Concetti di teoria dei campioni ad uso degli studenti di Statistica Economica e Finanziaria, A.A. 2016/2017

Giovanni Lafratta

Indice

Capitolo 1

Spazi, Disegni e Strategie Campionarie

Si introduce il concetto di popolazione finita.

Definizione 1 ***(Popolazione finita*** $\mathcal{P}_M$***)***

1. $\mathcal{P}_1 = \{1\}$ *è una popolazione finita;*
2. *per ogni* $M \in \mathbb{N} \setminus \{0\}$ *se* $\mathcal{P}_M$ *è una popolazione finita, allora*

$$\mathcal{P}_{M+1} = \mathcal{P}_M \cup \{M+1\}$$

   è una popolazione finita;

3. *le uniche popolazioni finite sono quelle formate sulla base delle clausole 1 e 2.*

Esempio 1 *Per* $M = 4$*, si ha*

$$\mathcal{P}_4 = \{1, 2, 3, 4\}.$$

Spazi e disegni di campionamento costituiscono strumenti utili nello studio della variabilità statisticamente derivante dal non aver osservato in modo esaustivo un dato fenomeno su una popolazione $\mathcal{P}_M$:

Definizione 2 ***(Spazio campionario*** $S_{\mathcal{P}_M}$***)*** *Sia* $\mathcal{P}_M$ *una popolazione finita, e, per ogni* $n \in \mathcal{P}_M$*, sia* S_n *la classe dei sottoinsiemi di* $\mathcal{P}_M$ *aventi cardinalità pari a* n*. Per spazio campionario su* $\mathcal{P}_M$ *si intenderà la classe*

$$S_{\mathcal{P}_M} = \bigcup_{n \in \mathcal{P}_M} S_n.$$

Per ***campione*** s *in* $\mathcal{P}_M$ *si intenderà ogni elemento* $s \in S_{\mathcal{P}_M}$.

Esempio 2 *Per* $M = 3$*, si ha*

$$S_{\mathcal{P}_M} = \{\{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \mathcal{P}_3\}.$$

Definizione 3 *(**Disegno campionario** p) Sia $S_{\mathcal{P}_M}$ uno spazio di campionamento. Una funzione $p : S_{\mathcal{P}_M} \to [0, +\infty[$ è un disegno campionario su $S_{\mathcal{P}_M}$ se e solo se risulta*

$$\sum_{s \in S_{\mathcal{P}_M}} p(s) = 1.$$

*Un disegno p si dirà **intenzionale** se e solo se*

$$\exists s \in S_{\mathcal{P}_M} : p(s) = 1.$$

Un disegno p può quindi essere interpretato come la funzione di probabilità di una variabile casuale S a valori in $S_{\mathcal{P}_M}$.

Esempio 3 *Si consideri la seguente funzione* $p : S_{\mathcal{P}_3} \to [0, +\infty[$*:*

$s \in S_{\mathcal{P}_3}$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\mathcal{P}_3$
$p(s)$	0	0	1/4	1/3	5/12	0	0

Essa è un disegno, perché $1/4 + 1/3 + 5/12 = 1$.

Definizione 4 *(**Probabilità d'inclusione I**) Per $i \in \mathcal{P}_M$, si definiscano gli insiemi*

$$A_i = \{s \in S_{\mathcal{P}_M} : i \in s\}.$$

*La probabilità π_i di inclusione **del primo ordine** dell'unità $i-esima$ è definita come segue:*

$$\pi_i = \sum\nolimits_{s \in A_i} p(s).$$

Esempio 4 *(**Continua l'esempio 3**) L'unità* 2 *è inclusa nei seguenti elementi (campioni) dello spazio campionario* $S_{\mathcal{P}_3}$ *:*

$$A_2 = \{\{2\}, \{1,2\}, \{2,3\}, \mathcal{P}_3\}$$

La probabilità d'inclusione di ordine 1 per l'unità 2 è quindi

$$\begin{aligned}
\pi_2 &= p(\{2\}) + p(\{1,2\}) + p(\{2,3\}) + p(\mathcal{P}_3) \\
&= 0 + \frac{1}{3} + 0 + 0 \\
&= \frac{1}{3}.
\end{aligned}$$

