Test D'ipotesi e Stime Intervallari

Vers. 1.0.6

Gianluca Mastrantonio

gianluca.mastrantonio@polito.it

Outline

Introduzione I

- Uno stimatore $W(\mathbf{X})$ non distorto, raggiunge il limite di Cramer-Rao se e solo se

$$a(\theta)(W(\mathbf{x}) - \theta) = \frac{\partial}{\partial\theta} logL(\theta|\mathbf{x})$$

  per qualche funzione $a(\theta)$
- se $W(\mathbf{X})$ è un “miglior stimatore non distorto” (best unbiased estimator) per θ, cioè se è non distorto e raggiunge Cramer-Rao, allora è unico.

Introduzione I

Stimare un parametro delle volte non è sufficiente, perchè più che al valore assunto dallo stimatore, potremmo essere interessati a dire qualcosa sul parametro vero (non noto). Facciamo qualche esempio (tra i dataset R, vedi **"Chunk introduzione"**):

- 10 pazienti testano due differenti soporiferi e viene misurata la differenza (soporifero 2 - soporifero 1) in ore di sonno (rispetto al gruppo di controllo). Assumiamo che la dirrefenza di effetto tra i due soporiferi sia $N(\mu, \sigma^2)$ e come stima di μ usiamo la media campionaria $\bar{x} = -1.58$ possiamo dire che soporifero 1 è migliore?
- Nell'esperimento di Michelson e Morley hanno preso diverse misurazioni della velocità della luce, anche in questo caso assumiamo che i dati provengano da una $N(\mu, \sigma^2)$, e stimiamo μ con la media campionaria $\bar{x} = 852.4$ (hanno sottratto 299000), sappiamo che il valore vero della velocità della luce è 792.458. Possimao dure che le misurazione avevano degli errori di cui non hanno tenuto conto?

Introduzione II

- Nello stesso esempio di prima, le 100 misurazioni sono state ottenure in 5 esperimenti diversi. Abbiamo che le medie campionarie nei vari esperimenti sono 909.0, 856.0, 845.0, 820.5, 831.5, mentre le varianze sono 11009.474, 3741.053, 6257.895, 3605.000, 2939.737. Possiamo affermare che gli esperimenti sono "uguali" o fatti nelle stesse condizioni, oppure i dati di ogni esperimento provengono da una normale diversa?
- Si sono studiati le intenzioni di voto sulla brexit facendo interviste telefoniche giornaliere, per un periodo di sei mesi, da gennaio 2016 a giugno 2016. Assumiamo che nella prima intervista fatta le intenzioni di voto provengano da una $B(n_1, p_1)$, mentre nell'ultima intervista da un $B(n_2, p_2)$. Le proporzioni osservate di leavers nei due periodi, che si possono usare come stime della probabilità di uan binomiale, sono $\hat{p_1} = 0.48$ e $\hat{p_1} = 0.38$. Possiamo affermare che queste sono diverse?

Introduzione III

Quindi nasce il problema di come verificare **ipotesi** sui parametri perchè anche se per esempio $p_1 = p_2$, le loro stime sono variabili aleatorie e quindi la probabilità che assumano lo stesso valore è molto bassa anche se l'ipotesi fosse vera. Potremmo utilizzare la loro distribuzione e vedere se sono "simili", ma generalmente queste dipendono dai parametri incogniti, e.g. $\bar{x} \sim N(\mu, \sigma^2/n)$, se X sono normali, e come vedete dipende dai aprametri incogniti.

Da un punto di vista più realista, c'è anche un'altro problema, che è difficile immaginare che due cose sia esattamente uguale, per esempio l'effetto dei due soporiferi non può essere esattamente lo stesso, ma ci sarà sempre una differenza, ma questo aspetto non lo toccheremo. L'idea alla base però è che vogliamo verificare se con i dati che abbiamo siamo in grado di distinguere le differenze tra i due o no.

La risposta degli statistici (o una delle risposte) a questo problema è la teoria dei test, e gli intervalli di confidenza.

Introduzione I

Prendiamo il caso più semplice, per esempio l'esperimento dei soporiferi, e assumiamo che i dati ottenuti con il primo soporiferi siano $(L_1, \ldots, L_{10})$ iid da una $N(\mu_1, \sigma^2/2)$, mentre quelli ottenuti con il secondo siano $(Z_1, \ldots, Z_{10})$ iid da una $N(\mu_2, \sigma^2/2)$, e assumiamo che σ^2 sia nota (per semplicità, ma poi vedremo il caso in cui non lo è). Vogliamo studiare la differenza di efficacia e visto che (L_j, Z_j) sono le misurazioni fatte sullp stesso soggetto, possiamo farne la differenza assumendo che L_j e Z_j siano indipendenti. Otteniamo

$$X_j = Z_j - L_j \sim N(\mu_2 - \mu_1, \sigma^2)$$

Indichiamo $\mu = \mu_2 - \mu_1$ e visto che i dati sono normali, lo stimatore di μ basato sulla massima verosimiglianza è la media campionaria (si dimostra che raggiunge Cramer Rao) e

$$\bar{x} \sim N(\mu, \sigma^2/n)$$

Introduzione II

Noi siamo interessati a dire qualcosa su μ visto che se $\mu > 0$ potremmo dire che le ore di sonno con soporifero 2 sono maggiori di soporifero 1, se $\mu < 0$ vale il viceversa, mentre se $\mu = 0$ i due medicinali sono identici (in media)

Il meglio che possiamo fare è ipotizzare qualcosa circa il valore di interesse, e vedere poi quanto è plausibile ciò che abbiamo osservato.

Diamo qualche definizione che ci sarà utile

Definizione - Ipotesi

Un'ipotesi (in ambito statistico) è una dichiarazione su un parametro della popolazione

Per esempio, riprendendo l'esempio precedente, io potrei fare l'ipotesi che $\mu = 0$. Lo scopo dei test d'ipotesi è di dare una regola di decisione basandosi sul campione osservato $\mathbf{x}$. Per esempio posso dire che la mia ipotesi è vera se $\bar{x} \in (-5, 5)$.

Definizione - H_0 e H_1

Due ipotesi complementari in un test d'ipotesi vengono chiamate **ipotesi nulla** (H_0) e **ipotesi alternativa** (H_1).

Riprendendo l'esempio di prima potrei assumere

$$H_0 : \mu = 0 \qquad H_1 : \mu > 0$$

I Test II

Attenzione: in un test si assume che le due ipotesi descrivano tutti i valori che il parametro può assumere, e quindi abbiamo che $\mu \in \Theta$, con $\Theta \equiv \mathbf{R}_0^+$. In una forma più generale, possiamo scrivere

$$H_0 : \theta \in \Theta_0 \qquad H_1 : \theta \in \Theta_0^c$$

I tre tipi classici di ipotesi sono

$\theta = c$ Ipotesi puntuale

$\theta < c$, oppure $\theta > c$ Ipotesi unilaterale

$\theta \neq c$ Ipotesi bilaterale

In statistica l'ipotesi nulla è l'ipotesi che rappresenta ciò che noi con il nostro esperimento vorremmo dimostrare essere falsa, e solo se abbiamo abbastanza "prove" che H_0 sia falsa (o abbastanza evidenze che H_0 sia falsa), allora rifiutiamo H_0.

Esempio - Test d'ipotesi

10 pazienti testano due differenti soporiferi e viene misurata la differenza (soporifero 2 - soporifero 1) in ore di sonno (rispetto al gruppo di controllo). Assumiamo che la dirrefenza di effetto tra i due soporiferi sia $N(\mu, \sigma^2)$. Le due ipotesi sono

$$H_0 : \mu = 0 \qquad H_1 : \mu \neq 0$$

Notate come l'ipotesi nulla è quella che vogliamo provare falsa perchè vorremmo dimostrare che c'è differenza tra i soporiferi.

Vediamo come costruire e valutare un test.

Costruzione del test I

Definizione - Test d'ipotesi e Regione di accetazione/rifiuto

Un test d'ipotesi è una regola che decide

- per quali campioni $\mathbf{X} \in \mathcal{X}$ si accetta H_0 come vera;
- per quali campioni $\mathbf{X} \in \mathcal{X}$ si rifiuta H_0, e quindi si assume H_1 vera;

La porzione dello spazio campionario per cui H_0 viene rifiutata si chiama **regione di rifiuto** ($R(\mathbf{X})$), il suo complementare è la **regione di accettazione** (($A(\mathbf{X})$))

Sebbene la regione di rifiuto sia definita generalmente in funzione di $\mathbf{X}$, lo spazio campionario di $\mathbf{X}$ può essere molto complicato e si preferisce usare una statistica, e scriviamo

$$R(\mathbf{X}) = \{\mathbf{X}; T(\mathbf{X}) \in C\}$$

dove $T(\mathbf{X})$ è una qualsiasi statistica.

