Laboratorio di R

Simulazioni Monte Carlo: variabili casuali

Variabile casuale. Dato uno spazio di probabilità $(\Omega, \mathcal{A}, \mathsf{P})$, una *variabile casuale* (o *aleatoria*), usualmente abbreviata con v.c., è una funzione $X(\omega) : \Omega \to \mathbf{R}$, avente come dominio Ω e codominio la retta reale $\mathbf{R}$, che soddisfa alla *condizione di misurabilità*, cioè tale che, per ogni $r \in \mathbf{R}$, l'insieme $\{\omega : X(\omega) \leq r\}$ appartiene ad $\mathcal{A}$. Questa definizione ci dice che non tutte le funzioni definite sullo spazio campionario Ω, che associano ad ogni evento elementare ω un numero reale sono delle variabili casuali. Fondamentalmente, la condizione di misurabilità richiede che una funzione per essere una variabile casuale debba essere tale per cui si possano anche calcolare le probabilità associate ai suoi valori.

Funzione di ripartizione. Per una data variabile casuale X, la sua *funzione di ripatizione* è definita, per ogni $x \in \mathbf{R}$, come

$$F_X(x) = P(X \leq x).$$

In altre parole, la funzione di ripartizione $F_X(x)$ associa ad ogni valore $x \in \mathbf{R}$ la probabilità che la variabile casuale X assuma valori minori od uguali ad x. La condizione di misuranilità ci garantisce che ogni variabile casuale abbia la sua funzione di ripartizione. Essenzialmente, questa funzione fornisce le probabilità associate ai valori che può assumere la variabile casuale. La funzione di ripartizione gode delle seguenti proprietà:

- per $x \to +\infty$ si ha che $F_X(x) \to 1$;
- per $x \to -\infty$ si ha che $F_X(x) \to 0$;
- è una funzione monotona non decrescente;
- nei punti di discontinuità è continua a destra.

Variabile casuale discreta. Una variabile casuale si dice *discreta* se l'insieme $x_1, x_2, \ldots$ dei valori che può assumere costituisce o un insieme finito o un insieme infinito numerabile.

Distribuzione di probabilità di una variabile casuale discreta. Per ciascun valore x che la variabile casuale discreta X può assumere, definiamo la probabilità che esso si realizzi come $f(x) = P(X = x)$. L'insieme di tutti i valori $f(x)$, per $x = x_1, x_2, \ldots$, prende il nome di *distribuzione* (o *funzione*) *di probabilità* di X. Per una variabile casuale discreta, la sua funzione di ripatizione $F_X(x) = P(X \leq x)$ si può ottenere, per ogni $x \in \mathbf{R}$, come la somma delle probabilità associate ai possibili valori $x_1, x_2, \ldots$ che sono minori od uguali ad x. Quindi, per la funzione di ripartizione di una variabile casuale discreta si ha che

$$F_X(x_i) - F_X(x_{i-1}) = P(X \leq x_i) - P(X \leq x_{i-1}) = P(X = x_i),$$

per tutti i possibili valori $x_1, x_2, \ldots$. In questo caso, la funzione $F_X(x)$ è una funzione a gradini: costante a tratti e con salti in corrispondenza dei valori $x_1, x_2, \ldots$.

Variabile casuale continua. Dato uno spazio di probabilità $(\Omega, \mathcal{A}, \mathsf{P})$, una variabile casuale $X(\omega) : \Omega \to \mathbf{R}$ con funzione di ripartizione $F_X(x)$ si dice *continua* se

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, \mathrm{d}t,$$

dove $f_X(x)$ si chiama *funzione di densità di probabilità* di X. In altre parole, una variabile casuale X si dice continua se possiamo esprimere la sua funzione di ripartizione come l'integrale di un'altra funzione chiamata densità. Si noti che l'insieme dei possibili valori che può assumere una variabile casuale continua è un insieme infinito non numerabile.