L'unità 1 *è inclusa nei seguenti elementi (campioni) dello spazio campionario* $S_{\mathcal{P}_3}$ *:*

$$A_1 = \{\{1\}, \{1,2\}, \{1,3\}, \mathcal{P}_3\}.$$

La probabilità d'inclusione di ordine 1 *per l'unità* 1 *è quindi*

$$\begin{aligned}\pi_1 &= p(\{1\}) + p(\{1,2\}) + p(\{1,3\}) + p(\mathcal{P}_3) \\ &= 0 + \frac{1}{3} + \frac{5}{12} + 0 \\ &= \frac{9}{12}.\end{aligned}$$

Per ogni $i \in \mathcal{P}_M$, si definisca ora la funzione $\delta_i : S_{\mathcal{P}_M} \longrightarrow \{0,1\}$ come segue:

$$\delta_i(s) = \begin{cases} 1 & \text{se } i \in s, \\ 0 & \text{altrimenti.} \end{cases}$$

La funzione δ_i prende il nome di *funzione indicatrice dell'unità i-esima*, grazie alla quale si può definire la variabile casuale

$$\Delta_i = \delta_i(S)$$

e si ha

$$\begin{aligned}\pi_i &= \sum\nolimits_{s \in S_{\mathcal{P}_M}} \delta_i(s)\, p(s) \\ &= E_p[\Delta_i].\end{aligned}$$

Definizione 5 *(**Cardinalità campionaria**) Dato il campione casuale S, si definisce come cardinalità campionaria la variabile*

$$|S| = \sum\nolimits_{i \in \mathcal{P}_M} \delta_i(S).$$

La cardinalità attesa è

$$\begin{aligned}E_p[|S|] &= \sum_{s \in S_{\mathcal{P}_M}} \sum_{i \in \mathcal{P}_M} \delta_i(s)\, p(s) \\ &= \sum_{i \in \mathcal{P}_M} \sum_{s \in A_i} p(s) \\ &= \sum_{i \in \mathcal{P}_M} \pi_i.\end{aligned}$$

Inoltre, se il disegno campionario p ammette l'estrazione solo di campioni aventi una data cardinalità $n \in \mathcal{P}_M$, ovvero se

$$\forall s \in S_{\mathcal{P}_M} : |s| \neq n \Rightarrow p(s) = 0,$$

allora
$$E_p\left[|S|\right]=\sum_{s\in S_n}|s|p\left(s\right)=n,$$
perché $\sum_{s\in S_n}p\left(s\right)=1$ e $|s|=n$ quando $s\in S_n$.

Definizione 6 *(**Probabilità d'inclusione II**) Per $i,j\in\mathcal{P}_M$, si definiscano gli insiemi*
$$A_{i,j}=\left\{s\in S_{\mathcal{P}_M}:\{i,j\}\subset s\right\}.$$
*La probabilità $\pi_{i,j}$ di inclusione **del secondo ordine** delle unità $\{i,j\}$ è definita come segue:*
$$\pi_{i,j}=\sum\nolimits_{s\in A_{i,j}}p\left(s\right).$$

Esempio 5 *(**Continua l'esempio 3**) Le unità $1,3$ sono incluse nei seguenti elementi (campioni) dello spazio campionario $S_{\mathcal{P}_M}$:*
$$A_{1,3}=\left\{\{1,3\},\mathcal{P}_3\right\}$$
La loro probabilità d'inclusione di ordine 2 è quindi
$$\begin{aligned}\pi_{1,3}&=p\left(\{1,3\}\right)+p\left(\mathcal{P}_3\right)\\&=\frac{5}{12}+0\\&=\frac{5}{12}.\end{aligned}$$

Per ogni $i,j\in\mathcal{P}_M$, si definisca ora la funzione $\delta_{i,j}:S_{\mathcal{P}_M}\longrightarrow\{0,1\}$ come segue:
$$\delta_{i,j}\left(s\right)=\begin{cases}1 & \text{se } \{i,j\}\subset s,\\ 0 & \text{altrimenti.}\end{cases}$$

La funzione $\delta_{i,j}$ è detta *funzione indicatrice della coppia di unità i e j* e si ha
$$\delta_{i,j}\left(s\right)=\delta_i\left(s\right)\delta_j\left(s\right).$$

Definita la variabile
$$\Delta_{i,j}=\delta_{i,j}\left(S\right)$$
risulta inoltre
$$\begin{aligned}\pi_{i,j}&=\sum\nolimits_{s\in S_{\mathcal{P}_M}}\delta_{i,j}\left(s\right)p\left(s\right)\\&=E_p\left[\Delta_{i,j}\right].\end{aligned}$$

Associando ad ogni elemento di $\mathcal{P}_M$ un numero reale Y_i, $i = 1, \ldots, N$, si ottiene la funzione Y come segue:

$$i \in \mathcal{P}_M \mapsto Y(i) = Y_i,$$

alla quale si fa riferimento parlando di **fenomeno su** $\mathcal{P}_M$. Dato il campione s, sia D_s la funzione definita su s e a valori in $\mathbb{R}$ tale che

$$i \in s \mapsto D_s(i) = Y_i.$$

Per definizione, D_s è quindi l'insieme delle coppie (i, Y_i) per $i \in s$:

$$D_s = \{(i, Y_i) : i \in s\}.$$

Ci si riferisce a D_s come all'insieme dei **dati statistici** relativi al campione s.