Costruzione del test II

$\mathbf{X}$ e $T(\mathbf{X})$ sono variabili aleatorie, e quindi l'appartenere alla regione di rifiuto o no è anch'essa una variabile aleatoria. Abbiamo allora 4 possibili casi:

	Accetto H_0	Accetto H_1
H_0 vera	✓	Errore di prima specie
H_1 vera	Errore di seconda specie	✓

Introduciamo la **funzione potenza**

Costruzione del test III

Definizione - Funzione potenza

la funzione potenza

$$\beta(\theta) = P(\mathbf{X} \in R; \theta)$$

è la probabilità di appartenere alla regione di rifiuto per un dato θ, che è uguale a

$$\beta(\theta) = \begin{cases} P(\text{Errore di prima specie}; \theta) & \text{se } \theta \in \Theta_0 \\ 1 - P(\text{Errore di seconda specie}; \theta) & \text{se } \theta \in \Theta_0^c \end{cases}$$

Facciamo un esempio molto semplice

Esercizio - Brexit

Riprendiamo l'esempio della brexit, assumendo che ogni intenzione di voto registrata, nel primo giorni in cui l'esperimento è stato fatto, provenga da una $B(p)$, dope p è la probabilità di votare leave, e verifichiamo le seguente ipotesi

$$H_0 : p = 0.5 \qquad H_1 : p \neq 0.5$$

e valutare la funzione potenza

Soluzione:

L'esercizio è libero e non ci dice che statistica usare. La cosa più ragionevole è usare

$$T(\mathbf{X}) = \sum_{i=1}^{n} X_i \sim \mathsf{Bin}(n, p)$$

dove $n = 4772$,poichè sappiamo la sua distribuzione campionaria. Se definiamo la regione di rifiuto come

$$R(\mathbf{X}) = \{\mathbf{X} : T(\mathbf{X}) \neq 0.5 * 4772\}$$

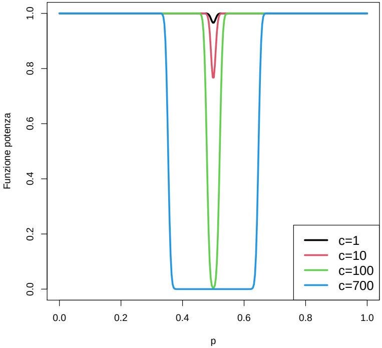
COn questa regione di rifiuto accettiamo H_0 solo se esattamente 2386 persone dichiarano leave, se anche il numero differisce di uno, non accettiamo. naturalmente questa non è una buona regione di rifiuto, e preferiamo qualcosa del tipo

$$R(\mathbf{X}) = \{\mathbf{X} : T(\mathbf{X}) \notin [2386 - c, 2386 + c]\}$$

con qualche c intero. Prendiamo per esempio $c = 10$ e visto che dal dataset (vedere **"Chunk funzione potenza"**) abbiamo $T(\mathbf{X}) = 2290$, con $c = 10$ rifiutiamo l'ipotesi nulla. Notate che per come è definita l'ipotesi alternativa, non sappiamo niente sul valore di p, ma solo che il test ci dice che non è 0.5.
Calcoliamo la funzione potenza

$$P(\mathbf{X} \in R; p) = P(\mathbf{X} \in R; p) = 1 - \sum_{i=2386-c}^{2386+c} \begin{pmatrix} 4772 \\ i \end{pmatrix} p^i (1-p)^{4772-i}$$

Costruzione del test VI



Nell'esempio precedente resta il problema di che valore c scegliere. Notate anche che per come è stato scritto il test, anche se rifiutiamo H_0 non sappiamo il valore di p, ma sappiamo solo che (probabilmente) non è 0.5.
Guardiamo la figura sopra, che rappresenta la funzione potenza per diversi valori di c. Vedete che un valore di c=1, la probabilità di rifiutare H_0, anche quando $p = 0.5$ è molto alta($\beta(0.5)$), più aumento c e più questa probabilità scende. Supponiamo adesso che il valore vero di p sia 0.4, con $c = 1$ la probabilità di rifiutare H_0 è alta (e questo è positivo per noi), più aumentiamo e più diminuisce, e con c=700 sarà difficile accettare H_1 anche se H_1 è vera.

In generale se diminuiamo la probabilità di errore di prima specie, aumentiamo quello di seconda, e non si possono minimizzare entrambi.

Ritorniamo quindi al problema originale. Come scegliere c? Quello che si fa è di decidere una probabilità di errore di prima specie massima che si è disposti a fare. Per esempio

Costruzione del test VIII

possiamo assumere che questa probabilità non sia maggiore di $\alpha = 0.05$ (valore standard), o $\alpha = 0.1$, e ricavare c usando quest'informazione.

Definizione - Livello/dimensione

Un test si dice di livello/dimensione α se $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$.

α viene anche chiamato **livello di significatività** del test.

Un test a livello alpha si può trovare solo se la statistica usate nel test è assolutamente continua , mentre se è discreta al più possiamo avere $\sup_{\theta \in \Theta_0} \beta(\theta) < \alpha$. Nell'esempio precedente il valore di c ottimale è 68.
Anche se non troppo preciso, in entrambi casi si diche che il test è a livello α.
Alcuni autori introducono delle differenze tra livello e dimensione del test, dove il livello ha $\leq$ e la dimensione $=$.

Costruzione del test IX

Se calcoliamo la funzione potenza per valori in Θ_0^c, abbiamo che $\beta(0.3) = 1$, $\beta(0.4) = 1$, $\beta(0.45) = 0.99$, $\beta(0.49) = 0.27$, questo significa che se p vero è molto lontano da 0.5, abbiamo una probabilità alta di rifiutare H_0, mentre se è molto vicino, allora è difficile rifiutare H_0 anche se non è vera. In generala il valore di $\beta()$ per $\theta \in \Theta_0^c$ aumenta con n, ma è un po' difficile vaderlo dall'esempio precedente, perchè se aumentiamo n dobbiamo cambiare anche c per avere un test a livello α.

Costruzione del test X

Definizione - Quantità pivotale

Una variabile aleatoria $Q(\mathbf{X}, \theta)$ è una quantità pivotale se la sua distribuzione non dipendende da θ

Delle volte la quantità pivotale viene indicata anche come statistica pivotale, ma quersto non è corresttissimo perchè Q non dipende solo dal campione ma anche dal parametro. Il caso classico è la statistica

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

nel caso il cui il campione sia iid da normali, mentre una quantità pivoltale è

$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0, 1)$$

Le quantità pivotali sono molto utili nei test (e negli intervalli di confidenza che faremo in futuro), e conviene quasi sempre calcolare la regione di rifiuto usando le quantità pivotali, perchè queste sono in general valide per ogni possibile valore del parametro.

Esercizio - Brexit - Quantità pivotali

Riprendiamo l'esempio precedente, e ricordiamo che la regione di rifiuto è

$$R(\mathbf{X}) = \{\mathbf{X} : T(\mathbf{X}) \in C\}$$

usata una quantità pivotale per determianre C. E indichiamo in maniera più generale le due ipotesi

$$H_0 : p = p_0 \qquad H_1 : p \neq p_0$$

Soluzione:

Una quantità pivotale per il caso binomiale non è facile trovarla, ma possiamo usare il TLC visto che

$$\bar{X} \approx N\left(p, \frac{p(1-p)}{n}\right)$$

e quindi $\frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$

Costruzione del test XII

In generale si sostituisce a $\frac{p(1-p)}{n}$ il valore $\frac{\bar{X}(1-\bar{X})}{n}$, ottenendo

$$\bar{X} \approx N\left(p, \frac{\bar{X}(1-\bar{X})}{n}\right)$$

che ci da, sempre approssimativamente

$$\frac{\bar{X}-p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \approx N(0,1)$$

Questo è valido perchè

$$\sqrt{\bar{X}(1-\bar{X})} \xrightarrow{P} \sqrt{p(1-p)}$$

allora è anche vero (per Slutsky) che

$$\frac{\bar{X}-p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \xrightarrow{d} Z \sim N(0,1)$$

