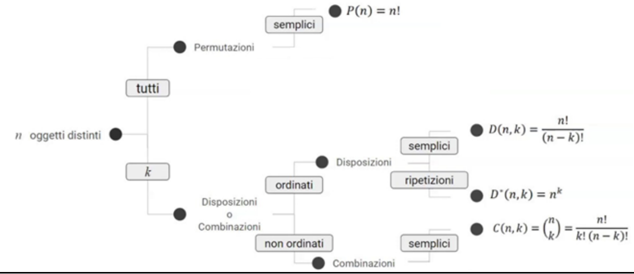


$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \Rightarrow \mathbb{P}[A \cap B] = \mathbb{P}[A|B]\mathbb{P}[B]$ $\quad \mathbb{P}[\bar{A}|B] = \frac{\mathbb{P}[\bar{A}\cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B]-\mathbb{P}[A\cap B]}{\mathbb{P}[B]} = 1 - \mathbb{P}[A|B]$

Serie $\mathbb{P}[A] = \prod_{i=1}^n \mathbb{P}[A_i] = \prod_{i=1}^n (1-p_i)$

Parallelo $\mathbb{P}[B] = 1 - \mathbb{P}[\bar{B}] = 1 - \prod_{i=1}^n \mathbb{P}[\bar{A}_i] = \prod_{i=1}^n (p_i)$

Legge della probabilità Totale $\mathbb{P}[A] = \sum_i^n \mathbb{P}[A \cap C_i] = \sum_i^n \mathbb{P}[C_i]\mathbb{P}[A|C_i]$

Bayes $\mathbb{P}[\boldsymbol{C_m}|\boldsymbol{A}] = \frac{\mathbb{P}[A|C_m]\mathbb{P}[C_m]}{\sum_i^n \mathbb{P}[C_i]\mathbb{P}[\boldsymbol{A}|\boldsymbol{C_i}]}$

Varianza $Var[X] = \mathbb{E}[X^2] - [\mathbb{E}[X]]^2$

Moda è il punto (o i punti) in cui la funzione di probabilità (per variabili aleatorie discrete) o la funzione di densità (per variabili aleatorie continue) assume il valore massimo.

Quantile di livello α è il minimo valore per cui $F(m) = \mathbb{P}[X \le m] \ge \frac{1}{2}$

Variabile aleatoria DISCRETA

Funzione di ripartizione (cumulative function) $F(x) = \mathbb{P}[X \le x]$

Funzione di probabilità

$\mathbb{P}[X = x_i]$

<u>Dalla funzione di ripartizione alla funzione di probabilità</u>

$\xrightarrow{\mathbb{F}(x)-\mathbb{F}(x^-)} \mathbb{P}[X = x_i]$

Media

$$\mathbb{E}[X] = \sum_i x_i p_i$$

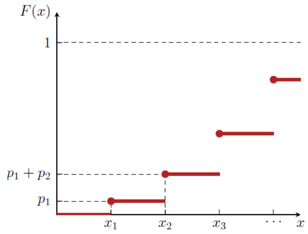
Varianza

$$Var[X] = \sum_i x^2 p_i - [\mathbb{E}[X]]^2$$

In particolare

$Var[a] = 0$

$Var[aX+b] = a^2 Var[X]$



Variabile aleatoria CONTINUA

Funzione di ripartizione (cumulative function) $F(x) = \mathbb{P}[X \le x]$

Funzione di densità

$$\mathbb{P}[X \in A] \Leftrightarrow \int_A f(x) \Leftrightarrow \int_A f(x) = 1$$

<u>Dalla funzione di ripartizione alla funzione di densità</u>

$\xrightarrow{\mathbb{D}(F(x))} \mathbb{P}[X \in A]$

Media

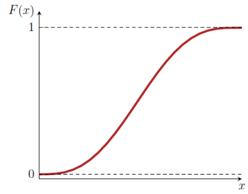
$\mathbb{E}[X] = \int_{\mathbb{R}} x \cdot f(x)\ dx$

In particolare

$\mathbb{E}[a] = a \quad \mathbb{E}[aX+b] = a\mathbb{E}[X] + b \quad \mathbb{E}[X - \mathbb{E}[X]] = 0$