Esempio 6 ***(Continua l'esempio 3)*** *Posto*

i	*1*	*2*	*3*
Y_i	*5*	*7*	*5*

risultano, ad esempio,

$$D_{\{1\}} = \{(1, 5)\},$$

$$D_{\{1,3\}} = \{(1, 5), (3, 5)\}$$

e

$$D_{\{1,2,3\}} = \{(1, 5), (2, 7), (3, 5)\}.$$

Definizione 7 ***(Statistica campionaria)*** *Sia Y un fenomeno su $\mathcal{P}_M$ e sia D l'insieme dei possibili dati statistici disponibili:*

$$D = \{D_s : s \in S_{\mathcal{P}_M}\}.$$

Data una funzione reale t definita su D:

$$D_s \in D \mapsto t(D_s) \in \mathbb{R},$$

si definisce statistica campionaria la variabile casuale

$$T = t(D_S).$$

Si noti che i valori assunti da una statistica T dipendono dal campione osservato. Di conseguenza, $t(D_S)$ dipende dalla funzione di probabilità p, il disegno campionario applicato.

Esempio 7 *La variabile $T = t(D_S) = \sum_{i \in S} Y_i$ è una statistica.*

Una statistica rispondente alla definizione 7 è destinata a fungere da stimatore di quantità deterministiche dipendenti da Y, note come **parametri della popolazione**, delle quali sono esempi la variabile media della popolazione $\overline{Y} = M^{-1}\sum_{i\in\mathcal{P}_M} Y_i$, o la variabile $\max(Y) = \sup_{i\in\mathcal{P}_M} Y_i$.

Definizione 8 ***(Distribuzione di una statistica campionaria)*** *Per un fissato piano di campionamento $p(s)$, la distribuzione campionaria dello stimatore $T = t(D_S)$ è definita dalla seguente funzione di probabilità p_T. Sia C_T l'insieme di tutti i possibili valori che T può assumere presso i vari campioni. Per $x \in C_T$, si definisca l'insieme*

$$A_x = \{s \in S_{\mathcal{P}_M} : t(D_s) = x\}.$$

Allora

$$p_T(x) \equiv \Pr\{T = x\} = \sum\nolimits_{s\in A_x} p(s).$$

Esempio 8 ***(Continua l'esempio 3)*** *Si vuole la distribuzione campionaria di $T = t(D_S) \equiv \sum_{i\in S} Y_i$. I valori che T può assumere presso i relativi campioni sono dati come segue:*

s	D_s	$t(D_s)$
$\{1\}$	$\{(1,5)\}$	5
$\{2\}$	$\{(2,7)\}$	7
$\{3\}$	$\{(3,5)\}$	5
$\{1,2\}$	$\{(1,5),(2,7)\}$	12
$\{1,3\}$	$\{(1,5),(3,5)\}$	10
$\{2,3\}$	$\{(2,7),(3,5)\}$	12
$\mathcal{P}_3$	$\{(1,5),(2,7),(3,5)\}$	17

Al singolo valore di T (si ha $C_T = \{5, 7, 10, 12, 17\}$) va associata una probabilità pari alla somma delle probabilità dei campioni in corrispondenza dei quali tale valore viene generato:

x	A_x	$\Pr\{T = x\}$
5	$\{\{1\},\{3\}\}$	$0 + 1/4$
7	$\{\{2\}\}$	0
10	$\{\{1,3\}\}$	$5/12$
12	$\{\{1,2\},\{2,3\}\}$	$1/3 + 0$
17	$\{\{1,2,3\}\}$	0

In conclusione, gli strumenti con cui raccogliere informazioni su un fenomeno Y presso una popolazione $\mathcal{P}_M$ sono sostanzialmente due: un disegno di campionamento e una statistica. Ciò giustifica la seguente:

Definizione 9 ***(Strategia campionaria per Y su $\mathcal{P}_M$)*** *Sia Y un fenomeno su $\mathcal{P}_M$. Una strategia per Y su $\mathcal{P}_M$ è una qualunque coppia (p, T), con p disegno campionario su $S_{\mathcal{P}_M}$ e T statistica per Y.*

La realizzazione di una indagine richiede, quindi, la scelta di una strategia.

