Distribuzioni e densità

Discreto

- **Distribuzione binomiale o processo di Bernoulli**

La distribuzione binomiale è una distribuzione di probabilità discreta che descrive quanti successi k si hanno su n prove.
Esempi di casi di distribuzione binomiale sono i risultati di una serie di lanci di una stessa moneta o di una serie di estrazioni da un'urna (con reintroduzione), ognuna delle quali può fornire due soli risultati: il successo con probabilità p e il fallimento con probabilità $q = 1 – p$

- **Distribuzione geometrica**

È un processo di Bernoulli illimitato che descrive la probabilità di avere il primo successo al tentativo k-esimo.
La probabilità che un dado (equilibrato, a 6 facce) debba venire lanciato *esattamente* 10 volte prima di fornire un "4" è data dalla distribuzione geometrica.
È caratterizzata da assenza di memoria.

- **Distribuzione binomiale negativa o distribuzione di Pascal**

La distribuzione di Pascal descrive il numero di fallimenti precedenti il successo n-esimo in un processo di Bernoulli ovvero quanti insuccessi k si ottengono prima di ottenere n successi.
Ad esempio, lanciando una moneta fino ad ottenere 3 volte testa, la distribuzione di Pascal descrive le probabilità per il numero di risultati croce visti nel frattempo.

- **Distribuzione ipergeometrica**

La distribuzione ipergeometrica descrive l'estrazione senza reinserimento di alcune palline, perdenti o vincenti, da un'urna.
Ad esempio, estraendo 5 palline da un'urna che ne contiene 3 bianche e 7 nere, il numero di palline bianche estratte è descritto dalla distribuzione ipergeometrica.

- **Distribuzione di Poisson**

La distribuzione di Poisson (o poissoniana) esprime le probabilità per il numero di eventi che si verificano successivamente ed indipendentemente in un dato intervallo di tempo, sapendo che mediamente se ne verifica un numero λ. Inoltre, viene utilizzata come approssimazione della legge binomiale in quanto è più di facile utilizzo quando si maneggiano grandi numeri e una probabilità molto piccola.
Ad esempio, si utilizza una distribuzione di Poisson per misurare il numero di chiamate ricevute in un call-center in un determinato arco temporale, come una mattinata lavorativa. Questa distribuzione è anche nota come legge degli eventi rari (non per l'avvenimento in sé ma per la probabilità molto bassa). Il tempo invece fra una chiamata e l'altra è modellabile tramite una esponenziale; dunque il tempo di attesa per l'occorrenza k-esima è modellato da una erlangiana di parametri k e lambda

Continuo

- **Distribuzione Gamma**

  La distribuzione Gamma comprende, come casi particolari, anche le distribuzioni esponenziale e chi quadrato.
  Viene utilizzata come modello generale dei tempi di attesa nella teoria delle code, soprattutto qualora siano importanti effetti che rimuovano "l'assenza di memoria" della distribuzione esponenziale. Nella statistica bayesiana è comune sia come distribuzione a priori che come distribuzione a posteriori. In particolare, i tempi di arrivo nei processi di Poisson hanno distribuzione gamma, e la distribuzione chi-quadro

- **Distribuzione esponenziale**

  La distribuzione esponenziale descrive la "durata di vita" di un fenomeno che non invecchia (ossia la distribuzione esponenziale è priva di memoria).
  Un esempio è la durata di vita di una particella radioattiva prima di decadere oppure la durata della richiesta di un servizio; dunque essa è in relazione al tempo di attesa del primo successo, in fenomeni aleatori con distribuzione geometrica.
  È caratterizzata da assenza di memoria.

