MODELLI STATISTICI, RICHIAMI

Corso di Tecniche di Simulazione, a.a. 2005/2006

Francesca Mazzia

Dipartimento di Matematica
Università di Bari

11 Aprile 2006

Modelli statistici nella simulazione

- Nel costruire un modello di simulazione bisogna introdurre eventi di tipo probabilistico.
- Nei sistemi a coda i tempi di interarrivo e i tempi di servizio sono spesso probabilistici.
- Nei problemi di scorte i tempi fra le richieste e il tempo fra l'ordine e l'arrivo delle scorte possono essere probabilistici.
- In modelli di affidabilità il tempo di fallimento è probabilistico;
- In tutti questi casi l'analista simulatore deve generare eventi casuali e
- preferirebbe usare modelli statistici noti se si riesce trovare la distribuzione giusta.

Sistemi a coda

- La distribuzione dei tempi fra gli arrivi e la distribuzione del numero di arrivi per periodo di tempo è importante.
- I tempi di servizio possono essere costanti o probabilistici.
- Se i tempi di servizio sono completamente casuali si utilizza spesso la distribuzione esponenziale per la simulazione.
- Se però i tempi di servizio sono costanti con qualche fluttuazione di tipo casuale, la distribuzione normale è piu utile per descrivereli.

Sistemi a coda

- In alcuni casi il fenomeno sembra seguire una distribuzione di probabilità normale, ma la variabile casuale assume valori maggiori o minori di un valore fissato. In questo caso si usa la distribuzione normale troncata.
- Le distribuzioni gamma e di Weibull sono anche usate per modellare tempi di interarrivo e tempi di servizio.
- La distribuzione esponenziale è un caso particolare delle distribuzioni gamma e di Weibull.

Sistemi a coda

- Le differenze fra la distribuzione esponenziale, gamma e di Weibull, coinvolge la posizione della moda della pdf, e la forma per valori piccoli e grandi.
- La distribuzione esponenziale ha la sua moda nell'origine, la distribuzione gamma e di Weibull hanno la moda in qualche punto $x \geq 0$ funzione dei parametri selezionati.
- il taglio di una distribuzione gamma è lungo come quello della distribuzione esponenziale, quello della distribuzione di Weibull va a zero più rapidamente o meno rapidamente. Questo significa che una distribuzione di Weibull può essere più adatta se i tempi di servizio sono più grandi.

Scorte

- In questi modelli ci sono tre tipi di variabili aleatorie:
  - il numero di unità richieste per ordine e per periodo.
  - il tempo fra le richieste.
  - il tempo fra l'ordine e l'arrivo delle scorte.
- in modelli semplici la richiesta è costante nel tempo e il tempo fra l'ordine e l'arrivo delle scorte è zero o costante. In modelli realistici, e quindi nella simulazione questo non è vero.

Scorte

- La distribuzione per il tempo fra l'ordine e l'arrivo delle scorte più adatta è quella gamma o di Weibull.
- Per quanto riguarda le richieste di solito si usano la distribuzione geometrica, di Poisson e la binomiale negativa;

Dati limitati

- Spesso la simulazione inizia prima di avere abbastanza informazioni sul comportamento dei dati di input.
- Tre distribuzioni si applicano a dati incompleti o limitati: la distribuzione uniforme, triangolare e beta.
- La distribuzione uniforme può essere usata quando i tempi di interarrivo sono casuali, ma non si hanno informazioni sulla distribuzione;
- La distribuzione triangolare si usa quando si possono fare delle assunzioni sui valori minimi e massimi e sui valori modali.
- La distribuzione beta fornisce una varietà di forme di distribuzioni nell'intervallo unitario, che possono essere estese a qualsiasi intervallo.

Distribuzioni discrete

- Sono utilizzate per definire fenomeni aleatori in cui la variabile aleatoria assume solo valori interi.
- Descriviamo quattro distribuzioni:
  - La distribuzione di Bernoulli e i tentativi di Bernoulli;
  - La distribuzione binomiale;
  - La distribuzione geometrica;
  - La distribuzione di Poisson;

Tentativi di Bernoulli

- Consideriamo un esperimento di n tentativi, ognuno dei quali può essere un successo o un fallimento.
- Poniamo $X_j = 1$ se il $j$simo esperimento è stato un successo; $X_j = 0$ se il $j$simo esperimento è stato un fallimento;
- Gli n tentativi di Bernoulli sono chiamati un processo di Bernoulli se i tentativi sono indipendenti, ogni tentativo ha solo due possibili risultati, e la probabilità di successo rimane costante.