Capitolo 2

Campionamento casuale semplice

2.1 Conteggio di sottoinsiemi di una popolazione finita

Si cosideri $S_n = \{s \in S_{\mathcal{P}_M} : |s| = n\}$. Si vuole calcolare $|S_n|$: quanti sono i sottoinsiemi di cardinalità n di un insieme di cardinalità M?

Cardinalità di un prodotto cartesiano

Siano $A_1, \ldots, A_k$ insiemi di cardinalità finita $|A_i|$. Il loro prodotto cartesiano è definito come segue:

$$\mathop{\times}_{i \in \{1,\ldots,k\}} A_i = \{(a_1, \ldots, a_k) : \forall i \in \{1, \ldots, k\} : a_i \in A_i\}.$$

La sua cardinalità è:

$$\left| \mathop{\times}_{i \in \{1,\ldots,k\}} A_i \right| = \prod_{i \in \{1,\ldots,k\}} |A_i|.$$

Ennuple ordinate di elementi distinti di una popolazione finita

Il prodotto cartesiano della popolazione $\mathcal{P}_M$ moltiplicata per se stessa n volte ha dunque cardinalità pari a M^n. Il prodotto ha per elementi tutte le ennuple ordinate (vettori) $(a_1, \ldots, a_n)$ che si possono formare scegliendo per a_i una qualunque unità nella popolazione $\mathcal{P}_M$. Questo significa che, all'interno di una ennupla, la stessa unità può ripetersi ed essere quindi presente in più di una delle n posizioni disponibili. Quale sarebbe il numero delle ennuple se si imponesse che ogni unità non può occupare più di una posizione in una data ennupla?

Anche l'insieme delle ennuple ordinate di elementi distinti può essere interpretato come un prodotto cartesiano tra n insiemi: l'i-esimo insieme, sia A_i, sarà la collezione delle possibili scelte da $\mathcal{P}_M$ per l'i-esima componente della ennupla, avendo cura che le unità selezionate siano distinte tra loro. La prima volta ($i = 1$) posso scegliere su tutto $\mathcal{P}_M$: $A_1 = \mathcal{P}_M$ avente cardinalità M. La seconda volta ($i = 2$) da $\mathcal{P}_M \setminus \{a_1\}$, con cardinalità

$M-1$. La terza volta ($i=3$) da $\mathcal{P}_M \setminus \{a_1, a_2\}$, con cardinalità $M-2$. L'n-esima volta da $\mathcal{P}_M \setminus \{a_1, \ldots, a_{n-1}\}$, un insieme avente cardinalità $M-(n-1)$. In generale si hanno

$$M\,(M-1)\,(M-2)\cdots(M-(n-2))\,(M-(n-1))$$

$n-$ple ordinate di elementi distinti da $\mathcal{P}_M$.

Tale numero si può esprimere come

$$\frac{M!}{(M-n)!}$$

dove $n! = n\,(n-1)\cdots(2)\,(1)$. Infatti

$$\begin{aligned}\frac{M!}{(M-n)!} &= \frac{M\,(M-1)\cdots(M-(n-1))\,(M-n)\cdots(1)}{(M-n)\cdots(1)}\\ &= M\,(M-1)\cdots(M-(n-1))\,.\end{aligned}$$

Sottoinsiemi di ampiezza n

Un sottoinsieme di ampiezza n dalla popolazione $\mathcal{P}_M$ è un campione di n unità disposte senza alcun ordine particolare.

Sia ora $s \in S_n$. Quante $n-$ple ordinate posso ottenere da s ?

Esattamente

$$\frac{n!}{(n-n)!} = n!.$$

La collezione di **tutte** le $n-$ple di elementi distinti da $\mathcal{P}_M$ sono invece

$$\frac{M!}{(M-n)!}.$$

Quindi, se $|S_n|$ è il numero di sottoinsiemi di ampiezza n, deve risultare

$$|S_n|\;\; n! = \frac{M!}{(M-n)!}$$

da cui

$$\begin{aligned}|S_n| &= \frac{M!}{n!\,(M-n)!}\\ &= \binom{M}{n},\end{aligned}$$

numero di combinazioni di M elementi presi a n a n.