Funzione di densità di probabilità di una variabile casuale continua. Per una variabile casuale continua, le probabilità associate ai possibili valori di X possono essere calcolate utilizzando la funzione di ripartizione, ovvero integrando la funzione di densità di probabilità. In particolare,

$$\begin{aligned} P(a < X \leq b) &= P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a) \\ &= \int_{-\infty}^{b} f_X(x)\,\mathrm{d}x - \int_{-\infty}^{a} f_X(x)\,\mathrm{d}x = \int_{a}^{b} f_X(x)\,\mathrm{d}x. \end{aligned}$$

Si noti che per una variabile casuale continua, la probabilità che questa assuma un determinato valore x è pari a zero, per ogni $x \in \mathbf{R}$. Ovviamente, una funzione di densità di probabilità $f_X(x)$ è una funzione non negativa, tale che $\int_{-\infty}^{+\infty} f_X(x)\,\mathrm{d}x = 1$.

Valore atteso e varianza. Sia X una variabile casuale discreta con distribuzione di probabilità $f(x) = P(X = x)$. Allora il suo *valore atteso* è definito come

$$\mathrm{E}(X) = \sum_x x f(x),$$

mentre la sua *varianza* è definita come

$$\mathrm{Var}(X) = \sum_x (x - \mathrm{E}(X))^2 f(x).$$

Sia invece X una variabile casuale continua con funzione di densità $f(x)$, allora il suo *valore atteso* è dato da

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x,$$

mentre la sua *varianza* è data da

$$\mathrm{Var}(X) = \int_{-\infty}^{+\infty} (x - \mathrm{E}(X))^2 f(x)\mathrm{d}x.$$

Distribuzione di Bernoulli. Una variabile casuale discreta X si dice avere *disctibuzione Bernoulliana* se la sua distribuzione di probabilità è data da

$$f(x) = P(X = x) = p^x (1 - p)^{1-x}, \quad x = 0, 1,$$

dove $0 \leq p \leq 1$. Una variabile casuale con questa distribuzione assume valore 1 con probabilità p e valore 0 con probabilità $1-p$. Per indicare che una variabile casuale X ha distribuzione Bernoulliana si scrive brevemente $X \sim \mathrm{Beroulli}(p)$. In molti contesti applicativi, si usa l'espressione variabile casuale di Bernoulli per indicare una variabile casuale che descrive il verificarsi o meno di un evento a cui è associata una probabilità di successo pari a p. Una variabile casuale discreta X avente distribuzione Bernoulliana ha valore atteso pari a p, e varianza pari a $p(1 - p)$.

Distribuzione Binomiale. Una variabile casuale discreta X si dice avere *distribuzione Binomiale* se la sua distribuzione di probabilità è data da

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \ldots, n,$$

dove $0 \leq p \leq 1$ e n è un qualsiasi numero naturale $1, 2, \ldots$. Per indicare brevemente che una variabile casuale X ha distribuzione Binomiale si scrive $X \sim$ Binomiale(n, p). Si noti che se $n = 1$, la distribuzione Binomiale coincide con la distribuzione Bernoulliana, ovvero che quest'ultima è un caso particolare della distribuzione Binomiale. La distribuzione Binomiale si incontra ogniqualvolta si ha il seguente schema sperimentale. Si considerino n prove dicotomiche (*Bernoulliane*) che possono risolversi con un successo o con un insuccesso, tutte aventi la stessa probabilità di successo pari a p. Allora la variabile casuale che conta il numero di successi in queste n prove ha distribuzione Binomiale. Una variabile casuale discreta X avente distribuzione Binomiale, ha valore atteso pari a np, e varianza pari a $np(1 - p)$.

Distribuzione di Poisson. Una variabile casuale discreta X si dice avere *distribuzione di Poisson* di parametro λ se la sua distribuzione di probabilità $f(x) = P(X = x)$ è data da

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots,$$

dove λ è un parametro reale maggiore di 0. Un variabile casuale con questa distribuzione può assumere i valori interi $0, 1, 2, \ldots$ (ovvero può assumere valori in un insieme infinito numerabile), e può essere utile per descrivere, ad esempio, il numero di arrivi in un aeroporto, il numero di chiamate ad un centralino telefonico e così via. Brevemente, per indicare che una variabile casuale X ha distribuzione di Poisson si scrive $X \sim$ Poisson(λ). Relativamente alla distribuzione di Poisson si possono richiamare i seguenti risultati.

