Esercizi di Teoria Statistica delle Decisioni (F. De Santis)
Parte 3: Decisioni statistiche
a.a. 2024-2025

A – Analisi forma estensiva (AfE) – Esercizi (miscellanea)

1. Sia $\mathbf{X}_n = (X_1, ..., X_n)$ un campione in cui le v.a X_i sono, condizionatamente a θ, i.i.d. con valore atteso $\mathbb{E}_\theta[X] = (1-\theta)/\theta$, $\mathbb{V}_\theta[X] = (1-\theta)/\theta^2$ e funzione di massa di probabilità

$$p_\theta^{X_i}(x) = (1-\theta)^x \theta, \qquad x = 0, 1, 2, \dots \qquad \theta \in [0, 1].$$

   (a) Verificare che la famiglia delle densità Beta(α, β), con $\alpha, \beta > 0$, è coniugata al modello in esame (per v.a. geometriche) e determinare i parametri della distribuzione a posteriori di Θ, $\pi(\theta|\mathbf{x}_n)$.

   (b) Determinare l'espressione delle azioni ottime $a_\pi^*(\mathbf{z}_n)$ che si ottengono utilizzando la perdita quadratica e la perdita 0-1 nel problema di stima puntuale del parametro.

   (c) Supponendo che il campione abbia dimensione $n = 5$, che $\alpha = \beta = 2$ e che $s_n = \sum_{i=1}^n x_i = 5$, calcolare il valore numerico di delle due azioni.

   (d) Calcolare il valore numerico della perdita finale attesa che si ottiene per l'azione ottima nel caso si usi la perdita quadratica.

   (e) Determinare $\pi^j(\theta)$, la distribuzione a priori di Jeffreys per θ e la corrispondente distribuzione a posteriori.

   (f) Determinare nuovamente (ma usando ora la distribuzione di Jeffreys) l'espressione delle azioni ottime $a_{\pi^j}^*(\mathbf{z}_n)$ che si ottengono utilizzando la perdita quadratica e la perdita 0-1 e i corrispondenti valori numerici ottenuti con i dati di cui al punto precedente.

Soluzione.

(a) $\ell(\theta) = (1-\theta)^{s_n}\theta^n$, con $s_n = \sum_{i=1}^n x_i$. Pertanto $\pi(\theta|\mathbf{z}_n) \propto \theta^{\overline{\alpha}-1}(1-\theta)^{\overline{\beta}-1}$ con $\overline{\alpha} = \alpha + n$ e $\overline{\beta} = \beta + s_n$, ovvero è una densità Beta$(\overline{\alpha}, \overline{\beta})$.

(b) Perdita quadratica: $a_\pi^\star(\mathbf{z}_n) = \mathbb{E}[\Theta|\mathbf{z}_n] = \frac{\overline{\alpha}}{\overline{\alpha}+\overline{\beta}}$.
Perdita $0-1$: $a_\pi^\star(\mathbf{z}_n) = \mathbb{M}\text{o}[\Theta|\mathbf{z}_n] = \frac{\overline{\alpha}-1}{\overline{\alpha}+\overline{\beta}-2}$.

(c) $\mathbb{E}[\Theta|\mathbf{z}_n] = \frac{1}{2}$; $\mathbb{M}\text{o}[\Theta|\mathbf{z}_n] = \frac{1}{2}$.

(d) $\rho_\pi(a^\star, \mathbf{z}_n) = \mathbb{V}[\Theta|\mathbf{z}_n] = \frac{\overline{\alpha}\overline{\beta}}{(\overline{\alpha}+\overline{\beta}+1)(\overline{\alpha}+\overline{\beta})^2} = 0.017$.

(e) Si verifica che $I_n(\theta) = \frac{n}{\theta^2(1-\theta)}$. Pertanto $\pi^J(\theta) \propto \frac{1}{\theta(1-\theta)^{1/2}}$, che si ottiene come caso limite dalla densità a priori Beta per $\alpha \to 0$ e $\beta = \frac{1}{2}$. La densità a posteriori è quindi quella di una Beta$(\overline{\alpha}_j, \overline{\beta}_j)$, con $\overline{\alpha}_j = s_n$ e $\overline{\beta}_j = \frac{1}{2} + n$. In questo caso: $\mathbb{E}[\Theta|\mathbf{z}_n] = 0.476$, $\mathbb{M}\text{o}[\Theta|\mathbf{z}_n] = 0.470$ e $\mathbb{V}[\Theta|\mathbf{z}_n] = 0.021$

2. Sia $X_1, \dots, X_n|\theta \sim$ Uniforme$[0, \theta]$ i.i.d. $(\theta > 0)$. Considerare $\pi(\theta) \propto 1/\theta$.

   (a) Verificare che $t = x_{(n)} = \max\{x_1, \dots, x_n\}$ è una statistica sufficiente per il modello.

   (b) Verificare che
$$\pi(\theta|t) = \frac{n}{\theta^{n+1}}\, t^n\, I_{[t,\infty]}(\theta).$$

   (c) Verificare che, se si usa la funzione di perdita quadratica per il problema di stima puntuale, si ottiene
$$a_\pi^*(\mathbf{x}_n) = \mathbb{E}[\Theta|t] = \frac{n}{n-1}t.$$

   (d) Calcolare il valore di $a_\pi^*(\mathbf{x}_n)$ per il campione $\mathbf{z}_n = (1/2, 1)$.

   (e) Utilizzando il campione osservato, determinare l'espressione di $\pi(\theta|t)$ e disegnarne il grafico.

   (f) Utilizzando il campione osservato, calcolare $\mathbb{P}(\Theta > 2|t)$.

Soluzione. FATTO: ES. 28 PARTE 1

3. Sia $X_1, \ldots, X_n|\theta \sim \text{Pois}(\theta)$ i.i.d. , $\theta > 0$.
   (a) Determinare $\pi^J(\theta)$ (Jeffreys prior).
   (b) Verificare che, utilizzando $\pi^J(\theta)$, allora $\pi(\theta|\mathbf{z}_n)$ è una densità Gamma$(s_n + 1/2, n)$.
   (c) Determinare l'azione ottima per il problema di stima puntuale $a_\pi(\mathbf{z}_n)$ che si ottiene con la funzione di perdita quadratica e l'espressione della perdita finale attesa corrispondente.
   (d) Considerare ora una distribuzione a priori Gamma$(1, 1)$. Determinare valore atteso e varianza predittivi (a priori e a posteriori) di $a_\pi(\mathbf{Z}_n)$.
   (e) Determinare l'approssimazione normale di $\pi(\theta|\mathbf{z}_n)$, l'azione ottima asintotica $\tilde{a}_\pi(\mathbf{z}_n)$ (associata alla perdita quadratica) e la perdita finale attesa asintotica $\rho_\pi(\tilde{a}_\pi, \mathbf{z}_n)$

Soluzione.

(a) L'informazione attesa di Fisher è $I_n(\theta) = \text{cr}(\theta)^{-1} = \frac{n}{\theta}$ e quindi $\pi^J(\Theta) \propto \frac{1}{\sqrt{\theta}}$.

(b) Osservando che $\ell(\theta) \propto \theta^{s_n} e^{-n\theta}$, con $s_n = \sum_{i=1}^n x_i$, abbiamo che $\Theta|\mathbf{z}_n \sim \text{Ga}(\overline{\alpha}_j, \text{rate} = \overline{\beta}_j)$, con $\overline{\alpha}_j = s_n + \frac{1}{2}$ e $\overline{\beta}_j = n$.

(c) $a^\star_\pi(\mathbf{z}_n) = \mathbb{E}[\Theta|\mathbf{z}_n] = \frac{\overline{\alpha}_j}{\overline{\beta}_j} = \frac{s_n+\frac{1}{2}}{n}$; $\rho_\pi(a^\star_\pi, \mathbf{z}_n) = \mathbb{V}[\Theta|\mathbf{z}_n] = \frac{\overline{\alpha}_j}{\overline{\beta}_j^2} = \frac{s_n+\frac{1}{2}}{n^2}$.

(d) Ricordare che, in generale, $\mathbb{E}_X[X] = \mathbb{E}_\Theta\left[\mathbb{E}_{X|\theta}(X)\right]$. In questo caso, utilizzando per Θ la distribuzione indicata, $\mathbb{E}_X[X] = \mathbb{E}_\Theta[\Theta] = 1$. Abbiamo quindi: $\mathbb{E}_X\left[\frac{\sum_{i=1}^n X_i + \frac{1}{2}}{n}\right] = \frac{n\mathbb{E}_X[X_i]+\frac{1}{2}}{n} = \frac{n+\frac{1}{2}}{n} = \frac{2n+1}{2n}$.

(e) $\Theta|\mathbf{z}_n \,\dot\sim\, \text{N}\left(\tilde{\theta}, I_B^{-1}\right)$, dove si verifica che $\tilde{\theta} = \mathbb{M}\text{o}[\Theta|\mathbf{z}_n] = \frac{\overline{\alpha}_j - 1}{\overline{\alpha}_j + \overline{\beta}_j - 2}$ e $I_B = \frac{(\overline{\alpha}_j+\overline{\beta}_j-1)^3}{(\overline{\alpha}_j-1)(\overline{\beta}_j-1)}$. La soluzione si trova sostituendo i valori di $\overline{\alpha}_j$ e $\overline{\beta}_j$ nelle precedenti formule.

4. Siano $X_1, ..., X_n|\theta \sim \text{Geom}(\theta)$ i.i.d., con valore atteso $\mathbb{E}_\theta[X] = \frac{1}{\theta}$, $\mathbb{V}_\theta[X] = \frac{1-\theta}{\theta^2}$ e funzione di massa di probabilità

$$p_\theta^X(x) = \theta(1-\theta)^{x-1}, \qquad x = 1, 2, 3, \ldots \qquad \theta \in [0, 1].$$

   (a) Determinare la funzione di verosimiglianza e la stima di massima verosimiglianza $d_{mv}(\mathbf{z}_n)$ del parametro θ.
   (b) Verificare che la famiglia delle densità Beta(α, β), con $\alpha, \beta > 0$, è coniugata al modello in esame e determinare i parametri $(\overline{\alpha}, \overline{\beta})$ della distribuzione a posteriori di Θ, $\pi(\theta|z)$.
   (c) Determinare le azioni ottime per la stima puntuale del parametro che si ottengono con le funzioni di perdita quadratica e 0-1 e verificare che, per valori elevati di n, per entrambe le azioni ottime si ha $a^*_\pi(\mathbf{z}_n) \simeq d_{mv}(\mathbf{z}_n)$.
   (d) Determinare l'incremento di perdita finale attesa che si ha usando la stima di massima verosimiglianza al posto dell'azione ottima nel caso di perdita quadratica.

Soluzione.

(a) $\ell(\theta) = \theta^n(1-\theta)^{n-s_n}$, con $s_n = \sum_{i=1}^n x_i$. Dall'equazione $\frac{d}{d\theta}\ln\ell(\theta) = 0$ e dallo studio di $\frac{d^2}{d\theta^2}\ln\ell(\theta)$ si verifica che $d_{mz}(\mathbf{z}_n) = \frac{n}{s_n} = \frac{1}{\overline{x}_n}$.

(b) $\pi(\theta|\mathbf{z}_n) \propto \ell(\theta)\cdot\pi(\theta) \propto \theta^{\overline{\alpha}-1}(1-\theta)^{\overline{\beta}-1}$, con $\overline{\alpha} = \alpha + n$ e $\overline{\beta} = \beta + s_n - n$.

(c) Perdita quadratica: $a^\star_\pi(\mathbf{z}_n) = \mathbb{E}[\Theta|\mathbf{z}_n] = \frac{\overline{\alpha}}{\overline{\alpha}+\overline{\beta}} = \frac{\alpha+n}{\alpha+\beta+s_n}$.
Perdita $0-1$: $a^\star_\pi(\mathbf{z}_n) = \mathbb{M}\mathrm{o}[\Theta|\mathbf{z}_n] = \frac{\overline{\alpha}}{\overline{\alpha}+\overline{\beta}} = \frac{\alpha+n-1}{\alpha+\beta+s_n-2}$. Per n elevato entrambe le stime sono approssimativamente uguali a $\frac{n}{s_n} = d_{mv}$.

(d) Per ogni $a \in \mathcal{A}$ si ha che $\rho_\pi(a,\mathbf{z}_n) = \mathbb{V}[\Theta|\mathbf{z}_n] + (a - a^\star_\pi)^2$. Quindi: $\rho_\pi(d_{mv},\mathbf{z}_n) = \mathbb{V}[\Theta|\mathbf{z}_n] + (d_{mv} - \mathbb{E}[\Theta|\mathbf{z}_n])^2$. L'incremento di perdita rispetto all'azione ottima è quindi
$(d_{mv} - \mathbb{E}[\Theta|\mathbf{z}_n])^2 = \left(\frac{n}{s_n} - \frac{\alpha+n}{\alpha+\beta+s_n}\right)^2 = \ldots$ (calcoli).

5. Sia $X_1,\ldots,X_n|\theta \sim \mathrm{N}(\mu_0,\theta)$ (μ_0 noto) e $\Theta \sim \mathrm{InvGa}(\alpha,\beta)$. Determinare quanto richiesto di seguito.

   (a) Distribuzione a posteriori di Θ.

   (b) Azione ottima $a^*_\pi(\mathbf{z}_n)$ per la stima puntuale del parametro che si ottiene con funzione di perdita quadratica e corrispondente valore della perdita finale attesa.

   (c) Azione ottima $a^*_\pi(\mathbf{z}_n)$ per la stima puntuale del parametro che si ottiene con funzione di perdita 0-1 e con perdita assoluta.

   (d) Informazione bayesiana I_B e l'approssimazione normale per la densità a posteriori di Θ.

   (e) Azione ottima asintotica $\tilde{a}^*_\pi(\mathbf{z}_n)$ per la stima puntuale del parametro che si ottiene con funzione di perdita quadratica e corrispondente valore di ρ_π.

   (f) Espressione di $\widetilde{C}(\mathbf{z}_n)$, l'insieme HPD di livello $1-\gamma$ per θ (usare approssimazione normale).

Soluzione.

(a) $\ell(\theta) \propto \theta^{-n/2}\exp\left\{-\frac{nS_0^2}{2\theta}\right\}$, con $S_0^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\mu_0)^2$. Quindi: $\pi(\theta|\mathbf{z}_n) \propto \frac{1}{\theta^{\overline{\alpha}+1}}\exp\left\{-\frac{\overline{\beta}}{\theta}\right\} = \ldots$, con $\overline{\alpha} = \alpha + \frac{n}{2}$ e $\overline{\beta} = \beta + \frac{nS_0^2}{2}$. Pertanto $\Theta|\mathbf{z}_n \sim \mathrm{InvGa}(\overline{\alpha}, \texttt{rate} = \overline{\beta})$.

(b) Perdita quadratica: sfruttando l'integrale della gamma inversa si verifica analiticamente che
$a^\star_\pi(\mathbf{z}_n) = \mathbb{E}[\Theta|\mathbf{z}_n] = \frac{\overline{\beta}}{\overline{\alpha}-1} = \ldots$ (sostituire i valori trovati sopra).
$\rho_\pi(a^\star_\pi,\mathbf{z}_n) = \mathbb{V}[\Theta|\mathbf{z}_n] = \ldots = \frac{\overline{\beta}^2}{(\overline{\alpha}-1)^2(\overline{\alpha}-2)}$ (sostituire i valori e calcolare).

(c) Perdita $0-1$: dallo studio delle derivare prima e seconda di $\ln\pi(\theta|\mathbf{z}_n)$ si verifica facilmente che
$a^\star_\pi(\mathbf{z}_n) = \mathbb{M}\mathrm{o}[\Theta|\mathbf{z}_n] = \tilde{\theta} = \frac{\overline{\beta}}{\overline{\alpha}+1} = \ldots$ (sostituire i valori trovati sopra).
Perdita assoluta: $a^\star_\pi(\mathbf{z}_n)$ coincide con la mediana della densità a posteriori.

(d) $\Theta|\mathbf{z}_n \mathrel{\dot\sim} \mathrm{N}\left(\tilde{\theta}, I_B^{-1}\right)$. Dallo studio delle derivate di $\ln\pi(\theta|\mathbf{x}_n)$ si verifica che:
$I_B = \frac{(\overline{\alpha}+1)^3}{\overline{\beta}^2}$. Pertanto: $\Theta|\mathbf{z}_n \mathrel{\dot\sim} \mathrm{N}\left(\frac{\overline{\beta}}{\overline{\alpha}+1}, \frac{\overline{\beta}^2}{(\overline{\alpha}+1)^3}\right)$ (sostituire i valori e calcolare).

6. Sia $X_1,\ldots,X_n|\theta \sim \mathrm{N}(\theta,1)$ e $\Theta \sim \mathrm{N}(\mu_0, n_0^{-1})$. Considerare il sistema di ipotesi $H_0 : \theta \in \Omega_0 = [\theta_L,\theta_U]$ vs. $H_1 : \theta \in \Omega_1 = [\theta_L,\theta_U]^C = (-\infty,\theta_L)\cup(\theta_U,+\infty)$.

   (a) Determinare la distribuzione a posteriori di Θ.

   (b) Determinare l'espressione del fattore di Bayes $B_{01}(\overline{x}_n)$ per il confronto tra le ipotesi.

   (c) Determinare $a^*_\pi(\mathbf{z}_n)$ e $\rho_\pi(a^*_\pi,\mathbf{z}_n)$ (perdita quadratica, con $b_0 = b_1 = 1$) e calcolarne i valori numerici per $\overline{x}_n = 3/2$, $\mu_0 = 1$, $n_0 = 1/2$, $n = 3$.

(d) Determinare il valore di $B_{01}(\overline{x}_n)$ assumendo i dati del precedente punto e supponendo che $\theta_L = 1$, $\theta_U = 2$ e dire quale ipotesi viene scelta sulla base del solo valore del fattore di Bayes (valore-soglia $= \frac{b_1}{b_0}\frac{\pi_1}{\pi_1}$).

(e) Dire quale ipotesi viene scelta se si confrontano le perdite finali attese, ponendo $b_0 = b_1 = 1$ nella funzione di perdita.

(f) Determinare l'espressione del fattore di Bayes $B_{01}(\overline{x}_n)$ per il confronto tra le ipotesi $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$.

(g) Determinare il valore di $B_{01}(\overline{x}_n)$ assumendo $\overline{x}_n = 3/2$, $\mu_0 = 1.4$, $\theta_0 = 1.6$, $n_0 = 1/2$, $n = 10$ e dire quale ipotesi viene scelta sulla base del solo valore del fattore di Bayes (valore-soglia $= 1$).

(h) Dire quale ipotesi viene scelta se si confrontano le perdite finali attese e si pone $b_0 = b_1 = 1$.

Soluzione.

(a) Per il modello in esame abbiamo $\Theta|\mathbf{z}_n \sim \mathrm{N}\left(\mu_p, \sigma_p^2\right)$, con $\mu_p = \frac{n_0\mu_0 + n\overline{x}_n}{n_0+n}$ e $\sigma_p^2 = \frac{1}{n_0+n}$ (in quanto $\sigma^2 = 1$).

(b) $B_{01}(\mathbf{z}_n) = \frac{\mathbb{P}(H_0|\mathbf{z}_n)/\mathbb{P}(H_1|\mathbf{z}_n)}{\mathbb{P}(H_0)/\mathbb{P}(H_1)} = \dfrac{\frac{\Phi\left(\frac{\theta_U-\mu_p}{\sigma_p}\right)-\Phi\left(\frac{\theta_L-\mu_p}{\sigma_p}\right)}{1-\left[\Phi\left(\frac{\theta_U-\mu_p}{\sigma_p}\right)-\Phi\left(\frac{\theta_L-\mu_p}{\sigma_p}\right)\right]}}{\frac{\Phi\left(\frac{\theta_U-\mu_0}{\sigma_0}\right)-\Phi\left(\frac{\theta_L-\mu_0}{\sigma_0}\right)}{1-\left[\Phi\left(\frac{\theta_U-\mu_0}{\sigma_0}\right)-\Phi\left(\frac{\theta_L-\mu_0}{\sigma_0}\right)\right]}}$, con $\sigma_0 = \frac{1}{\sqrt{n_0}}$.