2.2 Disegno casuale semplice di ampiezza data

Data una popolazione $\mathcal{P}_M$, esistono M disegni casuali semplici senza ripetizione, uno per ogni $n \in \mathcal{P}_M$, ciascuno di essi rende equiprobabili i campioni di ampiezza n ed assegna ai rimanenti probabilità zero:

$$p_{C,M,n}(s) = \begin{cases} \frac{1}{\binom{M}{n}} & \text{se } s \in S_n \\ 0 & \text{se } s \notin S_n \end{cases}$$

Esempio 9 *Per* $\mathcal{P}_4 = \{1,2,3,4\}$, *si consideri* $p_{C,4,3}$. *I campioni di ampiezza 3 da* $\mathcal{P}_4$ *sono* $\binom{4}{3} = \frac{4!}{(4-3)!3!} = 4$:

$$\{1,2,3\}, \{1,2,4\}, \{2,3,4\}, \{1,3,4\}.$$

Si ha

$$\begin{aligned} p_{C,4,3}(\{1,2,3\}) &= p_{C,4,3}(\{1,2,4\}) \\ &= p_{C,4,3}(\{2,3,4\}) \\ &= p_{C,4,3}(\{1,3,4\}) \\ &= \frac{1}{4} \end{aligned}$$

Per tutti gli altri elementi s *di* $S_{\mathcal{P}_4}$, *ad esempio per* $\{1,3\}, \{1\}, \{3,4\}$, *si ha* $p_{C,4,3}(s) = 0$.

2.2.1 Probabilità di inclusione

- **Primo ordine** - generica unità $i \in \mathcal{P}_M$:

$$\pi_i = \sum\nolimits_{s \in \left\{s \in S_{\mathcal{P}_M} : i \in s\right\}} p_{C,M,n}(s)$$

  - **Ma**, se $s \notin S_n$, $p_{C,M,n}(s) = 0$, così possiamo limitare la somma agli s in S_n :

$$\pi_i = \sum\nolimits_{s \in \{s \in S_n : i \in s\}} p_{C,M,n}(s)$$

  - Inoltre, se $s \in S_n$, $p_{C,M,n}(s) = \binom{M}{n}^{-1}$, quindi

$$\pi_i = \binom{M}{n}^{-1} \sum\nolimits_{s \in \{s \in S_n : i \in s\}} 1$$

  - Bisogna quindi contare i campioni di ampiezza n che includono i:
    - Si costruiscano tutti i campioni di $n-1$ unità da $\mathcal{P}_M$ che **non** contengono i : essi sono in numero di

$$\binom{M-1}{n-1}$$

- * Se ad ognuno di tali campioni aggiungiamo l'unità $i - esima$, otteniamo tutti i campioni di ampiezza n che contengono i.

- QUINDI:

$$\begin{aligned}\pi_i &= \binom{M}{n}^{-1} \sum\nolimits_{s\in\{s\in S_n : i\in s\}} 1 \\ &= \binom{M}{n}^{-1} \left|\{s\in S_n : i\in s\}\right| \\ &= \frac{\binom{M-1}{n-1}}{\binom{M}{n}}\end{aligned}$$

ovvero

$$\begin{aligned}\pi_i &= \frac{(M-1)!}{(M-n)!\,(n-1)!}\frac{n!\,(M-n)!}{M!} \\ &= \frac{n}{M}\end{aligned}$$

- **Secondo ordine** - generica coppia $(i, j) \in \mathcal{P}_M \times \mathcal{P}_M$, con $i \neq j$. Si ha

$$\pi_{i,j} = \sum\nolimits_{s\in\left\{s\in S_{\mathcal{P}_M} : i\in s, j\in s\right\}} p_{C,M,n}(s)$$

- **Ma**, se $s \notin S_n$, $p_{C,M,n}(s) = 0$, così possiamo limitare la somma agli s in S_n :

$$\pi_i = \sum\nolimits_{s\in\{s\in S_n : i\in s, j\in s\}} p_{C,M,n}(s)$$

- Inoltre, se $s \in S_n$, $p_{C,M,n}(s) = \binom{M}{n}^{-1}$, quindi

$$\pi_i = \binom{M}{n}^{-1} \sum\nolimits_{s\in\{s\in S_n : i\in s, j\in s\}} 1$$

- Bisogna quindi contare i campioni di ampiezza n che includono sia i sia j:
  * Si costruiscano tutti i campioni di $n-2$ unità da $\mathcal{P}_M$ che **non** contengono né i né j: essi sono in numero di

$$\binom{M-2}{n-2}$$

  * Se ad ognuno di tali campioni aggiungiamo le unità i e j, otteniamo tutti i campioni di ampiezza n che contengono i e j.