- Il valore atteso e la varianza di una variabile casuale X con distribuzione di Poisson coincidono e, in particolare, si ha $\mathrm{E}(X) = \mathrm{Var}(X) = \lambda$.
- La somma di m variabili casuali indipendenti aventi distribuzione di Poisson di parametri, rispettivamente, $\lambda_1, \ldots, \lambda_m$ ha distribuzione di Poisson di parametro $\lambda_1 + \cdots + \lambda_m$.
- Sia $X \sim$ Binomiale(n, p). Se $n \to \infty$ e $p \to 0$, in modo che np rimanga costante, allora la distribuzione della variabile casuale X tende alla distribuzione di Poisson con parametro $\lambda = np$.

Distribuzione Geometrica. Una variabile casuale discreta X si dice avere *distribuzione geometrica* di parametro p se la sua distribuzione di probabilità $f(x) = P(X = x)$ è data da

$$f(x) = p(1 - p)^{x-1}, \qquad x = 1, 2, \ldots,$$

dove $0 < p < 1$. La distribuzione geometrica può essere utilizzata per descrivere la probabilità del numero di prove necessarie affinchè si verifichi per la prima volta un successo, in una sequenza di prove Bernoulliane. In questo caso, p rappresenta la probabilità che si verifichi un successo in una singola prova. Brevemente, per indicare che una variabile casuale X ha distribuzione geometrica si scrive $X \sim$ Geometrica(p). Relativamente ad una variabile casuale X avente distribuzione geometrica si possono richiamare i seguenti risultati.

- Il valore atteso di X è pari a $\mathrm{E}(X) = 1/p$.
- La varianza di X è pari a $\mathrm{Var}(X) = (1 - p)/p^2$.
- Vale la seguente relazione tra probabilità condizionata e probabilità marginale

  $$P(X = x + a | X > a) = P(X = x),$$

  dove $a = 1, 2, \ldots$. Questa relazione è nota come la proprietà di *assenza di memoria*.

Distribuzione Normale. Una variabile casuale continua X si dice avere *distribuzione Normale* (o *distribuzione Gaussiana*) se la sua funzione di densità di probabilità è data da

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \qquad -\infty < x < \infty,$$

dove μ e σ^2 sono due parametri fissati, il primo con valori $-\infty < \mu < \infty$ ed il secondo con valori $\sigma^2 > 0$. Spesso questa densità si scrive utilizzando una scrittura tipograficamente diversa ma matematicamente equivalente, ovvero si scrive come

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\}, \qquad -\infty < x < \infty.$$

Una variabile casuale con questa distribuzione assume valori sull'intero asse reale. Brevemente, per indicare che una variabile casuale ha distribuzione Normale si scrive $X \sim N(\mu, \sigma^2)$. La distribuzione Normale è di gran lunga la più importante distribuzione, sia da un punto di vista teorico che pratico, in quanto approssima la distribuzione di moltissimi caratteri quantitativi che si riscontrano nella realtà. Relativamente alla distribuzione Normale, si richiamano i seguenti risultati.

- Una variabile casuale X con distribuzione Normale ha valore atteso $\mathrm{E}(X) = \mu$.
- Una variabile casuale X con distribuzione Normale ha varianza $\mathrm{Var}(X) = \sigma^2$.
- Data una variabile casuale X con distribuzione $N(\mu, \sigma^2)$, la variabile standardizzata $Z = (X - \mu)/\sigma$ ha distribuzione $N(0, 1)$. Essendo analiticamente impossibile risolvere l'integrale della distribuzione Normale, i valori della funzione di ripartizione della distribuzione Normale standardizzata sono ottenuti per via numerica e resi disponibili per mezzo di tavole.