Varianza

$$Var[X] = \int_{\mathbb{R}} x^2 f(x)\,dx - [\mathbb{E}[X]]^2$$



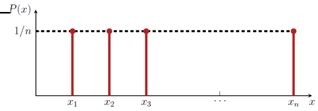
Valore atteso di una funzione di una v.a. $Y = g(X)$ con X v.a. che assume valori x_i con probabilità $p(x_i)$ <u>Discreta:</u> $\mathbb{E}[Y] = \sum_i g(x_i)p(x_i)$ <u>Continua:</u> $Var[X] = \int_{\mathbb{R}} g(x)f(x)\,dx$

Famiglie di variabili aleatorie DISCRETE

Distribuzione Discreta Uniforme

$X \sim U\{x_i, \dots, x_n\}$ con $p_i = \frac{1}{n}$

`dunif(x, min = 0, max = 1)`



Distribuzione Ipergeometrica

$X \sim Ig(N, K, n)$

Dove:

$N = n°\ di\ elementi\ totali$

$n = numero\ di\ estrazioni\ alla\ volta$

$K = sottoinsieme\ di\ N\ con\ la\ caratteristica\ cercata$

$k = n°\ di\ succcessi$

$$\mathbb{P}[X = k] = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \qquad \mathbb{E}[X] = n\frac{K}{N} \qquad Var[X] = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$$

`dhyper(x, m, n, k)` con $m = K, x = k, n = N - K, k = n, (k - x = n - k, m + n = N)$

Fare particolare attenzione alla richiesta: almeno, esattamente, più di…

Distribuzione Binomiale

$X \sim Bin(n, p)$

$$\mathbb{P}[X = x] = \binom{n}{x} p^x (1-p)^{n-x} \qquad \mathbb{E}[X] = np \qquad Var[X] = np(1-p)$$

`dbinom(x, size, prob)` con $x = vect, size = \#prove, prob = prob.\ success$ di ogni prova

Uso di R

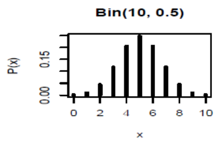
$\mathbb{P}[\boldsymbol{X} = \boldsymbol{x}]$ = `dbinom(x,n,p)`

$\mathbb{P}[\boldsymbol{X} \le \boldsymbol{x}]$ = `pbinom(x,n,p)`

$\mathbb{P}[\boldsymbol{X} \ge \boldsymbol{x}]$ = `1-pbinom(x-1,n,p)`

$\mathbb{P}[\boldsymbol{X} < \boldsymbol{x}]$ = `pbinom(x-1,n,p)`

$\mathbb{P}[\boldsymbol{X} > \boldsymbol{x}]$ = `1-pbinom(x,n,p)`



Distribuzione Bernoulliana

$X \sim Ber(p)$

$$\mathbb{P}[X = x] = p^x(1-p)^{1-x} \qquad \mathbb{E}[X] = p \qquad Var[X] = p(1-p)$$

Distribuzione di Poisson

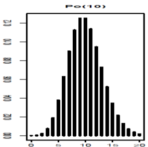
$X \sim Po(\lambda)$ `dpois(x, lambda)`

$$\mathbb{P}[X = k] = \frac{\lambda^k}{k!}e^{-\lambda} \qquad \mathbb{E}[X] = \lambda \qquad Var[X] = \lambda$$

Dove:

$k = eventi\ nell'arco\ di\ tempo$

È usata come modello per il conteggio di manifestazioni di un certo fenomeno d'interesse.



Approssimazione Poisson per la Binomiale

Se la Bin ha $\underbrace{n \to \infty}_{n \ge 100}$ e $\underbrace{p \to 0}_{0 \le p \le 0.05}$ allora si sceglie $\lambda = \mathbb{E}[X] = np$. "Legge degli eventi rari".

Distribuzione Geometrica

$X \sim Geo(p)$ `dgeom(x, prob)`

$$\mathbb{P}[X = x] = (1-p)^{x-1}p \qquad \mathbb{E}[X] = \frac{1}{p} \qquad Var[X] = \frac{1-p}{p^2}$$

Mancanza di memoria $\qquad$ *Funz. di sopravvivenza*

$$\mathbb{P}[X > m + n \mid X > m] = \frac{\mathbb{P}[X > m + n]}{\mathbb{P}[X > m]} = \mathbb{P}[X > n] \qquad \mathbb{P}[X > k] = (1-p)^k$$

Conta il numero di ripetizioni indipendenti per osservare il primo successo di un esperimento binario.

