Lab 2

October 6, 2025

1 Esercitazione 2

Nota: Se ci sono dei parametri e non vi dico quanto valgono, potete usare qualsiasi valore voi vogliate

1.1 Importance Sampling

Assumete di avere

$$X \sim TN(\mu, \sigma^2, l, u)$$

dove $TN(\mu, \sigma^2, l, u)$ è una normale troncata, definita in $[l, u]$ con $l < u$, i.e., $X \in [l, u]$, con densità

$$f(x) = \frac{\phi(x|\mu, \sigma^2)}{\Phi(u|\mu, \sigma^2) - \Phi(l|\mu, \sigma^2)}$$

dove

$$\phi(x|\mu, \sigma^2)$$

è la densità di una normale di media μ e varianza σ^2 valutata in x, e

$$\Phi(y|\mu, \sigma^2)$$

e la cumulata di una normale di media μ e varianza σ^2 valutata in y.

1. Scrivete una funzione che ne calcoli la densità, e poi mostratela graficamene
2. Usate Monte Carlo per valutare se la densità integra a 1
3. Utilizzate l'importance sampling per stimare il valore atteso di X, usando come densità "g()" un uniforme.
4. Utilizzate l'importance sampling per stimare il valore atteso di X, usando come densità "g()" una beta scalata con opportuni parametri.

La beta scalata è la distribuzione di una variabile $Z = (u - l)Y + l$, con $Y \sim B(a, b)$

1.2 Accept-Reject Sampling

Assumete la stessa distribuzione del punto precedente

$$X \sim TN(\mu, \sigma^2, l, u)$$

1. campionate dalla normale troncata usando l'algoritmo accept-reject. Disegnate i campioni accettati di un colore diverso di quelli rifiutati
2. verificare che i campioni seguono la distribuzione giusta
3. verificate quale è il valore del rate d'accettazione e che corrisponda a quello vero

1.3 Marginale e congiunta

Assumiamo

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^A \\ \mathbf{X}^B \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \begin{pmatrix} \Sigma_A & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_B \end{pmatrix} \right)$$

dove $\mathbf{X}^A$ e $\mathbf{X}^B$ sono vettori di dimensione 2. Definiamo

$$\mu = \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_A & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_B \end{pmatrix}$$

Sappiamo che

$$\mathbf{X}^A | \mathbf{X}^B \sim N_2(\mathbf{M}, \mathbf{V})$$

con

$$\mathbf{M} = \mu_A + \Sigma_{AB}\Sigma_B^{-1}(\mathbf{X}^B - \mu_B)$$

e

$$\mathbf{V} = \Sigma_A - \Sigma_{AB}\Sigma_B^{-1}\Sigma_{BA}$$

1. Simulare n campioni dalla normale multivariata
2. Simulare n campioni dalla marginale di μ_A
3. Simulare n campioni dalla normale multivariata simulando prima campioni dalla marginale di $\mathbf{X}_B$ e poi dalla condizionata di $\mathbf{X}_A|\mathbf{X}_B$
4. Far vedere che per tutte e 3 i punti precedenti i campioni di $\mathbf{X}_A$ provengono sempre dalla marginale

1.3.1 P.S.1

Per il punto 4 potete 1. usare la funzione smoothScatter (che è la stima di densità per una coppia di variabili), 2. oppure far vedere che la probabilità che $\mathbf{X}^A \in \mathbf{A}$ sia simile per i 3 metodi

$$P(X \in \mathbf{A}) = \int_{\mathbb{R}^2} 1_{\mathbf{A}}(x) f(x) dx$$

3. oppure calcolare la comulata

$$P(X \in (-\inf, b] \times (-\inf, d])$$

in un set di valori b e d equispaziati e usare il comando image

1.3.2 P.S.2

Mentre per μ potete prendere i valori che volete, Σ va scelta con cura visto che è una matrice simmetrica e definita positiva. Potete ottenrere una Σ valida definendo una matrice

$$\mathbf{D}$$

di dimensione 4 a rango pieno, e

$$\Sigma = \mathbf{D}\mathbf{D}^T$$

Oppure potete ottenere un campione da una $IW(\nu, \Psi)$ (Inverse Wishart) che è una distribuzione per matrici di covarianza usando il pacchecctto MCMCpack e la funzione riwish. Nell'inverse wishart $\nu \geq 4$ (dimensione di Σ) e Ψ è una matrice delle stesse dimensioni di Σ definita positiva (potete usare l'indentità)