Costruzione del test XIII

Il pivot è Z, che è basata su $\bar{X}$ e quindi la regione di rifiuto la possiamo scrivere come

$$R(\mathbf{X}) = \{\mathbf{X} : \bar{X} \in C\}$$

dove un set C ragionevole è un set che non contiene il valore 0.5 (H_0), quindi assumiamo qualcosa del tipo

$$R(\mathbf{X}) = \{\mathbf{X} : \bar{X} < c_1 \cup \bar{X} > c_2\}$$

Abbiamo detto che decidiamo (c_1, c_2) fissando la probabilità dell'errore di prima specie (o significatività del test) α: $P(\bar{X} < c_1) + P(\bar{X} > c_2; \theta) = \alpha$.
Attenzione siccome stiamo calcolando la probabilità dell'errore di prima specie, assumiamo che H_0 sia vera.
Usiamo il pivot

$$P(\bar{X} < c_1; \theta) + P(\bar{X} > c_2; \theta) = P\left(\frac{\bar{X} - p_0}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} < c_1^*; \theta\right) + P\left(\frac{\bar{X} - p_0}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} > c_2^*; \theta\right) =$$

Costruzione del test XIV

$$P(Z < c_1^*) + P(Z > c_2^*) = \alpha$$

Attenzione

$$\frac{\bar{X} - p_0}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}}$$

è una normale standard solo se H_0 è vera.
Visto che la normale è simmetrica e che il valore $Z = 0$ corrisponde a $\bar{X} = p$, e allora ragionevole splittare α in due parti uguali per le due componenti della prob

$$P(Z < c_1^*) = P(Z > c_2^*) = \alpha/2$$

allora $c_2^* = z_{\alpha/2}$ e $c_1^* = -z_{\alpha/2}$, dove
$P(Z < differentiationundertheintegralsign - z_{\alpha/2}) = \alpha/2.$
Quindi abbiamo che, intermine di statistia $\bar{X}$ la regione di rifiuto è

$$R(\mathbf{X}) = \{\mathbf{X} \in \mathcal{X}; \bar{X} < p_0 - z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \cup \bar{X} > p_0 + z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\}$$

$\square$

Nella maggior parte dei casi è più facile scrivere la regione di accettazione che nel caso sopra è

$$A(\mathbf{X}) = \{\mathbf{X} \in \mathcal{X}; p_0 - z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq \bar{X} \leq p_0 + z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\}$$

oppure

$$A(\mathbf{X}) = \{\mathbf{X} \in \mathcal{X}; z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq \bar{X} - p_o \leq z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\}$$

o

$$A(\mathbf{X}) = \{\mathbf{X} \in \mathcal{X}; |\bar{X} - p_0| \leq z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\}$$

Queste regioni valgono per qualsiasi valore di p_0 e qualsiasi valore α o n.

Costruzione del test XVI

Per adesso abbiamo visto solo ipotesi nulle puntuali. Vediamo come si trattano ipotesi unilaterale

Esercizio - Michelson e Morley

Riprendiamo l'esperimento di Michelson e Morley, e assumiamo che i dati siano iid da una normale $N(\mu, \sigma^2)$. Per semplicità assumiamo che σ^2 sia nota (il caso del test con due parametri lo vedremo in seguito).
L'idea è valutare se nell'esperimento ci fossero stati errori sistematici o se sono riusciti a calcolare la velocità della luce.

- (1) ci chiediamo se
$$H_0 : \mu = \mu_0 \qquad H_1 : \mu \neq \mu_0$$
con $\mu_0 = 792.458$ (velocità della luce meno 299000).
- (2) successivamente vogliamo verificare se è stata sottostimata o sovrastimata
$$H_0 : \mu \leq \mu_0 \qquad H_1 : \mu > \mu_0$$
Come statistica usiamo la media campionaria e probabilità di errore di prima specie α

Soluzione:

(1) Per il primo punto i calcoli sono uguali al caso Brexit, ma adesso non sono approssimati:

$$A(\mathbf{X}) = \left\{ \mathbf{X} \in \mathcal{X}; |\bar{X} - \mu_0| \leq z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \right\}$$

Abbiamo che $\bar{X} = 852.4$, $\sigma^2 = 6242.66$ (ipotizzo coem valore vero quello campionario), $n = 100$ e $z_{\alpha/2} = 1.960$, ci porta a rifiutare H_0. Quindi l'esperimento aveva degli errori sistematici e non stavano veramente misurando la velocità della luce.

(2) Per il secondo punto dobbiamo trovare la regione di accettazione/rifiuto in modo tale che la probabilità di errore di prima specie è al più α per ogni $\mu \in \Theta_0$. Come sempre dobbiamo assumere che H_0 sia vera (per calcolare α) e indichiamo con $\mu^* \in \Theta_0$ il valore vero (non noto). Una possibile regione di rifiuto potrebbe essere

$$R(\mathbf{X}) = \{\mathbf{X} \in \mathcal{X}; \bar{X} > \mu_0 + c\}$$

Costruzione del test XIX

cioè rifiutiamo H_0 se ci allontaniamo dal suo limite destro. Calcoliamo la probabilità di rifiutare quando $\mu \in \Theta_0$

$$P\left(\bar{X} > \mu_0 + c; \mu\right)$$

e vogliamo che il sup di questa funzione deve essere α. Abbiamo

$$P\left(\bar{X} > \mu_0 + c; \mu\right) = P\left(\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} > \frac{\mu_0 - \mu}{\sqrt{\frac{\sigma^2}{n}}} + c^*; \mu\right) = P\left(Z > \frac{\mu_0 - \mu}{\sqrt{\frac{\sigma^2}{n}}} + c^*; \mu\right)$$

che raggiunge il massimo quando $\mu = \mu_0$. Quindi la probabilit'a di cadere nella regione di rifiuto cambia, in base al valore di μ, ma il suo massimo si raggiunge a $\mu = \mu_0$, allora, se facciamo i calcoli, come se μ fosse uguale a μ_0 sappiamo che per qualsiasi altro valore, la probabilità di errore di prima specie può solo che essere minore. Quindi, nel caso peggiore abbiamo che

$$P\left(Z > c^*; \mu\right) = \alpha$$

da cui deriva che $c^* = z_\alpha = \frac{c}{\sqrt{\frac{\sigma^2}{n}}}$ e la regione di rifiuto è allora

$$R(\mathbf{X}) = \left\{ \mathbf{X} \in \mathcal{X}; \bar{X} > \mu_0 + z_\alpha \sqrt{\frac{\sigma^2}{n}} \right\} \equiv \left\{ \mathbf{X} \in \mathcal{X}; \bar{X} - \mu_0 > z_\alpha \sqrt{\frac{\sigma^2}{n}} \right\}$$

Notate come il caso puntuale e unilaterale producono regioni simili. Anche in questo caso rifiutiamo H_0 e quindi hanno sovrastimato la velocità della luce.

$\square$

Come esercizio, vedete cosa succede se provate a testare l'ipotesi

$$H_0 : \mu = \mu_0 \qquad H_1 : \mu < \mu_0$$

cioè escludete dai possibili risultati valori maggiori della velocità della luce. In questo caso non potete rifiutare H_0 anche se il valore è chiaramente diverso, perchè non ammettete tra i possibili valori $\mu > \mu_0$.
Nell'esempio precedente abbiamo implicitamente calcolato la funzione potenza

$$\beta(\mu) = P\left(Z > \frac{\mu_0 - \mu}{\sqrt{\frac{\sigma^2}{n}}} + z_\alpha; \mu\right)$$

e come questa è uguale a α per $\mu = \mu_0$ e è crescente in funzione di μ, quindi più μ è elevato e più facile è rifiutare H_0. Notate anche come, a parità di $\mu \in \Theta_0^c$, aumentando la numerosità campionaria (avendo quindi più informazioni), abbiamo una probabilità più alta di rifiutare H_0 se H_1 è vera.

Rapporto di verosimiglianza I

Abbiamo visto che cerchiamo test che abbiamo significatività α. In caso avessimo due test diversi, entrambi a livello α, potremmo chiederci quale dei due è migliore.

Definizione - Test uniformemente più potente

Indichiamo come $\mathcal{C}$ come una classe di test del tipo

$$H_0 : \theta \in \Theta_0 \qquad H_1 : \theta \in \Theta_0^c$$

Un test nella classe $\mathcal{C}$ con funzione potenza $\beta(\theta)$ si dice uniformemente più potente (UMP) se $\beta(\theta) \geq \beta'(\theta)$ per ogni $\theta \in \Theta_0^c$ e test in $\mathcal{C}$ con funzione potenza $\beta'(\theta)$

Trovare test UMP per classi $\mathcal{C}$ molto generali è difficile e in generale ci si restringe nella classe dei test con livello α. Sotto questa classe e in caso specifici, si può trovare un test UMP.