$\boldsymbol{p-, d-, r-, q-}$ in R

p per definire la funzione di ripartizione

d per definire la funzione di densità

r per definire un risultato di n lanci con m variabili aleatorie

q per definire il quantile di livello α ed è spesso usato con la normale

Famiglie di variabili aleatorie CONTINUE

$Mediana = \mathbb{F}(x) = \frac{1}{2}$

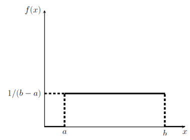
Distribuzione Continua Uniforme

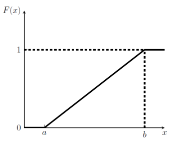
$X \sim U(a, b)$

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a,b) \\ 0, & altrove \end{cases}$$

$$\mathbb{F}(x) = \int_a^b \frac{1}{b-a}dx = \frac{x-a}{b-a}$$

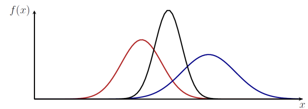
$$\mathbb{E}[X] = \frac{a+b}{2} \qquad Var[X] = \frac{(b-a)^2}{12}$$





Distribuzione Normale

$X \sim N(\mu, \sigma^2)$ `dnorm(x, mean = 0, sd = 1)`



$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0$$

$\mathbb{E}[X] = \mu \qquad Var[X] = \sigma^2$

Standardizzazione

$$X \sim N(0,1) \Leftrightarrow Z = \frac{X-\mu}{\sigma}$$

$$\mathbb{P}[a < X < b] = \mathbb{P}\left[\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Uso della tavola

$\mathbb{P}[Z \le z] = \Phi(z)$

$\mathbb{P}[Z < -z] = 1 - \mathbb{P}[Z \le z] = 1 - \Phi(z)$

$\mathbb{P}[Z \ge z] = 1 - \mathbb{P}[Z < z] = 1 - 1 - \Phi(z)$

$\mathbb{P}[Z \ge -z] = \mathbb{P}[Z \le z] = \Phi(z)$

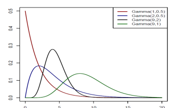
Approssimazione Normale per la Binomiale

Se la Binomiale ha n grande allora: $Bin(n, p) \approx N(\underbrace{np}_{\mu}, \overbrace{\underbrace{np(1-p)}_{\sigma^2}}^{\ge 10})$

Distribuzione Gamma

$X \sim Ga(\alpha, \lambda)$ `pgamma(q, rate = α, shape = λ)` o `pgamma(q, scale =` $^1/_\alpha$`, shape = λ)`

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, & x \in (0, \infty) \\ 0, & altrove \end{cases}$$



Dove:

$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx = (\alpha-1)!\ \ \lambda = arco\ di\ tempo;\ \ \alpha = step\ indipendenti$

$$\mathbb{E}[X] = \frac{\alpha}{\lambda} \qquad Var[X] = \frac{\alpha}{\lambda^2}$$

$Se\ \alpha = 1 \rightarrow\ X \sim Exp(\lambda)$

Distribuzione Esponenziale

$X \sim Exp(\lambda)$ `dexp(x, rate = λ)`

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \in (0, \infty) \\ 0, & altrove \end{cases} \qquad \mathbb{F}(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \ge 0 \end{cases}$$

$$\mathbb{E}[X] = \frac{1}{\lambda} \qquad Var[X] = \frac{1}{\lambda^2}$$

Dove: $\lambda = rate = tasso\ di\ arrivi$ $\qquad$ Per il tempo d'attesa.

Mancanza di memoria

$\mathbb{P}[X > t + a \mid X > t] = \frac{\mathbb{P}[X > t + s]}{\mathbb{P}[X > t]} = \mathbb{P}[X > s]$ infatti $\mathbb{P}[X > x] = e^{-\lambda x}, x > 0$

Processo di Poisson

$X_t \sim P(\lambda, t)$

Successione di variabili dipendenti da t. Contare il numero di manifestazioni in un intervallo di tempo t diverso da quello standard.