Tentativi di Bernoulli

- $p(x_1, x_2, ..., x_n) = p_1(x_1)p_2(x_2) \cdots p_n(x_n)$.

$$p_j(x_j) = p(x_j) = \begin{cases} p, & x_j = 1, j = 1, 2, ...n \\ q, & x_j = 0, j = 1, 2, ...n \\ 0, & \quad altrimenti \end{cases}$$

- Se $n = 1$ la distribuzione $p(x)$ è chiamata distribuzione di Bernoulli.
- La media e la varianza sono:
- $E(X_j) = 0 \cdot q + 1 \cdot p = p$.
- $V(X_j) = (0^2 \cdot q + 1^2 \cdot p) - p^2 = p(1 - p)$

Distribuzione binomiale

- La variabile aleatoria X che denota il numero di successi in n tentativi di Bernoulli ha una distribuzione binomiale data da $p(x)$,

$$p(x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} p^x q^{n-x}, & x = 0, 1, 2, ...n \\ 0, & altrimenti \end{cases}$$

- La probabilità di avere x successi seguiti da $n - x$ fallimenti in n tentativi di Bernoulli è data da $p^x q^{n-x}$. Vi sono $\begin{pmatrix} n \\ x \end{pmatrix}$ risultati che hanno lo stesso numero di fallimenti e successi.

Distribuzione binomiale

- La media e la varianza si calcolano considerando X come la somma di n variabili aleatorie di Bernoulli indipendenti, ognuna con media p e varianza $p(1-p) = pq$, cioè $X = X_1 + X_2 + \cdots + X_n$.
- $E(X) = p + p + \cdots + p = np$
- $V(X) = pq + pq + \cdots + pq = npq$.

Esempio

- Un processo di produzione costruisce chips per computer con una media di 2% non conformi. Ogni giorno un capione aleatorio di 50 chips è prelevato dal processo. Se il campione contiene più di due chips non conformi il processo viene fermato.
- Determiniamo la probabilità che il processo sia fermato.
- Consideriamo il processo come $n = 50$ tentativi di Bernoulli, ognuno con $p = 0.02$, X avrà una distribuzione binomiale con $p = 0.02$ e $q = 1 - p = 0.98$.

Esempio

- $P(X > 2) = 1 - P(X \leq 2)$
- $P(X \leq 2) = \sum_{x=0}^{2} \begin{pmatrix} 50 \\ x \end{pmatrix} (0.02)^x (0.98)^{50-x} \approx 0.92$
- quindi $P(X > 2) \approx= 0.08$
- $E(X) = np = 50 * 0.02 = 1,\ V(X) = npq = 0.98$

Distribuzione geometrica

- È legata ad una sequenza di tentativi di Bernoulli.
- La variabile di interesse è X definita come il numero di tentativi per avere il primo successo:

$$p(x) = \begin{cases} q^{x-1}p, & x = 1, 2, ... \\ 0, & altrimenti \end{cases}$$

- L'evento $X = x$ si verifica quando vi sono $x - 1$ fallimenti seguiti da un successo.
- $E(X) = \frac{1}{p}$, $V(X) = \frac{q}{p^2}$

Esempio

- Il 40% di stampanti assemblate sono respinte dopo l'ispezione. Trovare la probabilità che la prima stampante accettata sia la terza ispezionata.
- Ogni ispezione può essere considerata un tentativo di bernoulli con $q = 0.4$, $p = 0.6$.
- $p(3) = (0.4)^2 0.6 = 0.096$

Distribuzione di Poisson

- La distribuzione di Poisson descrive molti fenomeni aleatori abbastanza bene ed è matematicamente abbastanza semplice.
- La funzione massa di probabilità è, posto $\alpha > 0$:

$$p(x) = \begin{cases} \frac{e^{-\alpha}\alpha^x}{x!}, & x = 0, 1, 2, ... \\ 0, & altrimenti \end{cases}$$

- Una proprietà importante è che sia la media che la varianza sono uguali ad α.
- La funzione di distribuzione cumulativa è data da:

$$F(x) = \sum_{i=0}^{x} \frac{e^{-\alpha}\alpha^i}{i!}$$

Esempio

- In un problema di scorte, è importante studiare l'accumulo delle richieste che vi sono nei tempi che intercorrono fra quando viene fatto un ordine e l'arrivo della merce:

$$L = \sum_{i=1}^{T} D_i$$

- D_i sono le richieste nel periodo $i$simo, T il numero di periodi, L sono le richieste complessive.
- Si desidera che la probabilità di terminare le scorte non ecceda il 5% durante il tempo di attesa dell'arrivo delle scorte.
- Se le richieste hanno una distribuzione di Poisson, determinare quando effettuare gli ordini è facile.