– QUINDI:

$$\begin{aligned}
\pi_{i,j} &= \binom{M}{n}^{-1} \sum\nolimits_{s \in \{s \in S_n : i \in s, j \in s\}} 1 \\
&= \binom{M}{n}^{-1} |\{s \in S_n : i \in s, j \in s\}| \\
&= \frac{\binom{M-2}{n-2}}{\binom{M}{n}}
\end{aligned}$$

ovvero

$$\begin{aligned}
\pi_{i,j} &= \frac{(M-2)!}{(M-n)!\,(n-2)!} \frac{n!\,(M-n)!}{M!} \\
&= \frac{n\,(n-1)}{M\,(M-1)}.
\end{aligned}$$

Capitolo 3

Statistiche di Horvitz-Thompson

3.1 Stimatore del totale di una popolazione

Definizione 10 *Sia $\mathcal{P}_M$ una popolazione finita sulla quale è rilevabile un fenomeno Y e p un disegno di campionamento su $\mathcal{P}_M$. Per $i \in \mathcal{P}_M$ sia π_i la probabilità di inclusione dell'unità i-esima secondo il piano p. Si definisce* ***stimatore di Horvitz-Thompson*** *per il totale di Y la seguente statistica*

$$\widehat{Y}_{Total,HT}(S) = \sum_{i \in S} \frac{Y_i}{\pi_i}.$$

Esempio 10 *Per $s = \{2, 1\}$ da $\mathcal{P}_3$ nell'esempio 6, si ha $D(s) = \{(2, 5), (1, 7)\}$, da cui*

$$\begin{aligned} \widehat{Y}_{Total,HT}(\{2,1\}) &= \sum_{i \in \{2,1\}} \frac{Y_i}{\pi_i} \\ &= \frac{Y_2}{\pi_2} + \frac{Y_1}{\pi_1} \\ &= \frac{7}{\pi_2} + \frac{5}{\pi_1} \end{aligned}$$

Se si applica un piano casuale semplice (di ampiezza 2) risulta, per ogni i,

$$\pi_i = \frac{n}{M} = \frac{2}{3}$$

così

$$\begin{aligned} \widehat{Y}_{Total,HT}(\{2,1\}) &= \frac{3}{2}(7+5) \\ &= 18. \end{aligned}$$

Teorema 1 *Sia $\mathcal{P}_M$ una popolazione finita sulla quale è rilevabile un fenomeno Y e p un disegno di campionamento su $\mathcal{P}_M$. Per $i, j \in \mathcal{P}_M$ siano π_i la probabilità di inclusione del* *primo ordine dell'unità* i*-esima e* $\pi_{i,j}$ *quella di secondo ordine per la coppia* (i,j) *secondo il piano* p*. Vale allora la seguente proprietà di correttezza:*

$$E_p\left[\widehat{Y}_{Total,HT}\right] = \sum_{i\in\mathcal{P}_M} Y_i$$

Inoltre,

$$Var_p\left[\widehat{Y}_{Total,HT}\right] = \sum_{i\in\mathcal{P}_M} \frac{1-\pi_i}{\pi_i} Y_i^2 + \sum_{i\in\mathcal{P}_M} \sum_{j\in\mathcal{P}_M\setminus\{i\}} \left(\frac{\pi_{i,j}}{\pi_i\pi_j} - 1\right) Y_i Y_j.$$

Dimostrazione. Dimostriamo la correttezza. Possiamo esprimere $\widehat{Y}_{Total,HT}$ come segue:

$$\widehat{Y}_{Total,HT}(S) = \sum_{i\in\mathcal{P}_M} \frac{Y_i}{\pi_i}\delta_i(S),$$

$\widehat{Y}_{Total,HT}$ è quindi una combinazione lineare delle variabili casuali $\Delta_1,\ldots,\Delta_M$ i cui coefficienti sono dati da $\frac{Y_1}{\pi_1},\ldots,\frac{Y_M}{\pi_M}$. Il suo valore atteso rispetto a p è quindi pari alla combinazione lineare dei valori attesi delle δ_i :

$$E_p\left[\widehat{Y}_{Total,HT}\right] = E_p\left[\sum_{i\in\mathcal{P}_M} \frac{Y_i}{\pi_i}\Delta_i\right]$$

da cui, ricordando che

$$\pi_i = \sum_{s\in S_{\mathcal{P}_M}} \delta_i(s)\, p(s) \equiv E_p[\Delta_i],$$

si ha

$$\begin{aligned} E_p\left[\widehat{Y}_{Total,HT}\right] &= \sum_{i\in\mathcal{P}_M} \frac{Y_i}{\pi_i}\pi_i \\ &= \sum_{i\in\mathcal{P}_M} Y_i. \end{aligned}$$