$X_t \sim P(\lambda, t) \approx T \sim Exp(\lambda)$ Tempo che passa fra l'arrivo di due z successivi

<u>*Trasformazioni di v.a.*</u> DISCRETE

$Y = g(x), X \sim F(x)$

$$P_Y(y) = \sum_{x_i : g(x_i) = y} P_X(x_i)$$

Trasformazioni di v.a. CONTINUE

$Y = g(x), X \sim f_X(x)$ Biunivoca e derivabile altrimenti somma di tratti biunivoci e derivabili

$$Y \sim f_Y(y) = f_X(g^{-1}(y))\left|\frac{dg^{-1}(y)}{dy}\right|$$

Trasformazione con Metodo della funzione di Ripartizione

$Y \in D_Y : F_Y(y) = \mathbb{P}[Y \le y]$
$= \mathbb{P}[g(X) \le y] = \mathbb{P}[X \in I(y)]$

Simulazione Distribuzione Poisson

$Y_1, Y_n \sim Exp(\lambda)\ \ S_n = \sum_{i=1}^n Y_i\ \ N_t = n\ \ k = \min\{n: \sum_{i=1}^n Y_i > \lambda\}$, allora il valore simulato da $Po(\lambda)$ sarà $x = k - 1$.

<u>*SIMULAZIONI DI VARIABILI ALEATORIE*</u>

Metodo dell'Inversione CONTINUE

$U = F(x) \sim U(0,1) \qquad x = F^{-1}(u)$

$U(0, kg(\hat{x}))$ o $U(0,1)$; 3. Se $u \le f(\hat{x})$ o se

Es.: $f(x) = (1 - |1 - x|1_{(0,2)}(x)$, scegli

In R:
```
u1 <- runif (1000, 0, 2)
x <- u1[ u2 < (1-abs(1-u1))]
u2 <- runif(1000)
```

Metodo accettazione/rifiuto

1. Si genera $\hat{x}$ da $g(\cdot)$; 2. Si genera u da $u \le (f(\hat{x}))/kg(\hat{x})$ allora si accetta $\hat{x}$.

$g(x) = 1/2$, la densità di $U(0,2)$ e $k = 2$.

Scegliere k basso così serve un numero meno elevato di generazione di $g(\cdot)$

Legge (forte) dei grandi numeri Se si ha una sequenza di v.a. i.i.d. con la stessa μ e la stessa σ^2 finita, allora all'aumentare della numerosità campionaria, n, la media campionaria converge (q.c.) a μ.

Monte Carlo $\theta = \int_a^b g(x)$ $\quad \Theta = \frac{1}{n}\sum_{i=1}^{n} g(x_i \xrightarrow{gen.} U(a,b)) = p$ `mean(rDistr(nProve,a,b)<=estrP)` $\hat{p} = \frac{\#\{x_i \in A\}}{n} =$ `mean(x>5)` con `x=rDistr(...)` $\widehat{Var} = \frac{\hat{p}(1-\hat{p})}{n}$

DISTRIBUZIONI CONGIUNTE	**Distribuzioni congiunte DISCRETE**	**Distribuzioni congiunte CONTINUE**
Ripartizione congiunta $F(x,y) = \mathbb{P}[X \le x, Y \le y]$ **Ripartizione singola** $F_x = \mathbb{P}[X \le x, Y < \infty]$ **Indipendenza** $p/f(x,y) = p/f_X(x) \cdot p/f_Y(y)$ **Condizionata** $p/f_{X\|Y} = \frac{p/f(x,y)}{p/f_Y(y)}$ $\; p/f_{Y\|X} = \frac{p/f(x,y)}{p/f_X(x)}$	**Probabilità congiunta** $p(x,y) = \mathbb{P}[X = x, Y = y]$ (tableX2) **Probabilità marginale** $p_X(x) = \sum_{y:p(x,y)>0} p(x,y)$ **Valore atteso** $\mathbb{E}[g(X,Y)] = \sum_y \sum_x g(x,y)p(x,y)$	**Densità congiunta** $\mathbb{P}[X \in A, Y \in B] = \int_B \left(\int_A f(x,y)dx\right)dy$ **Densità marginale** $f_X(x) = \int_{\mathbb{R}} f(x,y)dy$ **Valore atteso** $\mathbb{E}[g(X,Y)] = \iint_{\mathbb{R}^2} g(x,y)f(x,y)dxdy$

$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ *Se indip.* $\mathbb{E}[X \cdot Y] = \mathbb{E}[X]\mathbb{E}[Y]$ $\quad Cov[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ Se $Cov[X,Y] = 0$ X e Y incorrelate $\quad$ Se X e Y indipendenti $\rightarrow Cov[X,Y] = 0$ (non vale il contrario)

