**Laboratorio di inferenza statistica (2024-2025)**

**Lab. 9 – Confronto distribuzioni esatte, empiriche (MC) e approssimazioni normali. Intervalli**
**8 maggio 2025**

1. Confrontare distribuzione empirica (MC), distribuzione esatta (usando `dgamma`) e approssimazione normale (usando `dnorm`) della funzione di densità dello stimatore $S_n^2$ di $\theta_2$ nel modello $\mathrm{N}(\theta_1, \theta_2)$, assumendo $n = 10$, $\theta_1 = 1$, $\theta_2 = 4$. Considerare $M = 10000$ per la procedura di Monte Carlo. Ripetere per $n = 50$.

   Ricorda che $S_n^2 \sim \mathrm{Gamma}\left(\frac{n-1}{2}, \frac{n-1}{2\theta_2}\right)$ e $S_n^2 \dot\sim \mathrm{N}\left(\theta_2, \frac{2\theta_2^2}{n}\right)$.

2. Confrontare distribuzione empirica (MC), distribuzione esatta (usando `dinvgamma`) e approssimazione normale (usando `dnorm`) della funzione di densità dello stimatore $1/\overline{X}_n$ di $\theta_2$ nel modello $\mathrm{EN}(\theta)$, assumendo $n = 10$, $\theta = 2$. Considerare $M = 10000$ per la procedura di Monte Carlo. Ripetere per $n = 50$.

   Ricorda che $\frac{1}{\overline{X}_n} \sim \mathrm{InvGa}(n, \texttt{rate} = n\theta)$ e $\frac{1}{\overline{X}_n} \dot\sim \mathrm{N}\left(\theta, \frac{\theta^2}{n}\right)$.

   [Sugg.: necessario installare il pacchetto "invgamma": vedi soluzioni].

3. Confrontare distribuzione empirica (MC) e approssimazione normale della funzione di densità dello stimatore di massima verosimiglianza del parametro $g(\theta) = e^{-\theta}$ (modello Poisson di parametro $\theta$), $n = 10$, $\theta = 2$. Considerare $M = 10000$ per la procedura di Monte Carlo. Ripetere per $n = 50$.

   Ricorda che $e^{-\overline{X}_n} \dot\sim \mathrm{N}\left(e^{-\theta}, e^{-2\theta}\frac{\theta}{n}\right)$.

4. **Determinazione di intervalli di confidenza, modello normale**

   I seguenti dati indicano il numero di azioni di diverso tipo scambiate in un'unità di tempo:

   $$(63, 65, 94, 37, 83, 95, 70, 96, 47, 29, 52, 38, 47, 79, 66, 25, 48, 80, 52, 49)$$

   Assumere la normalità dei dati.

   (a) Trovare le principali statistiche descrittive per i dati a disposizione
   (b) Problema N1. Calcolare l'intervallo di confidenza per $\theta$ assumendo $\sigma^2 = 400$ (noto). Scrivere prima il codice R e poi usare `z.test`.
   (c) Problema N2. Calcolare l'intervallo di confidenza di livello $1 - \alpha = 0.90$ per il valore attesosupponendo che la varianza sia incognita. Scrivere prima il codice R e poi usare `t.test`
   (d) Problema N4. Calcolare l'intervallo di confidenza di livello $1-\alpha = 0.90$ per la varianza, assumendo incognito il valore atteso. Scrivere il codice R per ottenere l'intervallo.

5. **Simulazione della probabilità di copertura di un intervallo di confidenza**

   Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione $\mathrm{N}(\theta, \sigma^2)$. Un intervallo di confidenza di livello $1 - \alpha$ per il parametro incognito $\theta$ è definito come segue:

   $$C(\mathbf{X}) = \left[\overline{X}_n - t_{n-1,1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}}, \overline{X}_n + t_{n-1,1\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}}\right], \qquad \alpha \in (0,1).$$

   Siamo interessati a valutare la probabilità di copertura effettiva dell'intervallo e a confrontarla con il livello nominale $1 - \alpha$.

   (a) Definire la matrice `x.matrice` di $N = 1000$ righe e $n = 10$ colonne in cui ciascuna delle righe contenga i valori osservati di un campione casuale da una popolazione normale con $\theta = 2$ e $\sigma^2 = 1$.
   (b) Determinare i due vettori `C.inf` e `C.sup` che contengano, rispettivamente, gli estremi inferiori e superiori degli $N$ intervalli di confidenza al livello 0.95 per $\theta$ ottenuti considerando l'espressione di $C(\mathbf{X})$.
   (c) Quanti sono, tra quelli generati, gli intervalli per i quali `C.sup` risulta minore o uguale a 2.5; quanti quelli per i quali `C.inf` risulta maggiore di 2.5?
   (d) Calcolare la frequenza relativa di intervalli che contengono il valore $\theta = 2$ tra gli $N$ generati (il valore determinato è una approssimazione della copertura effettiva dell'intervallo proposto).

(e) Ripetere quanto fatto per una griglia di valori per $\theta$ nell'intervallo [0,10] e tracciare una approssimazione per la curva che definisce la funzione $\text{Cop}_\theta(C)$.

6. **Intervalli ET e HD modello EN**

Per il modello EN($\theta$) gli intervalli di confidenza basati sulla quantità pivotale $Q(\mathbf{X}_n, \theta) = \theta T \sim \text{Ga}(n, 1)$, con $T(\mathbf{X}_n) = \sum_{i=1}^n X_i$, hanno la forma

$$C(\mathbf{X}_n) = \left[ \frac{q_1}{T(\mathbf{X}_n)}, \frac{q_2}{T(\mathbf{X}_n)} \right],$$

con $q_1$ e $q_2$ quantili della v.a. $\text{Ga}(n, 1)$ opportunamente scelti. Considerare il livello di confidenza pari a $1 - \alpha = 0.95$ e supporre che, per un campione osservato di dimensione $n = 10$, si abbia $T(\mathbf{x}_n) = 3$.

(a) Calcolare l'insieme ET (equal-tails), ovvero porre $q_1 = q_{\frac{\alpha}{2}}$ e $q_2 = q_{1-\frac{\alpha}{2}}$

(b) Calcolare l'insieme HD (highest-density), ovvero trovare $q_1^\star$ e $q_2^\star$ che minimizzano la lunghezza dell'intervallo di livello $1 - \alpha$, usando la funzione `HDInterval`.

(c) Verificare che la lunghezza dell'insieme HD è minore di quella dell'intervallo ET.