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
- Campioni si nasce

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 $x \in \{1; 2; 3; 4; \ldots\}$ dà luogo ad una distribuzione discreta.
Detta $p(\mathbf{x})$ la funzione che esprime le probabilità, in termini di momenti si ha:

$$\mu = \sum_{i=1}^{\infty} \left[x_i \cdot p(x_i) \right]$$

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

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

La distribuzione di Bernoulli e binomiale

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

- il risultato di ogni prova può essere solo "successo[1]" 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 quindi n-x insuccessi è data dall'equazione:

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

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·p·q   p·q·p   q·p·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.

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

Questa distribuzione, per la presenza del termine binomiale, assume anche il nome di **distribuzione binomiale**.
Se si sommano tutte le probabilità si ottiene:

$$\sum_{x=0}^{n} \Pr(x) = \sum_{x=0}^{n} \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} = [p + (1-p)]^n = 1$$

come deve essere per una distribuzione di probabilità.
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:

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

Questo concetto è alla base del **paradosso di Borel**: se un esperimento può essere ripetuto infinite volte nelle stesse condizioni, a furia di provare, una 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 verifiche 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 a 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% (ovvero su tutta la popolazione). Infatti, si deve tenere ben presente che il controllo al 100% non viene effettuare 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:

$$\mu = n \cdot p$$
$$\sigma^2 = n \cdot p \cdot (1-p)$$

Un tipico esperimento di Bernulli e quello di mischiare le carte di un mazzo nuovo per vedere se tagliando il mazzo di carte mischiate esca o no un asso di picche. Ogni test effettuato su un mazzo risponde ai requisiti di un esperimento semplice di Bernulli. Si ottengono così delle tabelle di distribuzione delle probabilità come la seguente:

x	Pr(x)
0	$p^0 \cdot q^n$
1	$p^1 \cdot q^{n-1}$
...	...
n	$p^n \cdot q^0$

Se si considera il caso della possibilità di ottenere un sette lanciando due dadi si ha:

- **N**=36
- ***n***=6
- ***p***≈16,67%

da cui:

N	36	popolazione	
p	16,67%	probabilità	
p·N=	6	casi favorevoli	
n	6	prove	
μ=	1,000	σ²=	0,833467
σ=	0,913	γ=	0,00

x	P(x)	x·P(x)	(x-μ)²·P(x)	f(x)
0	33,5%	0,000	0,334952	23,98%
1	40,2%	0,402	0,000000	43,70%
2	20,1%	0,402	0,200907	23,99%
3	5,4%	0,161	0,214395	3,97%
4	0,8%	0,032	0,072380	0,20%
5	0,1%	0,003	0,010297	0,00%
6	0,0%	0,000	0,000536	0,00%

NOTA: quando ***n***>69 è un problema determinare ***n***! mediante le calcolatrici, per ovviare a questo problema si può utilizzare la forma approssimata di:

$$n! \approx n^n \cdot e^{-n} \cdot \sqrt{2 \cdot \pi \cdot n}$$

oppure ricorrere alla distribuzione di Poisson.

La distribuzione ipergeometrica

Si supponga di disporre di una popolazione finita di **N** elementi. Un certo numero **C**=***p***·**N** di questi (**C≤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:

$$\Pr(x) = \frac{\begin{pmatrix} C \\ x \end{pmatrix} \cdot \begin{pmatrix} N-C \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} = \frac{\begin{pmatrix} p \cdot N \\ x \end{pmatrix} \cdot \begin{pmatrix} q \cdot N \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

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 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.
Se si sommano tutte le probabilità si ottiene:

$$\sum_{x=0}^{n} \Pr(x) = \frac{\sum_{x=0}^{n} \binom{N \cdot p}{x} \cdot \binom{N \cdot q}{n-x}}{\binom{N}{n}} = \frac{\binom{N \cdot p + N \cdot q}{n}}{\binom{N}{n}} = 1$$

come deve essere per una distribuzione di probabilità.
Al crescere di **N** la distribuzione ipergeometrica tende a quella binomiale. Infatti, se si pone il limite N→∞ si ottiene:

$$\lim_{N \to \infty} \frac{\binom{p \cdot N}{x} \cdot \binom{q \cdot N}{n-x}}{\binom{N}{n}} =$$

$$= \lim_{N \to \infty} \frac{n!}{x! \cdot (n-x)!} \cdot \frac{N \cdot p \cdot (N \cdot p - 1) \cdot \ldots \cdot (N \cdot p - x + 1)}{N \cdot (N-1) \cdot \ldots \cdot (N - x + 1)} \cdot \frac{N \cdot q \cdot (N \cdot q - 1) \cdot \ldots \cdot (N \cdot q - n + x + 1)}{N \cdot (N-1) \cdot \ldots \cdot (N - x + 1)} =$$