Covarianza $Cov[X,Y] = Cov[Y,X]$ $\quad Cov[X,X] = Var[X]$ $\quad Cov[aX,Y] = aCov[X,Y]$ $\quad Cov[X,a] = 0$ $\quad Cov[\sum_i X_i, \sum_j Y_j] = \sum_i \sum_j Cov[X_i, Y_j]$ $\quad$ Se X_i sono a 2 a 2 indipendenti allora $Cov[X_i, X_j] = 0 \; con \; i \neq j$ e $Var[\sum_i X_i] = \sum_i Var[X_i]$

Correlazione $Cor[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X] \cdot Var[Y]}}$ $\; -1 \le Cor[X,Y] \le 1$

Somma di v.a. INDIPENDENTI	**Media campionaria**	**Disuguaglianza di Chebyshev**	**Convergenza di variabili aleatorie**
n v.a. $Bin(1,p) = Ber(p) = Bin(n,p)$ n v.a. $Po(\lambda_i) = Po(\sum_i \lambda_i)$ n v.a. $Exp(\lambda) = Ga(n,\lambda)$ n v.a. $N(\mu_i, {\sigma^2}_i) = N(\sum_i \mu_i, \sum_i {\sigma^2}_i)$	$\overline{X_n} = \frac{\sum_i X_i}{n}$ $\mathbb{E}[\overline{X_n}] = \mu$ $Var[\overline{X_n}] = \sigma^2/n$	Sia X una variabile aleatoria con valore atteso $\mathbb{E}[X]$ e varianza $Var[X] < \infty$, allora $\forall \varepsilon > 0$ **Limite superiore** $\mathbb{P}[\|X - \mathbb{E}[X]\| > \varepsilon] \le Var[X]/\varepsilon^2$ **Limite inferiore** $\mathbb{P}[\|X - \mathbb{E}[X]\| < \varepsilon] \ge 1 - \left(\frac{Var[X]}{n}\right)/\varepsilon^2$	Sia $X_1, \dots, X_n$ sequenza di v.a. con f.r. F_n e X una v.a. con f.r. F. $X_n \xrightarrow{p} X$ se $\mathbb{P}[\|X_n - X\| > \varepsilon] \rightarrow 0, \forall \varepsilon > 0$ $X_n \xrightarrow{d} X$ se in ogni punto di x di continuità per F $F_n \rightarrow F(x)$ $X_n \rightarrow X \; q.c.$ se $\mathbb{P}[X_n \rightarrow X] = 1$

Teorema del limite centrale

Sia $X_1, \dots, X_n$ sequenza di v.a. i.i.d. con valore atteso $\mathbb{E}[X] = \mu$ e varianza $Var[X] = \sigma^2 < \infty$, allora:

1. Se si hanno μ e σ riferite a n oggetti: $\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$ $\quad \overline{X_n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
2. Se si hanno μ e σ riferite a 1 *solo* oggetto: $\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \rightarrow N(0,1)$ $\quad \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$

ESERCIZIO Distribuzione stazionaria (invariante)

Data la matrice di transizione P:

```
colonna = 5 #ncols
p <- matrix((c(0, 0.5, …), nrow = 5, ncol = colonna, byrow = TRUE)
```

1. Trovare P^T e sottrarre la matrice identità $(P^T - Id_n)$

```
A <- t(P) - diag(rep(1, Colonna))
```

2. Moltiplicare la matrice per il vettore degli stati $\pi = (\pi_1 \dots \pi_n)$

```
A[colonna,] <- rep(1, colonna)
```

3. Prendere le prime $(n = \#nrows) - 1$ righe e porle uguali a 0 e la n-esima equazione è la somma di tutti i π posta uguale a 1.
4. Si risolve il sistema e si ottiene il vettore della distribuzione stazionaria

```
ris <- solve(A, c(0, 0, 0, 0, 1))
ris
[1] (0 0 0 0.4 0.6)
```

Catene di Markov

$P = matrice \; di \; transizione$; $\; P^2 = matrice \; di \; transizione \; a \; due \; passi$

Distribuzione marginale $\pi^{(n)} = \pi^{(0)} \cdot P^n \; con \; \pi^{(0)} = vettore \; probabilità \; iniziali$

Passeggiata aleatoria è una catena di Markov con spazio degli stati S = $\mathbb{Z}$ che ad ogni istante si muove di un passo a dx (p) o a sx $(1-p)$. • se $p > 1/2$ il sistema $\rightarrow +\infty$; e se $p < 1/2$ il sistema $\rightarrow +\infty$ $(X_n \; diverge)$ • se $p = 1/2$ andamento meno prevedibile

Con barriere non assorbenti $S = \{-L, \dots, L\}$ quindi $(X_n \; NON \; diverge)$ rimbalzo.