Esempio

- Supponiamo che le D_i abbiano una distribuzione di Poisson con $\alpha = 10$ e che si desidera una protezione al 95%.
- Vogliamo trovare il valore più piccolo di x tale che la probabilità che L non ecceda x sia maggiore di 0.95
- $F(x) = \sum_{i=0}^{x} \frac{e^{-10}10^i}{i!} \geq 0.95$
- il risultato è $x = 15$

Distribuzioni continue

- Distribuzione Uniforme
- Distribuzione Esponenziale
- Distribuzione Gamma
- Distribuzione di Erlang
- Distribuzione Normale
- Distribuzione di Weibull
- Distribuzione Triangolare
- Distribuzione Lognormale

Distribuzione Uniforme

- Una variabile aleatoria X si dice uniformemente distribuita in $[a, b]$ se la sua pdf è data da:

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

- la cdf è data da:

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 0, & x \geq b \end{cases}$$

- $P(x_1 < X < x_2) = F(x_2) - F(x_1) = \frac{x_1 - x_2}{b-a}$.
- La media è $E(X) = (a+b)/2$.
- la varianza è $V(X) = (b-a)^2/12$

Distribuzione Uniforme

- Ha un ruolo importante nella simulazione.
- Numeri casuali distribuiti uniformemente fra 0 e 1 permettono di generare eventi casuali.
- Sono stati costruiti numerosi metodi per generare numeri casuali distribuiti uniformemente.
- Sono utilizzati anche per generare numeri casuali appartenenti ad altre distribuzioni.

Esempio

- Si sta sviluppando la simulazione di un deposito.
- Circa ogni 3 minuti arriva una chiamata all'operatore del carrello verticale, per andare in una certa posizione.
- l'assunzione iniziale fatta per il tempo fra le chiamate (arrivi) è di avere una distribuzione uniforme con media 3 minuti.
- Poichè i dati sono limitati assumiamo che la varianza sia massima.
- Utilizziamo la distribuzione uniforme con $a = 0$ e $b = 6$.

Distribuzione Esponenziale

- Una variabile aleatoria X ha una distribuzione esponenziale se, dato un parametro $\lambda > 0$ la sua pdf è:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & altrimenti \end{cases}$$

- La media è : $E(X) = 1/\lambda$.
- La varianzia è : $V(X) = 1/\lambda^2$.
- La cdf è:

$$F(x) = \begin{cases} 0 & x < 0 \\ \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

Distribuzione Esponenziale

- La distribuzione esponenziale è stata usata per modellare:
- tempi di interarrivo quando gli arrivi sono completamente casuali;
- tempi di servizio con grande variabilià;
- (In questi casi λ è un rapporto: arrivi per ora o servizi per minuto.)
- tempi di vita di componenti come le lampadine;
- (In questo caso λ è l'incidenza dei guasti.)

Esempio

- La vita di una lampada industriale, in migliaia di ore, è distribuita esponenzialmente con incidenza di guasti $\lambda = 1/3$ (un guasto ogni 3000 ore, in media).
- La probabilità che la lampadina duri più della sua vita media di 3000 ore è data da
  $P(X > 3) = 1 - P(X \leq 3) = 1 - F(3) = (1 - (1 - e^{-3/3}) = 0.368.$
- Questa probabilità è sempre la stessa indipendentemente dal valore di λ.
- La probabilità che la lampada duri fra 2000 e 3000 ore è data da:
  $P(2 \leq X \leq 3) = F(3) - F(2) = 1 - e^{-3/3} - (1 - e^{-2/3}) = 0.145$

Distribuzione Esponenziale

- Una proprietà importante della distribuzione esponenziale è che è senza memoria: $P(X > s + t | X > s) = P(X > t), \qquad s > 0, t > 0$
- Se X rappresenta la vita di una componente ed è distribuita esponenzialmente allora la probabilità che la componente duri per almeno $t + s$ ore, supposto che sia durata s ore è la stessa della probabilità iniziale che duri t ore.
- La probabilità condizionata $P(X > s + t | X > s) = \frac{P(X>s+t)}{P(X>s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}$
- La distribuzione esponenziale è l'unica distribuzione continua che ha questa proprietà (La distribuzione geometrica è l'unica distribuzione discreta con questa proprietà).