Simulazioni Monte Carlo. I metodi di simulazione Monte Carlo forniscono un potentissimo strumento per la risoluzione di problemi complessi. Fondamentalmente, si può sempre vedere una simulazione Monte Carlo come un metodo per approssimare un valore atteso attraverso una semplice media aritmetica. Tratteremo questo argomento implementando alcuni semplici esempi con il software R.

Esercizio 1 (distribuzione normale)
Simulare 50 realizzazioni da una distribuzione normale di media 5 e varianza 9 e, quindi, per il campione ottenuto:
a) calcolare la media e la varianza;
b) costruire l'istogramma di frequenza;
c) costruire la funzione di ripartizione empirica;
d) paragonare la funzione di ripartizione empirica con la corrispondente funzione di ripartizione teorica;
e) trovare i percentili corrispondenti a 0.08 e 0.20.
Simulare 10000 realizzazioni da una distribuzione normale di media 5 e varianza 9 e, quindi, per il campione ottenuto, ripetere i punti a), b), c), d) ed e).

Esercizio 2 (distribuzione uniforme)
Simulare 10000 realizzazioni da una distribuzione uniforme nell'intervallo $(2;\ 5)$ e, quindi, per il campione ottenuto:
a) calcolare la media e la varianza;

b) costruire l'istogramma di frequenza;
c) costruire la funzione di ripartizione empirica;
d) paragonare la funzione di ripartizione empirica con la corrispondente funzione di ripartizione teorica;
e) trovare i percentili corrispondenti a 0.08 e 0.20.

Esercizio 3 (trasformate di variabili casuali)
Simulare 10000 realizzazioni dalle seguenti tre variabili casuali:
i) $Y_1 = u_1(X_1) = a \exp(X) + b$;
ii) $Y_2 = U_2(X) = c \cdot [1 - \exp(-X/d)]$;
iii) $Y_3 = u_3(X) = X - (k/2) \cdot X^2$;
dove $a = 1/3$, $b = 4$, $c = 4$, $d = 2$ e $k = 2$, ed X una variabile casuale con distribuzione normale di media 5 e varianza 9. Quindi, per i tre campioni ottenuti:
a) calcolare la media e la varianza;
b) costruire l'istogramma di frequenza;
c) costruire la funzione di ripartizione empirica.

Esercizio 4 (trasformata integrale di probabilità)
Simulare 10000 realizzazioni da una variabile casuale Y con distribuzione esponenziale di parametro $\lambda = 1$, con funzione di densità $f(y) = \lambda e^{-\lambda y}$, $y \geq 0$, e funzione di ripartizione

$$F(y) = 1 - e^{-\lambda y},$$

utilizzando il metodo della trasformazione integrale di probabilità. (Se U è una variabile casuale uniformemente distribuita nell'intervallo $(0;\ 1)$, e $F(y)$ è una funzione di ripartizione continua, allora $Y = F^{-1}(U)$ è una variabile casuale continua con funzione di ripartizione $F(y)$. Perciò, nel caso della distribuzione esponenziale, basta considerare la trasformata $Y = F^{-1}(U) = -(1/\lambda)\ln(1-U)$, dove U è una variabile casuale uniformemente distribuita nell'intervallo $(0;\ 1)$.)
Sulla base dei 10000 valori simulati:
a) verificare che $\mathrm{E}(Y) = 1/\lambda$ e $\mathrm{Var}(Y) = 1/\lambda^2$;
b) costruire l'istogramma di frequenza;
c) costruire la funzione di ripartizione empirica.

Soluzione Monte Carlo

```
t0 <- proc.time()
prove <- 100000; lambda <- 1.0; u <- 0; y <- 0

for (i in 1:prove) {
  u[i] <- runif(1, min=0, max=1)
  y[i] <- -(1/lambda)*log(1-u[i])
}

umedia <- mean(u); umedia
ymedia <- mean(y); ymedia
yvar <- var(y)*((prove-1)/prove); yvar
# R fornisce la varianza campionaria (dividendo per n-1)
ysqm <- sqrt(yvar); ysqm

t1 <- proc.time() - t0; t1

hist(u, 20)
```

```
hist(u, 20, prob=TRUE)
plot(ecdf(u),do.points=FALSE,verticals=TRUE)

max(y)
hist(y, seq(0, 13, 0.5), prob=TRUE)
lines(density(y,bw=0.1))
rug(y)

plot(ecdf(y),do.points=FALSE,verticals=TRUE)
x <- seq(0, 13, 0.05)
lines(x, pexp(x, rate=lambda), lty=3)
```