(c) Con i valori numerici dati si trova (vedi codice R sotto): $a^\star_\pi = \mathbb{E}[\Theta|\mathbf{z}_n] = 1.42$ e $\rho_\pi(a^\star_\pi, \mathbf{z}_n) = \mathbb{V}[\Theta|\mathbf{z}_n] = 0.66$

(d) In generale $a^\star_\pi(\mathbf{z}_n) = \begin{cases} a_0 & \text{se} \quad B_{01}(\mathbf{z}_n) > \overline{k} \\ a_1 & \text{se} \quad B_{01}(\mathbf{z}_n) < \overline{k} \end{cases}$, con $\overline{k} = \frac{b_1}{b_0}\frac{\pi_1}{\pi_0}$ e $\pi_= \mathbb{P}[H_i] = \mathbb{P}[\Theta \in \Omega_i]$, $i = 1, 2$.

Si verifica (vedi codice sotto) che $B_{01}(\mathbf{z}_n) = 2.40$ e che $k = 0.35$. Si sceglie quindi a_0, ovvero H_0.

(e) $\rho_\pi(a_0, \mathbf{z}_n) = b_1\mathbb{P}[H_1|\mathbf{z}_n] = 0.54$; $\rho_\pi(a_1, \mathbf{z}_n) = b_1\mathbb{P}[H_0|\mathbf{z}_n] = 0.46$ (vedi codice sotto).

(f) $B_{01}(\overline{x}_n) = \frac{p^{\overline{X}_n}_{\theta_0}(\overline{x}_n)}{m_{\overline{X}_n}(\overline{x}_n)} = \frac{\phi\left(\overline{x}_n; \theta_0, \frac{1}{n}\right)}{\phi\left(\overline{x}_n; \mu_0, \frac{1}{n}+\frac{1}{n_0}\right)}$, dove $\phi(x; a, b)$ indica il valore della densità di una v.a. $\mathrm{N}(a, b)$ in x.

(g) $B_{01} = 4.37$ (vedi codice sotto) e quindi si sceglie H_0.

(h)

Per il calcolo dei valori numerici, procedere come segue:

```
CODICE
xmed = 3/2; mu0=1; n0=1/2; n=3; thL = 1; thU = 2;  sig0=1/sqrt(n0)
# parametri a posteriori
mup=(n0*mu0+n*xmed)/(n0+n); sig2p=1/(n0+1); sigp=sqrt(sig2p)

# stima puntuale
a.star=mup
rho.star=varp

# prob a priori e prior odds
PH0=pnorm((thU-mu0)/sig0)-pnorm((thL-mu0)/sig0); PH1=1-PH0
prior.odds=PH0/PH1

# prob a posteriori e posterior odds
PH0.p=pnorm((thU-mup)/sigp)-pnorm((thL-mup)/sigp); PH1.p=1-PH0.p
post.odds=PH0.p/PH1.p

# BF per ipotesi intervallare
B01=post.odds/prior.odds
B01

# BF per ipotesi nulla puntuale e ipotesi alternativa bilaterale
xmed = 1.5; mu0 = 1.4; th0 = 1.6; n0 = 1/2; n = 10
N=dnorm(xmed,mean=th0,sd=1/sqrt(n))
D=dnorm(xmed, mean=mu0,sd=sqrt(1/n+1/n0))
B01.bil=N/D
B01.bil
```

B – Analisi della forma normale (AfN)

B1 – Stima puntuale - AfN

7. Sia $X_1, \ldots, X_n|\theta$ i.i.d. - Si considerino i modelli: Ber(θ), Pois(θ), Exp(θ), N(θ, σ^2) (σ^2 noto), N(μ_0, θ) (μ_0 noto), Unif$[0, \theta]$. Per ciascuno dei modelli indicati rispondere alle seguenti domande.
   (a) Determinare (o fornire l'espressione di) $d_{mv}(\mathbf{Z}_n)$, stimatore di massima verosimiglianza del parametro θ.
   (b) Determinare l'espressione di $R(d_{mv}, \theta)$ (utilizzando la perdita quadratica) e studiare la consistenza dello stimatore $d_{mv}(\mathbf{Z}_n)$.
   (c) Determinare (o fornire direttamente) l'approssimazione asintotica della distribuzione campionaria di $d_{mv}(\mathbf{Z}_n)$.
   (d) Verificare (argomentando) se $d_{mv}(\mathbf{Z}_n)$ coincide con lo stimatore UMVUE $d_u(\mathbf{Z}_n)$ di θ.

Soluzione.

(a) Le risposte ai quesiti (a)-(c) sono riassunti nella seguente tabella.
Per la stima di massima verosimiglianza, procedere con lo studio di $\ell(\theta)$. Tutti i modelli, tranne l'ultimo, sono modelli regolari.

Modello	d_{mv}	$R(\theta, d)$	$d_{mv}(\mathbf{Z}_n)\vert\theta \dot\sim \ldots$
Ber(θ)	$\overline{X}_n$	$\frac{\theta(1-\theta)}{n}$	$\mathrm{N}\left(\theta, \frac{\theta(1-\theta)}{n}\right)$
Pois(θ)	$\overline{X}_n$	$\frac{\theta}{n}$	$\mathrm{N}\left(\theta, \frac{\theta}{n}\right)$
Esp(θ)	$\overline{X}_n$	$\frac{\theta^2}{n}$	$\mathrm{N}\left(\theta, \frac{\theta^2}{n}\right)$
N1; N(θ, σ^2)	$\overline{X}_n$	$\frac{\sigma^2}{n}$	$\mathrm{N}\left(\theta, \frac{\sigma^2}{n}\right)$ (esatta)
N3; N(μ_0, θ)	S_0^2	$\frac{2\theta^2}{n}$	$\mathrm{N}\left(\theta, \frac{2\theta^2}{n}\right)$
Unif$[0, \theta]$	$X_{(n)}$	$\frac{2\theta^2}{(n+1)(n+2)}$	non si applica

(b) Vedi tabella precedente. Ricordare che, nel caso di perdita quadratica, $R(\theta, d_{mv}) = \mathbb{V}_\theta[d_{mv}] + (\mathbb{E}_\theta[d_{mv}] - \theta)^2$. In tutti i casi, eccetto l'ultimo, d_{mv} è stimatore non distorto di θ e quindi $R(\theta, d_{mv}) = \mathbb{V}_\theta[d_{mv}]$.
Per lo stimatore del parametro del modello uniforme ricordare che: $\mathbb{E}_\theta[X_{(n)}] = \frac{n}{n+1}\theta$ e che $\mathbb{E}_\theta[X_{(n)}^2] = \frac{n}{n+2}\theta$.
Per tutti gli stimatori considerati abbiamo che $\lim_{n\to\infty} R(\theta, d_{mv} = 0)$, $\forall \theta \in \Omega$. Sono quindi tutti stimatori considetnti in R.

(c) Tutti gli stimatori considerati sono funzioni di statistiche sufficienti e complete. I primi 5 sono anche stimatori non distorti e quindi UMVUE.amiglie esponenziali. Questi sono funzioni di statistiche complete per una proprietà dei modelli che sono f
La completezza del sesto stimatore (per il modello uniforme) va invece dimostrata (vedi dispense Inferenza); è stimatore distorto di θ e quindi è UMVUE del proprio valore atteso, ma non di θ. È però asintoticamente non distorto (e ottimo) per Θ.

8. Sia $X_1, \ldots, X_n|\theta$ i.i.d. - Si considerino i modelli: Ber(θ), Pois(θ), Exp(θ), N(θ, σ^2) (σ^2 noto), N(μ_0, θ) (μ_0 noto), Unif$[0, \theta]$, Beta($1, \theta$). Per ciascuno dei modelli indicati rispondere alle seguenti domande.
   (a) Determinare lo stimatore dei momenti $d_m(\mathbf{Z}_n)$ di θ.
   (b) Verificare in quali casi $d_m \neq d_u$
   (c) Calcolare (escluso il modello Beta) l'espressione di $R(d_m, \theta)$ (utilizzare la perdita quadratica).
   (d) Per il modello Unif$[0, \theta]$ determinare la distribuzione asintotica di $d_m(\mathbf{Z}_n)$.

Soluzione.

(a) - Trattandosi di modelli uniparametrici, lo stimatore dei momenti $d_m(\mathbf{Z}_n)$ si trova, per tutti i modelli eccetto il modello N3, risolvendo l'equazione $\mathbb{E}_\theta[X] = \overline{X}_n$. Si verifica quindi facilmente che, per i modelli Ber, Pois, N1, Esp, $d_m(\mathbf{Z}_n) = d_{mv}(\mathbf{Z}_n) = \overline{X}_n$.
- Per il modello $N2$ l'equazione $\mathbb{E}_\theta[X] = \overline{X}_n$ non contiene θ e quindi non può essere usata per determinare lo stimatori. Se si impone l'uguaglianza dei momenti centrali, ovvero $\mathbb{E}_\theta[(X-\theta)^2] = \frac{1}{n} = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$, si ottiene che $d_m(\mathbf{Z}_n) = \widehat{\sigma}^2_n$, con $\widehat{\sigma}^2_n = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.
- Per il modello uniforme si ha che $d_m(\mathbf{Z}_n) = 2\overline{X}_n$.
- Per il modello Beta si trova che $d_m(\mathbf{Z}_n) = \frac{1-\overline{X}_n}{\overline{X}_n}$.

(b) I due stimatori coincidono in tutti i casi tranne i seguenti.
- Modello N3: si ha che $d_m(\mathbf{Z}_n) = \widehat{\sigma}^2_n \neq d_{mv}(\mathbf{Z}_n) = S_0^2$.
- Modello uniforme: si ha che $d_m(\mathbf{Z}_n) = 2\overline{X}_n \neq d_{mv}(\mathbf{Z}_n) = X_{(n)}$.
- Modello Beta: si trova che $d_m(\mathbf{Z}_n) = \frac{1-\overline{X}_n}{\overline{X}_n} \neq d_{mv}(\mathbf{Z}_n) = -\frac{n}{\ln \prod_{i=1}^n X_i}$.

(c) Per tutti i modelli considerati, eccetto N3, lo stimatore d_m è non distorto e quindi $R(\theta, d_m) = \mathbb{V}_\theta(d_m)$.
- Per i modelli Ber, Pois, Esp, N1 vedi la tabella nella soluzione dell'esercizio precedente.
- Per N3, $\mathbb{E}_\theta[\widehat{\sigma}^2_n] = \frac{n-1}{n}\theta$ e $\mathbb{V}_\theta[\widehat{\sigma}^2_n] = \frac{2(n-1)}{n^2}\theta^2$. Si ha quindi che $R(\theta, \widehat{\sigma}^2_n) = \frac{2n-1}{n^2}\theta^2$.
- Per il modello uniforme, $R(\theta, 2\overline{X}_n) = \mathbb{V}_\theta[2\overline{X}_n] = \frac{\theta^2}{3n}$.

(d) Per il teorema del limite centrale si ha che $d_m(\mathbf{Z}_n) = 2\overline{X}_n|\theta \ \dot\sim\ \mathrm{N}\left(\theta, \frac{\theta^2}{3n}\right)$.

9. Sia $X_1, \ldots, X_n|\theta \sim \mathrm{N}(\theta_1, \theta_2)$ i.i.d. (entrambi i parametri sono incogniti). Considerare θ_2 come parametro di interesse.

(a) Determinare (o fornire l'espressione di) $d_{mv}(\mathbf{Z}_n)$ per θ_2.

(b) Determinare la distribuzione campionaria di $d_{mv}(\mathbf{Z}_n)$ (o di una sua trasformata notevole).

(c) Determinare l'approssimazione asintotica della distribuzione campionaria di $d_{mv}(\mathbf{Z}_n)$.

(d) Determinare l' stimatore UMVUE di θ e verificare che $d_u(\mathbf{Z}_n) = \frac{n}{n-1} d_{mv}(\mathbf{Z}_n)$.

(e) Determinare $R(d_{mv}, \theta_2)$ e $R(d_u, \theta_2)$ (perdita quadratica).

Soluzione.

(a) $\ell(\theta_1, \theta_2) \propto \frac{1}{\theta_2^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\theta_2}\sum_{i=1}^n (x_i - \theta_1)^2\right\}$. Risolvendo il sistema delle log-verosimiglianze, ovvero $\left\{\begin{array}{lcl} \frac{\partial}{\partial\theta_1}\ln \ell(\theta_1,\theta_2) & = & 0 \\ \frac{\partial}{\partial\theta_2}\ln \ell(\theta_1,\theta_2) & = & 0 \end{array}\right.$ si verifica la stima di massima verosimiglianza di (θ_1, θ_2) è la coppia $(\overline{x}_n, \widehat{\sigma}^2_n)$, con $\widehat{\sigma}^2_n = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Quindi lo stimatore di massima verosimiglianza per θ_2 è $\widehat{\sigma}^2_n$.

(b) Dai risultati del campionamento da popolazioni normali sappiamo che $\widehat{\sigma}^2_n|\theta_2 \sim \mathrm{Ga}\left(\frac{n-1}{2, \mathtt{rate}=\frac{n}{2\theta_2}}\right)$, ovvero che $\frac{n\widehat{\sigma}^2_n}{\theta_2} \sim \chi^2_{n-1}$.

(c) Per la distribuzione campionaria degli stimatori di massima verosimiglianza dei modelli regolari vale l'approssimazione normale. In questo caso possiamo dire che $\widehat{\sigma}^2_n|\theta_2 \ \dot\sim\ \mathrm{N}\left(\theta_2, \frac{2(n-1)}{n^2}\theta_2^2\right)$.

(d) $d_U^\star(\mathbf{Z}_n) = S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$, in quanto: (a) non distorto per θ_2; (b) funzione di $(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ statistica sufficiente e completa.

(e) $R(\theta_2, \widehat{\sigma}^2_n) = \frac{2n-1}{n^2}\theta_2^2$; $R(\theta, S_n^2) = \mathbb{V}_\theta[S_n^2] = \frac{2}{n-1}\theta_2^2$.

10. Sia $\mathbf{Z}_n = (X_1, \ldots, X_n)$ un campione casuale di dimensione n proveniente da una popolazione $\mathrm{N}(0, \theta)$ con parametro incognito θ e funzione di densità delle X_i

$$p_\theta(x) = \frac{1}{\sqrt{2\pi\theta}} \exp\{-\frac{1}{2\theta}x^2\}, \qquad x \in \mathbb{R}, \qquad \theta > 0.$$

(a) Verificare che lo stimatore di massima verosimiglianza è $d_{mv}(\mathbf{Z}_n) = \frac{1}{n}\sum_{i=1}^n X_i^2$

(b) Studiare la distorsione di d_{mv}.

(c) Determinare l'UMVUE di θ (argomentare).

(d) Verificare che, $\forall\theta > 0$, si ha

$$\frac{n\, d_{mv}(\mathbf{Z}_n)}{\theta} \sim \chi_n^2.$$

(e) Considerare la funzione di decisione $d_1(\mathbf{Z}_n) = \frac{n}{n+2} d_{mv}(\mathbf{Z}_n)$ e confrontarla con $d_1(\mathbf{Z}_n)$ in termini di funzione di rischio normale $R(\cdot, \theta)$ (perdita quadratica).

(f) Quali conclusioni si possono ottenere dal confronto effettuato, in termini di analisi pre-ottimale condotta con $R(\cdot, \theta)$?

Soluzione.

(a) $\ell(\theta) \propto \theta^{-\frac{n}{2}} \exp\left\{-\frac{n}{2\theta}S_0^2,\right\}$ con $S_0^2 = \frac{1}{n}\sum_{i=1}^n X_i^2$. Dallo studio delle derivate prima e seconda di $\ln \ell(\theta)$ si trova che $d_{mv} = S_0^2$.

(b) Sappiamo che $\frac{nS_0^2}{\theta} \sim \chi_n^2$ [vedi anche punto (d)]. Tale v.a. ha quindi valore atteso uguale a n. Si verifica pertanto che $\mathbb{E}_\theta[S_0^2] = \theta$, $\forall\theta \in \mathbb{R}^+$.

(c) S_0^2 è UMVUE di θ in quanto: (a) stimatore non distorto di θ; (b) funzione di statistica sufficiente e completa (S_0^2 stessa).

(d) $X_i|\theta \sim \mathrm{N}(0, \theta) \;\Rightarrow\; Z_i = \frac{X_i - 0}{\sqrt{\theta}} \sim \mathrm{N}(0, 1) \;\Rightarrow\; Z_i^2 = \frac{X_i^2}{\theta} \sim \chi_1^2 \;\Rightarrow\; \sum_{i=1}^n Z_i^2 = \frac{nd_{mv}}{\theta} \sim \chi_n^2$.

(e) Si verifica che $R(\theta, d_1) = \frac{2}{n+2}\theta^2 < \frac{2}{n}\theta^2 = R(\theta, d_{mv})$, $\forall\theta \in \mathbb{R}^+$.

(f) $d_{mv} \equiv d_U^\star$ è inammissibile.

B2 – Stima intervallare (AfN) – Esercizi[1]

(Si riporta prima di ogni esercizio la numerazione relativa alla dispense Esercizi Inferenza statistica (vedere da pag. 52 in poi), alla quale si rimanda anche per le soluzioni. La notazione è leggermente diversa).

11. (Esercizio 1 disp.). Sia $\mathbf{Z}_n = (X_1, \ldots, X_n)$ un campione casuale dalla popolazione $\mathrm{N}(0, \theta)$.

    (a) Verificare che $d_u(\mathbf{Z}_n) = (S_0^2) = \sum_{i=1}^n X_i^2/n$ è lo stimatore UMVUE di θ.

    (b) Determinare una quantità pivotale $Q(\mathbf{z}_n, \theta)$ per θ basata su $d_u(\mathbf{Z}_n)$.

    (c) Determinare un intervallo di confidenza $C(\mathbf{z}_n)$ di livello $1-\alpha$ per θ in funzione di $d_u(\mathbf{Z}_n)$.

    (d) Determinare $R(\theta, C)$ utilizzando la funzione di perdita lineare.

Soluzione.

(a) Ricordando che $\frac{nS_0^2}{\theta} \sim \chi_n^2$, possiamo verificare facilmente che S_0^2 è: (a) stimatore non distorto di θ; (b) statistica sufficiente e completa (si tratta di famiglia esponenziale).

(b) $Q(\mathbf{z}_n, \theta) = \frac{nS_0^2}{\theta}$ soddisfa le tre proprietà che caratterizzano le quantità pivotali.

(c) $C = [L(\mathbf{Z}_n), U(\mathbf{Z}_n)] = \left[\frac{nS_0^2}{\chi^2_{n;1-\frac{\alpha}{2}}}, \frac{nS_0^2}{\chi^2_{n;\frac{\alpha}{2}}},\right]$

(d) $\mathcal{L}(\mathbf{Z}_n) = U(\mathbf{Z}_n) - L(\mathbf{Z}_n) = \ldots = c_n S_0^2$. Pertanto $R(\theta, C) = \mathbb{E}_\theta[\mathcal{L}(\mathbf{Z}_n)] - (1-\alpha) = c_n \mathbb{E}_\theta[S_0^2] = c_n\theta - (1-\alpha)$.

12. Sia $\mathbf{Z}_n = (X_1, \ldots, X_n)$ un campione casuale dalla popolazione $\mathrm{N}(\theta_1, \theta_2)$. Sia $\theta = \theta_2$ il parametro di interesse.

