

SUFFICIENZA

LA STATISTICA... CHE BASTA A SÈ STESSA...

DOMENICO VISTOCCO

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

`domenico.vistocco@unina.it`

La Statistica che non
deve chiedere mai!



SUFFICIENZA

DI COSA PARLIAMO?

Fenomeno di interesse $\longleftrightarrow \{Y \sim f(y; \theta); \theta \in \omega(\theta)\}$

- n osservazioni $y_1, \dots, y_n$ dalle quali si vogliono trarre conclusioni circa il valore di θ (parametro o vettore di parametri)
- il campione osservato è una lista di numeri ardua da interpretare (pensa al caso in cui n è grande)
- è utile sintetizzare l'informazione presente nei dati determinando poche caratteristiche chiave dei valori campionari (media, mediana, varianza, osservazione più grande, osservazione più piccola, ...)

Definition: Statistica T_n . Funzione di v.c. osservabili, e quindi a sua volta osservabile, che non contiene alcun parametro incognito

DI COSA PARLIAMO? (LESSICO ED OBIETTIVO)

Trovare un'opportuna sintesi (statistica):

1. efficace: capace di raggiungere un determinato obiettivo
2. “efficiente” (non ridondante): capacità di raggiungerlo con la minima allocazione di risorse

UN PRIMO PUNTO DI VISTA SU T_n

- qualsiasi statistica T_n definisce una riduzione o sintesi dei dati
- se si parte dal valore t_n che la v.c. $T_n = T(Y_1, \dots, Y_n)$ assume, due campioni $\mathbf{y}$ e $\mathbf{x}$ sono considerati equivalenti se $T(\mathbf{y}) = T(\mathbf{x})$ anche se gli effettivi valori campionari potrebbero non essere equivalenti

UN SECONDO PUNTO DI VISTA SU T_n

- la riduzione in termini di una particolare statistica può essere pensata come una partizione dello spazio campionario $\mathcal{Y}_n$
- una qualsiasi statistica T_n induce (definisce) una partizione di $\mathcal{Y}_n$: sia t_0 un valore di T_n ; quel sottoinsieme di $\mathcal{Y}_n$ consistente di tutti i punti $(y_1, \dots, y_n)$ per i quali $T(Y_1, \dots, Y_n) = t_0$ è un sottoinsieme di quelli compresi nella partizione (analogamente per altri valori di T_n)
- spesso (ma non necessariamente) una statistica corrisponde ad una condensazione di $\mathcal{Y}_n$

UN ESEMPIO DI PARTIZIONE DI $\mathcal{Y}_n$

$n = 3$ da una popolazione bernoulliana:

Y_1	Y_2	Y_3	$T_n = \sum_{i=1}^n Y_i$
0	0	0	0
0	0	1	1
0	1	0	
1	0	0	
0	1	1	2
1	0	1	
1	1	0	
1	1	1	3

DEFINIZIONE FORMALE DI PARTIZIONE

Immagine di $\mathcal{Y}_n$ sotto $T(\mathbf{y})$:

$$\mathcal{T} = \{t \mid t = T(\mathbf{y}) \text{ per alcuni } \mathbf{y} \in \mathcal{Y}_n\}$$

La statistica $T(\mathbf{y})$ partiziona lo spazio campionaria in insiemi A_t :

$$A_t = \{\mathbf{y} \mid T(\mathbf{y}) = t\}$$

Piuttosto che riportare l'intero campione $\mathbf{y}$, la statistica riporta solo che $T_n = T(\mathbf{y}) = t$ o, equivalentemente, $\mathbf{y} \in A_t$

UNA STATISTICHE, TANTE STATISTICHE

- statistiche diverse possono indurre la stessa partizione: ogni funzione biunivoca di T_n induce la stessa partizione di T_n
- sebbene non si tratti di una caratteristica necessaria, una delle ragioni dell'uso di una statistica è la capacità di ridurre le informazioni disponibili
- due statistiche T_{1n} e T_{2n} cui corrisponde la stessa partizione hanno la stessa utilità (si dicono equivalenti)

