left\{ \frac{(\bar{X} - \bar{Y}) - \Delta}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} > z_{1-\alpha} \right\}$$

è una regione critica per $H_0 : \mu_X - \mu_Y = \Delta$ o $H_0 : \mu_X - \mu_Y \leq \Delta$ contro $H_1 : \mu_X - \mu_Y > \Delta$,

$$\mathcal{G}_2 = \left\{ \frac{(\bar{X} - \bar{Y}) - \Delta}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} < -z_{1-\alpha} \right\}$$

è una regione critica per $H_0 : \mu_X - \mu_Y = \Delta$ o $H_0 : \mu_X - \mu_Y \geq \Delta$ contro $H_1 : \mu_X - \mu_Y < \Delta$ e

$$\mathcal{G}_3 = \left\{ \frac{\left|\bar{X} - \bar{Y} - \Delta\right|}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} > z_{1-\frac{\alpha}{2}} \right\}$$

è una regione critica per $H_0 : \mu_X - \mu_Y = \Delta$ contro $H_1 : \mu_X - \mu_Y \neq \Delta$.

Caso di varianze uguali ma incognite: t-test. Supponiamo ora che i quattro parametri $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$ siano tutti incogniti e le varianze siano uguali, cioè

$$\sigma_X^2 = \sigma_Y^2 = \sigma^2.$$

Come punto di partenza, stimiamo il comune valore della varianza σ^2 mediante una combinazione lineare delle due varianze campionarie S_X^2 e S_Y^2, con pesi che dipendono dalla numerosità dei due campioni m, n. Più precisamente, stimiamo σ^2 con la cosiddetta *varianza pooled* S_p^2 data da

$$S_p^2 = \frac{S_X^2(m-1) + S_Y^2(n-1)}{m+n-2}$$

La varianza pooled è uno stimatore di σ^2 non distorto perché combinazione lineare convessa di stimatori non distorti di σ^2. Inoltre, siccome S_X^2 e S_Y^2 sono indipendenti con $S_X^2(m-1)/\sigma^2 \sim$

χ^2_{m-1} e $S^2_Y(n-1)/\sigma^2 \sim \chi^2_{n-1}$ e $S^2_p(m+n-2)/\sigma^2 = (S^2_X(m-1)/\sigma^2 + S^2_Y(n-1)/\sigma^2)$ allora segue che

$$\frac{S^2_p(m+n-2)}{\sigma^2} \sim \chi^2_{m+n-2}.$$

Infine, in questo caso, la statistica test per verificare $H_0 : \mu_X - \mu_Y = \Delta$ (o $H_0 : \mu_X - \mu_Y \geq \Delta$ oppure $H_0 : \mu_X - \mu_Y \leq \Delta$) è

$$\frac{\bar{X} - \bar{Y} - \Delta}{\sqrt{S^2_p\left(\frac{1}{m} + \frac{1}{n}\right)}}$$

Se Δ è il vero valore di $\mu_X - \mu_Y$, allora

$$\frac{\bar{X} - \bar{Y} - \Delta}{\sqrt{S^2_p\left(\frac{1}{m} + \frac{1}{n}\right)}} \sim t_{m+n-2}$$

in quanto essa può essere vista come rapporto fra una va gaussiana e la radice quadrata di una va chiquadrato, indipendenti.

Vi lascio come esercizio la derivazione delle regioni critiche a seconda del problema di confronto fra medie specificato.

Caso di varianze incognite e diverse. Il problema di determinare un test di livello esattamente pari ad α per decidere se due popolazioni gaussiane indipendenti abbiano la stessa media è ancora aperto per il caso di varianze incognite e diverse e campioni piccoli. Questo problema è noto come *Problema di Behrens-Fisher* e nessuna delle strategie elaborate per risolverlo è ottimale.