    (a) Verificare che $d(\mathbf{Z}_n) = S_n^2 = \sum_{i=1}^n (X_i - \overline{X}_n)^2/(n-1)$ è lo stimatore UMVUE di θ.

    (b) Determinare una quantità pivotale $Q(\mathbf{z}_n, \theta)$ per θ basata su $d_u(\mathbf{Z}_n)$.

    (c) Determinare un intervallo di confidenza $C(\mathbf{Z}_n)$ di livello $1-\alpha$ per θ in funzione di $d_u(\mathbf{Z}_n)$.

    (d) Determinare $R(\theta, C)$ utilizzando la funzione di perdita lineare.

Soluzione.

(a) Ricordando che $\frac{(n-1)S_n^2}{\theta} \sim \chi_{n-1}^2$, possiamo verificare facilmente che S_n^2 è: (a) stimatore non distorto di θ_2; (b) statistica sufficiente e completa (si tratta di famiglia esponenziale).

(b) $Q(\mathbf{z}_n, \theta_2) = \frac{(n-1)S_n^2}{\theta_2}$ soddisfa le tre proprietà che caratterizzano le quantit'a pivotali.

(c) $C = [L(\mathbf{Z}_n), U(\mathbf{Z}_n)] = \left[\frac{(n-1)S_n^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}, \frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{\alpha}{2}}},\right]$

(d) $\mathcal{L}(\mathbf{Z}_n) = U(\mathbf{Z}_n) - L(\mathbf{Z}_n) = \ldots = c_n S_n^2$. Pertanto $R(\theta_2, C) = \mathbb{E}_{\theta_2}[\mathcal{L}(\mathbf{Z}_n)] - (1-\alpha) = c_n \mathbb{E}_\theta[S_n^2] = c_n\theta_2 - (1-\alpha)$.

13. (Esercizio 17 disp.). Sia $\mathbf{Z}_n = (X_1, \ldots, X_n)$ un campione casuale da una popolazione esponenziale di parametro θ, per la quale $\mathbb{E}[X_i] = \theta$ e $\mathbb{V}[X_i] = \theta^2$. Fissato un valore $\alpha \in (0,1)$, si consideri il seguente l'intervallo aleatorio:

$$C(\mathbf{Z}_n) = \left[\frac{2\sum_{i=1}^n X_i}{\chi^2_{1-\frac{\alpha}{2},2n}}, \frac{2\sum_{i=1}^n X_i}{\chi^2_{\frac{\alpha}{2},2n}}.\right]$$

[1]Dove non diversamente indicato, si considera per gli intervalli di stima $C(\mathbf{z}_n)$ la funzione di perdita lineare: $L(\theta, C(\mathbf{z}_n)) = b\mathcal{L}(\mathbf{z}_n) - I_{C(\mathbf{z}_n)}(\theta)$, $b > 0$, dove $\mathcal{L}_C(\mathbf{z}_n)$ indica la lunghezza di $C(\mathbf{z}_n)$. Per gli intervalli con probabilità di copertura costante e pari a $1-\alpha$ si ha quindi che: $R(\theta, C) = \mathbb{E}[L(\theta, C(\mathbf{Z}_n))] = b\mathbb{E}[\mathcal{L}_C(\mathbf{Z}_n)] - (1-\alpha)$.

(a) Verificare che [2] $Q(\mathbf{Z}_n, \theta) = \frac{2\sum_{i=1}^n X_i}{\theta} \sim \chi^2_{2n}$.

(b) Verificare che $C(\mathbf{Z}_n)$ è l' intervallo di confidenza di livello $1-\alpha$ per θ ottenuto da $Q(\mathbf{Z}_n, \theta)$.

(c) Determinare valore atteso e varianza di $\mathcal{L}_C(\mathbf{Z}_n)$.

(d) Determinare $R(\theta, C)$ (usare perdita lineare).

(e) Supponendo di avere osservato un campione di $n = 10$ osservazioni dalla popolazione considerata, in cui la somma campionaria è pari a 1740, calcolare gli estremi dell'intervallo di confidenza di livello 0.95 e la lunghezza osservata.

Soluzione.

(a) Dalle proprietà delle v.a. Gamma (parametrizzazione rate) sappiamo che, per campioni i.i.d.: $X_i \sim \text{Exp}(\theta) = \text{Ga}(1, \theta) \Rightarrow \sum_{i=1}^n X_i \sim \text{Ga}(n, \theta) \Rightarrow \frac{2\sum_{i=1}^n X_i}{\theta} \sim \text{Ga}\left(n, \frac{1}{2}\right) = \chi^2_{2n}$. Quindi: $Q(\mathbf{Z}_n, \theta) = \frac{2\sum_{i=1}^n X_i}{\theta} \sim \chi^2_{2n}$, da cui l'intervallo proposto.

(b) Segue dal punto precedente.

(c) $\mathcal{L}[C(\mathbf{Z}_n)] = \ldots = c_n \overline{X}_n$. Pertanto $R(\theta, C) = \mathbb{E}_\theta[\mathbb{L}_C(\mathbf{Z}_n)] - (1-\alpha) = c_n\theta - (1-\alpha)$.

(d) Procedere come segue.

```
sn=1740; alpha=0.05; n=19;
q1=qchisq(alpha/2,df=2*n); q2=qchisq(1-alpha/2,df=2*n)
N=2*sn; L=N/q2; U=N/q1
C=c(L,U); Lungh=U-L
C
Lungh
```

Si ottiene $C = [61.164, 152.107]$ e $\mathcal{L}[C(\mathbf{z}_n)] = 90.943$.

14. (Esercizio 14 disp.). Sia $X_1, \ldots X_n$ un campione casuale da una popolazione uniforme in $[0, \theta]$, $\theta > 0$. Si consideri la statistica campionaria $X_{(n)} = \max\{X_1, \ldots, X_n\}$, per la quale è noto che, $\forall \theta > 0$, $\mathbb{E}[X_{(n)}] = \frac{n}{n+1}\theta$, $\mathbb{V}[X_{(n)}] = \frac{n}{(n+1)^2(n+2)}\,\theta^2$.

(a) Verificare che, fissato $\alpha \in (0,1)$, l'insieme aleatorio $C(\mathbf{Z}_n) = \left[X_{(n)}, \frac{1}{\alpha^{1/n}}\, X_{(n)}\right]$, è un intervallo di confidenza per θ di livello $1-\alpha$.

(b) Verificare che $R(\theta, C) = b\mathbb{E}_\theta[\mathcal{L}_C(\mathbf{Z}_n)] - (1-\alpha)$, dove $\mathbb{E}_\theta[\mathcal{L}_C(\mathbf{Z}_n)] = \frac{n}{n+1}\left[\frac{1-\alpha^{1/n}}{\alpha^{1/n}}\right]\theta$.

Soluzione.

(a) La statistica sufficiente per il modello è $X_{(n)}$, che ha distribuzione campionaria $p_\theta(y) = ny^{n-1}/\theta^n$, $y \in [0, \theta]$. È semplice verificare che $Q(\mathbf{Z}_n, \theta) = X_{(n)}/\theta$ ha distribuzione $p_\theta(y) = nt^{n-1}$, $t \in [0,1]$ e che quindi è una quantità pivotale per θ. Più precisamente, $Q(\mathbf{Z}_n, \theta) = \frac{X_{(n)}}{\theta} \sim \text{Beta}(n, 1)$. Pertanto $\forall \theta > 0$, si ha $1 - \alpha = \mathbb{P}\left(q_{n,\alpha} \leq \frac{X_{(n)}}{\theta} \leq 1\right) = \mathbb{P}\left(X_{(n)} \leq \theta \leq \frac{X_{(n)}}{q_{n,\alpha}}\right)$, dove $q_{n,\alpha}$ è il quantile di livello α della v.a. Beta$(n,1)$. In questo esempio è semplice ottenere un'espressione esplicita di $q_{n,\alpha}$ in funzione di α. Infatti: $1 - \alpha = \int_{q_{n,\alpha}}^1 nt^{n-1}dt = t^n|_{q_{n,\alpha}}^1 = 1^n - (q_{n,\alpha})^n = 1 - (q_{n,\alpha})^n$, da cui si ottiene che $q_{n,\alpha} = \alpha^{1/n}$.

(b) La lunghezza dell'intervallo è $\mathcal{L}(\mathbf{X}_n) = (\alpha^{-1/n} - 1)X_{(n)}$ e e quindi, usando la perdita $\mathbb{L}(\theta, C) = \mathcal{L}(\mathbf{X}_n) - (1-\alpha)$ si ha che $R(\theta, C) = \frac{bn}{n+1}\left[\frac{1-\alpha^{1/n}}{\alpha^{1/n}}\right]\theta - (1-\alpha)$.

[2] Considerare $p_\theta(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$, $x \geq 0$, $\theta > 0$ e verificare che $\sum_{i=1}^n X_i|\theta \sim \text{Ga}(n, \theta)$, dove la f.ne di densità di una va. Ga(α, β) è $p(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}}$, $x \geq 0$, $\alpha, \beta > 0$. Verificare quindi che $Q(\mathbf{z}_n, \theta) = \frac{2\sum_{i=1}^n X_i}{\theta} \sim \chi^2_{2n}$ (da usare come quantità pivotale).

15. (Esercizio 30 disp.). Sia $\mathbf{Z}_n = (X_1, \dots, X_n)$ un campione casuale proveniente dalla popolazione con funzione di densità di probabilità $p_\theta(x) = \frac{3}{\theta^3}x^2,\ 0 < x < \theta,\ \theta > 0$.

   (a) Verificare che $\mathbb{E}[X] = \frac{3}{4}\theta,\ \mathbb{V}[X] = \frac{3}{80}\theta^2$. Verificare che la funzione di decisione $d(\mathbf{Z}_n) = \frac{4}{3}\ \overline{X}_n$ è uno stimatore non distorto di θ.

   (b) Determinare $R(\theta, d_n)$ e studiarne il comportamento al crescere di n (usare perdita quadratica).

   (c) Determinare la distribuzione asintotica dello stimatore $d(\mathbf{Z}_n)$.

   (d) Determinare una quantità pivotale per θ basata sulla distribuzione asintotica di d e determinare da questa un intervallo di confidenza (asintotico) C di livello $1-\alpha$ (usare $\mathbb{L}(\theta, C) = \mathcal{L}_C(\mathbf{Z}_n)$).

   (e) Sfruttando la distribuzione asintotica di d, determinare $R(\theta, C)$.

Soluzione.

(a) $\mathbb{E}_\theta[d] = \frac{4}{3}\mathbb{E}_\theta[X] = \frac{4}{3} \cdot \frac{3}{4}\theta = \theta$, $\forall\theta > 0$. Lo stimatore è quindi non distorto per θ.

(b) Lo stimatore è non distorto e quindi $R(\theta, d) = \mathbb{V}_\theta\left[\frac{4}{3}\overline{X}_n\right] = \frac{\theta^2}{15n}$.

(c) Per il teorema del limite centrale abbiamo che $d(\mathbf{Z}_n) = \frac{4}{3}\overline{X}_n \dot\sim \mathrm{N}\left(\theta, \frac{\theta^2}{15n}\right)$.

(d) $\tilde{Q}(\mathbf{Z}_n, \theta) = \sqrt{15n}\frac{[d(\mathbf{Z}_n)-\theta]}{\theta}$ è una q.p. asintotica, da cui l'intervallo $C(\mathbf{Z}_n) = d(\mathbf{Z}_n) \ \pm\ \frac{d(\mathbf{Z}_n)}{\theta} = \frac{4}{3}\overline{X}_n \ \pm\ \frac{4}{3\sqrt{15n}}u_{1-\frac{\gamma}{2}}\overline{X}_n$

(e) $R(\theta, C) = \frac{8}{3}\frac{1}{\sqrt{15n}}\frac{3}{4}\theta = \frac{2}{\sqrt{15n}}\theta$.

16. (Esercizio 32 disp.). Sia $\mathbf{Z}_n = (X_1, \dots, X_n)$ un campione casuale proveniente dalla popolazione con funzione di densità di probabilità $f_X(x;\theta) = \frac{1}{2\sqrt{\theta}}\frac{1}{\sqrt{x}},\ 0 < x < \theta,\ \theta > 0$.

   (a) Verificare che $\mathbb{E}[X] = \frac{1}{3}\theta,\ \mathbb{V}[X] = \frac{4}{45}\theta^2$.

   (b) Determinare $R(\theta, d)$ per lo stimatore di θ definito da $d(\mathbf{Z}_n) = 3\overline{X}_n$ (usare perdita quadratica) e studiare il comportamento di $R(\theta, d)$ al crescere di n.

   (c) Determinare la distribuzione asintotica di $d(\mathbf{Z}_n)$ e l'intervallo di confidenza asintotico per θ.

   (d) Determinare $d(\mathbf{z}_0)$ e $C(\mathbf{z}_0)$ con $1-\gamma = 0.95$, supponendo di avere osservato un campione $\mathbf{z}_0$ di dimensione $n = 36$ per il quale $\sum_{i=1}^n x_i = \frac{1}{3}$.

Soluzione.

(a) $\mathbb{E}_\theta[X] = \int_0^\theta x\frac{1}{2\sqrt{\theta}}\frac{1}{\sqrt{x}}dx = \frac{1}{2\theta^{\frac{1}{2}}}\left[\frac{x^{\frac{3}{2}}}{\frac{3}{2}}\right]_0^\theta = \frac{\theta}{3}$; $\mathbb{E}_\theta[X^2] = \int_0^\theta x^2\frac{1}{2\sqrt{\theta}}\frac{1}{\sqrt{x}}dx = \frac{1}{2\theta^{\frac{1}{2}}}\left[\frac{x^{\frac{5}{2}}}{\frac{5}{2}}\right]_0^\theta = \frac{\theta^2}{5}$;
$\mathbb{V}_\theta[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{\theta^2}{5} - \left(\frac{\theta}{3}\right)^2 = \frac{4}{45}\theta^2$.

(b) Si verifica facilmente $d(\mathbf{Z}_n)$ (stimatore dei momenti) è non distorto e quindi: $R(\theta, d) = \mathbb{V}_\theta[d(\mathbf{Z}_n)] = \mathbb{V}_\theta(3\overline{X}_n) = 9\frac{\mathbb{V}(X)}{n} = 9\frac{4}{45}\frac{1}{n}\theta^2 = \frac{4\theta^2}{5n}$ Poiché $R(\theta, d)$ tende a 0 per $n \to \infty$ possiamo concludere che lo stimatore dei momenti è consistente.

(c) Ricordando che, per il teorema del limite centrale, $\overline{X}_n \dot\sim \mathrm{N}\left(\mathbb{E}[X], \frac{\mathbb{V}[X]}{n}\right)$, si ricava facilmente che $d(\mathbf{Z}_n) = 3\overline{X}_n \dot\sim \mathrm{N}\left(3\mathbb{E}[X], 9\frac{\mathbb{V}[X]}{n}\right) \dot\sim \mathrm{N}\left(\theta, \frac{4\theta^2}{5n}\right)$.
$C = d(\mathbf{x}_n)\ \pm\ u_{1-\frac{\gamma}{2}}d(\mathbf{x}_n)\sqrt{\frac{4}{5n}}$.

(d) $d(\mathbf{z}_0) = 3\overline{x}_n = 3\frac{\sum_{i=1}^n x_i}{n} = 3\frac{1}{3}\frac{1}{36} = \frac{1}{36} = 0.028$.
Per l'intervallo sostituire i valori.

17. (Esercizio 35 disp.). Sia $\mathbf{Z}_n = (X_1, \ldots, X_n)$ un campione casuale proveniente da una distribuzione uniforme nell'intervallo $[0, \theta]$.

    (a) Verificare che $d(\mathbf{Z}_n) = 2\overline{X}_n$ è non distorto per θ ma che non è l'UMVUE.

    (b) Determinare distribuzione asintotica di $d(\mathbf{Z}_n)$ e, con quest'ultima, un intervallo di confidenza asintotico $C(\mathbf{Z}_n)$per θ di livello $1-\alpha$.

    (c) Dato un campione osservato $\mathbf{z}_0$ di dimensione $n = 20$ in cui $x_{(1)} = 0.2, \overline{x}_n = 0.4, S_n^2 = 0.1, x_{(n)} = 0.9$ determinare $d(\mathbf{z}_0)$ e $C(\mathbf{z}_0)$ ponendo $1-\alpha = 0.90$

Soluzione.

(a) Si tratta dello stimatore dei momenti. Infatti, poichè il primo momento di X è $\mu_1(\theta) = \mathbb{E}[X] = \frac{\theta}{2}$ e il primo momento campionario è $m_1(\mathbf{X}_n) = \overline{X}_n$, l'equazione dei momenti($\mu_1(\theta) = m_1(\mathbf{X}_n)$) diventa: $\frac{\theta}{2} = \overline{X}_n$, da cui si ottiene lo stimatore dei momenti: $d(\mathbf{Z}_n) = 2\overline{X}_n$. Osservando che, per ogni $\theta > 0$ si ha $\mathbb{E}_\theta[d(\mathbf{Z}_n)] = \mathbb{E}_\theta[2\overline{X}_n] = \theta$, possiamo dire che lo stimatore è non distorto per θ ma, non essendo funzione della statistica sufficiente $X_{(n)}$, non si tratta di UMVUE.

(b) Poichè $\mathbb{V}[d(\mathbf{Z}_n)] = \mathbb{V}[2\overline{X}_n] = \frac{\theta^2}{3n}$, per il teorema del limite centrale possiamo affermare che, per un qualsiasi θ, $\frac{d(\mathbf{Z}_n)-\theta}{\sqrt{\mathbb{V}[d(\mathbf{Z}_n)]}} \dot{\sim} \mathrm{N}(0,1)$ ovvero che $d(\mathbf{Z}_n) \dot{\sim} \mathrm{N}\left(\theta, \frac{\theta^2}{3n}\right)$. Da quanto appena determinato discende che, per un qualsiasi $\alpha \in (0,1)$, l'intervallo di confidenza asintotico per θ è $d(\mathbf{Z}_n) \pm z_{1-\frac{\alpha}{2}} \sqrt{\widehat{\mathbb{V}_\theta[d(\mathbf{Z}_n)]}}$, ovvero $2\overline{X}_n \pm z_{1-\frac{\alpha}{2}} \frac{2}{\sqrt{3n}} \overline{X}_n$.

(c) $C = [0.630, 0.969]$. Si ottiene da:

```
n=20
xmed=0.4
alpha=0.1
L=2*xmed-qnorm(1-alpha/2)*2*xmed/sqrt(3*n)
U=2*xmed+qnorm(1-alpha/2)*2*xmed/sqrt(3*n)
c(L,U)
```

18. (Esercizio 23 disp.). Sia $Z = (X_1, \ldots, X_n)$ un campione casuale da una popolazione di Poisson di parametro θ. Per la quantità $g(\theta) = e^{-\theta}$:

    (a) determinare lo stimatore di massima verosimiglianza di $g(\theta)$, $d_{mv}(\mathbf{z}_n)$;

    (b) verificare che $d_{mv}(\mathbf{Z}_n) \dot{\sim} \mathrm{N}\left(e^{-\theta}, e^{-2\theta}\frac{\theta}{n}\right)$;

    (c) determinare una stima per la varianza asintotica di $d_{mv}(\mathbf{Z}_n)$ e un intervallo di confidenza approssimato $C(\mathbf{Z}_n)$;

    (d) determinare $d_{mv}(\mathbf{z}_0)$ e $C(\mathbf{z}_0)$ per il seguente campione osservato: $x_1 = \ldots = x_5 = 2, x_6 = \ldots = x_{10} = 1, x_{11} = x_{15} = 3, x_{16} = \ldots = x_{25} = 4$ (porre $1-\alpha = 0.95$)

Soluzione.