IL PRINCIPIO DI BASE DI UNA “BUONA” SINTESI

- Ci possono essere modi diversi di sintetizzare i dati

Definition: Principio di sufficienza. I metodi di sintesi non devono disperdere informazioni rilevanti circa il parametro incognito θ

- Una statistica sufficiente è una particolare statistica capace di condensare $\mathcal{Y}_n$ in modo da non perdere informazioni su θ
- Le sole informazioni sul parametro θ della distribuzione $Y \sim f(y; \theta)$ dalla quale si è effettuato il campionamento sono contenute nel campione $Y_1, \dots, Y_n$
- quando si dice che una statistica è sufficiente per θ , ovvero che non perde informazioni su θ si intende che essa conserva tutte le informazioni su θ che sono contenute nel campione

SUFFICIENZA: INTUIZIONE

- Una statistica T_n è sufficiente per un parametro θ se è in grado di condensare (riassumere) tutta l'informazione che il campione fornisce su θ
- Se $T_n = T(Y_1, \dots, Y_n)$ assume in sé tutta l'informazione che il campione fornisce sul parametro θ , allora noto il valore t_n che la statistica T_n assume, il campione $y_1, \dots, y_n$ non ha null'altro da dire sul parametro, ossia non è in grado di fornire ulteriori informazioni su θ
- se T_n è sufficiente, quando è noto il valore t da essa assunto, perde importanza l'informazione puntuale fornita dalle singole componenti della n -pla campionaria $(y_1, \dots, y_n)$: ai fini di θ una statistica sufficiente sostituisce perfettamente la n -pla osservata, riassumendo in sé tutto ciò che è contenuto nella medesima a proposito di θ

SUFFICIENZA: DEFINIZIONE FORMALE

Definition: Sufficienza. Sia $Y_1, \dots, Y_n$ un campione casuale da $f(y; \theta)$. Sia inoltre $T(y_1, \dots, y_n)$ una funzione delle osservazioni $y_1, \dots, y_n$ e non di θ e sia $T(Y_1, \dots, Y_n)$ la corrispondente v.c..

T_n è allora una statistica, e T_n è sufficiente per θ se la distribuzione condizionata di $Y_1, \dots, Y_n$, dato il valore di T_n , non dipende da θ .

UN ESEMPIO “PARTICOLARE”

$n = 2$ da una popolazione bernoulliana, cinque statistiche esemplificative:

Y_1	Y_2	$T_{1n} = Y_{(1)}$	$T_{2n} = \sum_{i=1}^n Y_i$	$T_{3n} = Y_1 \times Y_2$	$T_{4n} = (Y_{(1)}, Y_{(2)})$	$T_{5n} = (Y_1, Y_2)$
0	0	0	0	0	(0,0)	(0,0)
0	1	0	1	0	(0,1)	(0,1)
1	0	0	1	0	(0,1)	(1,0)
1	1	1	2	1	(1,1)	(1,1)

- T_{1n} e T_{3n} sono equivalenti, così come T_{2n} e T_{4n}
- T_{1n} e T_{3n} non sono sufficienti per θ , a differenza di T_{2n} e T_{4n}
- T_{5n} è sufficiente (conclusione banale: non produce alcuna riduzione nei dati)

UN ESEMPIO “PARTICOLARE”

$n = 2$ da una popolazione bernoulliana, cinque statistiche esemplificative:

Y_1	Y_2	$P(Y_1 = y_1 \cap Y_2 = y_2)$	$T_{1n}(\equiv T_{3n})$	$P(T_{1n} = t)$
0	0	$(1 - \pi)^2$	0	$(1 - \pi)^2 + 2(1 - \pi)\pi$
0	1	$(1 - \pi)\pi$	0	
1	0	$\pi(1 - \pi)$	0	
1	1	π^2	1	π^2

UN ESEMPIO “PARTICOLARE”

$n = 2$ da una popolazione bernoulliana, cinque statistiche esemplificative

$$\begin{aligned} P(Y_1 = y_1 \cap Y_2 = y_2 \mid T_{1n} = t) &= \frac{P(Y_1 = y_1 \cap Y_2 = y_2; T_{1n} = t)}{P(T_{1n} = t)} \\ &= \frac{P(Y_1 = y_1 \cap Y_2 = y_2)}{P(T_{1n} = t)} \end{aligned}$$