Rapporto di verosimiglianza II

Questi sono due esempi di teoremi, che non dovete conoscere per l'esame, che ci danno delle condizioni in casi particolari

Lemma - lemma di Neyman-Pearson

Consideriamo il test d'ipotesi

$$H_0 : \theta = \theta_0 \qquad H_1 : \theta = \theta_1$$

allora se

$$\begin{cases} \mathbf{x} \in R(\mathbf{x}) & \text{if } f(\mathbf{x}|\theta_1) > kf(\mathbf{x}|\theta_0) \\ \mathbf{x} \in R^c(\mathbf{x}) & \text{if } f(\mathbf{x}|\theta_1) < kf(\mathbf{x}|\theta_0) \end{cases}$$

con $k \geq 0$ e dove $f()$ è la pdf o pmf, con

$$\alpha = P(\mathbf{X} \in R(\mathbf{x}); \theta_0)$$

allora ogni test che soddisfa queste condizioni è UMP nella classe dei test di livello α.

Rapporto di verosimiglianza III

Teorema - Carlin-Rubin

Consideriamo il test d'ipotesi

$$H_0 : \theta \leq \theta_0 \qquad H_1 : \theta > \theta_0$$

e che ci sia una statistica sufficiente T per θ e che la famiglia delle pmf o pdf di T $\{g(t|\theta) : \theta \in \Theta\}$ soddisfa la seguente condizione che $\frac{g(t|\theta_2)}{g(t|\theta_1)}$ è monotone in t per ogni $\theta_2 > \theta_1$ e t nel set $\{t : g(t|\theta_2) > 0 \text{ or } g(t|\theta_1) > 0\}$.
Allora, se un test rifiuta H_0 se e solo se $T > t_0$, con $P_{\theta_0}(T > t_0)$ è UMP di classe α.

Purtroppo non è sempre possibile trovare test UMP, o dimostrare che lo siano, ma i due teoremi appena visti ci danno casi particolari quando è possibile.
Un'altra proprietà importante di un test, che viene generalmente richiesta è la seguente

Definizione - Test non distorto

Un test con funzione potenza $\beta(\theta)$ si dice non distorto se $\beta(\theta) \geq \beta(\theta')$ per ogni $\theta \in \Theta_0^c$ e $\theta' \in \Theta_0$.

Rapporto di verosimiglianza IV

Che dice semplicemente che deve essere più probabile rifiutare H_0 quando non è vera piuttosto che quando è vera.
Abbiamo visto alcuni test, ma in alcune occasione non è facile trovare delle statistiche che possano essere usate come test. Introduciamo adesso un metodo che ci permette di definire regioni di rifiuto in maniera "automatica".
Definiamo il **rapporto di verosimiglianza** (LR)

$$\lambda(\mathbf{x}) = \frac{\sup\limits_{\theta \in \Theta_0} L(\theta|\mathbf{x})}{\sup\limits_{\theta \in \Theta} L(\theta|\mathbf{x})}$$

Attenzione che nel denominatore c'è Θ e non Θ_0^c.
Se esiste una statistica sufficiente abbiamo il seguente teorema

Teorema - statistiche sufficienti per il LR

Se $T(\mathbf{X})$ è una statistica sufficiente per θ e $\lambda^*(t)$ e $\lambda(\mathbf{x})$ sono i LR per $T(\mathbf{X})$ e $\mathbf{X}$, allora $\lambda^*(T(\mathbf{x})) = \lambda(\mathbf{x})$

Dimostrazione:

Ricordiamo che per il teorema di fattorizzazione abbiamo che

$$f(\mathbf{x}|\theta) = h(\mathbf{x})g(T(\mathbf{x})|\theta)$$

dove $g()$ è una funzione che dipende solo dalla statistica sufficiente, e tra le altre, può essere anceh la pmf pdf di $T(\mathbf{X})$, allora abbiamo che

$$\lambda(\mathbf{x}) = \frac{\sup\limits_{\theta\in\Theta_0} L(\theta|\mathbf{x})}{\sup\limits_{\theta\in\Theta} L(\theta|\mathbf{x})} = \frac{\sup\limits_{\theta\in\Theta_0} f(\mathbf{x}|\theta)}{\sup\limits_{\theta\in\Theta} f(\mathbf{x}|\theta)} = \frac{\sup\limits_{\theta\in\Theta_0} h(\mathbf{x})g(T(\mathbf{x})|\theta)}{\sup\limits_{\theta\in\Theta} h(\mathbf{x})g(T(\mathbf{x})|\theta)} = \frac{\sup\limits_{\theta\in\Theta_0} g(T(\mathbf{x})|\theta)}{\sup\limits_{\theta\in\Theta} g(T(\mathbf{x})|\theta)} = \lambda^*(T(\mathbf{x}))$$

$\square$

Rapporto di verosimiglianza VI

Possiamo usare il LR per costruire dei test

Definizione - test del rapporto di verosimiglianza

Il test del rapporto di verosimiglianza è un test con una regione di rifiuto

$$R(\mathbf{x}) = \{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$$

con $0 \leq c \leq 1$

e test costruiti in questo modo sono UMP se le ipotesi dei teoremi che abbiamo visto prima sono rispettate.

Facciamo qualche esempio di test classico nel caso normale

Rapporto di verosimiglianza VII

Esempio - rapporto di verosimiglianza sonniferi

Prendiamo l'esempio dei sonniferi assumendo che la differenza sia iid da $N(\mu, \sigma^2)$.

Testiamo el seguente ipotesi con il test del rapporto di verosimiglianza:

- (1) $H_0 : \mu = \mu_0 \qquad H_1 : \mu \neq \mu_0$, con σ^2 noto
- (2) $H_0 : \mu = \mu_0 \qquad H_1 : \mu \neq \mu_0$, con σ^2 non noto
- (3) $H_0 : \sigma^2 = \sigma_0^2 \qquad H_1 : \sigma^2 \neq \sigma_0^2$, con μ non noto.
- (4) $H_0 : \sigma^2 \leq \sigma_0^2 \qquad H_1 : \sigma^2 > \sigma_0^2$, con μ non noto.

"Chunk Test Ipotesi"

Soluzione:

(1) La congiunta (e quindi la verosimiglianza) è

$$f(\mathbf{x}|\theta) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2}\right)$$

Rapporto di verosimiglianza VIII

Abbiamo quindi che il rapporto di verosimiglianza è

$$\lambda(\mathbf{x}) = \frac{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{2\sigma^2}\right)}{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{2\sigma^2}\right)} = \frac{\exp\left(-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu_0)^2}{2\sigma^2}\right)}{\exp\left(-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{2\sigma^2}\right)}$$

$$\lambda(\mathbf{x}) = \exp\left(-\frac{n(\bar{x}-\mu_0)^2}{2\sigma^2}\right)$$

quindi la regione di rifiuto è

$$R(\mathbf{x}) = \left\{\mathbf{x} : \exp\left(-\frac{n(\bar{x}-\mu_0)^2}{2\sigma^2}\right) \leq c\right\}$$

che con opportune trasformazioni si può esprimere come

$$R(\mathbf{x}) = \{\mathbf{x} : |\bar{x}-\mu_0| \geq c^*\}$$

Rapporto di verosimiglianza IX

per un opportuno c^*. Adesso dobbiamo trovare c^* in modo tale che il livello del test sia α, ma una situazione simile l'abbiamo già trovata precedentemente e abbiamo che

$$c^* = z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}.$$

e quindi

$$R(\mathbf{x}) = \left\{\mathbf{x} : |\bar{x} - \mu_0| \geq z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right\}$$

Se prendiamo i dati reali, abbiamo $\bar{x} = -1.58$, $n = 10$ e assumiamo $\mu_0 = 0$ (nessun differenza tra i sonniferi) e supponiamo $\sigma^2 = 10$. In questo caso accettiamo H_0.