Con barriere assorbenti $S = \{-L, \dots, L\}$ quindi $(X_n \; converge \; a \; \pm L)$ blocco.

Catena regolare Se esiste n per cui P^n ha tutti gli elementi strettamente positivi (> 0)

Distribuzione stazionaria $\pi P^n = \pi, \sum_{i=1}^{M} \pi_i = 1$

Catena stazionaria Con π distribuzione stazionaria iniziale di $X_0, \pi P^n = \pi, \forall n$ e tutte le distribuzioni marginali della catena sono uguali a π.

R CODE `rep(5,times=100)` ripetizioni dell'elemento per le `times` specificate
`x <- c(4,12,31,7,4,14)` crea vettore con gli elementi specificati

Combinatoria `choose(n,k)`

Campionamento `urna <- c('b1','b2','n1')`
Senza ripetizioni: `sample(urna,nEstrazion)`
Con ripetizioni: `sample(urna,nEstrazioni,replace=T)`
`do()` *del pacchetto* `mosaic` *è per le replicazioni:* `> library(mosaic)` `> k <- do(1000)*sum(sample(mazzo, 10, replace=T)=="R")` `> table(k)/1000` `> plot(table(k)/1000)`

Dado truccato con ripetizioni `sample(urna,nEstrazioni,replace=T,prob=c(0.8,rep(0.1,2))`

Distribuzioni di probabilità `> y<-c(0,1,4,7)` `> p<-c(0.1,0.2,0.5,0.2)`
`> plot(y, p, type="h", ylim=c(0,0.6), xlim=c(-1,8))` *Istogramma (h)* `> points(y, p, pch=10)` *(Punti sui bastoncini)*

Funzione di ripartizione `> Fr<-cumsum(p)` `> Fr0<-c(0, Fr, 1)` `> y0<-c(-1, y, 8)` `> plot(y0, Fr0, type="s", ylab="f. di ripartizione", xlab="y")` `> points(y, Fr, pch=20)`

Valore atteso `> mu<-sum(y*p)` *Varianza* `> s2<-sum(y^2*p)-mu^2` *Simulare valori con* `> sample(y, 10, prob=p, replace=T)`

Rappresentazione v.a. Discreta `> par(mfrow=c(1,2))` `> plot(-1:4,dbinom(-1:4,3,1/3),type="h",xlab="",ylab="binomial probs")` `> plot(-1:4,pbinom(-1:4,3,1/3),type="s",xlab="",ylab="binomial cum. probs")` `> par(mfrow=c(1,1))`

Rappresentazione v.a. Continua `> par(mfrow=c(1,2))` `> curve(dnorm(x, 5, 0.8),xlim=c(2,8),xlab="",ylab="normal density")` `> curve(pnorm(x, 5, 0.8),xlim=c(2,8),xlab="", ylab="normal cum. distr.")` `> par(mfrow=c(1,1))`

Simulazione di Poisson Simulare un valore dalla variabile X ~Po(λ=3), sfruttando la sua relazione con la variabile esponenz. `set.seed(1)` (per fissare il seme del generatore numeri casuali) `Y=0 N=0 while(sum(Y)<=1){ Y=c(Y,-log(1-runif(1))/3) N=N+1 }` `plot( cumsum(Y),0:N, type="s")` (grafico per il processo di Poisson simulato) `x = N-1` (valore simulato dalla variabile di Poisson)

Simulazione di variabili aleatorie con il metodo dell'inversione Sia X una variabile aleatoria con funzione di ripartizione $F(x) = x^2, x \in (0,1)$. a. Si scrivano alcuni comandi di R per simulare 1000 valori dalla X utilizzando il metodo di inversione. Se $F(x) = u$, allora $x = F^{-1}(u) = \sqrt{u}$. Perciò si possono generare 100 valori da un'uniforme in $(0,1)$ e poi trasformarli tramite la F^{-1} per ottenere i corrispondenti valori da X: `u<-runif(1000)` `x<-sqrt(u)` b. Si costruisca l'istogramma dei valori simulati e si sovrapponga il grafico della densità della X. Cosa si osserva? La densità della X si ottiene derivando la funzione di ripartizione: $f(x) = 2x, x \in (0,1)$. Perciò: `hist(x,prob=TRUE,main="valori simulati",ylim=c(0,2.1),xlim=c(-0.2,1.2))` `curve(2*x, xlim=c(0,1), add=TRUE)` `abline(h=0)` C'è accordo fra l'istogramma dei valori simulati e la densità della X, segno che i valori sono stati simulati correttamente.