NOTA:

$P(Y_1 = y_1 \cap Y_2 = y_2 \mid T_{1n} = t) = 0$ se $T_{1n} \neq t$ e vale $P(Y_1 = y_1 \cap Y_2 = y_2)$ se $T_{1n} = t$

(ciò è ovvio in quanto noti y_1 e y_2 è anche noto T_{1n} , e quindi l'evento (y_1, y_2) implica il valore di T_{1n})

UN ESEMPIO “PARTICOLARE”

$n = 2$ da una popolazione bernoulliana, cinque statistiche esemplificative

Y_1	Y_2	$P(Y_1 = y_1 \cap Y_2 = y_2)$	$T_{1n}(\equiv T_{3n})$	$P(T_{1n} = t)$
0	0	$(1 - \pi)^2$	0	$(1 - \pi)^2 + 2(1 - \pi)\pi$
0	1	$(1 - \pi)\pi$	0	
1	0	$\pi(1 - \pi)$	0	
1	1	π^2	1	π^2

- $P(Y_1 = 0 \cap Y_2 = 0 \mid T_{1n} = 0) = \frac{(1-\pi)^2}{(1-\pi)^2 + 2(1-\pi)\pi} = \frac{(1-\pi)}{(1-\pi) + 2\pi}$
- $P(Y_1 = 0 \cap Y_2 = 1 \mid T_{1n} = 0) = \frac{(1-\pi)\pi}{(1-\pi)^2 + 2(1-\pi)\pi} = \frac{\pi}{(1-\pi) + 2\pi}$
- $P(Y_1 = 1 \cap Y_2 = 0 \mid T_{1n} = 0) = \frac{\pi}{(1-\pi) + 2\pi}$
- $P(Y_1 = 1 \cap Y_2 = 1 \mid T_{1n} = 1) = \frac{\pi^2}{\pi^2} = 1$

UN ESEMPIO “PARTICOLARE”

$n = 2$ da una popolazione bernoulliana, cinque statistiche esemplificative

Y_1	Y_2	$P(Y_1 = y_1 \cap Y_2 = y_2)$	$T_{2n} (\equiv T_{4n})$	$P(T_{2n} = t)$
0	0	$(1 - \pi)^2$	0	$(1 - \pi)^2$
0	1	$(1 - \pi)\pi$	1	$2(1 - \pi)\pi$
1	0	$\pi(1 - \pi)$	1	
1	1	π^2	2	π^2

- $P(Y_1 = 0 \cap Y_2 = 0 \mid T_{2n} = 0) = \frac{(1-\pi)^2}{(1-\pi)^2} = 1$
- $P(Y_1 = 0 \cap Y_2 = 1 \mid T_{2n} = 1) = \frac{(1-\pi)\pi}{2(1-\pi)\pi} = \frac{1}{2}$
- $P(Y_1 = 1 \cap Y_2 = 0 \mid T_{2n} = 1) = \frac{\pi(1-\pi)}{2(1-\pi)\pi} = \frac{1}{2}$
- $P(Y_1 = 1 \cap Y_2 = 1 \mid T_{2n} = 2) = \frac{\pi^2}{\pi^2} = 1$

UN ESEMPIO “PARTICOLARE”

$n = 2$ da una popolazione bernoulliana, cinque statistiche esemplificative

Y_1	Y_2	$P(Y_1 = y_1 \cap Y_2 = y_2)$	T_{5n}	$P(T_{5n} = t)$
0	0	$(1 - \pi)^2$	$(0, 0)$	$(1 - \pi)^2$
0	1	$(1 - \pi)\pi$	$(0, 1)$	$(1 - \pi)\pi$
1	0	$\pi(1 - \pi)$	$(1, 0)$	$\pi(1 - \pi)$
1	1	π^2	$(1, 1)$	π^2

- $P(Y_1 = y_1 \cap Y_2 = y_2 \mid T_{5n} = t) = 1$

UN ESEMPIO “PARTICOLARE”: TIRIAMO LE “SOMME”