Esercizio 5 (distribuzione normale bivariata)
Simulare 1000 realizzazioni da ciascuna delle due variabili casuali X_1 e X_2, ciascuna con distribuzione normale standard. Per il campione di 1000 coppie $(x_{i1,x_{i2}})$, $i = 1, \ldots, 1000$:
a) calcolare la media e la varianza per ciascuna delle due variabili;
b) costruire l'istogramma di frequenza per ciascuna delle due variabili;
c) costruire il grafico a dispersione delle coppie di punti $(x_{i1,x_{i2}})$, $i = 1, \ldots, 1000$;
d) calcolare la covarianza ed il coefficiente di correlazione tra le due variabili.
Simulare 1000 realizzazioni dalla variabile casuale doppia (Y_1, Y_2) definita da:

$$Y_1 = 5 + 2X_1, \qquad Y_2 = 2 + 4,5X_1 + 4X_2,$$

e quindi ripetere i punti a), b), c) e d). (Verificare che $\mathrm{E}(Y_1) = 5$, $\mathrm{E}(Y_2) = 2$, $\mathrm{Var}(Y_1) = 4$, $\mathrm{Var}(Y_2) = 36,25$, $\mathrm{Cov}(Y_1, Y_2) = 9$ e $\mathrm{Corr}(Y_1, Y_2) = 0.7474$.)

Esercizio 6 (pianificazione di progetto)
Si consideri la pianificazione di un progetto che consta di tre fasi: analisi di fattibilità, progettazione del prototipo; realizzazione del prototipo. Si supponga che la prima fase abbia una probabilità del 99% di essere superata, che la seconda fase abbia una probabilità del 95% di essere superata, e che la terza fase abbia una probabilità del 92% di essere superata. Si supponga inoltre che, ammesso che venga superata, una fase richiede un certo numero di giorni per essere portata a termine con successo. In particolare, si assuma che il tempo necessario per portare a termine la prima fase possa essere descritto da una variabile casuale continua T_1 avente distribuzione uniforme nell'intervallo $(2;\ 8)$, che l'analogo tempo per la seconda fase possa essere descritto da una variabile casuale continua T_2 con distribuzione uniforme nell'intervallo $(3;\ 10)$, e che per la terza fase l'analogo tempo T_3 abbia distribuzione uniforme nell'intervallo $(15;\ 30)$.
a) Qual'è la probabilità che tutte e tre le fasi del progetto vengano portate a termine con successo?
b) Assumendo che il progetto venga portato a termine con successo, qual'è il numero atteso di giorni necessario per espletare tutte e tre le fasi, e qual'è, più in generale, la distribuzione del numero di giorni necessario per espletare tutte e tre le fasi?
c) Assumendo che il progetto venga portato a termine con successo, qual'è la probabilità che il progetto venga portato a termine entro 30 giorni, ed entro quanti giorni abbiamo il 95% di probabilità che il progetto venga terminato?

Esercizio 7 (Bonus Malus)
Un soggetto si assicura per due anni con la compagnia di assicurazioni Reliable Insurance Ltd. secondo il seguente schema di punteggio bonus/malus: i) per il primo anno il punteggio è pari a 10; ii)