1.3.3 P.S.3

se

$$\Sigma^* = \mathbf{C}\mathbf{C}^T$$

e

$$\mathbf{Y} \sim N_d(\mathbf{0}_d, \mathbf{I}_d)$$

allora

$$\mu^* + \mathbf{C}\mathbf{Y} \sim N_d(\mu^*, \Sigma^*)$$

1.4 Inversa Generalizzata e Discretizzazioni di variabili continue

Definite

$$c = \frac{1}{m}$$

e assumete di avere un variabile

$$X \in \left\{\frac{0}{m} + \frac{c}{2}, \frac{1}{m} + \frac{c}{2}, \frac{2}{m} + \frac{c}{2}, \dots, \frac{m-1}{m} + \frac{c}{2}\right\} \equiv \mathcal{X}$$

discreta, che assume quindi m possibili valori. Definiamo la probabilità di X come

$$P\left(X = \frac{j}{m} + \frac{c}{2}\right) \propto f_{beta}\left(\frac{j}{m} + \frac{c}{2}|a, b\right) c$$

dove

$$f_{beta}\left(\frac{j}{m} + \frac{c}{2}|a, b\right)$$

è la densità di una beta di parametro (a, b) calcolata in $\frac{j}{m} + \frac{c}{2}$. Visto che le probabilità devono sommare a uno, devo avere per forza che

$$P\left(X = \frac{j}{m} + \frac{c}{2}\right) = \frac{f_{beta}\left(\frac{j}{m} + \frac{c}{2}|a, b\right)}{\sum_{h=0}^{m-1} f_{beta}\left(\frac{h}{m} + \frac{c}{2}|a, b\right)}$$

Queto è un modo per discretizzare uan variabile continua. Potete vedere più nel dettaglio cosa stiamo facendo nella figura sottostante in cui

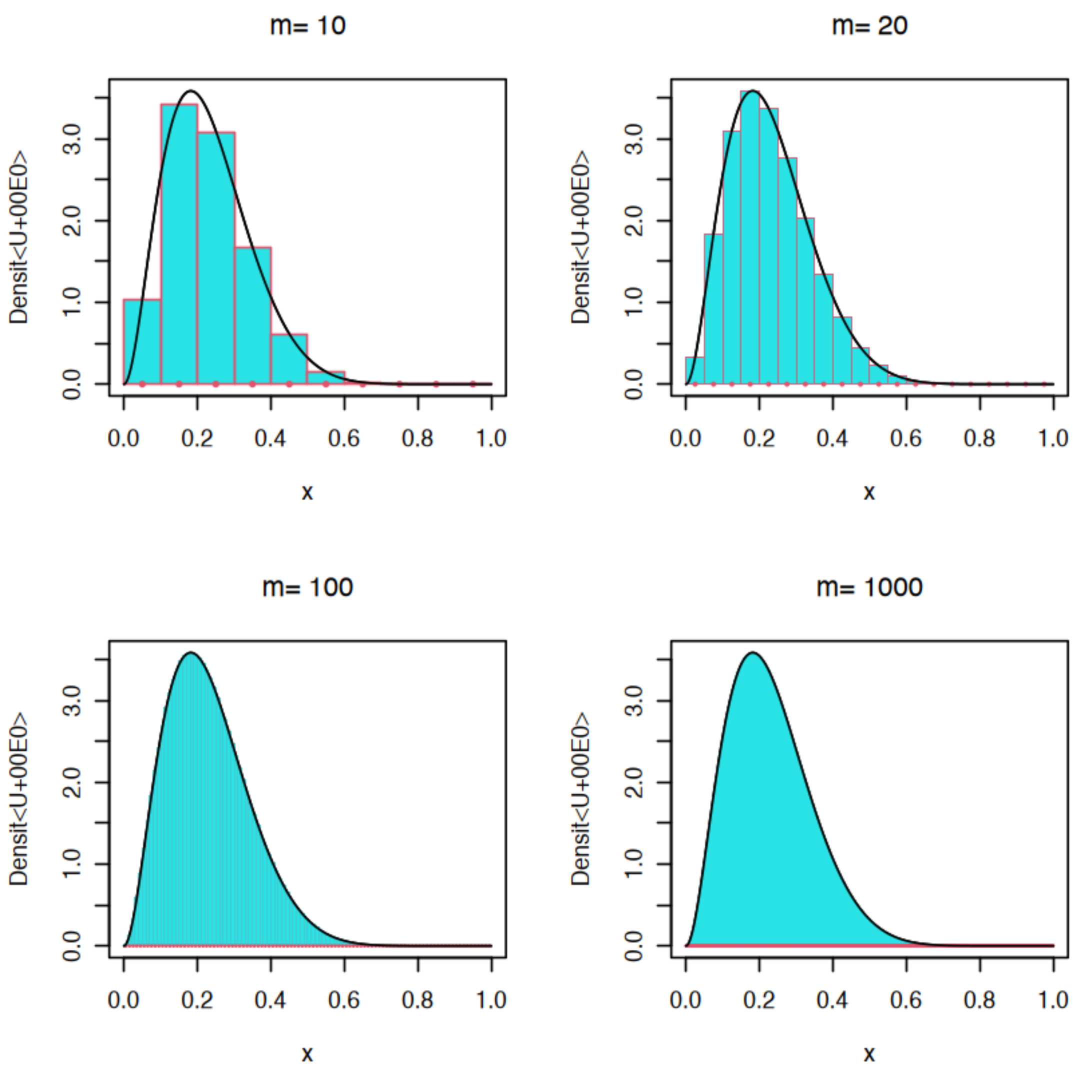
1. La linea nera è la densità di una beta
2. I punti rossi sono i valori di $\mathcal{X}$
3. L'area di ogni rettangolo è $f(x)$ calcolata nel punto rosso associato.

