

Minicorso Controllo Statistico di Processo

di Andrea Saviano

Parte 4

- Le frequenze cumulative, premessa
- Le distribuzioni discrete
- Le distribuzioni continue
- Distribuzioni di probabilità: come, dove e quando

Premessa



Una distribuzione di probabilità è un modello matematico che collega il valore di una variabile alla probabilità che tale valore si trovi all'interno della popolazione ovvero possano essere osservati. Ne consegue che l'esito di una misura può essere considerato una **variabile casuale**, poiché tale valore può assumere valori differenti all'interno della popolazione.

Si riconoscono due tipologie di distribuzione di probabilità:

- **discreta**, quando il parametro da misurare può assumere solo alcuni valori, il grafico delle probabilità si presenta come un istogramma.
- **continua**, quando la variabile da misurare è esprimibile mediante una scala continua, il grafico delle probabilità si presenta come una curva continua;



Formalmente, le distribuzioni di probabilità vengono espresse da una legge matematica detta:

- **funzione di densità di probabilità**, indicata con $f(x)$, per le distruzioni continue
- **funzione di probabilità**, indicata con $p(x)$, per le distruzioni discrete.

Le distribuzione discrete

Un evento in cui il campione è costituito da n dati sperimentali i quali possono assumere solo valori $x \in \{1; 2; 3; 4; \dots\}$ dà luogo ad una distribuzione discreta.

Detta $p(x)$ la funzione che esprime le probabilità, in termini di momenti si ha:

$$\mu = \sum_{i=1}^{\infty} [x_i \cdot p(x_i)]$$
$$\sigma^2 = \sum_{i=1}^{\infty} [(x_i - \mu)^2 \cdot p(x_i)]$$
$$\gamma = \frac{\sum_{i=1}^{\infty} [(x_i - \mu)^3 \cdot p(x_i)]}{\sigma^3}$$



La distribuzione di Bernoulli o binomiale

Si dice esperimento di Bernoulli una sequenza di n prove con le seguenti caratteristiche:

- il risultato di ogni prova può essere solo “successo¹” o “fallimento”;
- il risultato di ciascuna prova è **indipendente** dai risultati delle prove precedenti;
- la probabilità p di “successo”, e quindi la probabilità q di “fallimento”, sono **costanti in ciascuna prova**.

Allora:

$$n = p \cdot n + q \cdot n \Leftrightarrow q = 1 - p$$

La probabilità in un singolo tentativo che su n eventi ci siano x successi e $n-x$ insuccessi è data dall'equazione:

$$p(x) = p^x \cdot (1-p)^{n-x}$$

Per comprendere ciò, chiediamoci quale sia la probabilità che due persone su tre persone giungano puntuali ad una riunione. La probabilità totale è data dal prodotto della probabilità di due successi per quella di un insuccesso: $x=2$, $n=3$.

$$p \cdot p \cdot q = p^2 \cdot (1-p)^{3-2}$$

Le possibili combinazioni di 2 successi e un insuccesso sono $C_{3,2}=3$

$$p \cdot p \cdot q \quad p \cdot q \cdot p \quad q \cdot p \cdot p$$

allora, essendo $C_{n,x}$ tutte le combinazioni possibili dei successi e dei fallimenti, il numero x di successi in n prove ha più in generale una legge di distribuzione rappresentabile tramite l'equazione:

¹ Attenzione che il termine successo non significa che l'evento sia quello desiderato. Si può trattare, ad esempio, del successo nel trovare un pezzo difettoso tra molti che non lo sono.

$$p(x) = C_{n,x} \cdot p^x \cdot (1-p)^{n-x} = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

Questa distribuzione, per la presenza del termine binomiale, assume anche il nome di **distribuzione binomiale**.

Senza porre un limite al numero delle prove n , ci chiediamo ora se, in un processo di questo tipo, è mai possibile che non si abbia mai un successo. Un tale evento non è logicamente impossibile, tuttavia se $p > 0$ si ha:

$$\lim_{n \rightarrow +\infty} (1-p)^n = 0$$

Questo concetto è alla base del **paradosso di Borel**: se un esperimento può essere ripetuto infinite volte nelle stesse condizioni, a furia di provare, un qualsiasi combinazione di eventi anche di probabilità irrisoria non nulla si verificherà prima o poi con probabilità uno (evento certo). In altri termini: combinando all'infinito lettere a caso, un elaboratore finirebbe per scrivere di certo un libro degno del premio Nobel.

La distribuzione binomiale sin qui descritta è spesso utilizzata in controllo qualità, quando si debbano effettuare controlli su popolazioni molto estese (assimilabili quindi a quelle di dimensione infinita). In questo caso p rappresenta la **frazione di elementi non conformi presenti nella popolazione**, mentre x è il **numero di elementi non conformi osservati nel campione** casuale di n elementi prelevato dalla popolazione.

Il rapporto tra il numero osservato di elementi difettosi e la numerosità del campione si indica con il simbolo:

$$\hat{p} = \frac{x}{n}$$

ed è una stima del valore reale (ignoto) e si definisce: **frazione campionaria di elementi non conformi**.