(a) Lo stimatore di massima verosimiglianza di θ è $\overline{X}_n$. Per la proprietà di equivarianza si ha quindi che $d_{mv} = e^{-\overline{X}_n}$.

(b) Per il metodo Delta sappiamo che $\mathbb{V}_\theta[e^{-\overline{X}_n}] = g'(\theta)^2\ I_n(\theta)^{-1} = e^{-2\theta}\ \frac{\theta}{n}$. Pertanto: $\widehat{\mathbb{V}_\theta[e^{-\overline{X}_n}]} = e^{-2\overline{X}_n}\ \frac{\overline{X}_n}{n}$ e quindi $d_{mv}(\mathbf{Z}_n)\ \dot\sim\ \mathrm{N}\left(e^{-\theta}, e^{-2\theta}\frac{\theta}{n}\right)$. L'intervallo di confidenza asintotico è quindi: $e^{-2\overline{X}_n}\ \pm\ u_{1-\frac{\alpha}{2}}\ e^{-\overline{X}_n}\ \sqrt{\frac{\overline{X}_n}{n}}$.

(c) $C = [0.020, 0.100]$. Si ottiene da:

```
a=rep(2,5); b=rep(1,5); c=rep(3,5); d=rep(4,10)
dati=c(a,b,c,d)
n=length(dati)
xmed=mean(dati); xmed
d.mv=exp(-xmed); d.mv
alpha=0.05
L=d.mv - qnorm(1-alpha/2)*d.mv*sqrt(xmed/n)
U=d.mv + qnorm(1-alpha/2)*d.mv*sqrt(xmed/n)
c(L,U)
```

19. (I scritto 2012). Sia $Z = (X_1, \ldots, X_n)$ un campione casuale proveniente da una popolazione di Poisson di parametro incognito $\theta/2$, $\theta > 0$.

(a) Determinare $d_{mv}(\mathbf{Z}_n)$ e verificare che la sua distribuzione asintotica è: $\mathrm{N}\left(\theta, \frac{2\theta}{n}\right)$.

(b) Verificare che l'intervallo aleatorio $C(\mathbf{Z}_n) = \left[2\overline{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{4\overline{X}_n}{n}}, 2\overline{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{4\overline{X}_n}{n}}\right]$ è un intervallo di confidenza approssimato di livello $1-\alpha$ per θ.

(c) Determinare $R(\theta, d_{mv}) = \mathbb{E}[\mathcal{L}_C(\mathbf{Z}_n)^2]$ e il minimo valore di n^* per il quale $R(\theta, d_{mv}) \leq \ell_0$, dove $\mathcal{L}_C(\mathbf{Z}_n)$ indica la lunghezza (aleatoria) dell'intervallo considerato.

(d) Determinare n^* assumendo $\alpha = 0.05$, $\theta = 1$ e $\ell_0 = 0.5$.

Soluzione.

(a) Se si pone $\lambda = \frac{\theta}{2}$, lo stimatore di massima verosimiglianza di λ è $\overline{X}_n$ e quindi, per equivarianza degli stimatori MV si ottiene che lo SMV di θ è $d_{\mathbf{Z}_n} = 2\overline{X}_n$. (Si può anche risolvere direttamente, senza ricorrere alla parametrizzazione in λ).

(b) $\mathbb{E}_\theta[d(\mathbf{Z}_n)] = \theta$ e $\mathbb{V}_\theta[d_{mv}(\mathbf{Z}_n)] = \frac{2\theta}{n}$. Per il teorema del limite centrale abbiamo quindi che $d_{mv}(\mathbf{Z}_n)|\theta\ \dot\sim\ \mathrm{N}\left(\theta, \frac{2\theta}{n}\right)$.

(c) Deriva dal risultato generale $\tilde{C} = d_{mv}(\mathbf{Z}_n)\ \pm\ u_{1-\frac{\alpha}{2}}\ \sqrt{\mathbb{V}_\theta\,[\widehat{d_{mv}(\mathbf{Z}_n)}]}$.

(d) $\mathcal{L}_C(\mathbf{Z}_n) = 4\ u_{1-\frac{\alpha}{2}}\ \sqrt{\frac{\overline{X}_n}{n}}\ \Rightarrow\ \mathcal{L}_C(\mathbf{Z}_n)^2 = 16\ u^2_{1-\frac{\alpha}{2}}\ \frac{\overline{X}_n}{n}$. Pertanto $R(\theta, d_{mv}) = \frac{16}{n}\ u^2_{1-\frac{\alpha}{2}}\ \theta$. Si ha quindi $n^\star \approx \frac{8}{\ell_0}\ u^2_{1-\frac{\alpha}{2}}\ \theta$.

(e) Si trova $n^\star = 61.4 \approx 62$. Calcolare come segue:

```
alpha=0.05; theta=1; ell.0=0.5
n.star=8*qnorm(1-alpha/2)^2*theta/ell.0
n.star
```

(Si riporta prima di ogni esercizio la numerazione relativa alla dispense Esercizi Inferenza statistica (vedere da pag. 52 in poi), alla quale si rimanda anche per le soluzioni. La notazione è leggermente diversa).

20. (Esercizio 5 disp.). Si consideri un campione casuale $\mathbf{Z} = (X_1, \ldots, X_n)$ da una popolazione bernoulliana di parametro incognito θ e le due ipotesi $H_0 : \theta = 0.2$ vs. $H_1 : \theta = 0.4$.
    (a) Verificare che la generica regione di rifiuto del test d^* di Neyman-Pearson è $\mathcal{Z}_1^* = \{z \in \mathcal{Z} : \sum_{i=1}^n x_i \geq k\}$.
    (b) Determinare il valore della probabilità di errore di prima specie, α, che si ottiene ponendo $n = 4$ e $k = 2$.

> **Soluzione**.
>
> (a) Il rapporto delle verosimiglianze è $\lambda_{01}(\mathbf{z}_n) = \left(\frac{\theta_0}{\theta_1}\right)^{y_n} \left(\frac{1-\theta_0}{1-\theta_1}\right)^{n-y_n}$, con $y_n = \sum_{i=1}^n x_i$. In generale si rifiuta H_0 se $\lambda_{01}(\mathbf{z}_n) \leq k'$. Per $\theta_0 < \theta_1$, tale rapporto è decrescente in y_n. Quindi si rifiuta H_0 se $y_n \geq k$.
>
> (b) $\alpha = \mathbb{P}_{\theta_0}(R) = \mathbb{P}_{\theta_0}(Y_n \geq k) = \sum_{h=k}^n \mathbb{P}_{\theta_0}(Y_n = h)$, con $Y_n = \sum_{i=1}^n X_i$ v.a. binomiale di parametri (n, θ_0). Per $n = 4$ e $k = 2$ si ha $\alpha = \mathbb{P}_{\theta_0}(Y_n = 2) + \mathbb{P}_{\theta_0}(Y_n = 3) + \mathbb{P}_{\theta_0}(Y_n = 4) = ... = 0.1808$. Si può calcolare con
>
> ```
> sum(dbinom(2:4,size=4,prob=0.2))
> ```

21. (Esercizio 25 disp.). Si consideri un campione casuale $\mathbf{Z} = (X_1, \ldots, X_n)$ da una popolazione di Poisson di parametro incognito θ. Si consideri il sistema di ipotesi: $H_0 : \theta = \theta_0 = 2$ vs. $H_1 : \theta = \theta_1 = 1$.
    (a) Verificare che, per $n = 3$, la regione di rifiuto del test d^* di Neyman-Pearson risulta essere $\mathcal{Z}_1^* = \{z \in \mathcal{Z} : x_1 + x_2 + x_3 \leq k\}$.
    (b) Calcolare i valori di $R(d^*, \theta_0) = \alpha$ e della potenza del test, $1 - R(d^*, \theta_1) = 1 - \beta$, che si ottengono ponendo $k = 1$.

> **Soluzione**.
>
> (a) $\lambda_{01}(\mathbf{z}_n) = e^{-n(\theta_0 - \theta_1)} \left(\frac{\theta_0}{\theta_1}\right)_n^y$, con $y_n = \sum_{i=1}^n x_i$. In generale si rifiuta H_0 se $\lambda_{01}(\mathbf{z}_n) \leq k'$. Per $\theta_0 > \theta_1$ la funzione λ_{01} è crescente in y_n e quindi si rifiuta H_0 se $y_n \leq k$. Per $n = 3$ abbiamo che $y_n = x_1 + x_2 + x_3$ e quindi quanto richiesto risulta verificato.
>
> (b) $Y_3 = X_1 + X_2 + X_3 \mid \theta \sim$ Pois(3θ). Quindi:
> $\alpha = \mathbb{P}_{\theta_0}[Y_3 \leq 1] = \mathbb{P}_{\theta_0}[Y_3 = 0] + \mathbb{P}_{\theta_0}[Y_3 = 1] = e^{-6} + 6e^{-6} = 0.017$.
> Analogamente
> $1 - \beta = \mathbb{P}_{\theta_1}[Y_3 \leq 1] = \mathbb{P}_{\theta_1}[Y_3 = 0] + \mathbb{P}_{\theta_1}[Y_3 = 1] = e^{-3} + 3e^{-3} = 0.199$.
> Verifica con
>
> ```
> th.0=2
> sum(dpois(0:1,3*th.0))
> th.1=1
> sum(dpois(0:1,3*th.1))
> ```

22. (Esercizio 9 disp.). Si consideri un campione casuale $\mathbf{Z} = (X_1, \ldots, X_n)$ da una popolazione con funzione di densità $f_X(x; \theta) = \theta e^{-\theta x}, \quad x > 0, \quad \theta > 0$. Si considerino, per il parametro incognito θ, le ipotesi $H_0 : \theta = 1$ vs. $H_1 : \theta = 2$.

(a) Verificare che la generica regione di rifiuto del test d^* di Neyman-Pearson risulta essere: $\mathcal{Z}_1^* = \{z \in \mathcal{Z} : n\overline{x}_n = \sum_{i=1}^n x_i \leq k\}$.

(b) Determinare il valore di $R(\theta_0, d^\star)$ (probabilità di errore di prima specie) che si ottiene ponendo $n = 36$ e $k = 27$. [Sugg.: Si ricordi che la somma di v.a. esponenziali indipendenti ha distribuzione gamma].

(c) Ripetere il punto precedente utilizzando l'approssimazione normale per la distribuzione campionaria di $\overline{X}_n$, ricordando che $E_\theta(X) = 1/\theta$ e che $\mathrm{V}_\theta(X) = 1/\theta^2$.

Soluzione.

(a) $\ell(\theta) = \theta^n e^{-y_n\theta}$, con $y_n = \sum_{i=1}^n x_i$. Quindi $\lambda_{10}(\mathbf{z}_n) = \left(\frac{\theta_1}{\theta_0}\right)^n e^{-y_n(\theta_1-\theta_0)}$ che, per $\theta_1 > \theta_0$, è decrescente in y_n. Pertanto $\lambda_{10}(\mathbf{z}_n) \geq 1 \iff \overline{x} \leq k$.

(b) $X_i|\theta \sim \mathrm{Ga}(1, \texttt{rate} = \theta) \Rightarrow \sum_{i=1}^n X_i|\theta \sim \mathrm{Ga}(n, \texttt{rate} = \theta)$
$\alpha = R(\theta_0, d^\star) = \mathbb{P}_{\theta_0}[R] = \mathbb{P}_{\theta_0}[\sum_{i=1}^n X_i \leq k = 2/3] = \mathbb{F}_{\theta_0}(2/3)$, dove $\mathbb{F}_{\theta_0}(\cdot)$ indica la funzione di ripartizione della v.a. gamma di parametri $(n = 36, \theta_0 = 1)$. Il valore numerico, pari a 0.055, si ottiene con

```
k=27
n=36
pgamma(k,shape=n,rate=1)
```

(c) $\sum_{i=1}^n X_i|\theta_0 \;\dot\sim\; \mathrm{N}\left(\frac{n}{\theta_0}, \frac{n}{\theta_0^2}\right) = \mathrm{N}(36, 36)$ e $\mathbb{P}_{\theta_0}(\sum_{i=1}^n X_i \leq 27) \approx \Phi\left(\frac{27-36}{6}\right) = 0.066$, valore che si ottiene da `pnorm((k-36)/6)`.

23. (Esercizio 13 disp.). Sia $\mathbf{Z} = (X_1, \ldots, X_n)$ un campione casuale di dimensione n dalla popolazione di Poisson di parametro incognito θ. Si consideri il sistema di ipotesi: $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$.

(a) Verificare che, assumendo $\theta_0 > \theta_1$, la generica regione di accettazione del test d^* di Neyman-Pearson risulta essere $\mathcal{Z}_0^* = \{\mathbf{z}_n : \overline{x} > k\}$.

(b) Verificare che, per n sufficientemente elevato, la variabile aleatoria $\overline{X}_n$ ha, approssimativamente, distribuzione $\mathrm{N}(\theta, \theta/n)$ e calcolare $\eta_{d^*}(\theta_1)$ (potenza del test) che si ottiene ponendo $\theta_1 = 5$, $n = 25$ e $k = 5.5$.

Soluzione.

(a) $\ell(\theta) = c \cdot \theta^{y_n} e^{-n\theta}$ e quindi $\lambda_{10}(\mathbf{z}_n) = \left(\frac{\theta_1}{\theta_0}\right)^{y_n} e^{-n(\theta_1-\theta_0)}$ che, per $\theta_0 > \theta_1$ è funzione decrescente di $y_n = \sum_{i=1}^n x_i$. Si rifiuta H_0 se $\lambda_{10}(\mathbf{z}_n) \geq k' \Leftrightarrow \overline{x}_n \leq k$ e quindi si accetta se $\overline{x}_n K k$.

(b) Per il teorema del limite centrale $\overline{X}_n|\theta \;\dot\sim\; \mathrm{N}\left(\theta, \frac{\theta}{n}\right)$, in quanto $\mathbb{E}_\theta[X] = \mathbb{V}_\theta[X] = \theta$.
$\eta_{d^\star}(\theta_1) = 1 - \beta = \mathbb{P}_{\theta_1}[R] = \mathbb{P}_{\theta_1}[\overline{X}_n < k] = \mathbb{P}\left(U < \frac{\sqrt{n}(k-\theta_1)}{\sqrt{\theta_1}}\right) = \Phi\left(\frac{\sqrt{n}(k-\theta_1)}{\sqrt{\theta_1}}\right) = 0.868$, che si ottiene da

```
k=5.5; n=25; th1=5
prob. errore 1° tipo
A=sqrt(n)*(k-th1)/sqrt(th1)
pnorm(A)
```

24. (Esercizio 15 disp.). Sia $\mathbf{Z} = (X_1, \ldots, X_n)$ un campione casuale di dimensione n da una popolazione bernoulliana di parametro incognito θ. Si consideri il sistema di ipotesi: $H_0 : \theta = \theta_0 = 0.5$ vs. $H_1 : \theta = \theta_1 = 0.35$.

(a) Verificare che la regione di rifiuto del test d^* di Neyman-Pearson risulta essere $\mathcal{Z}_1^* = \{z \in \mathcal{Z} : \sum_{i=1}^n x_i < k\}$.

(b) Utilizzando l'approssimazione asintotica per la distribuzione di $\sum_{i=1}^n X_i$, calcolare $R(\theta, d^\star)$ per $n = 144$ e $k = 60$.

Soluzione.

(a) $\ell(\theta) = \theta^{y_n}(1-\theta)^{n-y_n}$, con $y_n = \sum_{i=1}^n x_i$. Pertanto $\lambda_{10}(\mathbf{z}_n) = \left(\frac{\theta_1}{\theta_0}\right)^{y_n}\left(\frac{1-\theta_1}{1-\theta_0}\right)^{n-y_n}$ che, per $\theta_0 < \theta_1$, è una funzione decrescente di y_n e di $\overline{x}_n$. Quindi si rifiuta H_0 se $\lambda_{10}(\mathbf{z}_n) > k' \Leftrightarrow \sum_{i=1}^n x_i < k$.

(b) Per il teorema del limite centrale si ha che $\sum_{i=1}^n X_i|\theta \,\dot{\sim}\, \mathrm{N}\left(\theta, \frac{\theta(1-\theta)}{n}\right)$ e quindi

$R(\theta_0, d^\star) = \alpha = \mathbb{P}_{\theta_0}[R] = \mathbb{P}_{\theta_0}[\overline{X}_n < k/n] = \mathbb{P}\left(U < \frac{\sqrt{n}(k-\theta_0)}{\sqrt{\theta_0(1-\theta_0)}}\right) = \Phi\left(\frac{\sqrt{n}(k-\theta_0)}{\sqrt{\theta_0}}\right) = 0.022$

e

$R(\theta_1, d^\star) = \beta = \mathbb{P}_{\theta_1}[A] = \mathbb{P}_{\theta_1}[\overline{X}_n > k/n] = 1 - \mathbb{P}\left(U < \frac{\sqrt{n}(k-\theta_1)}{\sqrt{\theta_1(1-\theta_1)}}\right) = 1 - \Phi\left(\frac{\sqrt{n}(k-\theta_1)}{\sqrt{\theta_1}}\right) = 0.046.$

Verificare con

```
th0=0.5; th1=0.35; k=60; n=144
A0=sqrt(n)*(k/n-th0)/sqrt(th0*(1-th0))
A1=sqrt(n)*(k/n-th1)/sqrt(th1*(1-th1))
alpha=pnorm(A0)
alpha
beta=1-pnorm(A1)
beta
```

25. (Esercizio 27 disp.). Si consideri un campione casuale $\mathbf{Z}$ di n osservazioni estratto da una popolazione X con funzione di densità: $p_\theta^X(x) = \theta\,(1-x)^{\theta-1},\ x \in (0,1),\ \theta > 0$,

(a) Verificare che lo stimatore di massima verosimiglianza per θ è $d_{mv}(\mathbf{z}_n) = -\frac{n}{\sum_{i=1}^n \log(1-X_i)}$.

(b) Si verifichi che, se $\theta_0 < \theta_1$, il test d^* di Neyman-Pearson per il confronto tra le ipotesi $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ ha regione di accettazione $\mathcal{Z}_0^* = \{\mathbf{z}_n :\ d_{mv}(\mathbf{z}_n) < k.\}$.

(c) Verificare che per n sufficientemente elevato, $d_{mv}(\mathbf{Z}_n) \,\dot{\sim}\, \mathrm{N}(\theta, \frac{\theta^2}{n})$.

(d) Determinare la generica espressione della potenza del test che si ottiene utilizzando l'approssimazione asintotica della distribuzione di $d_{mv}(\mathbf{Z}_n)$.

(e) Calcolare il valore approssimato della potenza del test $\eta_{d^*}(\theta_1)$ ottenuto ponendo $k = 1.464$, $n = 225$, $\theta_0 = 1.3$ e $\theta_1 = 1.6$.

(f) Ottenere l'espressione generica dell'intervallo di confidenza approssimato per θ di livello 0.80 e l'intervallo osservato in corrispondenza di un campione di dimensione $n = 225$ in cui $d_{mv}(\mathbf{z}_n) = 1.5$.

Soluzione.

(a) $\ell(\theta)c \cdot \theta^n t^\theta$, con $t = \prod_{i=1}^n (1 - x_i)$ (si noti che $t < 0$). Quindi $\ell'(\theta) = n \ln \theta + \theta t = \Leftrightarrow \ \theta = -\frac{n}{t}$, che è un punto di massimo per $\ell(\theta)$ poichè $\ell''(\theta) = -\frac{n}{\theta^2} < 0$ per ogni $\theta > 0$.

(b) $\lambda_{10}(\mathbf{z}_n) = \left(\frac{\theta_1}{\theta_0}\right)^n t^{\theta_1 - \theta_0}$ che, per $\theta_0 < \theta_1$, è crescente in t e quindi in $d_{mv} = -\frac{n}{t}$. Si rifiuta quindi H_0 se $d_{mv} > k$ ovvero si accetta H_0 se $d_{mv} < k$.