Provate ad aumentare m e vedrete che la versione discreta è simile alla continua

[13]:
```
# DISCRETIZZAZIONE DI UNA VARIABILE CONTINUA BETA
# Questo codice mostra come discretizzare una distribuzione Beta continua

# Parametri della distribuzione Beta
a = 3     # Primo parametro della Beta
b = 10    # Secondo parametro della Beta

# Configurazione grafico: 2x2 subplot per confrontare diversi valori di m
par(mfrow=c(2,2))
```

```r
# Loop per diversi livelli di discretizzazione
for(m in c(10,20,100,1000))
{
    # PASSO 1: Definizione della griglia discreta
    c = 1/m                                 # Larghezza di ogni intervallo di␣
↪discretizzazione
    xdom = c(0:(m-1))/m + c/2               # Punti centrali degli intervalli discreti
                                            # Formula: j/m + c/2 per j = 0,1,...,m-1

    # PASSO 2: Calcolo degli estremi degli intervalli
    x_start = xdom - c/2                    # Estremi sinistri: [j/m, (j+1)/m)
    x_end   = xdom + c/2                    # Estremi destri: [j/m, (j+1)/m)

    # PASSO 3: Calcolo delle probabilità discrete
    prob_x = dbeta(xdom, a, b)              # Valuto la densità Beta nei punti centrali
    prob_x = prob_x/sum(prob_x)             # Normalizzo per ottenere probabilità␣
↪discrete
                                            # P(X = xdom[j]) = f_beta(xdom[j]) / Σ␣
↪f_beta(xdom[k])

    # PASSO 4: Calcolo altezze per visualizzazione
    altezza = prob_x/c                      # Altezza = probabilità / larghezza␣
↪interval
                                            # Questo garantisce che area = probabilità

    # PASSO 5: Creazione della sequenza fine per la densità continua
    xseq = seq(0, 1, by = 0.001)          # Griglia fine per plot continuo

    # PASSO 6: Visualizzazione grafica
    # Densità Beta continua (linea nera)
    plot(xseq, dbeta(xseq,a,b), type="l", main=paste("m=", m), lwd=1,
         xlab="x", ylab="Densità", ylim=c(0, max(dbeta(xseq,a,b))))

    # Rettangoli per la distribuzione discreta (viola con bordo rosso)
    rect(x_start, rep(0, m), x_end, altezza,
         border=2,              # Bordo rosso
         col=5,                 # Riempimento viola
         lwd= c*10)             # Spessore bordo proporzionale a larghezza␣
↪intervallo

    # Punti centrali della discretizzazione (punti rossi)
    points(xdom, rep(0, m),
           pch=20,              # Simbolo punto pieno
           cex= c*5,            # Dimensione proporzionale a larghezza intervallo
           col=2)               # Colore rosso
```

```r
    # Ridisegno la densità continua sopra i rettangoli
    lines(xseq, dbeta(xseq,a,b), col=1, lwd=1)
}

# Ripristino layout grafico normale
par(mfrow=c(1,1))

# INTERPRETAZIONE:
# - Linea nera: densità Beta(a,b) continua
# - Rettangoli viola: approssimazione discreta con m intervalli
# - Punti rossi: punti centrali xdom dove valutiamo la densità
# - Area di ogni rettangolo = P(X = xdom[j]) nella versione discreta
# - All'aumentare di m, la discretizzazione approssima meglio la distribuzione␣
 ↪continua
```



1. Disegnate la cumulata della beta, e della X (stima) nello stesso grafico (provate diversi valori di m, per esempio m = 10, m = 50, m = 1000) e notate le differenze.
2. Usando il metodo dell'uniforme e l'inversa generalizzata, simulate dalla distribuzione di X, con m piccolo (e.g $m = 5$), e verificate, tramite monte carlo, che il valore di

$$P\left(X \leq \frac{j}{m} + \frac{c}{2}\right)$$

sia effettivamente

$$\sum_{l=0}^{j} \left(\frac{f_{beta}\left(\frac{l}{m} + \frac{c}{2}|a, b\right)}{\sum_{h=0}^{m-1} f_{beta}\left(\frac{h}{m} + \frac{c}{2}|a, b\right)} \right)$$

3. fate la stessa cosa del punto 2, ma simulate usando direttamente la distribuzione di X
4. Per $m = 5$, $m = 50$ e $m = 1000$ calcolare il quantile a livello 0.1, 0.2, 0.3, ..., 0.8, 0.9, e confrontateli con quelli della beta agli stessi livelli

Per il punto 2, dovete calcolare con monte carlo

$$\int_{\mathcal{X}} 1_{(-\infty, \frac{j}{m} + \frac{c}{2}]}(x) f(x) d\lambda(x)$$

e confrontarlo con $P\left(X \leq \frac{j}{m}\right)$ per tutti i valori di j ammissibili. I campioni da $f(x)$ devono essere ottenuti con il metodo dell'inversa generalizzata