se in un dato anno il soggetto causa almeno un sinistro, il punteggio per l'anno successivo aumenta di due punti; iii) se in un dato anno il soggetto non causa nessun sinistro, il punteggio per l'anno successivo diminuisce di un punto. Si assuma che la probabilità che in un dato anno il soggetto causi almeno un sinistro sia pari a 0.01, e che in questo caso l'ammontare complessivamente richiesto alla compagnia per risarcimento danni per quell'anno possa essere descritto da una variabile casuale continua con distribuzione uniforme nell'intervallo $(0;\ 10000)$. Si assuma inoltre che il premio annuo da pagare alla compagnia sia di 10 euro per ogni punto di bonus/malus.
a) Si disegni il diagramma ad albero che rappresenta i possibili sviluppi del contratto di assicurazione e determinare le probabilità associate ad ognuno degli eventi elementari dello spazio campionario (cioè ad ognuno dei nodi terminali del diagramma ad albero).
b) Si determinino la distribuzione di probabilità del punteggio di bonus/malus alla fine del secondo anno, e la distribuzione di probabilità del premio che si dovrebbe pagare per il terzo anno.
c) Relativamente ai due anni assicurati, calcolare la probabilità che la compagnia subisca una perdita.
d) Relativamente ai due anni assicurati, determinare il valore atteso della differenza tra i premi complessivamente incassati dalla compagnia e gli ammontari complessivamente richiesti per risarcimento danni.
e) Qual'è la probabilità che la compagnia subisca una perdita maggiore di 1000 euro? Relativamente ai due anni assicurati, qual'è il *Value at Risk* (VaR) all'1% per la compagnia di assicurazioni?
f) Si ripetano i punti a), b), c), d) ed e) assumendo che la probabilità che in un dato anno il soggetto causi almeno un sinistro sia pari a 0.2, e che in questo caso l'ammontare complessivamente richiesto alla compagnia per risarcimento danni per quell'anno possa essere descritto da una variabile casuale continua con distribuzione esponenziale di parametro $\lambda = 0.006$.

Soluzione Monte Carlo

```
t0 <- proc.time()
prove <- 100000; sinistro <- c("NO","SI"); BManno3 <- 0;
premioanno3 <- 0; premi2anni <- 0; rimborsi <- 0; differenza <- 0
# lambda <- 0.006

for (i in 1:prove) {
  anno1 <- sample(sinistro, size=1, prob=c(0.99,0.01))
  if (anno1=="NO")
    anno2 <- sample(sinistro, size=1, prob=c(0.99,0.01))
  if (anno1=="SI")
    anno2 <- sample(sinistro, size=1, prob=c(0.99,0.01))

  BManno3[i] <- 10+2*(anno1=="SI")-(anno1=="NO")+2*(anno2=="SI")-(anno2=="NO")
  premioanno3[i] <- 10*BManno3[i]
  premi2anni[i] <- 100+10*(10+2*(anno1=="SI")-(anno1=="NO"))

  rimborsi[i] <- runif(1,min=0,max=10000)*(anno1=="SI")
              +runif(1,min=0,max=10000)*(anno2=="SI") # uniforme
  # rimborsi[i] <- rexp(1,rate=lambda)*(anno1=="SI")
  #               +rexp(1,rate=lambda)*(anno2=="SI") # esponenziale

  differenza[i] <- premi2anni[i] - rimborsi[i]
}

differenzamedia <- mean(differenza); differenzamedia
vardifferenza <- var(differenza)*((prove-1)/prove); vardifferenza
  # R fornisce la varianza campionaria (dividendo per n-1)
sqmdifferenza <- sqrt(vardifferenza); sqmdifferenza

hist(differenza, 20)
```

```
quantile(differenza, probs=c(0, 0.0001, 0.001, 0.01, 0.0195, 0.02, 0.5, 1))

sum(differenza < 0); prove
probperdita <- sum(differenza < 0)/prove; probperdita

sum(differenza < -1000); prove
probperdita1000 <- sum(differenza < -1000)/prove; probperdita1000
t1 <- proc.time() - t0; t1
```