- $T_{5n} = (Y_1, Y_2)$ è in via naturale una statistica sufficiente per π
- $T_{2n} = Y_1 + Y_2$ sintetizza la coppia (Y_1, Y_2) ma salvaguarda comunque tutte le informazioni circa π
- $T_{4n} = (Y_{(1)}, Y_{(2)})$ effettua una sintesi della stessa coppia (Y_1, Y_2) meno ampia della precedente T_{2n} e salvaguarda a sua volta tutte le informazioni circa π
- $T_{1n} = Y_{(1)}$ riducendo la coppia (Y_1, Y_2) alla sola componente $Y_{(1)}$ provoca una perdita di informazioni circa π e quindi non risulta sufficiente
- lo stesso avviene per $T_{3n} = Y_1 \times Y_2$ che pur sfruttando sia Y_1 che Y_2 induce la stessa partizione di T_{1n}
- T_{1n} e T_{3n} sono equivalenti (inducendo la stessa partizione di $\mathcal{Y}_n$) e assumono gli stessi valori
- T_{2n} e T_{4n} sono equivalenti ma assumono differenti valori (e dimensioni)

UN ESEMPIO “GENERALE”

campione da una popolazione bernoulliana, $Y_i \sim Ber(\pi)$, $T = \sum_{i=1}^n Y_i$

$$\begin{aligned} g(y_1, \dots, y_n \mid T_n = t) &= \frac{\psi(y_1, \dots, y_n)}{h(T_n = t)} \\ &= \frac{\prod_{i=1}^n \pi^{y_i} (1 - \pi)^{1-y_i}}{\binom{n}{t} \pi^t (1 - \pi)^{n-t}} \\ &= \frac{\pi^{\sum_{i=1}^n y_i} (1 - \pi)^{n - \sum_{i=1}^n y_i}}{\binom{n}{t} \pi^t (1 - \pi)^{n-t}} \\ &= \frac{\pi^t (1 - \pi)^{n-t}}{\binom{n}{t} \pi^t (1 - \pi)^{n-t}} = \frac{1}{\binom{n}{t}} \end{aligned}$$

il numero totale di 1 nel campione bernoulliano contiene tutta l'informazione su π che è nei dati: altre caratteristiche dei dati, come ad esempio il valore effettivo di una qualunque Y_i , non apportano informazioni aggiuntive

UN ALTRO ESEMPIO “GENERALE”

popolazione normale, $Y_i \sim N(\mu, \sigma^2)$, σ^2 nota, $T_n = \bar{Y}_n = \frac{\sum_{i=1}^n Y_i}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$\begin{aligned}\psi(y_1, \dots, y_n) &= \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{(y_i - \mu)^2}{2\sigma^2}\right\} \\&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\sum_{i=1}^n \frac{(y_i - \mu)^2}{2\sigma^2}\right\} \\&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\sum_{i=1}^n \frac{(y_i - \bar{y}_n + \bar{y}_n - \mu)^2}{2\sigma^2}\right\} \\&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\sum_{i=1}^n \frac{(y_i - \bar{y}_n)^2 + n(\bar{y}_n - \mu)^2}{2\sigma^2}\right\}\end{aligned}$$

UN ALTRO ESEMPIO “GENERALE”

popolazione normale, $Y_i \sim N(\mu, \sigma^2)$, σ^2 nota, $T_n = \bar{Y}_n = \frac{\sum_{i=1}^n Y_i}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$\begin{aligned} g(y_1, \dots, y_n \mid T_n = t) &= \frac{\psi(y_1, \dots, y_n)}{h(T_n = t)} \\ &= \frac{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\sum_{i=1}^n \frac{(y_i - \bar{y}_n)^2 + n(\bar{y}_n - \mu)^2}{2\sigma^2}\right\}}{(2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-n \frac{(\bar{y}_n - \mu)^2}{2\sigma^2}\right\}} \\ &= n^{-\frac{1}{2}} (2\pi\sigma^2)^{-\frac{n-1}{2}} \exp\left\{-\sum_{i=1}^n \frac{(y_i - \bar{y}_n)^2}{2\sigma^2}\right\} \end{aligned}$$

CHI È “PIÙ SUFFICIENTE”?