(c) Per le proprietà degli stimatori di massima verosimiglianza (modelli regolari) sappiamo che $d_{mv}(\mathbf{Z}_n)|\theta \,\dot\sim\, \mathrm{N}\left(\theta, I_n(\theta)^{-1/2}\right)$. In questo caso, per quanto visto al punto precedente, $\ell''(\theta) = -\frac{n}{\theta^2}$ e quindi $I_n(\theta) = \frac{\theta^2}{n}$.

(d) $1 - \beta = R(\theta_1, d_{mv}) = \mathbb{P}_{\theta_1}[R] = \mathbb{P}_{\theta_1}[d_{mv}(\mathbf{Z}_n) > k] \approx 1 - \Phi\left(\frac{\sqrt{n}}{\theta_1}(k - \theta_1)\right)$.

(e) $R(\theta_1, d_{mv}) = 0.898$. Si ottiene con

```
k=1.464; th1=1.6; n=225
A=sqrt(n)*(k-th1)/th1
1-pnorm(A)
```

(f) $\tilde{C} = d_{mv} \ \pm \ u_{1-\frac{\alpha}{2}} \frac{d_{mv}}{\sqrt{n}}$ con $\alpha = 0.2$. Si ottiene $\tilde{C} = [1.371, 1.628]$, con

```
n=225; alpha=0.2; d.mv=1.5
L=d.mv - qnorm(1-alpha/2)*d.mv/sqrt(n)
U=d.mv + qnorm(1-alpha/2)*d.mv/sqrt(n)
c(L,U)
```

26. (Esercizio 33 disp.). Sia $\mathbf{Z} = (X_1, \ldots, X_n)$ un campione casuale proveniente dalla popolazione con funzione di densità di probabilità $p_\theta^X(x) = \theta e^{-\theta x}, \ x > 0, \ \theta > 0$ e con $\mathbb{E}[X] = \frac{1}{\theta}$, $\mathbb{V}[X] = \frac{1}{\theta^2}$. Si consideri il sistema di ipotesi $H_0 : \theta = \theta_0$, vs. $H_1 : \theta = \theta_1$ ($\theta_0 > \theta_1$).

(a) Verificare che la regione di accettazione del test d^* di Neyman-Pearson risulta essere l'insieme $\mathcal{Z}_0^* = \{\mathbf{z}_n : \overline{x}_n \leq k\}, \ k > 0$.

(b) Verificare che, utilizzando la distribuzione asintotica di $\overline{X}_n$, il valore di k per il quale $R(d^*, \theta_0)$ è pari ad α risulta essere $k_\alpha = \frac{1}{\theta_0} + \frac{1}{\theta_0\sqrt{n}} z_{1-\alpha}$.

(c) Determinare il valore di k_α per $\alpha = 0.05$, $\theta_0 = 2$ e $n = 25$.

(d) Determinare la potenza del test considerato, assumendo $\theta_1 = 1$.

(e) Stabilire se, per un campione osservato di dimensione $n = 25$ e con media campionaria pari a 1.5, l'ipotesi nulla viene accettata o rifiutata.

Soluzione.

(a) $\ell(\theta) = \theta^n e^{-y_n\theta}$, con $y_n = \sum_{i=1}^n x_i$. Quindi $\lambda_{10}(\mathbf{z}_n) = \left(\frac{\theta_1}{\theta_0}\right)^n e^{-y_n(\theta_1-\theta_0)}$ che, per $\theta_1 > \theta_0$, è decrescente in y_n. Pertanto $\lambda_{10}(\mathbf{z}_n) \geq 1 \iff \overline{x} \leq k$.

(b) Per il teorema del limite centrale $\overline{X}_n \dot\sim \mathrm{N}\left(\frac{1}{\theta}, \frac{1}{n\theta^2}\right)$. Pertanto $R(\theta_0, d^\star) = \mathbb{P}_{\theta_0}[d^\star(\mathbf{Z}_n) > k] \approx 1 - \Phi\left(\frac{\sqrt{n}(k-1/\theta_0)}{1/\theta_0}\right)$. Si ha quindi che $R(\theta_0, d^\star) = \alpha \Leftrightarrow \frac{\sqrt{n}(k-1/\theta_0)}{1/\theta_0} = z_{1-\alpha} \Leftrightarrow k_\alpha = \frac{1}{\theta_0} + \frac{1}{\theta_0\sqrt{n}} z_{1-\alpha}$.

(c) Si ottiene $k_\alpha = 0.664$ da

```
alpha=0.05; n=25; th0=2
k.alpha=1/th0+qnorm(1-alpha)/(th0*sqrt(n))
k.alpha
```

(d) $R(\theta_1, d^\star) = \mathbb{P}_{\theta_1}[d^\star(\mathbf{Z}_n) > k_\alpha] \approx 1 - \Phi\left(\frac{\sqrt{n}(k_\alpha-1/\theta_1)}{1/\theta_1}\right) = 0.953$. Si ottiene con

```
th1=1
A=sqrt(n)*(k.alpha-1/th1)/(1/th1)
1-pnorm(A)
```

(e) $\overline{x}_n = 1 > 0.664 = k_\alpha$ e quindi H_0 viene rifiutata al livello $\alpha = 0.05$.

27. (Esercizio 31 disp.). Sia $\mathbf{Z} = (X_1, \ldots, X_n)$ un campione casuale proveniente dalla popolazione con funzione di densità di probabilità $p_\theta^X(x) = \theta^2 x e^{-\theta x}$, $x > 0$, $\theta > 0$, con $\mathbb{E}[X] = \frac{2}{\theta}$, $\mathbb{V}[X] = \frac{2}{\theta^2}$. Si consideri il sistema di ipotesi $H_0 : \theta = \theta_0$, vs. $H_1 : \theta = \theta_1$ $(\theta_0 > \theta_1)$.

(a) Verificare che la regione di accettazione del test d^* di Neyman-Pearson risulta essere l'insieme $\mathcal{Z}_0^* = \{\mathbf{z}_n : \overline{x}_n < k\}$, $k > 0$.

(b) Verificare che, utilizzando la distribuzione asintotica di $\overline{X}_n$, il valore di k per il quale $R(d^*, \theta_0)$ (probabilità di errore di I specie) è pari ad α risulta essere $k_\alpha = \frac{2}{\theta_0} + \sqrt{\frac{2}{n\theta_0^2}} z_{1-\alpha}$.

(c) Determinare il valore di k_α per $\alpha = 0.05$, $\theta_0 = 2$ e $n = 25$.

(d) Determinare la potenza del test considerato per $\theta_1 = 1$.

(e) Stabilire se, per un campione osservato di dimensione $n = 25$ e con media campionaria pari a 1, l'ipotesi nulla viene accettata o rifiutata.

Soluzione.

(a) $\ell(\theta) = \theta^{2n} e^{-\theta y_n}$, con $y_n = \sum_{i=1}^n x_i$ e $\lambda_{10}(\mathbf{z}_n) = \left(\frac{\theta_1}{\theta_0}\right)^n e^{-(\theta_1-\theta_0)y_n}$. Il rapporto delle verosimiglianze è quindi una funzione crescente di y_n se $\theta_0 > \theta_1$. Si rifiuta H_0 se $y_n \geq k$ ovvero si accetta H_0 se $\overline{x}_n \leq k$.

(b) Per il teorema del limite centrale $\overline{X}_n \dot{\sim} \mathrm{N}\left(\frac{2}{\theta}, \frac{2}{n\theta^2}\right)$. Pertanto $R(\theta_0, d^\star) = \mathbb{P}_{\theta_0}[\overline{X}_n > k] \approx 1 - \Phi\left(\frac{\sqrt{n}(k-2/\theta_0)}{\sqrt{2}/\theta_0}\right)$. Si ha quindi che $R(\theta_0, d^\star) = \alpha \Leftrightarrow \frac{\sqrt{n}(k-2/\theta_0)}{\sqrt{2}/\theta_0} = z_{1-\alpha} \Leftrightarrow k_\alpha = \frac{2}{\theta_0} + \frac{\sqrt{2}}{\theta_0\sqrt{n}} z_{1-\alpha}$.

(c) Si ottiene $k_\alpha = 1.23$ con

```
# k_alpha
alpha=0.05
n=25
th0=2
k.alpha=2/th0+qnorm(1-alpha)*sqrt(2)/(th0*sqrt(n))
k.alpha
```

(d) $R(\theta_1, d^\star) = \mathbb{P}_{\theta_1}[\overline{X}_n > k_\alpha] \approx 1 - \Phi\left(\frac{\sqrt{n}(k_\alpha-2/\theta_1)}{\sqrt{2}/\theta_1}\right) = 0.996$. Si ottiene con

```
th1=1
A=sqrt(n)*(k.alpha-2/th1)/(sqrt(2)/th1)
1-pnorm(A)
```

(e) $\overline{x}_n = 1 < 1.23 = k_\alpha$ e quindi si accetta H_0 a livello $\alpha = 0.05$.

28. (Esempi notevoli). Per il sistema di ipotesi $H_0 : \theta \geq \theta_0$, vs. $H_1 : \theta < \theta_0$ e per i modelli indicati di seguito, implementare il test d^* di Karlin-Rubin, ovvero:

   i) verificare la monotonia del rapporto delle verosimiglianze rispetto alla statistica sufficiente T;
   ii) determinare la distribuzione di $T(\mathbf{X}_n)|\theta = \theta_0$;
   iii) determinare la regola di rifiuto di H_0 generica e quella per il test di ampiezza α (supponendo che, per i modelli di v.a. discrete, esista esattamente il quantile opportuno).

   Modelli:

   (a) $\mathrm{Bern}(\theta)$;
   (b) $\mathrm{Pois}(\theta)$;
   (c) $\mathrm{N}(\theta, \sigma^2)$, varianza nota;
   (d) $\mathrm{N}(\mu_0; \theta)$, valore atteso noto;
   (e) $\mathrm{Exp}(\theta)$;
   (f) $\mathrm{EN}(\theta)$.

Soluzione.

Quanto richiesto è riportato nella seguente tabella.

Modello	$T(\mathbf{Z}_n)$	$T(\mathbf{Z}_n)\|\theta_0 \sim \dots$
$\text{Ber}(\theta)$	$\sum_{i=1}^n X_i$	$\text{Ber}(n, \theta_0)$
$\text{Pois}(\theta)$	$\sum_{i=1}^n X_i$	$\text{Pois}\,(n\theta_0)$
N1; $\text{N}(\theta, \sigma^2)$	$\overline{X}_n$	$\text{N}\left(\theta, \frac{\sigma^2}{n}\right)$ (esatta)
N3; $\text{N}(\mu_0, \theta)$	S_0^2	$\text{Ga}\left(\frac{n}{2}, \texttt{rate} = \frac{n}{2\theta^2}\right)$
$\text{Esp}(\theta)$	$\sum_{i=1}^n X_i$	$\text{Ga}\,(n, \texttt{rate} = \theta)$
$\text{EN}(\theta)$	$\frac{1}{\sum_{i=1}^n X_i}$	$\text{InvGa}\,(n, \texttt{rate} = \theta)$

In tutti i casi si rifiuta H_0 se $T(\mathbf{z}_n) < k$); il test di ampiezza α si ottiene ponendo $k = q_\alpha(\theta_0)$, dove $q_\alpha(\theta_0)$ è il quantile di livello α della distribuzione di $T(\mathbf{Z}_n)|\theta = \theta_0$.

29. (Esempi notevoli). Ripetere il precedente esercizio per il sistema di ipotesi $H_0 : \theta \leq \theta_0, \text{ vs. } H_1 : \theta > \theta_0$

Soluzione.

Vale la tabella dell'esercizio precedente. Cambia solo la regola di rifiuto e i quantili da usare. In tutti i casi si ha $\mathcal{Z}_1(k) = (\mathbf{z}_n \in \mathcal{Z} : T(\mathbf{z}_n) > k)$; il test di ampiezza α si ottiene ponendo $k = q_{1-\alpha}(\theta_0)$, dove $q_{1-\alpha}(\theta_0)$ è il quantile di livello α della distribuzione di $T(\mathbf{Z}_n)|\theta = \theta_0$.

30. (Esempi notevoli). Per i modelli dell'Esercizio 30 [escluso (c)] e per i tre tipi di ipotesi principali, determinare la regola ri rifiuto del test di ampiezza α utilizzando il test di Wald basato sulla statistica $d_w = \frac{\sqrt{n}(d_{mv} - \theta_0)}{I_n(\theta_0)^{-1/2}}$.

Soluzione.

(a)

Modello	d_{mv}	$I_n^{-1}(\theta)$	$d_{mv}(\mathbf{Z}_n)\|\theta \dot\sim \dots$	$d_w(\mathbf{z}_n)$
$\text{Ber}(\theta)$	$\overline{X}_n$	$\frac{\theta(1-\theta)}{n}$	$\text{N}\left(\theta, \frac{\theta(1-\theta)}{n}\right)$	$\frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sqrt{\theta_0(1-\theta_0)}}$
$\text{Pois}(\theta)$	$\overline{X}_n$	$\frac{\theta}{n}$	$\text{N}\left(\theta, \frac{\theta}{n}\right)$	$\frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sqrt{\theta_0}}$
N1; $\text{N}(\theta, \sigma^2)$	$\overline{X}_n$	$\frac{\sigma^2}{n}$	$\text{N}\left(\theta, \frac{\sigma^2}{n}\right)$ (esatta)	$\frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sigma}$
N3; $\text{N}(\mu_0, \theta)$	S_0^2	$\frac{2\theta^2}{n}$	$\text{N}\left(\theta, \frac{2\theta^2}{n}\right)$	$\frac{\sqrt{2n}(S_0^2 - \theta_0)}{\theta_0}$
$\text{Esp}(\theta)$	$\overline{X}_n$	$\frac{\theta^2}{n}$	$\text{N}\left(\theta, \frac{\theta^2}{n}\right)$	$\frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\theta_0}$
$\text{EN}(\theta)$	$\frac{1}{\overline{X}_n}$	$\frac{\theta^2}{n}$	$\text{N}\left(\theta, \frac{\theta^2}{n}\right)$	$\frac{\sqrt{n}\left(\frac{1}{\overline{x}_n} - \theta_0\right)}{\theta_0}$

- Per i test di tipo $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$ si rifiuta H_0 se $d_w(\mathbf{z}_n) > u_{1-\alpha}$
- Per i test di tipo $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$ si rifiuta H_0 se $d_w(\mathbf{z}_n) < u_\alpha$
- Per i test di tipo $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ si rifiuta H_0 se $|d_w(\mathbf{z}_n)| > u_{1-\frac{\alpha}{2}}$

31. (Esercizio 36 disp.). Sia X una v.a. assolutamente continua con funzione di densità $p_\theta(x) = \theta\ x^{\theta-1},\ x \in (0,1),\ \theta > 0$.

   (a) Si consideri il sistema di ipotesi $H_0 :\ \theta \leq 1$ vs. $H_1 :\ \theta > 1$. ed il test $d(x)$ con regione di rifiuto basata su una singola osservazione ($n = 1$) e definita da: $Z_1 = \left\{x \in (0,1) : x > \frac{1}{2}\right\}$. Determinare la funzione di potenza del test e calcolarne il valore in $\theta = 2$.

   (b) Determinare la funzione di rischio $R(\theta, d)$ e tracciarne il grafico (in funzione di $\theta \in \Omega$).

Soluzione.

(a) Se $n = 1$ allora $\lambda_{21}(x) = \frac{\theta_2}{\theta_2} x^{\theta_2 - \theta_1}$ che, per $\theta_2 > \theta_1$ è crescente in x. Si rifiuta H_0 se $\lambda_{21} > 1 \Leftrightarrow x > k$. La funzione di potenza, per $k = \frac{1}{2}$ è $\eta(\theta) = \mathbb{P}_\theta(R) = \mathbb{P}_\theta\left(X > \frac{1}{2}\right) = \int_{\frac{1}{2}}^{1} \theta x^{\theta-1} dx = x^\theta|_{\frac{1}{2}}^{1} = 1 - \frac{1}{2^\theta}$, e quindi $\eta(2) = \frac{3}{4}$.

(b) $R(\theta, d) = \begin{cases} \alpha(\theta) = 1 - \frac{1}{2^\theta} & \theta \leq 1 \\ \beta(\theta) = \frac{1}{2^\theta} & \theta > 1 \end{cases}$ Per il grafico procedere analiticamente e confrontare con

```
rischio.fun=function(theta){
  (1-1/(2^theta))*(theta<=1)+(1/(2^theta))*(theta>1)
}
curve(rischio.fun(x),from=0,to=3,xlab=expression(theta),ylab="rischio")
```

32. (da Esercizio 28 disp. modificato).

33. Si consideri un campione casuale di n osservazioni estratte da una popolazione X con distribuzione di densità di probabilità: $p_\theta(x) = (\theta + 2)\, x^{\theta+1},\ 0 \leq x \leq 1,\ \theta > -2$.

    (a) Determinare $d_{mv}(\mathbf{Z}_n)$, stimatore di massima verosimiglianza di θ.

    (b) Verificare che la distribuzione campionaria asintotica di $d_{mv}(\mathbf{Z}_n)$ è $\mathrm{N}\left(\theta, \dfrac{(\theta+2)^2}{n}\right)$.

    (c) Per il confronto tra le ipotesi $H_0 : \theta = \theta_0$ vs. $H_1 : \theta < \theta_0$, scrivere la regola di rifiuto di H_0 basata sulla statistica test di Wald $d_w = \frac{d_{mv} - \theta_0}{\sqrt{I_n(\theta_0)^{-1}}}$.

    (d) Dire se si tratta o no di un test di Karlin-Rubin.

    (e) Sfruttando la normalità asintotica di $d_{MV}(\mathbf{Z}_n)$ determinare k_α in modo tale che il test sia di ampiezza α.

Soluzione.

(a) $\ell(\theta) = (\theta+2)^n t^\theta$, con $t = \prod_{i=1}^n x_i$. Si verifica agevolmente che $d_{mv}(\mathbf{z}_n) = -2 - \frac{n}{t}$ (si noti che $t < 0$ e quindi $d_{mv}(\mathbf{z}_n) > -2$).
Inoltre Si ha che $\frac{d^2}{d\theta^2} \ln \ell(\theta) = -\frac{n}{(\theta+2)^2} < 0,\ \forall\theta$.

(b) Poichè $\frac{d^2}{d\theta^2} \ln \ell(\theta) = -\frac{n}{(\theta+2)^2}$, da cui si evince che $I_n(\theta) = \frac{(\theta+2)^2}{n}$. Il modello è regolare e quindi vale l'approssimazione normale proposta.

(c) Si rifiuta H_0 a livello α se $d_w = \frac{\sqrt{n}(d_{mv} - \theta_0)}{(\theta_0+2)} < k$.

(d) Si tratta di un test di Karlin-Rubin in quanto costruito sullo stimatore di massima verosimiglianza, rispetto al quale il modello ha rapporto delle verosimiglianze monotono.

(e) Si rifiuta H_0 a livello α se $d_w = \frac{\sqrt{n}(d_{mv} - \theta_0)}{(\theta_0+2)} < u_\alpha$ ovvero se $d_{mv} < \frac{\theta_0+2}{\sqrt{n}} u_\alpha + \theta_0 = k_\alpha$, con u_α quantile di livello α della v.a. normale standardizzata.

34. Si consideri un campione casuale $\mathbf{Z}_n = (X_1, \ldots, X_n)$ generato da una distribuzione con funzione di densità $p_\theta(x) = \frac{3}{8} \frac{x^2}{\theta^3} I_{[0,2\theta]}(x) \qquad \theta > 0$.

    (a) Verificare che $\mathbb{E}[X] = \frac{3}{2}\theta$ e $\mathbb{V}[X] = \frac{3}{20}\theta^2$.