Calcoliamo la varianza di $\widehat{Y}_{Total,HT}$. A tal fine utilizziamo ancora l'espressione precedente per lo stimatore

$$\widehat{Y}_{Total,HT}(S) = \sum_{i\in\mathcal{P}_M} \frac{Y_i}{\pi_i}\delta_i(S).$$

La varianza di una combinazione lineare Z di $X_1,\ldots,X_M$ variabili casuali aventi coefficienti $c_1,\ldots,c_M$,

$$Z = \sum_{i=1}^{M} c_i X_i,$$

si ottiene come segue:

$$Var[Z] = \sum_{i=1}^{M} c_i^2 Var[X_i] + \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} c_i c_j Cov[X_i, X_j].$$

Nel nostro caso:

$$c_i = \frac{Y_i}{\pi_i},$$

$$X_i = \Delta_i,$$

$$Z = \widehat{Y}_{Total,HT}.$$

Rimangono quindi da calcolare varianze e covarianze delle variabili $\Delta_1, \ldots, \Delta_M$. La variabile Δ_i segue una distribuzione di Bernoulli $B(p)$ di parametro $p = \pi_i$, la cui varianza è

$$Var[\Delta_i] = \pi_i (1 - \pi_i).$$

Qual é la covarianza tra Δ_i e Δ_j? Ricordiamo che

$$Cov[X, Z] = E[XZ] - E[X] E[Z].$$

Nel nostro caso,

$$X = \Delta_i$$

e

$$Z = \Delta_j.$$

Sappiamo inoltre che risultano le identità

$$E_p[\Delta_i \Delta_j] = \pi_{i,j}$$

e

$$E_p[\Delta_i] = \pi_i.$$

Così,

$$\begin{aligned} Cov_p[\Delta_i, \Delta_j] &= E_p[\Delta_i \Delta_j] - E_p[\Delta_i] E_p[\Delta_j] \\ &= \pi_{i,j} - \pi_i \pi_j. \end{aligned}$$

Quindi, in conclusione,

$$
\begin{aligned}
Var_p\left[\widehat{Y}_{Total,HT}\right] &= Var_p\left[\sum_{i\in\mathcal{P}_M}\frac{Y_i}{\pi_i}\Delta_i\right]\\
&= \sum_{i\in\mathcal{P}_M}\frac{Y_i^2}{\pi_i^2}Var_p\left[\Delta_i\right]+\sum_{i\in\mathcal{P}_M}\sum_{\substack{j\in\mathcal{P}_M\\ j\neq i}}\frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}Cov_p\left[\Delta_i,\Delta_j\right]\\
&= \sum_{i\in\mathcal{P}_M}\frac{Y_i^2}{\pi_i^2}\pi_i\left(1-\pi_i\right)+\sum_{i\in\mathcal{P}_M}\sum_{\substack{j\in\mathcal{P}_M\\ j\neq i}}\frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}\left(\pi_{i,j}-\pi_i\pi_j\right)\\
&= \sum_{i\in\mathcal{P}_M}\frac{1-\pi_i}{\pi_i}Y_i^2+\sum_{i\in\mathcal{P}_M}\sum_{\substack{j\in\mathcal{P}_M\\ j\neq i}}\left(\frac{\pi_{i,j}}{\pi_i\pi_j}-1\right)Y_iY_j
\end{aligned}
$$

■

3.2 Stimatore della media di una popolazione

La media del fenomeno Y nella popolazione $\mathcal{P}_M$ può essere stimata a partire dallo stimatore Horvitz-Thompson del totale, essendo i due parametri legati dalla relazione

$$Y_{Total} = M\left(Y_{Mean}\right), \quad Y_{Total} = \sum_{i\in\mathcal{P}_M} Y_i.$$

Basta definire il seguente stimatore per il parametro media:

$$\widehat{Y}_{Mean,HT}\left(S\right) = \frac{1}{M}\widehat{Y}_{Total,HT}\left(S\right).$$