(2) Nel secondo caso abbiamo σ^2 non noto. IN questo caso si assume che in entrambe le ipotesi $\sigma^2 \in \mathbb{R}^+$ (il parametro appertiene a tutto il suo dominio)

$$H_0 : \mu = \mu_0, \sigma^2 \in \mathbb{R}^+ \qquad H_1 : \mu \neq \mu_0, \sigma^2 \in \mathbb{R}^+$$

Rapporto di verosimiglianza X

Avevamo gia visto che $\bar{x}$ e $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$ sono gli MLE nel caso non-restricted, mentre è facile verificare che il sup della verosimiglianza sotto H_0 è raggiunto in μ_0 e $\hat{\sigma}_0^2 = \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{n}$. Abbiamo quindi

$$\lambda(\mathbf{x}) = \frac{(2\pi\hat{\sigma}_0^2)^{-\frac{n}{2}} \exp\left(-\frac{n\hat{\sigma}_0^2}{2\hat{\sigma}_0^2}\right)}{(2\pi\hat{\sigma}^2)^{-\frac{n}{2}} \exp\left(-\frac{n\hat{\sigma}^2}{2\hat{\sigma}^2}\right)} = \left(\frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)^{-\frac{n}{2}} =$$

$$\left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)^{-\frac{n}{2}} = \left(1 + \frac{n(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)^{-\frac{n}{2}}$$

Abbiamo che la regione di rifiuto è

$$R(\mathbf{x}) = \left\{\mathbf{x} : \left(1 + \frac{n(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)^{-\frac{n}{2}} \leq c\right\}$$

e che può essere scritta come

$$R(\mathbf{x}) = \left\{\mathbf{x} : \left(1 + \frac{n(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)^{-\frac{n}{2}} \leq c\right\} \equiv \{\mathbf{x} : |\bar{x} - \mu_0| \geq c^*\}$$

A prima vista potrebbe sembrare simile al caso (1) ma adesso non possiamo usare la Z perchè non possiamo standardizzare, ma possiamo usare la T di student

$$\alpha = 1 - P(-c^* < \bar{x} - \mu_0 < c^*; \mu_0) = 1 - P\left(-c^* < \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{n}}} < c^*; \mu_0\right)$$

dove $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$, e con c^* indichiamo una generica costante (il cui valore cambia). Sappiamo anche che

$$\frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{n}}} \sim T_{n-1}$$

e quindi, come per il caso normale, possiamos scegliere $c^* = t_{n-1,\alpha/2}$, dove $t_{n-1,\alpha/2}$ è il quantile della T con n-1 gdl che ha comulata pari a $1 - \alpha/2$. Possiamo scrivere la regione di rifiuto come

$$R(\mathbf{x}) = \left\{ \mathbf{x} : |\bar{x} - \mu_0| \geq t_{n-1,\alpha/2}\sqrt{\frac{s^2}{n}} \right\}$$

Rapporto di verosimiglianza XII

o alternativamente

$$A(\mathbf{x}) = \left\{ \mathbf{x} : \mu_0 - t_{n-1,\alpha/2}\sqrt{\frac{s^2}{n}} < \bar{x} < \mu_0 + t_{n-1,\alpha/2}\sqrt{\frac{s^2}{n}} \right\}$$

Se n è abbastanza grande possiamo approssimare $t_{n-1,\alpha/2}$ con $z_{\alpha/2}$.
Utilizzando i dati reali abbiamo che possiamo rifiutare H_0, quindi c'è differenza tra i due trattamenti.

(3) Calcoliamo il rapporto di verosimiglianze in questo caso

$$\lambda(\mathbf{x}) = \frac{(2\pi\sigma_0^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{2\sigma_0^2}\right)}{(2\pi\hat{\sigma}^2)^{-\frac{n}{2}} \exp\left(-\frac{n}{2}\right)} = \left(\frac{\hat{\sigma}^2}{\sigma_0^2}\right)^{\frac{n}{2}} \exp\left(-\frac{n}{2}\frac{\hat{\sigma}^2}{\sigma_0^2}\right) \exp\left(\frac{n}{2}\right)$$

e la regione di rifiuto è

$$R(\mathbf{x}) = \left\{ \mathbf{x} : \exp\left(\frac{n}{2}\right)\left(\frac{\hat{\sigma}^2}{\sigma_0^2}\right)^{\frac{n}{2}} \exp\left(-\frac{n}{2}\frac{\hat{\sigma}^2}{\sigma_0^2}\right) \leq c \right\}$$

è facile vedere come $\left(\frac{\hat{\sigma}^2}{\sigma_0^2}\right)^{\frac{n}{2}} \exp\left(-\frac{n}{2}\frac{\hat{\sigma}^2}{\sigma_0^2}\right)$ raggiunge il massimo a $\frac{\hat{\sigma}^2}{\sigma_0^2} = 1$.
Determinare il c che ci da un livello α è complicato perchè dovremmo trovare due valori

$$\frac{\hat{\sigma}^2}{\sigma_0^2} \leq k_1 \qquad \frac{\hat{\sigma}^2}{\sigma_0^2} \geq k_2$$

che corrispondono a $\left(\frac{\hat{\sigma}^2}{\sigma_0^2}\right)^{\frac{n}{2}} \exp\left(-\frac{n}{2}\frac{\hat{\sigma}^2}{\sigma_0^2}\right) \leq c$ assicurando che il livello sia α e purtroppo si può fare solo a livello numerico. In genere quando non si possono trovare dei limiti giusti, ci si accontenta di approssimare il test di verosimiglianza chiedendo

$$P\left(\frac{\hat{\sigma}^2}{\sigma_0^2} \leq k_1; \sigma_0^2\right) = P\left(\frac{\hat{\sigma}^2}{\sigma_0^2} \geq k_2; \sigma_0^2\right) = \frac{\alpha}{2}$$

e abbiamo che

$$P\left(\frac{\hat{\sigma}^2}{\sigma_0^2} \leq k_1; \sigma_0^2\right) = P\left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n\sigma_0^2} \leq k_1; \sigma_0^2\right) = P\left(\frac{(n-1)s^2}{\sigma_0^2} \leq k_1^*; \sigma_0^2\right)$$

ma siccome sappiamo che $\frac{(n-1)s^2}{\sigma_0^2} \leq k_1^*$ è un χ^2_{n-1}, allora $k_1^* = w_{n-1,\alpha/2}$, dove $w_{n-1,1-\alpha/2}$ è il quantile della χ^2_{n-1}. COn ragionamenti simili troviamo che

$$P\left(\frac{\hat{\sigma}^2}{\sigma_0^2} \geq k_2; \sigma_0^2\right) = P\left(\frac{(n-1)s^2}{\sigma_0^2} \geq w_{n-1,\alpha/2}; \sigma_0^2\right)$$

e la regione di accettazione è allora

$$A(\mathbf{x}) = \left\{\mathbf{x} : w_{n-1,1-\alpha/2} < \frac{(n-1)s^2}{\sigma_0^2} < w_{n-1,\alpha/2}\right\}$$

naturalmente in questo caso, per via dell'approssimazione, potremmo andare nella zona di rifiuto campioni che hanno valore più elevato di altrei che sono in quella d'accettazione.
Il problema si può risolvere se invece di un'ipotesi nulla puntuale, si prende un'ipotesi nulla unilaterale, come il punto (4), visto che in questo caso $\frac{\hat{\sigma}^2}{\sigma_0^2}$ può solo essere maggiore di 1.
Prendendo $\sigma_0^2 = 10$, rifiutiamo l'ipotesi nulla, cosa abbastanza ragionevole visto che $s^2 = 1.51$.

Rapporto di verosimiglianza XV

(4) Il punto 4 lo lascio come esercizio. Per il calcolo del livello, ricordate che dovete considerare che il sup su $\sigma^2 \in H_0$ la funzione potenza deve essere minore o uguale a α.
$\square$

Esempio - rapporto di verosimiglianza sonniferi 2

Prendiamo l'esercizio di prima, con σ^2 noto e uguale a 10, e assumiamo di voler testare

$$H_0 : \mu \leq \mu_0 \qquad H_1 : \mu > \mu_0$$

con $\mu_0 = 0$ e livello α. Trovare la regione di rifiuto e determianre che n serve per poter accettere H_1 con probabilità 0.8 se l'effetto vero è 2 ore.
"Chunk Test Ipotesi"

Soluzione:

Come trovare la regione di rifiuto la lascio come esercizio, ma il risultato

$$R(\mathbf{x}) = \{\mathbf{x} : \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \geq z_\alpha\}$$

Per risolvere l'esercizio dobbiamo trovare la funzione potenza, che è

$$\beta(\mu) = P\left(\frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \geq z_\alpha; \mu\right)$$

e ricordiamo che se μ è il valore vero allora $\frac{\bar{x}-\mu_0}{\sqrt{\frac{\sigma^2}{n}}}$ non è una normal standard. Allora

$$\beta(\mu) = P\left(\frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \geq z_\alpha; \mu\right) = P\left(\frac{\bar{x} - \mu_0 - \mu}{\sqrt{\frac{\sigma^2}{n}}} \geq z_\alpha - \frac{\mu}{\sqrt{\frac{\sigma^2}{n}}}; \mu\right) =$$