    (b) Determinare, $d_m(\mathbf{Z}_n)$, lo stimatore dei momenti di θ e verificare che $d_m(\mathbf{Z}_n) \dot\sim \mathrm{N}\left(\theta, \frac{\theta^2}{15n}\right)$.

    (c) Basandosi sulla distribuzione asintotica di $d_m(\mathbf{Z}_n)$, determinare un intervallo di confidenza asintotico per θ di livello $1 - \alpha$ e determinarne la lunghezza $\mathcal{L}(\mathbf{Z}_n)$.

(d) Fissati due valori $\epsilon > 0$ e $\theta_D > 0$, trovare l'espressione di $e_n(\theta) = \mathbb{E}_\theta[\mathcal{L}(\mathbf{Z}_n)]$ e determinare n^*, il minimo valore di n affinché $e_n(\theta_D) < \epsilon$.

(e) Per il seguente sistema di ipotesi su θ $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$, si consideri il test con regione di rifiuto $Z_1 = \{z \in \mathcal{Z} : d_m(\mathbf{z}_n) > k\}$. Determinare il valore della probabilità di errore di I tipo assumendo $\theta_0 = 3$, $n = 15$ e $k = 3.1$ (usare approssimazione normale).

(f) In corrispondenza di un campione osservato in cui $\sum_{i=1}^n x_i = 65$ e $n = 15$, verificare se il test del punto precedente porta alla accettazione o al rifiuto di H_0, usando $k = 3.1$.

(g) Si consideri ora $theta_0 = 2.8$. Ripetere il test ma assumendo $\alpha = 0.2$.

(h) Determinare, sempre per $\alpha = 0.2$ e per il campione osservato in esame, i valori numerici degli estremi dell'intervallo di confidenza asintotico di θ del punto (3).

Soluzione.

(a) $\mathbb{E}_\theta[X] = \int_0^{2\theta} x \cdot \frac{3}{8}\frac{x^2}{\theta^3}\,dx = \frac{3}{8\theta^3}\frac{x^4}{4}\Big|_0^{2\theta} = \frac{3}{2}\theta$.
Per il calcolo della varianza, determinare prima il momento secondo di X e poi usare la relazione $\mathbb{V}_\theta[X] = \mathbb{E}_\theta[X^2] - \mathbb{E}_\theta[X]^2$.

(b) Imponendo la relazione $\mathbb{E}_\theta[X] = \overline{X}_n$ si ottiene che $d_m(\mathbf{Z}_n) = \frac{2}{3}\overline{X}$. Inoltre $\mathbb{V}_\theta\left(\frac{2}{3}\overline{X}_n\right) = \frac{4}{9}\cdot\frac{3\theta^2}{20n} = \frac{\theta^2}{15n}$.

(c) $\tilde{C} = d_m \ \pm\ u_{1-\frac{\alpha}{2}}\sqrt{\widehat{\mathbb{V}_\theta}[d_m]} = \frac{2}{3}\ \pm\ u_{1-\frac{\alpha}{2}}\frac{\frac{2}{3}\overline{X}_n}{\sqrt{15n}}$

(d) $\mathcal{L}_C(\mathbf{Z}_n) = \frac{4}{3\sqrt{15n}}u_{1-\frac{\alpha}{2}}\overline{X}_n$.
$e_n = \mathbb{E}_\theta[\mathcal{L}_C(\mathbf{Z}_n)] = \frac{4u_{1-\frac{\alpha}{2}}}{3\sqrt{15n}}\mathbb{E}\theta[\overline{X}_n] = \frac{4}{3\sqrt{15n}}u_{1-\frac{\alpha}{2}}\cdot\frac{3}{2}\theta = \frac{2}{\sqrt{15n}}u_{1-\frac{\alpha}{2}}\theta$.
$e_n(\theta_D) < \epsilon \ \Leftrightarrow\ n > \frac{4u^2_{1-\frac{\alpha}{2}}\theta_D^2}{15\epsilon^2}$. Pertanto $n^\star = \left\lceil\frac{4u^2_{1-\frac{\alpha}{2}}\theta_D^2}{15\epsilon^2}\right\rceil$.

(e) $\alpha = \mathbb{P}_{\theta_0}(d_m > k) = 1 - \Phi\left[\frac{\sqrt{15n}(k-\theta_0)}{\theta_0}\right] = 0.309$. Calcolare come segue:

```
th0=3; k=3.1; n=15
# prob errore 1 tipo
A=sqrt(15*n)*(k-th0)/th0
1-pnorm(A)
```

(f) $d_m = 2.88 < 3.1 = k$. Si accetta quindi H_0.

```
xmed=65/15
d.m=(2/3)*xmed
d.m
# 2.88
d.m>k
#  FALSE
```

(g) Struttando l'approssimazione asintotica possiamo usare la seguente regola di rifiuto di H_0: $\tilde{w}_{oss} = \sqrt{15n}\frac{d_m-\theta_0}{\theta_0} > u_{1-\alpha}$. Se $\alpha = 0.2$, $u_{1-\alpha} = u_{0.8} = 0.842$, mentre ù $\tilde{w}_{oss} = 0.476$. L'ipotesi nulla viene quindi accettata.

```
alpha=0.2
u=qnorm(1-alpha)
u
w.oss=sqrt(15*n)*(d.m-th0)/th0
w.oss
w.oss>u
```

(h) Si ottiene $C = [2.642, 3.135]$.

```
alpha=0.2
u=qnorm(1-alpha/2)
L=d.m-u*d.m/sqrt(15*n)
U=d.m+u*d.m/sqrt(15*n)
c(L,U)
```

35. Sia $Z = (X_1, ..., X_n)$ un campione casuale in cui $X_i|\theta \sim \text{Pois}(\theta)$.

(a) Si determini la distribuzione asintotica dello stimatore $\overline{X}_n$.

(b) Costruire un intervallo di confidenza asintotico per il parametro $\psi = e^{-\theta}$ usando il metodo Delta.

(c) Calcolare l'intervallo di confidenza asintotico con livello di confidenza 0.95 in corrispondenza di un campione casuale osservato di $n = 100$ osservazioni in cui la media campionaria è risultata pari a 0.2.

(d) In base al risultato ottenuto al punto precedente spiega quali conclusioni trarresti per verificare il seguente sistema di ipotesi $H_0 : \psi = 0.8$ vs. $H_1 : \psi \neq 0.8$ con un test d di dimensione approssimativamente uguale a 0.05.

(e) Per verificare il seguente sistema di ipotesi composte $H_0 :\ \theta \leq 1.8$ vs. $H_1 : \theta > 1.8$ riguardanti il parametro θ viene utilizzata la seguente regione di rifiuto $\mathcal{Z}_1 = \{z \in \mathcal{Z} : \sum_{i=1}^n x_i > k\}$. Determinare la funzione di potenza del test (in funzione di $\Phi(\cdot)$, funzione di ripartizione della v.a. normale standardizzata).

(f) Calcolare, in base al punto precedente, l ampiezza del test d, $\eta(\theta_0)$, che si ottiene se $k = 160$.

Soluzione.

(a) $\overline{X}_n|\theta \dot\sim \mathrm{N}\left(\theta, \frac{\theta}{n}\right)$.

(b) Dal metodo delta di ottiene che $\widehat{\psi} = e^{-\overline{X}_n}|\theta \dot\sim \mathrm{N}\left(e^{-\theta}, \frac{e^{-2\theta}\theta}{n}\right)$. Pertanto

$$\tilde{C} = e^{-\overline{X}_n} \pm 2u_{1-\frac{\alpha}{2}}\sqrt{\frac{e^{-2\overline{X}_n}\overline{X}_n}{n}}.$$

(c)
```
n=100
alpha=0.05
u=qnorm(1-alpha/2)
u
xmed=0.2
L=exp(-xmed)-u*sqrt(exp(-2*xmed)*xmed/n)
U=exp(-xmed)+u*sqrt(exp(-2*xmed)*xmed/n)
c(L,U)
```
Si ottiene $\tilde{C} = [0.746, 0.890]$.

(d) Poichè l'intevallo di confidenza di livello $1 - \alpha = 0.95$ contiene $\psi_0 = 0.8$, si accetta l'ipotesi nulla del test considerato di ampiezza $\alpha = 0.05$.

(e) $\sum_{i=1}^n X_i|\theta \dot\sim \mathrm{N}(n\theta, n\theta)$. Pertanto (usando l'approssimazione normale)

$$\eta(\theta) = \mathbb{P}_\theta[\textstyle\sum_{i=1}^n X_i > k] \approx 1 - \Phi\left(\frac{k-n\theta}{\sqrt{n\theta}}.\right)$$

(f)
```
n=100
th0=1.8
k=160
A=(k-n*th0)/sqrt(n*th0)
potenza=1-pnorm(A)
potenza
```
Si ottiene $\eta(\theta_0) = 0.931$.

C – AfE-AfN: miscellanea (da tracce esami 2016-2018)

36. Sia $X_1, \ldots, X_n|\theta \sim \text{Ber}(\theta)$ i.i.d..
    (a) Determinare (o fornire l'espressione di) $d_{mv}(\mathbf{Z}_n)$.
    (b) Determinare o fornire l'approssimazione asintotica della distribuzione campionaria di $d_{mv}(\mathbf{Z}_n)$.
    (c) Determinare l'intervallo di confidenza $C(\mathbf{Z}_n)$ di livello $1-\alpha$ per θ utilizzando la quantità pivotale asintotica ottenuta da $d_{mv}(\mathbf{Z}_n)$.
    (d) Determinare l'espressione della perdita $\mathbb{L}(\theta, C) = b[\mathcal{L}_\mathcal{C}(\mathbf{Z}_n)]^2$, $b > 0$, dove $\mathcal{L}_\mathcal{C}(\mathbf{Z}_n)$ indica la lunghezza aleatoria dell'intervallo $C(\mathbf{Z}_n)$.
    (e) Determinare la funzione di rischio normale $R(\theta, C)$.

Soluzione.

(a) $d_{mv} = \overline{X}_n$.

(b) $d_{mv}(\mathbf{Z}_n)|\theta \,\dot{\sim}\, \text{N}\left(\theta, \frac{\theta(1-\theta)}{n}\right)$.

(c) $\widetilde{Q}(\mathbf{Z}_n, \theta) = \frac{\sqrt{n}(\overline{X}_n - \theta)}{\sqrt{\theta(1-\theta)}}$; Pertanto $\tilde{C} = \overline{X} \pm u_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}$.

(d) $\mathbb{L}(\theta, C) = b\mathcal{L}_C(\mathbf{Z}_n)^2 = \frac{4}{n}u^2_{1-\frac{\alpha}{2}}\overline{X}_n(1-\overline{X}_n)$.

(e) $R(\theta, C) = \mathbb{E}_\theta[\mathbb{L}(\theta, C)] = b\mathcal{L}_C(\mathbf{Z}_n)^2 = \frac{4}{n}u^2_{1-\frac{\alpha}{2}}\mathbb{E}_\theta[\overline{X}_n(1-\overline{X}_n)]$. Il risultato si ottiene osservando che $\mathbb{E}_\theta[\overline{X}_n(1-\overline{X}_n)] = \mathbb{E}_\theta[\overline{X}_n - \overline{X}^2_n] = \mathbb{E}_\theta[\overline{X}_n] - \mathbb{E}_\theta[\overline{X}^2_n] = \mathbb{E}_\theta[\overline{X}_n] - \mathbb{V}_\theta[\overline{X}_n] - \mathbb{E}_\theta[\overline{X}_n]^2 = \theta - \frac{\theta(1-\theta)}{n} - \theta^2 = \frac{(n-1)\theta(1-\theta)}{n}$.

37. Sia $X_1, \ldots, X_n|\theta \sim \text{Ber}(\theta)$ i.i.d. e $\pi(\theta) = \text{Beta}(\theta|\alpha, \beta)$, $\alpha, \beta > 0$ e considerare le due funzioni di decisione:

$$d_1(\mathbf{z}_n) = d_{mv}(\mathbf{z}_n) \qquad \text{e} \qquad d_2(\mathbf{z}_n) = a^*(\mathbf{z}_n),$$

dove $a^*(\mathbf{Z}_n)$ indica la decisione ottima rispetto alla funzione di perdita quadratica nell'analisi in forma estensiva.

(a) Determinare $\rho(d_1, z)$ e $\rho(d_2, z)$ (rispetto alla perdita quadratica).
(b) Determinare la funzione di rischio normale $R(\theta, d_1)$.
(c) Verificare che la funzione di rischio normale di d_2 è

$$R(\theta, d_2) = \frac{n\theta(1-\theta)}{(\alpha+\beta+n)^2} + \left(\frac{n\theta+\alpha}{\alpha+\beta+n} - \theta\right)^2.$$

(d) Verificare che con $\alpha = \beta = \sqrt{n/4}$ si ottiene $d_2(\mathbf{Z}_n) = \frac{\sum_{i=1}^n X_i + \sqrt{n/4}}{n+\sqrt{n}}$ e $R(\theta, d_2) = \frac{n}{4(n+\sqrt{n})^2}$.
(e) Tracciare i grafici delle funzioni $R(\theta, d_1)$ e $R(\theta, d_2)$ al variare di $\theta \in (0,1)$, per $n = 10$.

Soluzione.

(a) Per il modello considerato $\Theta|\mathbf{z}_n \sim \text{Ga}(\overline{\alpha}, \texttt{rate} = \overline{\beta})$, dove $\overline{\alpha} = \alpha + s_n$ e $\overline{\beta} = \beta + n - s_n$, con $s_n = \sum_{i=1}^n x_i$. Si ha quindi che $d_2 = \mathbb{E}[\Theta|\mathbf{z}_n] = \frac{\overline{\alpha}}{\overline{\alpha}+\overline{\beta}} = \frac{\alpha+s_n}{\alpha+\beta+n}$. Inoltre, per risoltati noti, $d_{mv} = \overline{X}_n$.

Con perdita quadratica abbiamo quindi che, essendo $d_2 = a^\star$, $\rho(d_2, \mathbf{z}_n) = \mathbb{V}[\Theta|\mathbf{z}_n] = \frac{\overline{\alpha}\overline{\beta}}{(\overline{\alpha}+\overline{\beta}+1)(\overline{\alpha}+\overline{\beta})^2}$ e $\rho(d_1, \mathbf{z}_n) = \mathbb{V}[\Theta|\mathbf{z}_n] + (d_1 - d_2)^2$.

(b) $R(\theta, d_1) = \mathbb{V}_\theta[d_1] = \frac{\theta(1-\theta)}{n}$.

(c) Con funzione di perdita quadratica si ha che

$$R(\theta, d_2) = \mathbb{V}_\theta[d_2] + (\mathbb{E}_\theta[d_2] - \theta)^2 = \frac{n\theta(1-\theta)}{(\alpha+\beta+n)^2} + \left(\frac{\alpha+n\theta}{\alpha+\beta+n} - \theta\right)^2.$$

(d) Si verifica osservando che $\alpha + \beta = n + \sqrt{n}$ e svolgendo i semplici passaggi algebrici.

(e)
```
n=10
R.d2=n/(4*(n+sqrt(n))^2)
R.d1=function(theta){
  theta*(1-theta)/n
}
curve(R.d1(x),from=0, to=1, xlab=expression(theta), ylab="")
abline(h=R.d2,lty=2)
```

38. Sia $X_1, \ldots, X_n|\theta \sim \text{Exp}(\theta)$ i.i.d. (ovvero, $p_\theta(x) = \frac{1}{\theta}e^{-x/\theta},\ x > 0,\ \theta > 0$).

(a) Determinare (o fornire direttamente) la stima di massima verosimiglianza $d_{mv}(\mathbf{z}_n)$ per θ.

(b) Considerare il sistema di ipotesi $H_0 : \theta = \theta_0$, vs. $H_1 : \theta > \theta_0$. Determinare la generica regione di rifiuto $\mathcal{Z}_1^*(k),\ k > 0$, del test d_{kr} di Karlin-Rubin in funzione di $d_{mv}(\mathbf{z}_n)$.

(c) Determinare k_α in modo tale che il test abbia ampiezza α utilizzando l'approssimazione normale per la distribuzione di $d_{mv}(\mathbf{Z}_n)$.

(d) Utilizzare l'approssimazione normale per la distribuzione di $d_{mv}(\mathbf{Z}_n)$ e determinare la funzione $\tilde{R}(d_{kr}, \theta)$ per il test di ampiezza α.

(e) Tracciarne il grafico di $\tilde{R}(d_{kr}, \theta)$ in funzione di θ, per $n = 10$ e $\theta_0 = 1$.

Soluzione.

(a) Da risultati noti, $d_{mv} = \overline{X}_n$.

(b) Il rapporto delle verosimigliante è crescente in $d_{mv} = \overline{X}_n$, quindi $\mathcal{Z}_1^\star(k) = \{\mathbf{z}_n \in \mathcal{Z} : \overline{X}_n > k\}, \quad k > 0$.

(c) $d_{mv} = \overline{X}_n \dot\sim \mathrm{N}\left(\theta, \frac{\theta}{n}\right)$, e quindi dobbiamo determinare k_α tale che $\mathbb{P}_{\theta_0}[R] = \alpha$, ovvero tale che $\alpha = \mathbb{P}_{\theta_0}(\overline{X}_n > k) = \mathbb{P}\left(U > \frac{\sqrt{n}(k-\theta_0)}{\sqrt{\theta_0}}\right)$. Questo si ha se e solo se $1 - \Phi\left(\frac{\sqrt{n}(k-\theta_0)}{\sqrt{\theta_0}}\right) = \alpha$ ovvero se $\frac{\sqrt{n}(k-\theta_0)}{\sqrt{\theta_0}} = u_{1-\alpha}$ e quindi $k_\alpha = \sqrt{\frac{\theta_0}{n}} u_{1-\alpha} + \theta_0$.

(d) $R(\theta, d_{kr}) = \begin{cases} \eta(\theta) & \theta = \theta_0 \\ 1 - \eta(\theta) & \theta > \theta_0 \end{cases} = \begin{cases} \alpha & \theta = \theta_0 \\ 1 - \eta(\theta) & \theta > \theta_0 \end{cases}$

La funzione di potenza del test is esame è:

$\eta(\theta) = \mathbb{P}_\theta[R] = \mathbb{P}_\theta[\overline{X}_n > k_\alpha] = 1 - \Phi(\frac{\sqrt{n}(k_\alpha-\theta)}{\sqrt{\theta}})$. Sostituire k_α per avere l'espressione finale.

(e)
```
th0=1; n=10; alpha=0.05
u=qnorm(1-alpha)
k=sqrt(th0/n)*u+th0
rischio.fun=function(theta){
  1-pnorm((sqrt(n)*(k-theta))/theta)
}
curve(rischio.fun(x),from=1,to=5, xlab=expression(theta), ylab="potenza")
```

39. Sia $Z = (X_1, \ldots, X_n)$ un campione casuale di dimensione n proveniente da una popolazione $\mathrm{N}(0, \theta)$ con parametro incognito θ e funzione di densità delle X_i

$$p_\theta(x) = \frac{1}{\sqrt{2\pi\theta}} \exp\{-\frac{1}{2\theta}x^2\}, \qquad x \in \mathbb{R}, \qquad \theta > 0.$$

(a) Verificare che lo stimatore di massima verosimiglianza è $d_{mv}(\mathbf{Z}_n) = \frac{1}{n}\sum_{i=1}^n X_i^2$

(b) Verificare che, $\forall \theta > 0$, si ha

$$\frac{n d_{mv}(\mathbf{Z}_n)}{\theta} \sim \chi^2_n.$$

(c) Considerare la funzione di decisione $d_1(\mathbf{Z}_n) = \frac{n}{n+2} d_{mv}(\mathbf{Z}_n)$ e confrontarla con $d_1(\mathbf{Z}_n)$ in termini di funzione di rischio normale $R(\theta, \cdot)$ (perdita quadratica).