Distribuzione Gamma

- Una funzione usata per definire la distribuzione gamma è la funzione gamma, che è definita da:

$$\Gamma(\beta) = \int_0^\infty x^{\beta-1} e^{-x} dx$$

- integrando l'equazione precedente per parti si mostra che:

$$\Gamma(\beta) = (\beta - 1)\Gamma(\beta - 1)$$

- Se β è un intero allora $\Gamma(\beta) = (\beta - 1)!$
- La funzione gamma è una generalizzazione della nozione di fattoriale che si applica a tutti i numeri positivi.

Distribuzione Gamma

- Una variabile aleatoria X ha una distribuzione gamma se, dati un parametro $\beta > 0$ e un parametro θ la sua pdf è:

$$f(x) = \begin{cases} \frac{\beta\theta}{\Gamma(\beta)}(\beta\theta x)^{(\beta-1)}e^{-\beta\theta x}, & x \geq 0 \\ 0, & altrimenti \end{cases}$$

- β si chiama parametro di forma.
- θ si chiama parametro di scala.
- La media è : $E(X) = 1/\theta$.
- La varianzia è : $V(X) = 1/(\beta\theta^2)$.
- $F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \int_x^\infty \frac{\beta\theta}{\Gamma(\beta)}(\beta\theta t)^{(\beta-1)}e^{-\beta\theta t}dt, & x > 0 \end{cases}$

Distribuzione Gamma

- Se β è un intero la distribuzione gamma è legata alla distribuzione esponenziale.
- Se la variabile aleatoria X è la somma di β variabili aleatorie indipendenti distribuite esponenzialmente, ognuna con parametro $\beta\theta$, allora X ha una distribuzione gamma con parematri β e θ.
- Se $\beta = 1$ abbiamo una distribuzione esponenziale.

Distribuzione di Erlang

- Se $\beta = k$ nella distribuzione gamma, la distribuzione prende il nome di distribuzione di Erlang di ordine k con k un intero.
- La distribuzione di Erlang si verifica nei seguenti contesti: consideriamo k stazioni che devono essere attraversate da un cliente per completare un servizo.
- Un nuovo cliente non può entrare alla prima stazione se il cliente precedente non ha superato tutte le stazioni.
- Ogni stazione ha una distribuzione esponenziale del tempo di servizio con parametro $k\theta$.

Distribuzione di Erlang

- La media $E(X) = 1/\theta$.
- La varianza $V(X) = 1/(k\theta^2)$.
- La cdf è data da:

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \sum_{i=0}^{k-1} \frac{(k\theta x)^i e^{-k\theta x}}{i!}, & x > 0 \end{cases}$$

- la cdf è la somma di termini di Poisson con media $\alpha = k\theta x$.

Esempio

Un professore parte per le vacanze, ma vuole avere una luce sempre accesa in casa per scoraggiare i ladri. Il professore costruisce un dispositivo con due lampadine. Il dispositivo passerà da una lampadina all'altra se la prima si spegne. Le lampadine hanno una vita media di 1000 ore distribuita in modo esponenziale. Il professore sarà via per 90 giorni (2160 ore). Quale è la probabilità che la luce sia spenta quando il professore ritorna?

- La probabilità che il sistema funziona per almeno x ore è chiamata funzione di affidabilità $R(x) = 1 - F(x)$.
- In questo caso X è la somma di due variabili aleatorie con distribuzione esponenziale $\beta = k = 2$ lampadine e $k\theta = 1/1000$ per ora, quindi $\theta = 1/2000$ per ora.
- Si usa la $F(x)$ della distribuzione di Erlang.

$$F(2160) = 1 - \sum_{i=0}^{2} \frac{((2)(1/2000)2160)^i e^{-2(1/2000)2160}}{i!} = 0.636$$

Distribuzione Normale

- Una variabile aleatoria ha una distribuzione normale con media $\mu(-\infty < \mu < \infty)$ e varianza $\sigma^2 > 0$ se ha la pdf:

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

- una notazione tipca per la distribuzione Normale è $N(\mu, \sigma^2)$
- la cdf è data da:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2) dt.$$

- La $F(x)$ può essere calcolata solo numericamente.
- conviene quindi effettuare una trasformazione di variabili per evitare di dover calcolare la $F(x)$ per ogni valore di μ e σ^2.