- possono esserci diverse statistiche sufficienti in un dato problema
- come stabilire se una statistica sufficiente è migliore di un'altra statistica sufficiente?
- una statistica capace di offrire la massima riduzione possibile senza perdita di informazioni circa θ potrebbe essere considerata preferibile

Definition: Sufficienza minimale. Una statistica è sufficiente minimale se può essere espressa come funzione di ogni altra statistica sufficiente. Ovvero: una statistica T_n è una statistica sufficiente minimale se, per qualsiasi altra statistica sufficiente T'_n , T_n è una funzione di T'_n .

UNA SUFFICIENZA ALTERNATIVA...

- la definizione di statistica sufficiente potrebbe essere poco maneggevole
- è infatti necessario ipotizzare che una statistica T_n sia sufficiente, trovare la distribuzione di probabilità di T_n e controllare che il rapporto tra $\psi(y_1, \dots, y_n)$ e $h(t)$ non dipenda da θ

UNA SUFFICIENZA ALTERNATIVA...

Esistono differenti risultati (o forse differenti versioni di unico risultato) che permettono di trovare una statistica sufficiente semplicemente analizzando la $\psi(y_1, \dots, y_n)$ dal campione

Definition: Teorema / criterio fattorizzazione Neyman (1935).

Si supponga che $Y_1, \dots, Y_n$ sia un campione casuale da $f(y; \theta)$.

Allora $T_n = T(Y_1, \dots, Y_n)$ è una statistica sufficiente per θ se e solo se esistono due funzioni non negative tali che la funzione di verosimiglianza $\mathcal{L}(\theta, \mathbf{y})$ possa essere scritta come:

$$\mathcal{L}(\theta, \mathbf{y}) = K_1[t(y_1, \dots, y_n; \theta)] K_2[y_1, \dots, y_n]$$

dove $K_1[t(y_1, \dots, y_n; \theta)]$ dipende da θ e dalla n -pla osservata $(y_1, \dots, y_n)$ solo attraverso la sintesi t , mentre $K_2[y_1, \dots, y_n]$ non dipende da θ ma solo dalla n -pla osservata $(y_1, \dots, y_n)$ e può anche coincidere con una costante.

UNA SUFFICIENZA ALTERNATIVA...

- il concetto di sufficienza (e il criterio di fattorizzazione) può essere facilmente esteso al caso di più di un parametro
- se θ è un vettore di k parametri e la statistica $\mathbf{T}_n$ è un vettore di m statistiche scalari, allora $\mathbf{T}_n$ è sufficiente per θ se e solo se $g(Y_1, \dots, Y_n \mid \mathbf{T}_n)$ non dipende da θ
- una condizione equivalente è che $\mathcal{L}(\theta, \mathbf{y})$ possa essere scritta come:

$$\mathcal{L}(\theta, \mathbf{y}) = K_1[\mathbf{t}; \theta] K_2[\mathbf{y}]$$

- il numero m di elementi in $\mathbf{T}_n$ non deve essere necessariamente k , ovvero al numero di parametri
- se si è interessati alla sufficienza minimale si avrà di solito $m = k$ sebbene $m < k$ o $m > k$ possano occasionalmente presentarsi

UNA SUFFICIENZA ALTERNATIVA: ESEMPIO 1

popolazione Bernoulli: $Y \sim Ber(\pi)$, $T_n = \sum_{i=1}^n Y_i$

$$\begin{aligned}\mathcal{L}(\theta, \mathbf{y}) &= \prod_{i=1}^n f(y_i, \theta) \\ &= \prod_{i=1}^n \pi^{y_i} (1 - \pi)^{(1-y_i)} \\ &= \pi^{\sum_i y_i} (1 - \pi)^{(n - \sum_i y_i)}\end{aligned}$$

Ponendo:

$$t = \sum_i y_i; \quad K_1[t(y_1, \dots, y_n; \theta)] = \pi^{\sum_i y_i} (1 - \pi)^{(n - \sum_i y_i)}; \quad K_2[y_1, \dots, y_n] = 1$$

discende immediatamente che $T = \sum_{i=1}^n Y_i$ è una statistica sufficiente per π