- **Distribuzione di Weibull**

  La distribuzione di Weibull viene impiegata per descrivere sistemi con tasso di guasto variabile nel tempo, come estensione della distribuzione esponenziale che prevede tassi di guasto costanti nel tempo.
  Come la distribuzione esponenziale descrive la "durata di vita" di un fenomeno privo di memoria, così la distribuzione di Weibull può descrivere la durata di vita per un fenomeno la cui "probabilità di morire" può variare nel tempo in funzione di k. Viene utilizzata in molti ambiti che trattano appunto i guasti, come l'analisi dei guasti, l'analisi di sopravvivenza, l'ingegneria dell'affidabilità e il controllo della qualità. Viene utilizzata anche nelle previsioni meteorologiche, come generalizzazione della distribuzione di Rayleigh.

- **Distribuzione normale o distribuzione di Gauss**

  La distribuzione normale è spesso usata come prima approssimazione per descrivere variabili casuali a valori reali che tendono a concentrarsi attorno a un singolo valor medio.
  Il grafico della funzione di densità di probabilità associata è simmetrico e ha una forma a campana, nota come campana di Gauss (o anche come curva degli errori, curva a campana, ogiva).

- **Distribuzione di Erlang**

  La distribuzione di Erlang con parametro k unitario si semplifica in una distribuzione esponenziale. Infatti tale distribuzione può essere considerata come la somma di k variabili aleatorie indipendenti e identicamente distribuite secondo un'esponenziale di parametro λ. La distribuzione di Erlang è anche un caso particolare di una distribuzione Gamma la quale prevede, generalmente, un parametro reale. Ponendo invece $\lambda = 1 / 2$ la distribuzione diventa una distribuzione chi quadrato con 2 k gradi di libertà.
  La distribuzione di Erlang fu sviluppata per esaminare il numero di chiamate telefoniche che gli operatori di un centralino possono ricevere nello stesso istante.
  L'occorrenza di eventi indipendenti aventi un certo tasso medio viene modellizzata con un processo di Poisson. Poiché in un processo di Poisson gli intertempi tra un evento ed il successivo sono distribuiti esponenzialmente, il tempo di attesa per osservare la realizzazione di k eventi è distribuito secondo una Erlang (k , λ).
  La distribuzione di Erlang, che misura il tempo tra chiamate telefoniche, può essere utilizzata in combinazione col valore atteso delle durate delle chiamate in modo tale da avere informazioni riguardo al traffico di telefonate. Ciò viene utilizzato per calcolare la probabilità di perdere telefonate o di far attendere dei clienti al telefono, in base alle ipotesi di rifiuto di chiamate (Formula di Erlang B) o di messa in coda (Formula di Erlang C).

La legge di Erlang di ordine n descrive anche il tempo al guasto dei sistemi con n-1 riserve a commutazione perfetta. Nel caso della commutazione imperfetta, entra in gioco una variabile condizionante e dunque la formula della distribuzione del tempo di guasto è definita da una probabilità totale

- **Distribuzione iperesponenziale**

Una variabile casuale X con densità di probabilità iperesponenziale può essere considerata come una somma pesata di R variabili casuali con densità di probabilità esponenziale dove il peso è dato dalla probabilità di un determinato avvenimento. La sua distribuzione e densità sono assimilabili a delle formule di distribuzione totale.

Un esempio in cui può essere usata la distribuzione iperesponenziale è per definire il tempo di esecuzione di un processo preso a caso fra alcuni dati o per modellare gli arrivi a raffica

Il modello iperesponenziale può essere usato per modellare il tempo di servizio di un sistema in cui vi sono n componenti in parallelo (ma che non possono lavorare come tali) ognuno con un tempo di servizio e una probabilità alpha di essere scelti. Nel caso in cui la probabilità che uno dei processi venga scelto è vicino ad uno o che la media aumenti allora tale modello descrive un processo di raffica perché vuol dire che arrivano molte richieste

Il nome iperesponenziale deriva dal fatto che il rapporto fra deviazione standard e media è maggiore di uno (nella esponenziale è uno e nella ipoesponenziale è minore di uno); tale coefficiente è detto *coefficiente di variazione.*

- **Distribuzione ipoesponenziale**

Modella un processo in cui, anziché essere in parallelo, i dispositivi sono posti in serie. Il coefficiente di variazione (deviazione standard fratto media) è minore di uno. Tale distribuzione va a modellare ad esempio il Triple Modular Redundancy system in cui vi è un elemento di riserva; essa è una ipoesponenziale a tre stadi