(d) Quali conclusioni si possono ottenere dal confronto effettuato, in termini di analisi pre-ottimale condotta con la funzione di rischio?

(e) Si consideri il sistema di ipotesi: $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$, (assumere $\theta_0 < \theta_1$). Verificare che la regione di accettazione del test basato sul rapporto delle verosimiglianze risulta essere

$$\mathcal{Z}_0 = \{z \in \mathcal{Z} : \ d_{mv}(\mathbf{z}_n) \leq k\}. \tag{1}$$

(f) Verificare che il valore di k in (1) in corrispondenza del quale la probabilità di errore di I specie è pari ad α risulta essere

$$k = \frac{\theta_0}{n}\chi^2_{n;1-\alpha},$$

dove $\chi^2_{n;1-\alpha}$ è il percentile di livello $1 - \alpha$ della v.a. chi quadrato con n gradi di libertà.

(g) Determinare il valore di k che si ottiene ponendo $\alpha = 0.05$, $n = 3$ e $\theta_0 = 1$. Supponendo di avere osservato il campione $z = (1, 1/4, 2)$, stabilire se l'ipotesi nulla viene o meno accettata.
[**N.B.** Attenzione alla differenza tra **valori critici** e **percentili** della v.a. χ^2.]

Soluzione.

(a) Risultato noto, che comunque si verifica facilmente massimizzando la funzione di verosimiglianza di θ, che qui risulta essere $\ell(\theta) \propto \left(\frac{1}{\theta}\right)^{\frac{n}{2}} e^{-\frac{nS_0^2}{2\theta}}$, con $S_0^2 = \frac{1}{n}\sum_{i=1}^n X_i^2$.

(b) $X_i|\theta \sim \mathrm{N}(0,\theta) \Rightarrow U_i = \frac{X_i}{\sqrt{\theta}} \sim \mathrm{N}(0,1) \Rightarrow U_i^2 \sim \chi_1^2 \Rightarrow \sum_{i=1}^n U_i^2 = \frac{\sum_{i=1}^n X_i^2}{\theta} \sim \chi_n^2$.

(c) $d_1(\mathbf{Z}_n) = \frac{n}{n+2} d_{mv}(\mathbf{Z}_n) = \frac{\sum_{i=1}^n X_i^2}{n+2}$. Poichè $\mathbb{E}_\theta[d_{mv}] = \theta$ e $\mathbb{V}_\theta[d_{mv}] = \frac{2\theta^2}{n}$, si ha che $\mathbb{E}_\theta[d_1] = \frac{n\theta}{n+2}$ e $\mathbb{V}_\theta[d_1] = \frac{2n\theta^2}{(n+2)^2}$. Quindi, usando la funzione di perdita quadratica,

$R(\theta, d_{mv}) = \mathbb{V}_\theta[d_{mv}] = \frac{2\theta^2}{n}$ e $R(\theta, d_2) = \mathbb{V}_\theta[d_1] + B_\theta^2[d_1] = \frac{2n\theta^2}{(n+2)^2} + \left(\frac{n\theta}{n+2} - \theta\right)^2 = \frac{2\theta^2}{n+2}$.

Pertanto $R(\theta, d_1) < R(\theta, d_{mv})$, $\forall \theta > 0$, $\forall n \in \mathbb{N}$.

(d) d_{mv} è inammissibile.

(e) In generale i modelli hanno rapporto di verosimiglianza monotono in d_{mv}. Questo succede sicuramente se il modello, come in questo caso, è famiglia esponenziale. Se lo si vuole/deve verificare, si osservi che, presi (θ_0, θ_1) con $\theta_1 > \theta_0$, il rapporto delle verosimiglianze

$$\lambda_{10}(\mathbf{z}_n) = \left(\frac{\theta_0}{\theta_1}\right)^{\frac{n}{2}} e^{-\frac{nS_0^2}{2}\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)}$$

è funzione crescente di $d_{mv} = S_0^2$. Pertanto si accetta H_0 se $\lambda_{10}(\mathbf{z}_n) \le k' \Leftrightarrow d_{mv} \le k$.

(f) $\alpha = \mathbb{P}_{\theta_0}[S_0^2 > k] = \mathbb{P}_{\theta_0}\left(\frac{nS_0^2}{\theta_0} > \frac{nk}{\theta_0}\right) = 1 - \mathbb{F}(\frac{nk}{\theta})$, con $\mathbb{F}(\cdot)$ funzione di ripartizione della v.a. χ_n^2. Pertanto $k = \frac{\theta_0}{n}\chi^2_{n;1-\alpha}$.

(g) FIN QUI

40. Siano $X_1, ..., X_n|\theta \sim \mathrm{Geom}(\theta)$ i.i.d., con valore atteso $\mathbb{E}_\theta[X] = \frac{1}{\theta}$, $\mathbb{V}_\theta[X] = \frac{1-\theta}{\theta^2}$ e funzione di massa di probabilità

$$p_\theta(x) = \theta(1-\theta)^{x-1}, \qquad x = 1, 2, 3, \dots \qquad \theta \in [0,1].$$

Si consideri il parametro $\lambda = \frac{1}{\theta}$.

(a) Determinare $\widehat{\lambda}_{mv}(\mathbf{z}_n)$, stima di massima verosimiglianza di λ.

(b) Verificare che

$$\widehat{\lambda}_{mv}(\mathbf{Z}_n)|\theta \,\dot{\sim}\, \mathrm{N}\left(\lambda, \frac{\lambda(\lambda-1)}{n}\right).$$

(c) Sulla base della precedente distribuzione asintotica determinare un intervallo di confidenza di livello $1-\alpha$ per λ.

(d) Supponendo che il campione abbia dimensione $n = 20$ e che $\sum_{i=1}^n x_i = 50$, determinare gli estremi dell'intervallo di confidenza di livello 0.95.

(e) Sulla base dell'intervallo determinato, è possibile accettare l'ipotesi nulla $H_0 : \lambda = 1.3$ contro l'ipotesi alternativa $H_1 : \lambda \neq 1.3$ per un test di ampiezza $\alpha = 0.05$?

Soluzione.

(a) $\ell(\theta) = \theta^2(1-\theta)^{s_n - n}$, con $s_n = \sum_{i=1}^n x_i$. Si verifica facilmente che lo stimatore di massima verosimiglianza per θ è $1/\overline{x}_n$ e quindi che $\widehat{\lambda}_{mv}(\mathbf{Z}_n) = \overline{X}_n$.

(b) Per il teorema del limite centrale sappiamo che $\overline{X}_n \dot{\sim} \mathrm{N}\left(\frac{1}{\theta}, \frac{1-\theta}{n\theta^2}\right)$ da cui il risultato.

(c) $\tilde{C} = \widehat{\lambda}_{mv}(\mathbf{Z}_n) \pm u_{1-\frac{\alpha}{2}}\sqrt{\widehat{\mathbb{V}_\theta[\widehat{\lambda}_{mv}(\mathbf{Z}_n)]}} = \overline{X}_n \pm u_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}_n(\overline{X}-1)}{n}}$.

(d) $\tilde{C} = [1.65, 3.35]$

```
n=20
sn=50
smv=sn/n
sd.smv=sqrt(smv*(smv-1)/n)
L.n=smv-qnorm(0.975)*sd.smv
U.n=smv+qnorm(0.975)*sd.smv
c(L.n,U.n)
```

(e) Poichè $1.3 \notin \tilde{C}$ rifiuto H_0.

41. Sia $Z = (X_1, \ldots, X_n)$ un campione casuale proveniente da una popolazione di Poisson di parametro incognito $\theta/2$, $\theta > 0$.

(a) Determinare lo stimatore di massima verosimiglianza $d_{mv}(\mathbf{Z}_n)$ per il parametro θ e verificare che la sua distribuzione asintotica è: $\mathrm{N}\left(\theta, \frac{2\theta}{n}\right)$.

(b) Determinare l'intervallo di confidenza asintotico di livello $1-\alpha$ per θ (basarsi sulla distribuzione asintotica di d_{mv})

(c) Determinare l'espressione di $\mathcal{L}_{\tilde{C}}(\mathbf{Z}_n)$, ovvero la lunghezza (aleatoria) dell'intervallo considerato.

(d) Sia $R(\theta, C) = \mathcal{L}_{\tilde{C}}(\mathbf{Z}_n)^2$. Determinare $e_n = \mathbb{E}[R(\theta, \tilde{C})]$ e il minimo valore di n^* per il quale $e_n < \ell_0$ (utilizzare la distribuzione asintotica di d_{mv}).

(e) Determinare n^* assumendo $\alpha = 0.05$, $\theta = 1$ e $\ell_0 = 0.5$.

Soluzione.

(a) Si possono fare i calcoli oppure procedere come segue. Porre $\lambda = \frac{\theta}{2}$. Poichè $\widehat{\lambda}_{mv}) = \overline{X}_n$, per equivarianza abbiamo per θ che $d_{mv} = 2\overline{X}_n$. Inoltre, poichè $\mathbb{E}_\theta[d_{mv}] = \theta$ e $\mathbb{V}_\theta[d_{mv}] = \frac{2\theta}{n}$, abbiamo che (per il teorema del limite centrale) $d_{mv} \dot{\sim} \mathrm{N}\left(\theta, \frac{2\theta}{n}\right)$.

(b) $\tilde{C} = d_{mv} \pm u_{1-\frac{\alpha}{2}}\sqrt{\frac{2d_{mv}}{n}} = 2\overline{X}_n \pm u_{1-\frac{\alpha}{2}} 2\sqrt{\frac{\overline{X}_n}{n}}$.

(c) $\mathcal{L}_{\tilde{C}}(\mathbf{Z}_n) = 4u_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}_n}{n}}$.

(d) $e_n = \frac{8}{n}u_{1-\frac{\alpha}{2}}\theta$

(e) $n^\star = \lceil \frac{8}{\ell_0}u_{1-\frac{\alpha}{2}}\theta \rceil$; se $\alpha = 0.95$, $\theta = 1$ e $\ell_0 = 0.5$ abbiamo $n^\star = 32$.

42. Sia $Z = (X_1, \ldots, X_n)$ un campione casuale di dimensione n proveniente da una popolazione $\mathrm{N}(0, \theta)$ con parametro incognito θ e funzione di densità delle X_i

$$p_\theta(x) = \frac{1}{\sqrt{2\pi\theta}} \exp\{-\frac{1}{2\theta}x^2\}, \qquad x \in \mathbb{R}, \qquad \theta > 0.$$

Si consideri il sistema di ipotesi: $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$, (assumere $\theta_0 < \theta_1$). Verificare che la regione di rifiuto del test basato sul rapporto delle verosimiglianze risulta essere

$$\mathcal{Z}_0 = \{z \in \mathcal{Z} :\ S_0^2 < k_1 \ \vee S_0^2 > k_2\}, \qquad k_1 < k_2. \tag{2}$$

Soluzione. In questo caso $\ell(\theta) = \theta^{-n} \exp\left\{-\frac{n}{2\theta} d_{mv}\right\}$ con $d_{mv} = S_0^2$. Pertanto $\lambda_{01}^m = \frac{\ell(\theta_0)}{\ell(d_{mv})} = \left(\frac{S_0^2}{\theta_0}\right)^{\frac{n}{2}} \exp\left\{-\frac{n}{2}\frac{S_0^2}{\theta_0} + \frac{n}{2}\right\}$. La generica regola di rifiuto del test del rapporto delle verosimiglianze massimizzate rifiuta H_0 se $\lambda_{01}^m < k$, con $k > 0$. Dobbiamo quindi risolvere la disequazione $\left(\frac{S_0^2}{\theta_0}\right)^{\frac{n}{2}} \exp\left\{-\frac{n}{2}\frac{S_0^2}{\theta_0} + \frac{n}{2}\right\} < k$, rispetto alla variabile S_0^2.
Analiticamente si tratta di una disequazione del tipo $x^a \mathrm{e}^{-bx}$, $x \geq 0$, dove $a, b > 0$.
L'equazione associata non si risolve analiticamente (lasciato qui come esercizio) ma, dallo studio della funzione, siamo certi che esistano due soluzioni, k_1 e k_2. La funzione infatti assume valore zero in $x = 0$, cresce fino a un punto di massimo (in $x = a/b$) e poi decresce ea zero per $x \to \infty$ (vedi esempio con `a seguire`). La disequazione in x è quindi risolta per valori di x esterni all'intervallo delimitato da k_1 e k_2
Tornando al test, possiamo quindi dire che si rifiuta H_0 se $S_0^2 < k_1$ oppure se $S_0^2 > k_2$, con k_1 e k_2 da fissare in modo che il test abbia ampiezza (qui coincidente con probabilità di errore di primo tipo) uguale a α. In questo esercizio è comunque richiesta la sola struttura generale della regione di rifiuto.
Per visualizzare l'andamento della funzione $x^a \mathrm{e}^{-bx}$, $x \geq 0$ si usi il seguente codice, dove abbiamo fissato $a = 2$ e $b = 1$.

```
a=2
b=1
prova.fun=function(x){(x^a)*exp(-b*x)}
curve(prova.fun(x), from=0,to=10,xlab="x",ylab="")
```

43. Sia $X_1, \ldots, X_n|\theta \sim \mathrm{N}(\mu_0, \theta)$ i.i.d. (μ_0 nota) e $\Theta \sim \mathrm{InvGa}(\alpha, \beta)$, con $\alpha, \beta > 0$ e funzione di densità

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{\theta^{\alpha+1}} \exp\left\{-\frac{\beta}{\theta}\right\}, \quad \theta > 0.$$

(a) Verificare che la moda della distribuzione a priori risulta uguale a $\tilde{\theta} = \frac{\beta}{\alpha+1}$.
(b) Determinare la funzione di densità a posteriori di Θ e la moda a posteriori, $\mathrm{Mo}(\Theta|\mathbf{z}_n)$.
(c) Sia $d(\mathbf{Z}_n) = \mathrm{Mo}(\Theta|\mathbf{Z}_n)$. Verificare che $d(\mathbf{Z}_n)$ è uno stimatore distorto di θ ma non asintoticamente distorto.
(d) Determinare $R(\theta, d)$ assumendo la funzione di perdita quadratica e verificare se $d(\mathbf{Z}_n)$ è uno stimatore consistente per θ.
(e) Determinare l'espressione dello stimatore $d_u(\mathbf{Z}_n)$ che sia funzione di $d(\mathbf{Z}_n)$ e UMVUE per θ (giustificare il ragionamento).

Soluzione.

(a) Si ottiene facilmente derivando rispetto a θ la funzione $\ln \pi(\theta) = \ln c - (\alpha+1)\ln\theta - \frac{\beta}{\theta}$.

(b) Si verifica facilmente che $\pi(\theta|\mathbf{z}_n)$ è una densità gamma inversa di parametri $\overline{\alpha} = \alpha + \frac{n}{2}$ e $\overline{\beta} = \beta + \frac{n}{2}S_0^2$, con $S_0^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2$.

(c) $\text{Mo}[\Theta|\mathbf{z}_n] = \frac{\overline{\beta}}{\overline{\alpha}+1} = \frac{\beta+\frac{n}{2}S_0^2}{\alpha+\frac{n}{2}+1}$. Pertanto $\mathbb{E}_\theta\left(\text{Mo}[\Theta|\mathbf{z}_n]\right) = d(\mathbf{Z}_n) = \frac{\beta+\frac{n}{2}\theta}{\alpha+\frac{n}{2}+1} \to \theta$ per $n\to\infty$. La moda a posteriori è quindi uno stimatore distorto ma asintotricamente non distorto.

(d) $R(\theta,d) = \mathbb{V}_\theta[d] + (\mathbb{E}_\theta[d]-\theta)^2$. Ricordando che $\mathbb{V}_\theta[S_0^2] = \frac{n\theta^2}{2}$, si trova che $\mathbb{V}_\theta[d] = \frac{n\theta^2}{2(\alpha+n/2+1)^2}$ e che $(\mathbb{E}_\theta[d]-\theta)^2 = \left(\frac{\beta+\frac{n}{2}\theta}{\alpha+\frac{n}{2}+1} - \theta\right)^2$. Al crescere di n, abbiamo quindi che $R(\theta,d)\to 0$, $\forall\theta>0$, e quindi la consistenza di d risulta dimostrata.

(e) Notare che $d = a + bS_0^2$ con $a = \frac{\beta}{\alpha+\frac{n}{2}+1}$ e $b = \frac{n}{2\alpha+n+2}$. Poichè $\mathbb{E}_\theta[S_0^2] = \theta$, abbiamo che $d_u = \frac{d-a}{b}$, che risulta appunto non distorto, ovvero $\mathbb{E}_\theta[d_u] = \theta$, $\forall\theta>0$.

44. Sia $X_1,\dots,X_n|\theta \sim \text{N}(0,\theta)$ i.i.d. , $\theta>0$.

(a) Determinare $d_{mv}(\mathbf{z}_n)$.

(b) Determinare $R(d_{mv},\theta)$ (funzione di perdita quadratica).

(c) Verificare che

$$d_{mv}(\mathbf{Z}_n) \dot{\sim} \text{N}\left(\theta, \frac{2\theta^2}{n}\right).$$

(d) Considerare il test $H_0:\theta=\theta_0$ vs. $H_1:\theta>\theta_0$. Determinare la regione di rifiuto del test (asintotico) UMP_α basato sulla distribuzione asintotica di d_{mv}. Indicare tale test con $\tilde{d}_{KR}$.

(e) Determinare l'espressione della funzione di potenza del test $\tilde{d}_{KR}$.

Soluzione.

(a) Risultato standard per il modello normale con valore atteso noto e varianza incognita: $d_{mv} = S_0^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\mu_0)^2$; qui $\mu_0 = 0$ e quindi $d_{mv} = S_0^2 = \frac{1}{n}\sum_{i=1}^n X_i^2$.

(b) Poichè S_0^2 è stimatore non distorto di θ, abbiamo che $R(\theta,d_{mv}) = \mathbb{V}_\theta[S_0^2] = \frac{2\theta^2}{n}$ (la varianza di S_0^2 la ricaviamo ricordando qual è la sua distribuzione campionaria o quella di $\frac{nS^23_0}{\theta}$).

(c) Da risultati noti (teo. Blackwell-Rao e Lehmann-Scheffe) so che S_0^2 è stimatore UMVUE di θ, con varianza coincidente con limite inferiore di Cramer-Rao, che coincide con l'inverso dell'informazione attesa di Fisher. Pertanto possiamo affermare che $d_{mv} \dot{\sim} \text{N}\left(\theta, \frac{2\theta^2}{n}\right)$.

(d) Per il sistema di ipotesi considerato, il test UMP asintotico (quello di Karlin-Rubin) rifiuta H_0 se $d_{mv} = S_0^2 > k$. Imponendo che il test abbia probabilità di errore di primo tipo pari ad α, ovvero che $\mathbb{P}_{\theta_0}[S_0^2 > k] = \alpha$, si verifica facilmente che il test di ampiezza α rifiuta H_0 se $\frac{nS_0^2}{\theta_0} > \chi^2_{n;1-\alpha}$.

La funzione di potenza del test è quindi: $\eta(\theta) = \mathbb{P}_\theta\left(\frac{nS_0^2}{\theta} > \frac{\theta_0}{\theta}\chi^2_{n;1-\alpha}\right) = 1 - \mathbb{F}\left(\frac{\theta_0}{\theta}\chi^2_{n;1-\alpha}\right)$, dove $\mathbb{F}$ indica la funzione di ripartizione della v.a. χ^2_n.