Esercizio 8 (incasso giornaliero)
In un dato giorno, sia X_i l'ammontare complessivo di euro spesi dal cliente i-esimo in un certo negozio, e si assuma che X_i abbia distribuzione Esponenziale di parametro $\lambda = 0.04$ (con densità $f(x) = \lambda e^{-\lambda x}$, per $x > 0$, con valore atteso $1/\lambda$ e varianza $1/\lambda^2$) e che X_i sia indipendente da X_j, per $i \neq j$, dove X_j rappresenta la spesa del cliente j-esimo. Si assuma inoltre che in un dato giorno il numero N di clienti che entrano nel negozio si possa descrivere da una variabile casuale con distribuzione Geometrica di parametro $p = 0.03$ (con distribuzione $f_N(n) = p(1-p)^{n-1}$, per $n = 1, 2, \ldots$, con valore atteso $1/p$ e varianza $(1-p)/p^2$) e che N sia indipendente dalle X_i. Si consideri quindi la variabile casuale $Y = \sum_{i=1}^{N} X_i$ che rappresenta l'incasso giornaliero.
a) Si determinino il valore atteso e la varianza dell'incasso giornaliero Y.
b) Si determini la distribuzione dell'incasso giornaliero Y (si noti che la somma di n variabili casuali indipendenti con distribuzione Esponenziale di parametro λ, ha distribuzione Gamma di parametri n e λ, con densità $f(y) = (1/\Gamma(n))\lambda^n y^{n-1} e^{-\lambda y}$, dove per la funzione Gamma si ha che $\Gamma(n) = (n-1)!$).
c) Si determini la probabilità che l'incasso giornaliero sia superiore a 500 euro.

Soluzione Monte Carlo

```
t0 <- proc.time()
prove <- 100000; prob <- 0.03; lambda <- 0.04; incasso <- 0

for (i in 1:prove) {
  numclienti <- rgeom(1, prob) + 1
  scontrini <- rexp(numclienti, lambda)
  incasso[i] <- sum(scontrini)
}

incassomedio <- mean(incasso); incassomedio
varincasso <- var(incasso)*((prove-1)/prove); varincasso
  # R fornisce la varianza campionaria (dividendo per n-1)
sqmincasso <- sqrt(varincasso); sqmincasso

incasso500 <- incasso>500;
sum(incasso500)
prob500 <- sum(incasso500)/prove; prob500

hist(incasso, 20)
t1 <- proc.time() - t0; t1
```

Esercizio 9 (strategia d'investimento)
Un investitore acquista un'azione della società K-WAY Inc. ad un prezzo di 8 euro e controlla il prezzo dell'azione dopo la fase d'asta di apertura dei successivi quattro giorni (assumiamo che dopo

la fase d'asta di apertura il prezzo rimanga costante per tutto il giorno). Si supponga che il prezzo dell'azione possa essere pari a 1.2 volte il prezzo del giorno precedente con probabilità pari a 0.3 o essere pari a 0.8 volte il prezzo del giorno precedente con probabilità pari a 0.7. Si assuma inoltre la seguente strategia da parte dell'investitore. L'investitore decide di: acquistare una nuova azione ogni volta che il prezzo è compreso tra i 6 e gli 8 euro; vendere tutte le azioni in suo possesso non appena si verifica il secondo aumento del prezzo (anche non consecutivo); vendere tutte le azioni in suo possesso non appena il prezzo scende sotto i 6 euro; vendere (in qualsiasi circostanza) tutte le azioni in suo possesso al quarto giorno; mantenere le azioni in tutte le altre circostanze.
a) Si disegni il diagramma ad albero che rappresenta i possibili comportamenti dell'investitore e determinare le probabilità associate ad ognuno degli eventi elementari dello spazio campionario (cioè ad ognuno dei nodi terminali del diagramma ad albero).
b) Sia X la variabile casuale che rappresenta il prezzo di un'azione al momento della vendita e sia Y la variabile casuale che rappresenta il valore complessivo (ricavo) delle azioni in possesso dell'investitore al momento della vendita. Si calcolino $\mathrm{E}(X^2)$ e $\mathrm{E}(Y)$.
c) Calcolare la probabilità che con questa strategia di investimento l'investitore subisca una perdita (considerando costi e ricavi).
d) Considerando che l'investitore deve convertire in dollari (1 euro = 1.35 dollari) il ricavato dell'investimento, calcolare il valore atteso dell'ammontare di dollari in possesso dell'investitore dopo aver rivenduto le azioni.