Rapporto di verosimiglianza XVII

$$P\left(\frac{\bar{x}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \geq z_\alpha + \frac{\mu_0-\mu}{\sqrt{\frac{\sigma^2}{n}}}; \mu\right) = P\left(Z \geq z_\alpha + \frac{\mu_0-\mu}{\sqrt{\frac{\sigma^2}{n}}}; \mu\right) = 0.8$$

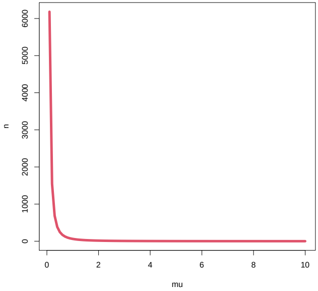
Per far si che la prob sia 0.8, allora

$$z_\alpha + \frac{\mu_0-\mu}{\sqrt{\frac{\sigma^2}{n}}} = z_{0.8} \Rightarrow z_\alpha\sqrt{\sigma^2} + \sqrt{n}(\mu_0-\mu) = z_{0.8}\sqrt{\sigma^2} =$$

$$n = \sigma^2\left(\frac{z_{0.8}-z_\alpha}{\mu_0-\mu}\right)^2 = 10\left(\frac{z_{0.8}-z_\alpha}{0-2}\right)^2 = 15.456$$

quindi è sufficiente $n = 16$.

Rapporto di verosimiglianza XVIII



Il valore di n cresce se μ si distanzia da μ_0, e se aumentiamo la probabilità, e per μ che tende a μ_0, serve un numero infinito di osservazioni (**"Chunk Test Ipotesi"**). $\square$

Esempio - LRT Poisson

Assumiamo che i dati, di dimensione n, siano iid da $P(\lambda)$. Trovare la regione di rifiuto di un test

$$H_0 : \lambda \leq \lambda_0 \qquad H_1 : \lambda > \lambda_0$$

Soluzione:

Calcoliamo il rapporto di verosimiglianza. Se lo stimatore di massima verosimiglianza è $\leq \lambda_0$, abbiamo che

$$\lambda(\mathbf{x}) = 1$$

altrimenti

$$\lambda(\mathbf{x}) = \frac{\prod_{i=1}^{n} \frac{\lambda_0^{x_i} e^{-\lambda_0}}{x_i!}}{\prod_{i=1}^{n} \frac{\bar{x}^{x_i} e^{-\bar{x}}}{x_i!}} = \frac{\frac{\lambda_0^{\sum_{i=1}^{n} x_i} e^{-n\lambda_0}}{\prod_{i=1}^{n} x_i!}}{\frac{\bar{x}^{\sum_{i=1}^{n} x_i} e^{-n\bar{x}}}{\prod_{i=1}^{n} x_i!}} = \left(\frac{\lambda_0}{\bar{x}}\right)^{\sum_{i=1}^{n} x_i} e^{-n(\lambda_0 - \bar{x})}$$

Sappiamo anche che $t = \sum_{i=1}^{n} x_i \sim P(\lambda^*)$, dove $\lambda^* = n\lambda$, è una statistica sufficiente per λ, e possiamo vedere che il rapporto di verosimiglianza di t è la stessa

$$\lambda(t) = \frac{\frac{(n\lambda_0)^t e^{-n\lambda_0}}{t!}}{\frac{t^t e^{-t}}{t!}}$$

siccome t è lo stimatore di massima verosimiglianza di λ^* generale (e quindi $t/n = \lambda$) e $n\lambda_0$ lo è sotto H_0. Quindi

$$\lambda(t) = \left(\frac{n\lambda_0}{t}\right)^t e^{-n\lambda_0 + t} = \left(\frac{\lambda_0}{\bar{x}}\right)^t e^{-n\lambda_0 + n\bar{x}} = \left(\frac{\lambda_0}{\bar{x}}\right)^t e^{-n(\lambda_0 - \bar{x})}$$

Quindi la regione di rifiuto deve dipendere dalla variabile aleatoria $\sum_{i=1}^{n} x_i$. Sappiamo che se $\lambda_0 > \sum_{i=1}^{n} x_i/n$ allora il rapporto di verosimiglianze è 1 quindi possono stare nella zona di rifiuto solo campioni con $\lambda_0 < \sum_{i=1}^{n} x_i/n$. Calcoliamo al derivata di $\log \lambda(\mathbf{x})$ rispetto a $\bar{x} = \sum_{i=1}^{n} x_i/n$, ricordando che $t = n\bar{x}$:

$$\frac{d \log \lambda(\mathbf{x})}{d\bar{x}} = \frac{d}{d\bar{x}}\left(n\bar{x}\log\lambda_0 - n\bar{x}\log\bar{x} - n\lambda_0 + n\bar{x}\right) = n\log\lambda_0 - (n\log\bar{x} + n) + n$$

che è uguale a

$$n\log\lambda_0 - n\log\bar{x} = n\log(\lambda_0/\bar{x})$$

che raggiunge il massimo a λ_0 e è decrescente per valori maggiori.
La regione di rifiuto è allora

$$R(\mathbf{x}) = \{\mathbf{x} : \bar{x} \geq c\} \equiv R(\mathbf{x}) = \{\mathbf{x} : \sum_{i=1}^{n} x_i \geq c\}$$

per determinare c dobbiamo imporre un livello α del test e vediamo la funzione potenza

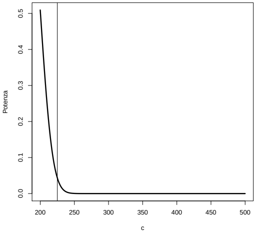
$$\beta(\lambda) = P(\sum_{i=1}^{n} x_i \geq c; \lambda) = P(Y \geq c; \lambda) = 1 - \sum_{j=0}^{c-1} \frac{(n\lambda)^j e^{-n\lambda}}{j!}$$

dove $Y \sim P(n\lambda)$ e come detto precedentemente $\lambda_0 \leq \bar{x}$.
Dovremmo trovare il c per cui

$$\sup_{\lambda \leq \lambda_0} \beta(\lambda) \leq \alpha$$

Non stiamo usando una quantità pivotale e quindi è complicato trovare una soglia per ogni possibile valore λ_0, ma possiamo determiarla per specifici valori computazionalmente.
Per esempio, prendiamo il codice **"chunk Test Ipotesi"**, possiamo trovare c nel caso del numero di scoperte, assumendo che provengono da una Poisson di parametro λ, con $\lambda_0 = 2$ e $\alpha = 0.05$. Troviamo che c= 225 produce il risultato cercato

Rapporto di verosimiglianza XXIII



Una soluzione è usare la relazione

$$P(X \leq x) = P(Y \geq a)$$

se $X \sim G(a, b)$ e $Y \sim P(x/b)$, con a intero, che lascio come esercizio (per la dimostrazione vedete l'esempio 3.3.1 del casella berger).

Rapporto di verosimiglianza XXIV

Un'altra possibile soluzione è usare il TLC, ricordando che

$$\bar{x} \overset{\text{Approx}}{\sim} N(\lambda, \frac{\lambda}{n})$$

per i motivi visti precedentemente con la bernulliana, possiamo anche scrivere

$$\bar{x} \overset{\text{Approx}}{\sim} N(\lambda, \frac{\bar{x}}{n})$$

e usando quanto visto precedentemente abbiamo che la regione di rifiuto è

$$R(\mathbf{x}) = \{\mathbf{x} : \frac{\bar{x} - \lambda_0}{\sqrt{\frac{\bar{x}}{n}}} \geq z_\alpha\}$$

Un'altra soluzione è usare

$$\bar{x} \overset{\text{Approx}}{\sim} N(\lambda, \frac{\lambda}{n})$$

che da luogo alla regione di rifiuto

$$R(\mathbf{x}) = \{\mathbf{x} : \bar{x} \geq \lambda_0 + z_\alpha \sqrt{\frac{\lambda_0}{n}}\} \equiv \{\mathbf{x} : \sum_{i=1}^{n} x_i \geq n(\lambda_0 + z_\alpha \sqrt{\frac{\lambda_0}{n}})\}$$

Con gli stessi valori che abbiamo usato per nel caso precedente (λ, con $\lambda_0 = 2$ e $\alpha = 0.05$), abbiamo che

$$n(\lambda_0 + z_\alpha \sqrt{\frac{\lambda_0}{n}}) = 223.261$$

che è molto vicino al valore di c nel caso esatto della Poisson, ma in questo caso abbiamo un intervallo valido per qualsiasi n, λ_0, α.