Usando il metodo di inversione, simulate 1000 valori casuali dalla densità $f(x) = 1/x^2 I_{(1,+\infty)}(x)$ e verifica sovrapponendo all'istogramma il grafico della densità $f(x)$. La f.r. è $F(x) = (1 - 1/x)I_{(1,+\infty)}(x)$. La sua inversa $x = 1/(1-u)$ dove u è un valore da una $U(0,1)$. `y<-1/(1-runif(1000))` `hist(y,prob=T,xlim=c(0.0001,500))` `curve(1/xˆ2,xlim=c(0.0001,500),add=T)`

Simulazione di variabili aleatorie con il metodo di accettazione/rifiuto Usando il metodo di accettazione/rifiuto, generate 1000 valori casuali dalla distribuzione con densità $f(x) = 2 - 2|1-2x|, x \in (0,1)$. a. Qual è la distribuzione $g(x)$ di riferimento? E' una distribuzione uniforme in (0,1). b. Quanto vale la costante k tale che $f(x) \le kg(x)$? $k = 2$ perché 2 è il massimo valore che assume $f(x)$. `x<-runif(10000)` `y<-runif(10000,0,2)` `x<-x[y < ( 2-2*abs(1-2*x) )]` `x<-x[1:1000]` c. Rappresentate graficamente i valori simulati tramite un istogramma e si sovrapponga il grafico della densità $f(x)$. `hist(x,prob=T)` `curve(2-2*abs(1-2*x),add=T,col='red')`

Usando il metodo di accettazione/rifiuto, simulate 1000 valori casuali dalla densità $f(x) = ce^{-x^{3/2}} I_{(0,+\infty)}(x), dove \; c$ è $la \; costante \; di \; normalizzazione$. Poiché $e^{-e^{-x^{3/2}}} < 2e^{-x}$ possiamo usare l'esponenziale come densità di riferimento. `curve(exp(-xˆ(3/2)),ylim=c(0,2),xlim=c(0,4))` `curve(2*exp(-x),add=T,col="red");` `x1<-rexp(10000)` `x2<-runif(10000,0,2*dexp(x1))` `x<-x1[x2<exp(-x1ˆ(3/2))]` `hist(x,prob=T)`

Metodo di Monte Carlo: Calcolo di una probabilità Usando il metodo Monte Carlo, approssimate la probabilità che una variabile aleatoria esponenziale di media 3 assuma valori compresi fra 2 e 3. `y<-rexp(10000,1/3)` `mean(y<3 & y>2)` b. Verificate che il vostro metodo funzioni confrontando il risultato con il valore esatto della probabilità. `pexp(3,1/3)-pexp(2,1/3)`

Metodo di Monte Carlo: Calcolo di un integrale Usando il metodo Monte Carlo, approssimate il seguente integrale: $\int_1^2 \log(x^2)dx$. L'integrale si può vedere come il valore atteso della trasformazione $log(X^2)$, dove X è una variabile aleatoria uniforme in (1,2). Perciò: `x<-runif(1000,1,2)` `mean(log(xˆ2))`

Distribuzioni congiunte Possibili valori di X `x<-0:1` Possibili valori di Y `y<-0:3` Matrice di probabilità congiunte `pxy<-matrix(c(0.2, 0.2, 0.2, 0.1, 0.05, 0.1, 0.05, 0.1),nrow=2)` Marginale di X `px<-apply(pxy,1,sum)` Marginale di Y `py<-apply(pxy,2,sum)` Valori Attesi `EX <- sum(x*px)` `EY<-sum(y*py)` Varianze `VX<-sum(x^2*px)-EX^2` `VY<-sum(y^2*py)-EY^2` Covarianza `CXY <- sum(outer(x,y)*pxy)-EX*EY` Correlazione `rXY <- CXY/sqrt(VX*VY)` Distribuzione di X condizionata a Y `pxy/py` (per ogni valore di Y=y, la colonna somma 1) e di Y condizionata a X `pxy/px` (per ogni valore di X=x, la riga somma 1)