Se invece abbiamo grandi campioni, una soluzione approssimata esiste e si basa sul fatto che la va $\dfrac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{S^2_X}{m} + \frac{S^2_Y}{n}}}$ per $m, n \to \infty$ ha f.d.r. gaussiana standard Φ. Quindi, a titolo d'esempio, per verificare $H_0 : \mu_X - \mu_Y \leq \Delta$ contro $H_1 : \mu_X - \mu_Y > \Delta$, usiamo la regione critica

$$\mathcal{G} = \left\{ (x_1, \ldots, x_n) : \frac{(\bar{X} - \bar{Y}) - \Delta}{\sqrt{\frac{S^2_X}{m} + \frac{S^2_Y}{n}}} > z_{1-\alpha} \right\}$$

9.2 F-test per confrontare le varianze

Per confrontare le varianze di due campioni gaussiani indipendenti, $X_1, \ldots, X_m$ *i.i.d.* $\sim \mathcal{N}(\mu_X, \sigma^2_X)$ e $Y_1, \ldots, Y_n$ *i.i.d.* $\sim \mathcal{N}(\mu_Y, \sigma^2_Y)$, abbiamo bisogno di introdurre una nuova distribuzione di probabilità.

Definizione 9.1 Siano U, W due variabili aleatorie indipendenti e chiquadrato con gradi di libertà rispettivamente, a, b: $U \sim \chi^2_a, W \sim \chi^2_b$. Allora $V = \dfrac{U/a}{W/b}$ è una variabile aleatoria continua e la sua densità si chiama *F di Fisher con a gradi di libertà al numeratore e b al denominatore*. Indichiamo questa densità con il simbolo $F_{a,b}$ e il suo quantile di ordine α con il simbolo $F_{a,b}(\alpha)$.

I quantili della f.d.r. di Fisher $F_{a,b}(\alpha)$ sono tabulati tipicamente per valori di $\alpha \geq 0.5$; mentre, otteniamo facilmente quelli di ordine $\alpha < 0.5$, osservando che, se $V \sim F_{a,b}$, allora $1/V \sim F_{b,a}$ e quindi $F_{a,b}(\alpha)$ e $F_{b,a}(1-\alpha)$ sono legati dalla relazione

$$F_{a,b}(\alpha) = \frac{1}{F_{b,a}(1-\alpha)} \qquad \forall \alpha \in (0,1) \tag{21}$$

Provate a verificare (21).

Caso di medie incognite. Cerchiamo ora una regione critica per il problema di ipotesi

$$H_0 : \sigma_X^2 = \sigma_Y^2 \text{ contro } H_1 : \sigma_X^2 \neq \sigma_Y^2 \tag{22}$$

nel caso che entrambe le medie μ_X e μ_Y siano incognite.

Siccome S_X^2 è stimatore non distorto di σ_X^2 e S_Y^2 è stimatore non distorto per σ_Y^2, la statistica test da cui partiamo è il rapporto S_X^2/S_Y^2. Rifiutiamo H_0 se S_X^2/S_Y^2 è troppo grande o troppo piccola. In altri termini, la forma della regione critica è

$$\mathcal{G} = \Big\{(x_1, \dots, x_m, y_1, \dots, y_n) : \ \frac{s_X^2}{s_Y^2} \leq k_1 \text{ oppure } \frac{s_X^2}{s_Y^2} \geq k_2\Big\}$$

Per determinare k_1, k_2, ricordiamoci che S_X^2, S_Y^2 sono indipendenti con $S_X^2(m-1)/\sigma_X^2 \sim \chi^2_{m-1}$ e $S_Y^2(n-1)/\sigma_Y^2 \sim \chi^2_{n-1}$. Pertanto, sotto H_0, cioè se $\sigma_X^2 = \sigma_Y^2$, la statistica S_X^2/S_Y^2 può essere letta come rapporto fra due chiquadrato indipendenti ciascuna divisa per i propri gradi di libertà. Segue che se H_0 è vera allora $S_X^2/S_Y^2 \sim F_{m-1,n-1}$.