45. Sia $X_1,\dots,X_n|\theta \sim \text{Ber}(\theta)$ i.i.d. e Sia inoltre $\mathbb{L}(\theta,C) = \mathcal{L}_C[\mathbf{z}_n]^2$ la funzione di perdita associata a un generico intervallo di stima per θ [dove $\mathcal{L}_C[\mathbf{z}_n]$ indica la lunghezza di $C(\mathbf{z}_n)$]. Sia

$$C(\mathbf{Z}_n) = \overline{X}_n \pm u_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}.$$

uno stimatore intervallare di θ.

Determinare le espressioni di $\mathbb{L}(\theta, C)$ e di $R(\theta, C)$ e $\lim_{n\to\infty} R(\theta, C)$.

Soluzione.

$\mathbb{L}(\theta, C)] = 4u^2_{1-\frac{\alpha}{2}} \frac{\overline{X}_n(1-\overline{X}_n)}{n}$. Pertanto $R(\theta, C) = \frac{4u^2_{1-\frac{\alpha}{2}}}{n}\mathbb{E}_\theta[\overline{X}_n - \overline{X}^2_n] = \frac{4u^2_{1-\frac{\alpha}{2}}}{n}\left(\theta - \frac{\theta(1-\theta)}{n} - \theta^2\right) \to 0$ per $n \to \infty$, $\forall \theta \in [0, 1]$.

46. Sia $X_1, \ldots, X_n|\theta \sim \mathrm{N}(\theta, 1)$ i.i.d. e $\Theta \sim \mathrm{N}(0, n_0^{-1})$. Considerare il sistema di ipotesi $H_0 : \theta = 0$ vs. $H_1 : \theta \neq 0$.

   (a) Calcolare l'espressione di $B_{10}(\overline{x}_n)$.

   (b) Considerare il test basato sulla regione di rifiuto

   $$\mathcal{Z}_1 = \{\mathbf{z}_n \in \mathcal{Z} : B_{10}(\overline{x}_n) > 1\}.$$

   Esprimere la regione di rifiuto in funzione di una relazione basata su $\overline{x}_n$ e un generico valore $k \in \mathbb{R}$.

   (c) Calcolare la funzione di potenza del test basato su $\overline{x}_n$.

   (d) Stabilire se il test così determinato, basato su $\overline{x}_n$, ha proprietà di ottimalità dal punto di vista dell'analisi in forma normale.

Soluzione.

(a) $B_{10} = \frac{m(\overline{x}_n)}{p_{\theta_0}(\overline{x}_n)}$. Ricordando che la densità marginale di $\overline{X}_n$ è quella di una v.a. $\mathrm{N}\left(0, \frac{n+n_0}{nn_0}\right)$ e che $\overline{X}_n|\theta_0 \sim \mathrm{N}\left(\theta_0, \frac{\sigma^2}{n}\right)$, si trova che $B_{10}(\overline{x}_n) = \left(\frac{n_0}{n+n_0}\right)^2 \exp\left\{-\frac{n^2}{2(n+n_0)}\overline{x}^2_n\right\}$, che è una funzione decrescente di $\overline{x}^2_n$. Pertanto $B_{01}(\overline{x}_n) > 1 \iff x^2_n < k \iff \overline{x}_n \in (-k', k')$ con $k' = \sqrt{k} > 0$.

(b) Ricordando che $\overline{X}_n|\theta \sim \mathrm{N}\left(\theta, \frac{1}{n}\right)$ si ha che $\eta(\theta) = \mathbb{P}_\theta(R) = \Phi[\sqrt{n}(k' - \theta)] - \Phi[-\sqrt{n}(k' + \theta)]$.

(c) Il test è UMPU in quanto basato sulla statistica $\overline{X}_n$ (sufficiente e completa).

47. Sia $X_1, \ldots, X_n|\theta \sim \mathrm{Pois}(\theta)$ i.i.d. e $\Theta \sim \mathrm{Gamma}(\alpha, \text{rate} = \beta)$, $\alpha, \beta > 0$. Sia inoltre $\mathrm{L}(\theta, C) = (\mathcal{L}[C(\mathbf{z}_n)])^2$ la funzione di perdita associata a un generico intervallo di stima per θ [dove $\mathcal{L}[C(\mathbf{z}_n)]$ indica la lunghezza di $C(\mathbf{z}_n)$]. Sia

$$C(\mathbf{Z}_n) = \mathbb{E}[\Theta|\mathbf{Z}_n] \pm u_{1-\frac{\alpha}{2}}\sqrt{\mathbb{V}]\Theta|\mathbf{Z}_n]}$$

uno stimatore intervallare di θ.

   (a) Determinare l'espressione esplicita (in funzione della statistica sufficiente) di $C(\mathbf{Z}_n)$ e di $\mathcal{L}_C[\mathbf{Z}_n]$.

   (b) Determinare l'espressione di $R(C, \theta)$.

   (c) Determinare $\lim_{n\to\infty} R(C, \theta)$.

   (d) Determinare l'espressione di $r_\pi(C)$.

Soluzione.

(a) Le espressioni richieste si trovano osservando che $\mathcal{L}_C[\mathbf{Z}_n] = 2u_{1-\frac{\alpha}{2}}\sqrt{\mathbb{V}(\Theta|\mathbf{z}_n)}$ e ricordando che $\mathbb{E}[\Theta|\mathbf{z}_n] = \frac{\overline{\alpha}}{\overline{\beta}} = \frac{\alpha+n\overline{x}_n}{\beta+n}$ e $\mathbb{V}[\theta|\mathbf{z}_n] = \frac{\overline{\alpha}}{\overline{\beta}^2} = \frac{\alpha+n\overline{x}_n}{(\beta+n)^2}$.

(b) $R(C, \theta) = \mathbb{E}_\theta[\mathcal{L}_C[\mathbf{Z}_n]^2] = 4u^2_{1-\frac{\alpha}{2}}\frac{\alpha+n\overline{X}_n}{(\beta+n)^2} = 4u^2_{1-\frac{\alpha}{2}}\frac{\alpha+n\theta}{(\beta+n)^2}$.

(c) $\lim_{n\to\infty} R(C, \theta) = 4u^2_{1-\frac{\alpha}{2}}\frac{\alpha+n\theta}{(\beta+n)^2} = 0$, $\forall \theta > 0$.

(d) $r_\pi(C) = 4u^2_{1-\frac{\alpha}{2}}\frac{\alpha+n\mathbb{E}[\Theta]}{(\beta+n)^2} = 4u^2_{1-\frac{\alpha}{2}}\frac{\alpha+n\frac{\alpha}{\beta}}{(\beta+n)^2}$.

48. Sia $X_1, \ldots, X_n|\theta \sim \mathrm{N}(\theta, 1)$ i.i.d. e $\Theta \sim \mathrm{N}(1, n_0^{-1})$. Considerare il sistema di ipotesi $H_0 : \theta \leq 0$ vs. $H_1 : \theta > 0$.

(a) Calcolare l'espressione di $d(\mathbf{z}_n) = \mathbb{E}[\Theta|\mathbf{z}_n]$.

(b) Considerare il test basato sulla regione di rifiuto

$$\mathcal{Z}_1 = \{\mathbf{z}_n \in \mathcal{Z} : d(\mathbf{z}_n) > k\}.$$

Determinare k in modo tale che il test sia di ampiezza α.

(c) Calcolare l'espressione della funzione di potenza del test di ampiezza α basato su $d(\mathbf{z}_n)$.

(d) Verificare analiticamente che il modello considerato ha rapporto delle verosimiglianze monotono.

(e) Stabilire se il test così determinato, basato su $(\mathbf{z}_n)$, ha proprietà di ottimalità dal punto di vista dell'analisi in forma normale.

Soluzione.

(a) $\mathbb{E}[\Theta|\mathbf{z}_n] = \frac{n_0\mu_0 + n\bar{x}_n}{n_0+n} = \frac{n\bar{x}_n}{n_0+n}$ poichè $\mu_0 = 0$.

(b) $d(\mathbf{z}_n) > k \iff \bar{x}_n > \frac{n_0+n}{n}k = k$ (possiamo usare lo stesso simbolo per k in quanto da determinare). Ci siamo quindi ricondotti all'usuale test di Karlin-Rubin, ovvero al test UMP. Il test ha ampiezza α se $\eta(\theta_0) = \alpha$ ovvero se (... vedi libro ...) $k = k_\alpha = \frac{\sigma}{\sqrt{n}} u_{1-\alpha} + \theta_0$.

(c) La funzione di potenza del test di ampiezza α è

$\eta(\theta) = \mathbb{P}_\theta[\overline{X}_n > k_\alpha] = \ldots = 1 - \Phi\left(\frac{\sqrt{n}}{\sigma}(\theta_0 - \theta) + u_{1-\alpha}\right).$

(d) Siamo nelle condizioni del teorema di Karlin-Rubin: il modello ha rapporto delle verosimiglianze monotono in $\overline{X}_n$ e le ipotesi sono unilaterali. La regola di rifiuto proposta individua quindi un test UMP di ampiezza α.

49. Sia $X_1, \ldots, X_n|\theta \sim \mathrm{Ber}(\theta)$ i.i.d. e $\pi(\theta) = \mathrm{Beta}(\theta|\alpha, \beta)$, $\alpha, \beta > 0$ e considerare le due funzioni di decisione:

$$d_1(\mathbf{z}_n) = d_{mv}(\mathbf{z}_n) \qquad \text{e} \qquad d_2(\mathbf{z}_n) = a^*(\mathbf{z}_n),$$

dove $a^*(\mathbf{Z}_n)$ indica la decisione ottima rispetto alla funzione di perdita quadratica nell'analisi in forma estensiva.

(a) Determinare l'espressione di a^*.

(b) Determinare $\rho(d_1, \mathbf{z}_n)$ e $\rho(d_2, \mathbf{z}_n)$ (rispetto alla perdita quadratica).

(c) Determinare la funzione di rischio normale $R(\theta, d_1)$.

(d) Determinare la funzione di rischio normale di $d_2(\mathbf{z}_n)$.

(e) Verificare che con $\alpha = \beta = \sqrt{n/4}$ si ottiene $d_2(\mathbf{Z}_n) = \frac{\sum_{i=1}^n X_i + \sqrt{n/4}}{n+\sqrt{n}}$.

(f) Stabilire se il test basato sulla regione di rifiuto

$$\mathcal{Z}_1 = \{d_2(\mathbf{z}_n) > k\}$$

è un test UMP per il sistema di ipotesi $H_0 : \theta \geq \theta_0$ vs. $H_1 : \theta < \theta_0$.

Soluzione.

(a) $a^\star = \mathbb{E}[\Theta|\mathbf{z}_n] = \frac{\alpha+n\overline{x}_n}{\alpha+\beta+n}$.

(b) Usare i seguenti fatti.

$\rho(d_2, \mathbf{z}_n) = \rho(a^\star, \mathbf{z}_n) = \mathbb{V}[\Theta|\mathbf{z}_n] = \ldots$ (vedi libro).

$\rho(d_1, \mathbf{z}_n) = \rho(\overline{X}_n, \mathbf{z}_n) = \rho(a^\star, \mathbf{z}_n) + (\overline{X}_n - a^\star)^2 = \ldots$ (calcoli).

(c) $R(\theta, d_1) = \mathbb{V}_\theta[\overline{X}] = \frac{\sigma^2}{n}$.

(d) $R(\theta, d_2) = \mathbb{V}_\theta[d_2] + [\mathbb{E}_\theta(d_2) - \theta]^2 = \ldots$ (calcoli).

(e) Sostituire α e β in $a^\star = \mathbb{E}[\Theta|\mathbf{z}_n] = \frac{\alpha+n\overline{x}_n}{\alpha+\beta+n}$.

(f) La regione di rifiuto proposta basata su d_2 si riconduce alla regione di rifiuto del test di Karlin-Rubin per il sistema di ipotesi considerato in quanto $d_2 > k \iff \overline{X}_n > k'$ (con k e k' arbitrari).

D - Altri esercizi (2024)

50. (da Esonero 1 23-24). Siano $X_1, \dots, X_n$ le risposte quantitative (aleatorie) a un trattamento medico somministrato a n pazienti e sia θ l'*effetto del trattamento* medio nella popolazione da cui provengono i dati. Supporre che, condizionatamente alla conoscenza di θ, le risposte al trattamento siano v.a. indipendenti con distribuzione **normale** centrata sul valore vero dell'effetto del trattamento e varianza unitaria. Si dispone di informazione pre-sperimentale e sperimentale sull'*effetto del trattamento*.
Informazione sperimentale: su 10 osservazioni sperimentali, la somma delle risposte al trattamento è uguale a 20 (in opportuna unità di misura).
Informazione pre-sperimentale: basandoci su dati storici è e lecito supporre che il vero valore del trattamento θ abbia distribuzione **normale** con valore atteso pari a 1/2 e varianza pari a 1/4.

   (a) Formalizzare opportunamente il problema per mezzo di un modello bayesiano, ovvero: indicare quali sono le distribuzioni di $X_i|\theta$ e di Θ da utilizzare e i valori da associare ai parametri noti e agli e iperparametri dei modelli considerati.

   (b) Scrivere le espressioni generiche (formule) e specifiche (sostituendo i valori numerici) della distribuzione a posteriori di Θ.

   (c) Fornire una stima puntuale e una intervallare di θ basata sulla distribuzione a posteriori di Θ [riportare espressioni generiche (formule) e specifiche (sostituendo i valori numerici)]. Per la stima intervallare utilizzare un livello di credibilità $\gamma = 0.95$ e ricordare che $z_{0.975} = 1.96$.

Soluzione.

(a) Si deve utilizzare il modello normale-normale bayesiano. Quindi $X_i|\theta \sim \mathrm{N}(\theta, \sigma^2)$, $i = 1, \dots n$ i.i.d. e $\Theta \sim \mathrm{N}\left(\mu_0, \frac{\sigma^2}{n}\right)$, con $n = 10$, $\overline{x}_n = 2$ $\sigma^2 = 1$ $n_0 = 4$ $\mu_0 = \frac{1}{2}$. Osservare che n_0 si trova dalla relazione $\frac{\sigma^2}{n_0} = \frac{1}{4}$.
Abbiamo quindi che $X_i|\theta \sim \mathrm{N}(0, 1)$ e $\Theta \sim \mathrm{N}\left(\frac{1}{2}, \frac{1}{4}\right)$.

(b) Si ottiene $\Theta|\mathbf{z}_n \sim \mathrm{N}\left(\overline{\mu}_p, \overline{\sigma}_p^2\right)$, con $\overline{\mu}_p = \frac{n_0\mu_0 + n\overline{x}}{n_0+n} = \frac{11}{7}$ e $\overline{\sigma}_p^2 = \frac{\sigma^2}{n_0+n} = \frac{1}{14}$.

(c) $\widehat{\theta}_B = \overline{\mu}_p = \frac{11}{7}$;
$C = \overline{\mu}_p \pm u_{1-\frac{\gamma}{2}}\overline{\sigma}_p = \frac{11}{7} \pm 1.96\frac{1}{\sqrt{14}} = [1.04, 2.09]$, usando $1 - \alpha = 0.95$.

51. (da Esonero 1 23-24). Sia $X|\theta \sim p_\theta^X(x) = \theta x^{\theta-1}$, $x \in (0, 1)$ e $\theta > 0$. Considerare: $H_0 : \theta = \theta_0$ e $H_1 : \theta = \theta_1$, con $\theta_0 > \theta_1 > 0$.

   (a) Determinare l'espressione di $B_{01}(x)$, fattore di Bayes per il confronto tra le ipotesi in esame basato su una singola osservazione x.

   (b) Verificare che $B_{01}(x) > 1 \iff x > A$ e determinare l'espressione di A (in funzione di θ_0 e θ_1).

   (c) Determinare la probabilità $\mathbb{P}_{\theta_0}[B_{01}(X) > 1]$ [**Sugg.**: usare la distribuzione di $X|\theta_0$].

   (d) Calcolare il valore numerico che si ottiene ponendo $\theta_0 = 2$ e $\theta_1 = 1$.

Soluzione.

(a) $B_{01}(x) = \frac{p^X_{\theta_0}(x)}{p^X_{\theta_1}(x)} = \frac{\theta_0}{\theta_1} x^{\theta_0 - \theta_1}$

(b) $B_{01}(x) > 1 \;\Leftrightarrow\; x^{\theta_0-\theta_1} > \frac{\theta_1}{\theta_0} \;\Leftrightarrow\; x > A = \left(\frac{\theta_1}{\theta_0}\right)^{\frac{1}{\theta_0-\theta_1}}$, dove l'ultima disuguaglianza è valida in quanto $\theta_0 > \theta_1$.

(c) $\mathbb{P}_{\theta_0}[B_{01}(X) > 1] = \mathbb{P}_{\theta_0}[X > A] = \int_A^1 \theta_0 x^{\theta_0 - 1} dx = \theta_0 \frac{x^{\theta_0}}{\theta_0}|_A^1 = x^{\theta_0}|_A^1 = 1 - A^{\theta_0} = 1 - \left(\frac{\theta_1}{\theta_0}\right)^{\frac{\theta_0}{\theta_0-\theta_1}}$

(d) Per $\theta_0 = 2$, $\theta_1 = 1$ si ha che $A = \frac{1}{2}$ e quindi $\mathbb{P}_{\theta_0}[B_{01}(X) > 1] = 1 - \frac{1}{4} = \frac{3}{4}$.

52. (da prova scritta febbraio 2024). Sia $\mathbf{Z}_n = (X_1, ..., X_n)$ un campione casuale con funzione di densità di probabilità $p_\theta(x) = \theta x^{\theta-1}$, $x \in [0, 1]$, $\theta > 0$.

(a) Determinare la funzione di verosimiglianza e derivare analiticamente l'espressione di $d_{mv}(\mathbf{z}_n)$, stima di massima verosimiglianza del parametro.

(b) Determinare $\pi^J(\theta)$, distribuzione a priori di Jeffreys per Θ.

(c) Determinare $\pi^J(\theta|\mathbf{z}_n)$.

(d) Determinare $\tilde{\theta}$ (moda a posteriori), I_B (informazione bayesiana) e approssimazione normale della distribuzione di $\Theta|\mathbf{z}_n$ basata su tali quantità.

(e) Determinare l'espressione di $\tau(\theta) = \mathbb{P}_\theta\left[X > \frac{1}{2}\right]$. Determinare $\widehat{\tau(\theta)}$ supponendo che $\tilde{\theta} = 2$.

Soluzione.

(a) $\ell(\theta) = c \cdot \theta^n \, t^\theta$, con $t = \prod_{i=1}^n x_i$. Pertanto si verifica che $d_{mv} = -\frac{n}{t}$ (si noti che $\ln t < 0$).

(b) Poichè $\frac{d^2}{d\theta^2}\ell(\theta) = -\frac{n}{\theta^2}$, si ha che $I_n(\theta) = \frac{n}{\theta^2}$ e quindi $\pi^J(\theta) = \frac{c}{\theta}$.

(c) $\pi^J(\theta|\mathbf{z}_n) = c \cdot \theta^n \, t^\theta \times \frac{1}{\theta} = c \cdot \theta^{n-1} e^{-(-\ln t)\theta}$. Quindi $\Theta|\mathbf{z}_n \sim \text{Ga}(n, \text{rate} = -\ln t)$.

(d) Dall'equazione $\frac{d}{d\theta}\ln \pi(\theta|\mathbf{z}_n) = 0$ si ottiene $\tilde{\theta} = -\frac{n-1}{t}$. Si trova inoltre che $I_B = \frac{t^2}{n-1}$ e quindi $\Theta|\mathbf{z}_n \mathrel{\dot\sim} \text{N}\left(-\frac{n-1}{t}, \frac{n-1}{t^2}\right)$.

(e) $\tau(\theta) = \int_{1/2}^1 p_\theta(x)dx = \int_{1/2}^1 \theta x^{\theta-1} dx = x^\theta|_{1/2}^1 = 1 - \frac{1}{2^\theta}$. Pertanto $\tau(\tilde{\theta}) = \frac{3}{4}$.