Distribuzione Normale

- Poniamo $Z = (X - \mu)/\sigma$.
- $F(x) = P(X \leq x) = P(Z \leq \frac{x-\mu}{\sigma})$
- $F(x) = \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} exp(-\frac{z^2}{2}) dz$.
- $F(x) = \int_{-\infty}^{(x-\mu)/\sigma} \phi(z) dz = \Phi(\frac{x-\mu}{\sigma})$
- $\phi(z)$ è la pdf di una distribuzione normale con media 0 e varianza 1.
- $N(0,1)$ viene chiamata distribuzione normale standard.
- La $F(x)$ di $N(0,1)$ è stata tabulata e viane usata opportunamente per calcolare i valori della $F(x)$ delle altre distribuzioni normali.

Distribuzione Normale

Proprietà

- $lim_{x \to -\infty} f(x) = 0$, $lim_{x \to +\infty} f(x) = 0$
- $f(\mu - x) = f(\mu + x)$
- Il valore massimo si trova in $x = \mu$, cioè la media e la moda sono uguali.

Esempio

- Il tempo richiesto per caricare una nave, X, è distribuito come $N(12, 4)$.
- La probabilità che la nave sarà caricata in meno di 10 ore è data da $F(10)$ con:

$$F(10) = \Phi(\frac{10-12}{2}) = \Phi(-1) = 1 - \Phi(1)$$

- in generale per le proprietà di simmetria $F(-x) = 1 - F(x)$
- La probabilità che ci vogliano più di 12 ore per caricare la nave è: $P(X > 12) = 1 - F(12) = 1 - \Phi(0) = 1 - 0.5 = 0.5$
- La probabilità che ci vogliano fra 10 e 12 ore per caricare la nave è: $P(10 \leq X \leq 12) = F(12) - F(10) = 0.5 - 0.1587 = 0.3413$

Esempio

- Il tempo di attesa in una coda prima di iniziare il self-service in una caffetteria, X, è distribuito come $N(15, 9)$.
- La probabilità che un cliente aspetta fra 14 e 17 minuti è:

$$\begin{aligned} &P(14 \leq X \leq 17) = F(17) - F(14) = \\ &= \Phi(\tfrac{17-15}{3}) - \Phi(\tfrac{14-15}{3}) \\ &= \Phi(0.667) - \Phi(-0.333) \end{aligned}$$

$$\Phi(0.667) - \Phi(-0.333) = \Phi(0.667) - 1 + \Phi(0.333) = 0.378$$

Distribuzione di Weibull

- Una variabile aleatoria ha una distribuzione di Weibull se la sua pdf è data da :

$$f(x) = \begin{cases} \frac{\beta}{\alpha}(\frac{x-\nu}{\alpha})^{\beta-1} exp(-(\frac{x-\nu}{\alpha})^{\beta}), & x \geq \nu \\ 0 & altrimenti \end{cases}$$

- I tre parametri della distribuzione di Weibull sono:
  $\nu(-\infty < \nu < \infty)$: parametro di posizione;
  $\alpha > 0$: parametro di scala;
  $\beta > 0$: parametro di forma;
- Quando $\beta = 1$ e $\nu = 0$ la distribuzione è una distribuzione esponenziale con parametro $\lambda = 1/\alpha$

Distribuzione di Weibull

- La media è:

$$E(X) = \nu + \alpha\Gamma(\frac{1}{\beta} + 1).$$

- La varianza è:

$$V(X) = \alpha^2(\Gamma(\frac{2}{\beta} + 1) - \Gamma(\frac{1}{\beta} + 1)^2).$$

- Il parametro ν non ha effetto sulla varianza, però aumenta o diminuisce la media.
- La cdf è:

$$F(x) = \begin{cases} 1 - exp(-(\frac{x-\nu}{\alpha})^\beta), & x \geq \nu \\ 0 & x < \nu \end{cases}$$

Esempio

Il tempo che impiega un aereo per decollare e liberare la pista in un aereoporto internazionale ha una distribuzione di Weibull con $\nu = 1.34$ minuti, $\beta = 0.5$ e $\alpha = 0.04$ minuti. Determina la probabilità che l'aereo impieghi più di 1.5 minuti per decollare.