Grafici tridimensionali La funzione `contour()` produce grafici simili alle mappe topografiche (curve di livello) Ha bisogno di 3 argomenti: un vettore di valori per l'asse x, un vettore per l'asse y e una matrice contenente i valori z corrispondenti ad ogni coppia di coordinate (x, y). `# f(x, y) = cos(y)/(1 + x^2).` `x=seq(-pi,pi,length=50)` `y=x` `f=outer(x,y,function(x,y) cos(y)/(1+x^2))` `contour(x,y,f)` `contour(x,y,f,nlevels=45,add=T).` La funzione `image()` lavora in modo simile a `contour()`, ma usa colori diversi a seconda dei valori di z. Produce mappe come quelle usate per rappresentare temperature diverse nelle previsioni del tempo. `# fa(x, y) = (cos(y)/(1+x^2)-cos(x)/(1+y^2))/2 = (f(x,y)-f(y,x))/2` `fa=(f-t(f))/2` `image(x,y,fa).` Infine possiamo produrre grafici tridimensionali usando il comando `persp()`. Le opzioni theta e phi controllano gli angoli da cui si vede il grafico. `persp(x,y,fa,theta=30,phi=20)` `persp(x,y,fa,theta=30,phi=70)` `par(mfrow=c(1,1))`

La distribuzione normale bivariata La libreria "`mnormt`" fornisce le funzioni necessarie per calcolare la densità e la funzione di ripartizione della normale bivariata. Il loro uso è simile a quello delle funzioni `dnorm()` e `pnorm()`. Bisogna solo specificare in modo opportuno un vettore di dimensione 2 per la media e una matrice 2 per 2 di varianza-covarianza. `par(mfrow=c(2,2))` `x=seq(-2,4,length=50)` `y=seq(-3,7,length=50)` `mu <- c(1,2)` `Sigma <- matrix(c(1,2,2,5), 2, 2)` `f <- outer(x,y, function(x,y)dmnorm(cbind(x,y), mu, Sigma))` `contour(x,y,f,nlevels=10)` `image(x,y,f)` `persp(x,y,f)` `persp(x,y,f,theta=30,phi=40)` `par(mfrow=c(1,1))`

TLC Consideriamo n variabili Bernoulli(p) i.i.d. e facciamo vedere come all'aumentare di n la funzione di ripartizione della media campionaria standardizzata si avvicini sempre più a quella di una N(0,1). `par(mfrow=c(2,2))` `p<-0.2` `nobs<-c(10,20,100,1000)` `for (n in nobs) { y<-0:n prob<-pbinom(y,n,p) z<-(y/n-p)*sqrt(n)/sqrt(p*(1-p)) ind<-(z>-3)&( z<3) z<-z[ind] prob<-c(0,prob[ind]) plot(stepfun(z, prob, f = 0),verticals=FALSE,pch=20, main=paste("n = ",n ,", p = ", p),ylab="F(z)",xlab="z") curve(pnorm(x),from=min(z),to=max(z),add=TRUE) }` `par(mfrow=c(1,1))`

LGN Consideriamo n = 10 replicazioni da Y, una v.c. con distribuzione Poisson(5) e calcoliamone la media aritmetica (campionaria): `set.seed(30)` `y<-rpois(10,5)` `mean(y)`

Markov `P%*%P` oppure `P %^%2` La libreria "`expm`" permette di calcolare le potenze di una matrice. Utilizzando la funzione `Markov()` della libreria di R `labstat`, possiamo simulare `x<-c("P","S","N")` `traj<-Markov("S",15,x,P).` Rappresentazione Traiettoria `plot(traj$t,unclass(factor(traj$X)),type="s", axes=FALSE, xlab="t",ylab="Che tempo fa")` `axis(1)` `axis(2,c(1,2,3),levels(factor(traj$X)))` `box()` Passeggiata aleatoria `con p=1/2` `set.seed(1)` `n<-50` `x<-rbinom(n,1,0.5)` `x[x==0]<--1` `y<-cumsum(x)` `plot(1:n,y,type="l",main=" passeggiata aleatoria",xlab="tempo",ylab="posizione")` `abline(h=0,lty=3)` Barriere assorbenti `L<--40` `t1<-min(which(y==L))` `y[t1:500]<-L` `plot(1:n,y,type="l",main="passeggiata aleatoria con barriera assorbente",xlab="tempo",ylab="posizione")` `abline(h=0,lty=3)`