Applicando questa logica a più controlli, il numero x di non conformi proveniente da prove sperimentali tenderà ad oscillare intorno ad un valore centrale (valor medio di k esperimenti su campioni di dimensioni n), tale valore risulta una buona stima del livello di difettosità della produzione e la distribuzione degli esiti dei singoli esperimenti tende ad addensarsi su tale valore con una legge di probabilità normale (teorema del limite centrale). Anticipando argomenti che si troveranno più avanti, questo significa che è possibile verificare una presenza di dati anomali come un aumento ingiustificato di pezzi difettosi utilizzando un campione discreto invece che un controllo al 100% (un controllo su tutta la popolazione). Si deve infatti tenere ben presente che il controllo 100% non viene effettuato per verificare la presenza di non conformi, ma per lo scopo esattamente contrario: riuscire a selezionare i "pochi" buoni dai "molti" cattivi.

In termini di momenti si ha:

$$\begin{aligned}\mu &= n \cdot p \\ \sigma^2 &= n \cdot p \cdot (1-p)\end{aligned}$$

La distribuzione ipergeometrica

Si supponga di disporre di una popolazione finita di N elementi. Un certo numero $C = p \cdot N$ di questi ($C \leq N$) ricade in un insieme d'interesse (ad esempio: sono i non conformi). Da questa popolazione viene estratto un campione casuale di n elementi in cui vengono rilevati x elementi appartenenti all'insieme d'interesse.

In questo caso la distribuzione delle probabilità è rappresentabile tramite l'equazione:

$$p(x) = \frac{\binom{C}{x} \cdot \binom{N-C}{n-x}}{\binom{N}{n}} = \frac{\binom{p \cdot N}{x} \cdot \binom{q \cdot N}{n-x}}{\binom{N}{n}}$$

In pratica si ha il rapporto tra il prodotto del numero di combinazioni che realizzano x successi e $n-x$ insuccessi e il numero totale di combinazioni, tale distribuzione prende il nome di **distribuzione ipergeometrica**.

Si nota che la questo è un modello molto adatto se, dato un lotto di N elementi dei quali una percentuale p si ritiene non conforme o difettosa, si debba giudicare il lotto selezionando un campione casuale di n elementi. In questo caso x sarà il numero di elementi non conformi trovati nel campione.

Ancora una volta, ripetendo questi controlli più volte, gli esiti x delle prove sperimentali tenderanno ad oscillare intorno ad un valore centrale con una legge di probabilità normale (teorema del limite centrale).

$$\mu = n \cdot p$$
$$\sigma^2 = n \cdot p \cdot q \cdot \frac{N-n}{N-1}$$

La distribuzione di Poisson

Consideriamo la distribuzione binomiale nel caso in cui n sia molto grande e p sia molto piccolo e che $n \cdot p = \lambda$. La situazione limite sarà:

$$\lim_{n \rightarrow +\infty} n \cdot p = \lambda$$

A questo punto è possibile scrivere nell'equazione della distribuzione binomiale secondo la forma:

$$p(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} = \frac{n!}{x!(n-x)!} \cdot \left(\frac{\lambda}{n}\right)^x \cdot \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Effettuiamo il limite per $n \rightarrow +\infty$

$$\lim_{n \rightarrow +\infty} p(x) = \lim_{n \rightarrow +\infty} \left[\frac{n \cdot (n-1) \cdot \dots \cdot (n-x+1)}{n^x} \cdot \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^x} \cdot \frac{\lambda^x}{x!} \right]$$

ora, svolgendo i limiti sui singoli moltiplicatori, si ha:

$$\left[\lim_{n \rightarrow +\infty} \frac{n \cdot (n-1) \cdot \dots \cdot (n-x+1)}{n^x} \right] \cdot \left[\lim_{n \rightarrow +\infty} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^x} \right] \cdot \frac{\lambda^x}{x!} = 1 \cdot \frac{e^{-\lambda}}{1} \cdot \frac{\lambda^x}{x!}$$

Per cui la **distribuzione delle probabilità poissoniana** è rappresentabile tramite l'equazione parametrica:

$$p(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

Come dimostrato si tratta di un caso particolare della distribuzione binomiale: quando cioè il numero di prove n è molto grande e contemporaneamente la probabilità p di successo in una singola prova è molto piccola; per questo motivo è detta **legge degli eventi rari**, poiché la probabilità che l'evento si verifichi è estremamente bassa. È chiamata anche legge dei piccoli numeri, in quanto la frequenza assoluta di questi eventi è espressa da un numero piccolo, anche in un numero elevato di prove.

In termini di momenti si ha:

$$\begin{aligned}\mu &= \lambda \\ \sigma^2 &= \lambda\end{aligned}$$

La distribuzione di Pascal, quella binomiale negativa e quella geometrica

Se si considera una serie di prove indipendenti, ciascuna con una probabilità di successo p e si indica con x la prova in cui si ottiene il successo numero r , allora la distribuzione delle probabilità è spesso rappresentabile tramite l'equazione:

$$P(x) = \binom{x-1}{r-1} \cdot p^r \cdot (1-p)^{x-r}$$

Un caso particolare si ha quando $r > 0$, ma non sia necessariamente un intero, in questo caso si parla di: **distribuzione binomiale negativa**.