Come esercizio potete verificare che è difficile trovare una regione di rifiuto per

$$H_0 : \lambda = \lambda_0 \qquad H_1 : \lambda \neq \lambda_0$$

ma diventa facile se si utilizza l'approssimazione normale. $\square$

p-value I

La procedura di test è molto utile ma tra i lati negativi è che da una risposta dicotomica accette/rifiuto. Un'altra possibile soluzione è di usare un **p-value**.

Definizione - p-value

Un p-value è una statistica $0 \leq p(\mathbf{x}) \leq 1$ per ogni campione $\mathbf{x}$, per cui valori piccoli di $p(\mathbf{x})$ danno supporto a H_1. Un p-value è valido se per ogni $\theta \in \Theta_0$ e ogni $0 < \alpha < 1$, abbiamo

$$P(p(\mathbf{X}) \leq \alpha) \leq \alpha$$

Con un p-value è facile costruire un test di livello α, visto che basta rifiutare H_0 ogni volta che il p-value è minore di α. Oltretutto, se il suo valore è vicino a α rifiutiamo H_0 con meno sicurezza di quando un p-value è prossimo a zero.
Un modo classico di definire un p-value (quello che useremo sempre) è dato dal prossimo teorema

Teorema - p-value

Se $W(\mathbf{X})$ è una statistica test tale che valori elevati della statistica danno evidenza su H_1 vera, allora se per ogni campione $\mathbf{x}$ definiamo

$$p(\mathbf{x}) = \sup_{\theta \in \Theta_0} P(W(\mathbf{X}) \geq W(\mathbf{x}))$$

allora $p(\mathbf{x})$ è un valido p-value.

Se prendiamo il classico test sulla media di una normale (varianza nota), per un'ipotesi bilaterale, la statistica test è

$$Z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}.$$

Se H_0 è unilaterale, per esempio $H_0 : \mu \geq \mu_0$, allora rifiutiamo H_0 se

$$Z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \leq -z_\alpha$$

p-value III

Invece di Z prendiamo $-Z$ e è facile verificare che valori elevati di $-Z$ sono più palusibili se H_1 è vera (danno suppporto a questa ipotesi), quindi seguendo il teorema, potremmo definire il p-value come

$$p(\mathbf{x}) = P(-Z \geq z_\alpha) \equiv P(Z \leq -z_\alpha)$$

e, ogni volta che questo p-value è minore di α rifiutiamo H_0. Se il p-value è molto piccolo siamo molto confidenti che H_0 sia falsa. Se l'ipotesi alternativa è $H_0 : \mu \leq \mu_0$ arriviamo a un simile risultato.
Il problema nasce quando l'ipotesi nulla è puntuale perchè la accettiamo H_0 se

$$-z_{\alpha/2} < Z < z_{\alpha/2}$$

e la statistica Z non rispetta l'ipotesi del teorema. Una possibile soluzione è di considerare la variabile aleatoria Z^2 e usare questa, che è un $\chi^2(1)$, oppure $|Z|$.

p-value IV

Molte volte non è così facile trovare una statistica da usare per ipotesi nulle puntuali, e la statistica standard da utilizzare (se non si trova di meglio) è

$$p(\mathbf{x}) = 2\min\left(P(W(\mathbf{X}) \geq W(\mathbf{x})), P(W(\mathbf{X}) \leq W(\mathbf{x}))\right)$$

dove $W(\mathbf{X})$ è una generica statistica test.
Nel caso in cui le statistiche siano multivariate la definizione diventa più complicata e non ce ne occupiamo qui

Intervalli di confidenza I

Come nel caso del p-value, anche congli intervalli di confidenza, o stime intervallari, vogliamo dire qualcosa in più sul parametro, invece di una sola riposta dicotomica. Una soluzione è di definire un'intervallo che con una data probabilità contenga il valore vero del parametro

Definizione - Stima/stimatore intervallare

Una **stima intervallare** per un parametro θ **continuo** è una coppia di funzioni $L(\mathbf{x})$ e $U(\mathbf{x})$ che soddisfano $L(\mathbf{x}) \leq U(\mathbf{x})$ e osservando un valore del campione l'inferenza che facciamo sul parametro è $L(\mathbf{x}) \leq \theta \leq U(\mathbf{x})$. L'intervallo random $[L(\mathbf{X}), U(\mathbf{X})]$ è una **stimatore intervallare**

L'intervallo è una variabile aleatoria e quindi l'evento $L(\mathbf{x}) \leq \theta \leq U(\mathbf{x})$ ha una probabilità associata. Fate attenzione che la probabilità non è che θ appartenga all'intervallo, visto che θ non è aleatorio, ma che l'intervallo contenga θ.

Intervalli di confidenza II

Definizione - Copertura

Per ogni stimatore intervallare $[L(\mathbf{X}), U(\mathbf{X})]$ di un parametro θ, la probabilità di copertura è la probabilità che l'intervallo copra il valore vero:

$$P(\theta \in [L(\mathbf{X}), U(\mathbf{X})]; \theta) \equiv P(L(\mathbf{X}) \leq \theta; \theta) \cap P(U(\mathbf{X}) \geq \theta; \theta)$$

Definizione - Confidenza

Per ogni stimatore intervallare $[L(\mathbf{X}), U(\mathbf{X})]$ di un parametro θ, la confidenza è l'inf della copertura vero:

$$\inf_{\theta \in \Theta} P(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$$

Quindi la copurtura ci dice la proabbilità per un dato θ, mentre la confidenza è la minima coprtura che abbiamo.

Esempio - Confidenza e Copertura

Assumiamo che i dati siano normali iid con media μ e varianza 1, e prendiamo il seguente intervallo

$$\mu \in [\frac{\sum_{i=1}^{n} |x_i|}{n} - z_{\alpha/2}\sqrt{\frac{1}{n}}, \frac{\sum_{i=1}^{n} |x_i|}{n} + z_{\alpha/2}\sqrt{\frac{1}{n}}]$$

A livello intuitivo, se μ è molto lontano da 0, abbiamo che $\frac{\sum_{i=1}^{n} |x_i|}{n} \approx \mu$, e quindi l'intervallo avrà una copertura elevata (vedremo dopo che è $\approx 1 - \alpha$). Se invece μ è molto più piccolo di 0 la copertura sarà 0, ne segue che la confidenza è zero.
Noi vogliamo intervalli con confidenza elevata e non copertura elevata, perchè non sappiamo il valore del parametro e dobbiamo avere una probabilità elevata per tutti i possibili valori.

Come con i test, possiamo definire un intervallo come vogliamo, ma ci sono metodi che sono facilmente utilizzabili. La prima cosa che possiamo fare è utilizzare una statistica che nella sua costruziono o nella sua distribuzione, dipende dal parametro.

Esempio - varianza di una normale

Ipotizziamo che gli n dati siano iid da una normale, e vogliamo trovare un intervallo per σ^2 di confidenza $1-\alpha$

Soluzione:

Possiamo usare la statistica

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

e imporre

$$P\left(a \leq \frac{(n-1)S^2}{\sigma^2} \leq b\right) = P\left(a \leq \chi^2_{n-1} \leq b\right) = 1-\alpha$$

e possiamo definire a e b in modo tale che l'equazione sopra sia valida. Una volta che abbiamo questi due valori possiamo scrivere

$$P\left(a \leq \frac{(n-1)S^2}{\sigma^2} \leq b\right) = P\left(\frac{(n-1)S^2}{b} \leq \sigma^2 \leq \frac{(n-1)S^2}{a}\right) = 1-\alpha$$

e quindi per costruzione l'intervallo $[\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a}]$ è un intervallo con confidenza α.
$\square$

Notate come, se volessimo fare un test su σ^2, con livello α, con $H_0 : \sigma = \sigma_0$, potremmo usare la stessa statistica dell'esempio sopra con regione di accettazione

$$A(\mathbf{x}) = \left\{\mathbf{x} : a \leq \frac{(n-1)S^2}{\sigma_0^2} \leq b\right\}$$

dove a e b sono gli stessi dell'esempio sopra. Questo non è un caso perchè il metodo standard è di invertire un test di livello α, con ipotesi puntuale, per ottenere un intervallo con confidenza $1-\alpha$. Questo perchè i σ_0^2 presenti nell'intervallo sono tutti quelli che sarebbero accettati con il test. Come per i test, le quantità pivotali sono molto utili perchè permettono di trovare intervalli molto facilmente.

