Simulazione 1

- La variabilità campionaria verrà illustrata nel modo seguente:

1. verrà considerata una variabile discreta (generica) che può assumere soltanto un piccolo numero di valori possibili (N = 4). Il modello probabilistico associato ai 4 valori sarà uniforme ($p = 1/N$).
2. verrà fornito l'elenco di tutti i possibili campioni di grandezza $n = 2$;
3. verrà calcolata la media di ciascuno dei possibili campioni di grandezza $n = 2$;
4. verrà esaminata la distribuzione delle medie di tutti i possibili campioni di grandezza $n = 2$.

- La media μ e la varianza σ^2 della popolazione verranno anch'esse calcolate.
- μ e σ^2 sono dei parametri, mentre la media aritmetica $\bar{x}$ e la varianza s^2 di ciascun campione sono delle statistiche.

- L'esperimento di questo esempio consiste in $n = 2$ estrazioni con rimessa di una pallina x_i da un'urna che contiene $N = 4$ palline.
- Le palline sono numerate nel modo seguente: $\Omega = \{2, 3, 5, 9\}$
- L'estrazione con rimessa corrisponde ad una popolazione di grandezza infinita (è sempre possibile infatti estrarre una nuova pallina dall'urna).

- Per ciascun campione di grandezza $n = 2$ viene calcolata la media dei valori delle palline estratte

$$\bar{x} = \sum_{i=1}^{2} x_i/2$$

- Per esempio, se le palline estratte sono $x_1 = 2$ e $x_2 = 3$, allora $\bar{x} = (2 + 3)/2 = 5/2 = 2.5$.

Tre distribuzioni

- Dobbiamo distinguere tre distribuzioni:
  1. la distribuzione della popolazione,
  2. la distribuzione di un particolare campione
  3. la distribuzione campionaria delle medie di tutti i possibili campioni.

Distribuzione della popolazione: la distribuzione di X (il valore della pallina estratta) nella popolazione. In questo caso la popolazione è infinita e ha la seguente distribuzione di probabilità:

x_i	p_i
2	$\frac{1}{4}$
3	$\frac{1}{4}$
5	$\frac{1}{4}$
9	$\frac{1}{4}$
somma	1

- Il valore atteso della popolazione è

$$\mu = \sum_{i=1}^{4} x_i p_i = 4.75$$

- La varianza della popolazione è

$$\sigma^2 = \sum_{i=1}^{4} (x_i - 4.75)^2 p_i = 7.1875$$

Distribuzione di un campione: la distribuzione di X in un particolare campione.

- Per esempio, se $x_1 = 2$ e $x_2 = 3$, allora la media di questo campione sarà $\bar{x} = 2.5$ e la varianza sarà $s^2 = 0.25$.

Distribuzione campionaria della media

Distribuzione campionaria della media: la distribuzione delle medie $\bar{x}$ di tutti i possibili campioni.

- Se $n = 2$, ci sono $4 \times 4 = 16$[1] possibili campioni. Possiamo dunque elencarli, insieme alle loro medie.

campione	$\bar{x}_i$	campione	$\bar{x}_i$
$\{2;3\}$	2.5	$\{3;2\}$	2.5
$\{5;2\}$	3.5	$\{2;5\}$	3.5
$\{9;2\}$	5.5	$\{2;9\}$	5.5
$\{5;3\}$	4.0	$\{3;5\}$	4.0
$\{9;3\}$	6.0	$\{3;9\}$	6.0
$\{9;5\}$	7.0	$\{5;9\}$	7.0
$\{2;2\}$	2	$\{3;3\}$	3
$\{5;5\}$	5	$\{9;9\}$	9

Ciascuna coppia di osservazioni $\{x_i; x_j\}$ per $i : 1, \ldots, 4$, e $j : 1, \ldots, 4$, ha la stessa probabilità $p = 1/4 \times 1/4 = 1/16$. Per costruire la distribuzione di probabilità della media campionaria sarà sufficiente contare le frequenze (relative) di ciascun valore $\bar{x}_i$.