Nella distribuzione binomiale viene fissata la dimensione del campione e s'ottiene il numero di successi (o insuccessi), nella distribuzione binomiale negativa si fissa il numero di successi (o insuccessi) necessari e s'ottiene la numerosità del campione richiesto.

Un altro caso particolare è quello in cui $r=1$, in questo caso si parla di: **distribuzione geometrica**.

Si tratta del numero di prove richieste per giungere al primo successo (insuccesso).

Le distribuzioni continue

Un processo di fabbricazione produce migliaia d'oggetti al giorno che devono avere una determinata resistenza. In media una certa percentuale di questi oggetti non resiste correttamente (non è conforme alle specifiche). Ogni ora si preleva un certo numero di oggetti per effettuare un'ispezione che, tramite una prova distruttiva, ne misuri la caratteristica.

In questo caso la variabile casuale x rappresenta un valore che si distribuisce in maniera continua all'interno di un intervallo.

Detta $f(x)$ la funzione che esprime le frequenze in termini di probabilità, si ha:

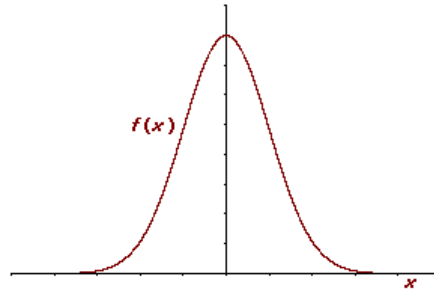
$$\begin{aligned}\mu &= \int_{-\infty}^{+\infty} x \cdot f(x) dx \\ \sigma^2 &= \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f(x) dx \\ \gamma^3 &= \frac{\int_{-\infty}^{+\infty} (x - \mu)^3 \cdot f(x) dx}{\sigma^3}\end{aligned}$$

La distribuzione normale o di Gauss

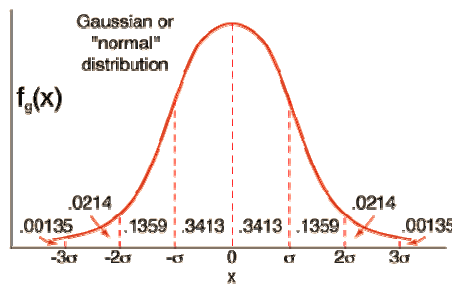
In un processo caratterizzato da una variabile casuale continua che è soggetta solo a variabilità di tipo naturale la distribuzione delle probabilità è rappresentabile tramite l'equazione:

$$P(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}}$$

La rappresentazione grafica di questo tipo di distribuzione è quello di una curva simmetrica, unimodale a campana.



Su tale figura la deviazione standard ha anche un significato geometrico, poiché i due flessi della campana si trovano alla distanza di una deviazione standard dalla media.



Un esempio tipico di applicazione di questo modello in controllo qualità è quello di un prodotto che debba rispettare una caratteristica di resistenza. Tramite un esperimento si rileva un valor medio μ e una deviazione standard σ . La specifica è del tipo $x \geq a$. la probabilità che il prodotto non soddisfi le specifiche è:

$$P\{x \geq a\} = 1 - P\{x < a\}$$

Parametrizzando la variabile casuale:

$$z = \frac{a - \mu}{\sigma}$$

è possibile ottenere il valore cercato, esaminando **le tavole di distribuzione della probabilità per la curva normalizzata standard**.

Talvolta è invece necessario effettuare il processo contrario. Data una certa probabilità di non conformi si ricerca il valore nominale che si deve avere come obiettivo. In questo caso dalle tavole si ricava il parametro z dal quale poi:

$$\mu = a - \sigma \cdot z$$

Infine, può essere necessario individuare un valore che assicuri una certa probabilità di non conformi (in pratica si cerca di introdurre una tolleranza in grado di assicurare una certa affidabilità).

In questo caso dalle tavole si ricava il parametro z dal quale poi:

$$a = \mu - \sigma \cdot z$$

La distribuzione lognormale

$$P(x) = \frac{1}{x \cdot \omega \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{[\ln(x) - \vartheta]^2}{2 \cdot \omega^2}}$$

La distribuzione esponenziale

$$P(x) = \lambda \cdot e^{-\lambda \cdot x}$$

La distribuzione gamma

$$P(x) = \frac{\lambda}{\Gamma(r)} \cdot (\lambda \cdot x)^{r-1} \cdot e^{-\lambda \cdot x}$$

La distribuzione di Weibull

$$P(x) = \frac{\beta}{\vartheta} \cdot \left(\frac{x}{\vartheta}\right)^{\beta-1} \cdot e^{-\left(\frac{x}{\vartheta}\right)^\beta}$$