Intervalli di confidenza VI

Per fare un'esempio più chiaro, in un test sulla media di una normale con varianza non nota, abbiamo che accettiamo se

$$-t_{n-1,\alpha/2} \leq \frac{\bar{x}-\mu_0}{\sqrt{\frac{s^2}{n}}} \leq t_{n-1,\alpha/2}$$

Adesso ci possiamo chiedere quali valori di μ_0 portano ad accettare questo test, e la risposta è

$$\bar{x} - t_{n-1,\alpha/2}\frac{s^2}{n} \leq \mu_0 \leq \bar{x} + t_{n-1,\alpha/2}\frac{s^2}{n}$$

quindi i dati danno sostegno a tutte le ipotesi (nulle) di μ in quest'intervallo. Abbiamo allora che

$$[\bar{x} - t_{n-1,\alpha/2}\sqrt{\frac{s^2}{n}}, \bar{x} + t_{n-1,\alpha/2}\sqrt{\frac{s^2}{n}}]$$

è un intervallo di confidenza di livello $1-\alpha$ per μ. Diamo un teorema per formalizzare quest'idea

Intervalli di confidenza VII

Teorema - Invertire Test

Per ogni $\theta_i \in \Theta$ indichiamo $A_{\theta_0}(\mathbf{x})$ come la regione di accettazione di un test puntuale con $H_0 : \theta = \theta_0$ di livello α, allora per ogni $\mathbf{x} \in \mathcal{X}$ definite il set $C_{\mathbf{x}}(\theta_0)$

$$C_{\mathbf{x}}(\theta_0) = \{\theta_0 : \mathbf{x} \in A_{\theta_0}(\mathbf{x})\}$$

allora $C_{\mathbf{x}}(\theta_0)$ è un **set** di confidenza $1 - \alpha$.
Oltretutto se $C_{\mathbf{x}}(\theta_0)$ è un **set** di confidenza $1 - \alpha$, allora per ogni $\theta_0 \in \Theta$, definiamo

$$A_{\theta_0}(\mathbf{x}) = \{\mathbf{x} : \theta_0 \in C_{\mathbf{x}}(\theta_0)\}$$

è una regione di accettazione di livello α per un test $H_0 : \theta = \theta_0$.

Il teorema parla di set, perchè in generale non è assicurato che $C_{\mathbf{x}}(\theta_0)$ sia un intervallo, ma potrebbe essere l'unione di intervalli, ma nei casi che abbiamo visto è sempre un intervallo.

Intervalli di confidenza VIII

Naturalmente ci sono diversi intervalli che si possono definire. Per esempio riprendendo il caso precedente

$$P\left(a \leq \frac{(n-1)S^2}{\sigma^2} \leq b\right) = P\left(a \leq \chi^2_{n-1} \leq b\right) = 1 - \alpha$$

potremmo scegliere a e b in diversi modo e comunque avere una probabilità di $1-\alpha$. In genere, si preferisce lasciare uan probabilità di $\alpha/2$ a distra e a sinistra, ma non è la scelta migliore, perchè porta ad avere intervalli più ampi del dovuto, e nel caso di test di accettare H_0 con valori di densità della statistica più bassi di valori sotto H_1. Il seguente teorema spiega come trovare il test con intervallo più corto.

Teorema - Lunghezza minima dell'intervallo

Se $f(x)$ è una densità unimodale, l'intervallo $[a, b]$ che soddisfa
(1) $\int_a^b f(x)dx = 1 - \alpha$
(2) $f(a) = f(b)$
(3) $a \leq x^* \leq b$, dove x^* è la moda
allora $[a, b]$ è l'intervallo più corto che soddisfa (1).

Statistiche da conoscere I

Questa è una lista delle statistiche da usare per situazioni specifiche.

- **Z-test** dati: $X_i \sim N(\mu, \sigma^2)$ - varianza nota - Ipotesi nulla della forma $H_0 : \mu = \mu_0$, $H_0 : \mu \leq \mu_0$, $H_0 : \mu < \mu_0$, $H_0 : \mu \geq \mu_0$, $H_0 : \mu > \mu_0$

$$Z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

- **T-test** dati: $X_i \sim N(\mu, \sigma^2)$ - varianza non nota - Ipotesi nulla della forma $H_0 : \mu = \mu_0$, $H_0 : \mu \leq \mu_0$, $H_0 : \mu < \mu_0$, $H_0 : \mu \geq \mu_0$, $H_0 : \mu > \mu_0$

$$T = \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}} \sim t(n-1)$$

con $S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})}{n-1}$

- **Chi-test** dati: $X_i \sim N(\mu, \sigma^2)$ - media non nota - Ipotesi nulla della forma $H_0 : \sigma^2 = \sigma_0^2$, $H_0 : \sigma^2 \leq \sigma_0^2$, $H_0 : \sigma^2 < \sigma_0^2$, $H_0 : \sigma^2 \geq \sigma_0^2$, $H_0 : \sigma^2 > \sigma_0^2$

$$V = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$$

- **Test delle proporzioni** dato: $X_i \sim Bin(p)$ - Ipotesi nulla della forma $H_0 : p = p_0$, $H_0 : p \leq p_0$, $H_0 : p < p_0$, $H_0 : p \geq p_0$, $H_0 : p > p_0$

$$Z = \frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{approx}{\sim} N(0,1)$$

  oppure

$$Z = \frac{\bar{X} - p_0}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \overset{approx}{\sim} N(0,1)$$

- **Z-test per differenze tra medie** dati: $X_i \sim N(\mu_x, \sigma_x^2)$ e $Y_j \sim N(\mu_y, \sigma_y^2)$, con $i = 1, \ldots, n$ e $j = 1, \ldots, m$, indipendenti a due a due - varianza note - Ipotesi nulla della forma $H_0 : \mu_x - \mu_y = \mu_0$, $H_0 : \mu_x - \mu_y \leq \mu_0$, $H_0 : \mu_x - \mu_y < \mu_0$, $H_0 : \mu_x - \mu_y \geq \mu_0$, $H_0 : \mu_x - \mu_y > \mu_0$

$$Z = \frac{\bar{X} - \bar{Y} - \mu_0}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim N(0,1)$$

Statistiche da conoscere III

- **T-test per differenze tra medie** dati: $X_i \sim N(\mu_x, \sigma_x^2)$ e $Y_j \sim N(\mu_y, \sigma_y^2)$, con $i = 1, \dots, n$ e $j = 1, \dots, m$, indipendenti a due a due - varianza non note, **ma supposte uguali** - Ipotesi nulla della forma $H_0 : \mu_x - \mu_y = \mu_0$, $H_0 : \mu_x - \mu_y \leq \mu_0$, $H_0 : \mu_x - \mu_y < \mu_0$, $H_0 : \mu_x - \mu_y \geq \mu_0$, $H_0 : \mu_x - \mu_y > \mu_0$

$$T = \frac{\bar{X} - \bar{Y} - \mu_0}{\sqrt{S_{pool}^2 \left(\frac{1}{n} + \frac{1}{m}\right)}} \sim t(n + m - 2)$$

con $S_{pool}^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2} = \frac{\sum_{i=1}^n (X_i - \bar{X}) + \sum_{i=1}^n (Y_i - \bar{Y})}{n+m-2}$

Statistiche da conoscere IV

- **F-test per differenze tra varianze** dati: $X_i \sim N(\mu_x, \sigma_x^2)$ e $Y_j \sim N(\mu_y, \sigma_y^2)$, con $i = 1, \ldots, n$ e $j = 1, \ldots, m$, indipendenti a due a due - media non nota - Ipotesi nulla della forma $H_0 : \sigma_x^2/\sigma_y^2 = c_0$, $H_0 : \sigma_x^2/\sigma_y^2 \leq c_0$, $H_0 : \sigma_x^2/\sigma_y^2 < c_0$, $H_0 : \sigma_x^2/\sigma_y^2 \geq c_0$, $H_0 : \sigma_x^2/\sigma_y^2 > c_0$

$$F = \frac{S_x^2}{c_0 S_y^2} \sim F(n-1, m-1)$$

che deriva dal fatto che

$$\frac{\sigma_y^2 S_x^2}{\sigma_x^2 S_y^2} \sim F(n-1, m-1)$$