- $P(X > 1.5) = 1 - P(X \leq 1.5) = 1 - F(1.5) = 0.135$

Distribuzione Triangolare

- Una variabile aleatoria ha una distribuzione triangolare se la sua pdf è data da :

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & a \leq x \leq b \\ \frac{2(c-x)}{(c-b)(c-a)} & b \leq x \leq c \\ 0 & altrimenti \end{cases}$$

  con $a < b < c$.
- la media è: $E(X) = (a + b + c)/3$

Distribuzione Triangolare

- La moda è in $x = b = 3E(X) - (a + c)$.
- $\frac{2a+x}{3} \leq E(X) \leq \frac{a+2c}{3}$
- La moda è usata più spesso della media per caratterizzare la distribuzione triangolare, la sua altezza è $2/(c - a)$.
- La varianza è poco usata.
- La cdf è

$$F(x) = \begin{cases} 0 & x \leq a \\ \frac{(x-a)^2}{(b-a)(c-a)} & a < x \leq b \\ 1 - \frac{(c-x)^2}{(c-b)(c-a)} & b < x \leq c \\ 1 & x > c \end{cases}$$

Esempio

Un sensore elettronico determina la qualità dei chips di memoria, rigettando quelli che non superano il test. Su richiesta un sensore fornisce il numero minimo e massimo di chip rifiutati in un'ora di produzione, durante le ultime 24 ore. La media è anche fornita. Senza ulteriori informazioni, il dipartimento di controllo della qualità ha assunto che il numero di chips rifiutati può essere approssimato con una distribuzione triangolare.

- Il numero minimo di rifiuti è zero, il numero massimo è 10 e la media è 4.
- Sapendo che $a = 0$, $c = 10$ e $E(X) = 4$ il valore di $b = 3E(X) - (a + c) = 2$.

Esempio

- L'altezza della moda è $2(10 - 0) = 0.1$.
- La mediana è il punto in cui 0.5 dell'area è a destra e 0.5 è a sinistra. L'area a destra della moda è $F(2) = 0.2$, quindi la mediana si trova fra b e c. Ponendo $F(x) = 0.5$ e risolvendo per x si ottiene $x = 3.7$.
- La moda, la media e la mediana non sono necessariamente uguali.

Distribuzione lognormale

- Una variabile aleatoria ha una distribuzione lognormale se la sua pdf è data da ($\sigma^2 > 0$):

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} exp(-(\frac{(lnx-\mu)^2}{2\sigma^2})), & x \geq 0 \\ 0 & altrimenti \end{cases}$$

- La media $E(X) = e^{\mu+\sigma^2/2}$.
- La varianza $V(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$.

Distribuzione lognormale

- Se Y ha una distribuzione $N(\mu, \sigma^2)$, allora $X = e^Y$ ha una distribuzione lognormale con parametri μ e σ^2.
- Se la media e la varianza di una distribuzione lognormale sono μ_L e σ_L^2 allora:

$$\mu = ln\left(\frac{\mu_L^2}{\sqrt{\mu_L^2 + \sigma_L^2}}\right)$$

e

$$\sigma^2 = ln\left(\frac{\mu_L^2 + \sigma_L^2}{\mu_L^2}\right)$$

Distribuzioni Empiriche

- Una distribuzione empirica può essere continua o discreta.
- Si usa quando è impossibile o non è necessario stabilire che la variabile casuale ha una distribuzione nota.
- Il vantaggio di usare distribuzioni note nella simulazione è la facilità con cui si possono cambiare i parametri per portare avanti una analisi di sensitività.

Esempio

- I clienti arrivano in un ristorante in gruppi da 1 a 8 persone. Il numero di persone negli ultimi 300 gruppi è stato osservato:

arrivi (minuti)	frequenza	frequenza relativa	frequenza relativa cumulativa
1	30	0.10	0.10
2	110	0.37	0.47
3	45	0.15	0.62
4	71	0.24	0.86
5	12	0.04	0.90
6	13	0.04	0.94
7	7	0.02	0.96
8	12	0.04	1.00