QUANDO USARE DETERMINATE CURVE

- Data una media, numero di chiamate ad un call center o di richieste ad un server, eventi indipendenti -> Poisson (approssimazione della binomiale utilizzata per gli eventi di probabilità piccola)
- Dato un database con un numero di elementi difettosi e altri no, la probabilità di pescare un numero di elementi difettosi fra un numero di pescaggi totali (senza reinserimento) -> ipergeometrica
- Dato un database con un con un numero di elementi difettosi e altri no, la probabilità di pescare un numero di elementi difettosi fra un numero di pescaggi totali (con reinserimento) -> binomiale
- Data una serie di prove, quante bisogna farne per avere k successi -> binomiale negativa
- Data una serie di prove, quante bisogna farne per avere il primo successo al tentativo k -> geometrica
- Dato un sistema con n dispositivi, qual è la probabilità che almeno m componenti funzionano (sistemi m-out-of-n) -> sommatoria di prove di Bernoulli dove p=R; si usa la sommatoria perché ALMENO m componenti devono funzionare (dunque la sommatoria va da m a n)
- In un sistema, se viene richiesto di calcolare l'affidabilità di percorsi particolari allora bisogna applicare la regola classica di affidabilità, altrimenti si può applicare il concetto di sistemi m-out-of-n
- Durata di un fenomeno casuale non soggetto ad usura -> esponenziale
- Durata di un fenomeno casuale soggetto ad usura -> Weibull: ha un tasso di usura diverso da quello costante ma con alfa uguale a 1 si ha appunto una esponenziale
- Tempo al k-esimo fenomeno dove tutti i k fenomeni sono descritti tramite una esponenziale -> erlang
- Tempo di guasto fra t e delta t -> è dato dalla funzione fx(x) per dt
- Affidabilità congiunta -> si usa per definire l'affidabilità di un sistema in cui le cause di guasto sono dipendenti (es. il guasto di un componente causa il guasto di un altro cioè una causa di guasto può essere comune). Essa non è esattamente complementare alla distribuzione congiunta relativa
- Distribuzione totale -> la formula della distribuzione (probabilità) totale è data dal prodotto delle probabilità condizionate per le probabilità dei rispettivi condizionanti

ESERCIZI

- Blocchi in serie o parallelo -> applico normalmente le formule di affidabilità
- Blocchi in serie o parallelo con dipendenza dei percorsi ->
  - Metodo dello spazio degli eventi -> sviluppo la probabilità dell'unione dei percorsi come somma delle singole probabilità meno le intersezioni; applico la chain rule e sviluppo tutte le intersezioni come probabilità condizionate
  - Metodo della probabilità totale -> si partiziona l'insieme universo in due parti in base al comportamento (funzionante o no) del componente che crea la dipendenza dei percorsi e si va a calcolare la probabilità pesata dei percorsi con tale componente funzionante e no.
- Indice di criticità -> corrisponde a calcolare la probabilità condizionata che, dato un sistema non funzionante, la colpa sia del componente non funzionante; per calcolare la probabilità che il sistema non funzioni basta prendere il complementare della probabilità calcolata col metodo dello spazio degli eventi o della probabilità totale.
- Conflitto di assegnazione -> si ha quando due risorse chiedono accesso contemporaneamente; la probabilità è data dall'unione delle probabilità che almeno due risorse richiedano accesso

- Probabilità che un fenomeno finisca entro un tempo t -> calcolo della funzione di distribuzione cumulativa in quel punto

DEFINIZIONI IMPORTANTI

- Rischio di guasto/tasso di guasto -> funzione che definisce la probabilità che un componente si rompa tra t e delta t sapendo che ha funzionato fino a t. Il rischio di guasto può essere costante e non; se è costante l'affidabilità si modella come una semplice esponenziale altrimenti al posto dell'esponente si trova l'integrale della funzione di rischio