Essendo $\widehat{Y}_{Total,HT}$ corretto per il totale, $\widehat{Y}_{Mean,HT}$ risulterà corretto per la corrispondente media. Per la varianza, invece, vale la relazione:

$$Var_p\left[\widehat{Y}_{Mean,HT}\right] = \frac{1}{M^2}Var_p\left[\widehat{Y}_{Total,HT}\right].$$

Capitolo 4

Materiali per l'esame

4.1 Esercitazioni

Esercizio 4.1.1 *È data la popolazione finita $\mathcal{P}_3$. Sul corrispondente spazio campionario è definito il seguente piano di campionamento:*

$$p(s) = \begin{cases} \frac{1}{6} & se\ s = \{1,2\} \\ \frac{1}{3} & se\ s = \{1,3\} \\ \frac{1}{2} & se\ s = \{2,3\} \\ 0 & altrimenti \end{cases}$$

Determinare quanto segue.

1. *Le probabilità d'inclusione del primo ordine.*

2. *La varianza della variable Δ_2.*

3. *Le probabilità d'inclusione del secondo ordine.*

4. *La covarianza tra le variabili Δ_1 e Δ_3.*

5. *È stato estratto il campione $s = \{2,3\}$ e si è rilevato $D_s = \{(2,17),(3,8)\}$. Stimare con la statistica di Horvitz-Thompson il totale di Y su $\mathcal{P}_3$.*

6. *È stato estratto il campione $s = \{1,3\}$ e si è rilevato $D_s = \{(1,6),(3,12)\}$. Sapendo che il valore atteso della statistica di Horvitz-Thompson per il totale di Y è pari a 50, determinare Y_2.*

Esercizio 4.1.2 *E' data la popolazione finita P_8. Sul corrispondente spazio campionario è definito un piano di campionamento casuale semplice di ampiezza 3.*

Determinare quanto segue.

1. *Le probabilità $\pi_{1,4}$ e $\pi_{7,7}$.*

2. *Il valore atteso della variabile* $\Delta_{2,8}$*;*

3. *La varianza della variabile* Δ_3*.*

4. *La correlazione tra le variabili* Δ_1 *e* Δ_7*.*

5. *È stato estratto il campione* $s = \{4, 6, 8\}$ *e si è rilevato* $D_s = \{(4, 10), (6, 20), (8, 15)\}$*. Stimare con la statistica di Horvitz-Thompson il totale di* Y *su* $\mathcal{P}_8$*.*

Esercizio 4.1.3 *E' data la popolazione finita P_8. Sul corrispondente spazio campionario è definito un piano di campionamento p, rispetto al quale risultano noti i seguenti momenti:*

$$Cov[\Delta_3, \Delta_5] = -1/16,\ Var[\Delta_3] = 3/16,\ E[\Delta_5] = 1/4.$$

Determinare quanto segue.

1. *La probabilità π_5.*
2. *La probabilità π_3 sapendo che $\pi_3 > 1/2$.*
3. *Il valore atteso della variabile $\Delta_{3,5}$.*

Esercizio 4.1.4 *E' data la popolazione finita P_{20}. Sul corrispondente spazio campionario è definito un piano di campionamento p. Rispetto a p, delle variabili $X = 3\Delta_7 + 1$ e $Y = -4\Delta_{11} - 5$ sono noti i seguenti momenti:*

$$Cov[X, Y] = -2/9,\ E[X] = 2,\ E[Y] = -6.$$

Determinare quanto segue.

1. *Le probabilità π_7, π_{11} e $\pi_{7,11}$.*

Esercizio 4.1.5 *E' data la popolazione finita P_9. Sul corrispondente spazio campionario è definito un piano di campionamento p, rispetto al quale sono noti i seguenti momenti:*

$$Corr[\Delta_1, \Delta_3] = -1/6,\ E[\Delta_1] = 4/25,\ E[\Delta_3] = 1/5.$$

Determinare quanto segue.

1. *Il valore atteso della variabile $\Delta_{1,3}$.*

Esercizio 4.1.6 *E' data la popolazione finita P_M, con $M > 1$. Sul corrispondente spazio campionario è definito un piano di campionamento casuale semplice $p_{C,M,n}$, con $n < M$, rispetto al quale sono note le seguenti probabilità di inclusione:*

$$\forall i, j \in P_M : \pi_{i,j} = 2/9,\ se\ i \neq j.$$

$$\forall i \in P_M : \pi_i = 1/2.$$

Determinare quanto segue.

1. *I valori di M e di n.*