UNA SUFFICIENZA ALTERNATIVA: ESEMPIO 2

popolazione Normale: $Y \sim N(\mu, \sigma^2)$

$$\begin{aligned}\mathcal{L}(\theta; \mathbf{y}) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n y_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n y_i - \frac{n\mu^2}{2\sigma^2} \right\}\end{aligned}$$

UNA SUFFICIENZA ALTERNATIVA: ESEMPIO 2

popolazione Normale: $Y \sim N(\mu, \sigma^2)$

$$\mathcal{L}(\theta; \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n y_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n y_i - \frac{n\mu^2}{2\sigma^2} \right\}$$

CASO 1: μ incognito, σ^2 noto

$$t = \sum_i y_i; \quad K_1[t(\mathbf{y}; \mu)] = \exp \left\{ \frac{\mu t}{\sigma^2} - \frac{n\mu^2}{2\sigma^2} \right\}; \quad K_2[\mathbf{y}] = \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n y_i^2 \right\}$$

$T = \sum_{i=1}^n Y_i$ (o equivalentemente $\bar{Y}_n$) è una statistica sufficiente per μ

UNA SUFFICIENZA ALTERNATIVA: ESEMPIO 2

popolazione Normale: $Y \sim N(\mu, \sigma^2)$

$$\mathcal{L}(\theta; \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n y_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n y_i - \frac{n\mu^2}{2\sigma^2} \right\}$$

CASO 2: μ noto, σ^2 incognito

$$t = \sum_i (y_i - \mu)^2; \quad K_1[t(\mathbf{y}; \sigma^2)] = \mathcal{L}(\sigma^2; \mathbf{y}); \quad K_2[\mathbf{y}] = 1$$

$T = \sum_{i=1}^n (Y_i - \mu)^2$ è una statistica sufficiente per σ^2

UNA SUFFICIENZA ALTERNATIVA: ESEMPIO 2

popolazione Normale: $Y \sim N(\mu, \sigma^2)$

$$\mathcal{L}(\theta; \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n y_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n y_i - \frac{n\mu^2}{2\sigma^2} \right\}$$

CASO 3: μ incognito, σ^2 incognito

$$t_1 = \sum_i y_i; \quad t_2 = \sum_i y_i^2; \quad \mathbf{t} = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \quad K_1[\mathbf{t}; \sigma^2] = \mathcal{L}(\theta; \mathbf{y}); \quad K_2[\mathbf{y}] = 1$$

Il vettore di statistiche $T = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}$ è sufficiente per $\theta = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$

UNA SUFFICIENZA ALTERNATIVA: ESEMPIO 2

- la singola statistica sufficiente per σ^2 (caso 2) non è una delle due statistiche che sono congiuntamente sufficienti per μ e σ^2 (caso 3)
- presa singolarmente nessuna delle due statistiche mostrate al caso 3, congiuntamente sufficienti per μ e σ^2 , è sufficiente per σ^2 (chiaramente, se considerate insieme, lo sono)
- la singola statistica sufficiente per μ (caso 1) è una della coppia mostrata per il caso 2

CHI È “PIÙ SUFFICIENTE”?

Definition: Sufficienza minimale. Una statistica è sufficiente minimale se può essere espressa come funzione di ogni altra statistica sufficiente. Ovvero: una statistica T_n è una statistica sufficiente minimale se, per qualsiasi altra statistica sufficiente T'_n , T_n è una funzione di T'_n .

- dire che T_n è una funzione di T'_n significa semplicemente dire che se $T'_n(\mathbf{x}) = T'_n(\mathbf{y})$, allora $T_n(\mathbf{x}) = T_n(\mathbf{y})$
- in termini delle partizioni indotte su $\mathcal{Y}_n$, se $\{B_{t'} : t' \in \mathcal{T}'\}$ sono gli insiemi delle partizioni indotte da T'_n su $\mathcal{Y}_n$, e $\{A_t : t \in \mathcal{T}\}$ sono quelli indotti da T_n , la definizione afferma che ogni $B_{t'}$ è un sottoinsieme di qualche A_t
- la partizione associata con una statistica sufficiente minimale è perciò la partizione più fine possibile per una statistica sufficiente
- una statistica sufficiente minimale consegue la massima riduzione possibile nei dati per una statistica sufficiente