Una regione critica $\mathcal{G}$ di ampiezza α per il problema (22) con medie incognite è

$$\mathcal{G} = \Big\{(x_1, \dots, x_m, y_1, \dots, y_n) : \ \frac{s_X^2}{s_Y^2} \leq F_{m-1,n-1}(\frac{\alpha}{2}) \text{ oppure } \frac{s_X^2}{s_Y^2} \geq F_{m-1,n-1}(1 - \frac{\alpha}{2})\Big\}$$

I test unilateri sul confronto di varianze nel caso di medie incognite si costruiscono allo stesso modo e sono di seguito riassunti:

$\mathbf{H_0}$	$\mathbf{H_1}$	**Si rifiuta $\mathbf{H_0}$ se**	**p-value**
$\sigma_X^2 \leq \sigma_Y^2$	$\sigma_X^2 > \sigma_Y^2$	$\frac{s_X^2}{s_Y^2} \geq F_{m-1,n-1}(1-\alpha)$	$1 - P\left(F_{m-1,n-1} \leq \frac{s_X^2}{s_Y^2}\right)$
$\sigma_X^2 \geq \sigma_Y^2$	$\sigma_X^2 < \sigma_Y^2$	$\frac{s_X^2}{s_Y^2} \leq F_{m-1,n-1}(\alpha)$	$P\left(F_{m-1,n-1} \leq \frac{s_X^2}{s_Y^2}\right)$
$\sigma_X^2 = \sigma_Y^2$	$\sigma_X^2 \neq \sigma_Y^2$	$s_X^2/s_Y^2 \geq F_{m-1,n-1}(1-\alpha/2)$ oppure $s_X^2/s_Y^2 \leq F_{m-1,n-1}(\alpha/2)$	$2\min\{p_1,p_2\}$ dove $p_1 = P\big(F_{m-1,n-1} \leq s_X^2/s_Y^2\big)$ e $p_2 = 1 - p_1$

Caso di medie note. Nel caso di medie entrambe note partiremo dagli stimatori $S^2_{0,X} = \sum_{j=1}^m (X_j - \mu_X)^2/m$ e $S^2_{0,Y} = \sum_{j=1}^n (Y_j - \mu_Y)^2/n$ e useremo la statistica test S^2_{0X}/S^2_{0Y} che sotto l'ipotesi $H_0 : \sigma^2_X = \sigma^2_Y$ ha distribuzione di F Fisher con m, n gradi di libertà[3]. Quindi le forme delle varie regioni critiche unilatere e bilatere saranno le stesse, ma, dove prima trovavamo S^2_X/S^2_Y, ora avremo S^2_{0X}/S^2_{0Y} e, dove prima c'erano i quantili della f.d.r. $F_{m-1,n-1}$, ora avremo i quantili della f.d.r. $F_{m,n}$.

Infine, i test di confronto di varianze di popolazioni gaussiane indipendenti sono detti *F-test*, dal nome della distribuzione campionaria coinvolta.

Osservazione 9.2 Dati due campioni gaussiani indipendenti, potremmo voler confrontare sia le medie che le varianze. Per esempio vorremmo poter stabilire se entrambi sono gaussiani con stessa media e varianza, cioè $H_0 : \sigma^2_X = \sigma^2_Y$ e $\mu_X = \mu_Y$, contro l'alternativa H_1 che siano gaussiani ma di parametri diversi. Una possibile strategia è di effettuare in sequenza un F-test sulle varianze seguito da un t-test sulle medie. Più precisamente:

Primo passo A un primo passo eseguiamo un F-test di livello prefissato α_1 per verificare $H_0 : \sigma^2_X = \sigma^2_Y$ contro $H_1 : \sigma^2_X \neq \sigma^2_Y$. Se accettiamo H_0, passiamo al secondo passo.

Secondo passo Eseguiamo un t-test sulle medie, quello per il caso di varianze incognite ma uguali, per verificare $H_0 : \mu_X = \mu_Y$ contro $H_1 : \mu_X \neq \mu_Y$. Se accettiamo a un livello prefissato α_2, allora concludiamo che non possiamo escludere che i due campioni siano regolati da un modello gaussiano con stesse medie e varianze.

Si può dimostrare che il test a due passi sopra descritto ha livello di significatività pari a $1 - (1 - \alpha_1)(1 - \alpha_2)$.

[3] Perché? Provate a verificarlo