La distribuzione campionaria della media ha la seguente distribuzione di probabilità:

$\bar{x}$	p_i
2.0	1/16
2.5	2/16
3.0	1/16
3.5	2/16
4.0	2/16
5.0	1/16
5.5	2/16
6.0	2/16
7.0	2/16
9.0	1/16
somma	1.00

[1] Con il calcolo combinatorio abbiamo $\frac{4!}{2!} = 12$ estrazioni possibili senza rimessa più le 4 coppie dovute alla rimessa $\{22, 33, 55, 99\}$

- La media della distribuzione campionaria della media è

$$\mu_{\bar{x}} = \sum \bar{x}_i p_i = 4.75$$

- La varianza della distribuzione campionaria della media è

$$\sigma^2_{\bar{x}} = \sum (\bar{x}_i - \mu_{\bar{x}})^2 p_i = 3.59375$$

- L'esercizio presente ha a che fare con una situazione particolare, quella in cui la distribuzione della popolazione è conosciuta.
- In pratica, la distribuzione della popolazione non è mai conosciuta.
- Questo esercizio ci permette però di notare come la distribuzione campionaria della media possieda due importanti proprietà.

La media $\mu_{\bar{x}}$ della distribuzione campionaria della media è uguale alla media della popolazione μ.

La varianza $\sigma^2_{\bar{x}}$ della distribuzione campionaria della media è uguale alla varianza della popolazione σ^2 divisa per la grandezza del campione n:

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = \frac{7.1875}{2} = 3.59375$$

Si noti che:

- la media e la varianza della distribuzione campionaria sono determinate dalla media e varianza della popolazione:

$$\mu_{\bar{x}} = \mu \qquad \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$$

- la varianza della distribuzione campionaria della media è più piccola della varianza della popolazione.

In seguito utilizzeremo le proprietà della distribuzione campionaria per fare delle inferenze a proposito dei parametri della popolazione anche quando la distribuzione della popolazione non è conosciuta.

Si noti inoltre che abbiamo distinto tra tre diverse distribuzioni.

1. **Distribuzione della popolazione:**
   $\Omega = \{2, 3, 5, 9\},\ \mu = 4.75,\ \sigma^2 = 7.1875$

2. **Distribuzione di un particolare campione:**
   $\Omega_i = \{2, 3\},\ \bar{x} = 2.5,\ s^2 = 0.25$

3. **Distribuzione campionaria della media:**
   $\Omega_{\bar{x}} = \{2.5; 3.5; 5.5; 4; 6; 7; 2.5; 3.5; 4; 6; 7; 2; 5; 3; 9\},$
   $\mu_{\bar{x}} = 4.75, \qquad \sigma^2_{\bar{x}} = 3.59375$

Distribuzione della popolazione La distribuzione che contiene tutte le possibili modalità della variabile aleatoria. Media e varianza di questa distribuzione si indicano con μ e σ^2.

Distribuzione del campione La distribuzione dei valori della popolazione che fanno parte di un particolare campione casuale di grandezza n. Le singole osservazioni si indicano con $x_1, x_2, \ldots, x_n$, e hanno media $\bar{x}$ e varianza s^2.

Distribuzione campionaria delle medie dei campioni La distribuzione di $\bar{x}_i$ per tutti i possibili campioni di grandezza n che si possono estrarre dalla popolazione considerata. Media e varianza della distribuzione campionaria della media si indicano con $\mu_{\bar{x}}$ e $\sigma^2_{\bar{x}}$.

• La distribuzione che sta alla base dell'inferenza statistica è la distribuzione campionaria.

Definizione: la distribuzione campionaria di una statistica è la distribuzione dei valori che quella statistica assume in tutti i campioni di grandezza n che possono essere estratti dalla popolazione.

• Si noti che, se in una simulazione consideriamo un numero di campioni minore di quello che teoricamente è possibile, la distribuzione risultante ci fornirà soltanto un'approssimazione alla vera distribuzione campionaria.