$$= \lim_{N \to \infty} \binom{n}{x} \cdot \underbrace{p \cdot \frac{\left(p - \frac{1}{N}\right)}{\left(1 - \frac{1}{N}\right)} \cdot \ldots \cdot \frac{\left(p - \frac{x}{N} + \frac{1}{N}\right)}{\left(1 - \frac{x}{N} + \frac{1}{N}\right)}}_{x \text{ volte}} \cdot \underbrace{\frac{q}{\left(1 - \frac{1}{N}\right)} \cdot \frac{\left(q - \frac{1}{N}\right)}{\left(1 - \frac{x}{N} - \frac{1}{N}\right)} \cdot \ldots \cdot \frac{\left(q - \frac{n}{N} + \frac{x}{N} + \frac{1}{N}\right)}{\left(1 - \frac{n}{N} + \frac{1}{N}\right)}}_{(n-x) \text{ volte}} =$$

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

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).

$$\mu = n \cdot p$$

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

Si ottengono così delle tabelle di distribuzione delle probabilità come la seguente:

x	**Pr(x)**
0	$\binom{n}{0} \cdot p^0 \cdot q^n$
1	$\binom{n}{1} \cdot p^1 \cdot q^{n-1}$
...	...
n	$\binom{n}{n} \cdot p^n \cdot q^0$

Se si considera il caso della possibilità di ottenere delle risposte positive ad un invito chiedendo di uscire a quattro ragazze su cento sapendo che il 40% risponderebbe positivamente:

- **N**=100
- ***n***=4

- p=40%

da cui:

N=	100			
p·N=	40	p=	40,0%	
q·N=	60	q=	60,0%	
n=	4			
μ=	1,600	σ²=	0,930909	
σ=	0,965			
x	P(x)	x·P(x)	(x-μ)²·P(x)	f(x)
0	12,4%	0,000	0,318356	10,45%
1	34,9%	0,349	0,125667	34,08%
2	35,2%	0,704	0,056333	37,94%
3	15,1%	0,454	0,296307	14,43%
4	2,3%	0,093	0,134245	1,87%

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:

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

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

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

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

$$\lim_{n\to+\infty}\Pr(x)=\lim_{n\to+\infty}\left[\frac{n\cdot(n-1)\cdot\ldots\cdot(n-x+1)}{n^x}\cdot\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^x}\cdot\frac{\lambda^x}{x!}\right]$$

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

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

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

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

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. É 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.
Se si sommano tutte le probabilità si ottiene:

$$\sum_{x=0}^{n} \Pr(x) = e^{-\lambda} \cdot \sum_{x=0}^{n} \frac{\lambda^x}{x!}$$

Se effettuiamo il limite della sommatoria per ***n***→+∞ si ha

$$\lim_{n\to+\infty} \sum_{x=0}^{n} \frac{\lambda^x}{x!} = e^{\lambda} \Rightarrow \sum_{x=0}^{n} \Pr(x) = e^{-\lambda} \cdot e^{\lambda} = 1$$

come deve essere per una distribuzione di probabilità.
Se si analizza la distribuzione delle probabilità poissoniana in termini di momenti si ha:

$$\mu = \lambda$$
$$\sigma^2 = \lambda$$

Se si considera il caso che non si ammetta una difettosità maggiore del 3% sul'interno di un lotto e si voglia verificare la bontà del lotto tramite un campione di cento pezzi:

- ***n***=100
- ***p***=3%
- λ=3

da cui[2]:

p	3,00%		probabilità		
n	100		prove		
μ=	3,000	σ²=	3,00000		
σ=	1,732	λ=	3,000		
	x	P(x)	x·P(x)	(x-μ)²·P(x)	f(x)
	0	4,978707%	0,000	0,448084	5,14%
	1	14,936121%	0,149	0,597445	11,83%
	2	22,404181%	0,448	0,224042	19,50%
	3	22,404181%	0,672	0,000000	23,03%
	4	16,803136%	0,672	0,168031	19,50%
	5	10,081881%	0,504	0,403275	11,83%
	6	5,040941%	0,302	0,453685	5,14%
	7	2,160403%	0,151	0,345665	1,60%
	8	0,810151%	0,065	0,202538	0,36%

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:

[2] S'interrompe nel caso in esame la tabella per valori inferiori all'1% circa.

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

Un caso particolare si ha quando ***r***>0, ma non sia necessariamente un intero, in questo caso si parla di: **distribuzione binomiale negativa**, si tratta del numero di prove richieste per giungere dopo ***x*** insuccessi (successi) al ***n*** successi (insuccessi) quando la probabilità di successo (insuccesso) è ***p***.
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 dopo ***x*** insuccessi (successi) al primo successo (insuccesso) quando la probabilità di successo (insuccesso) è ***p***.

$$\Pr(x) = p^x \cdot (1-p)$$

Le distribuzioni continue

Un processo di fabbricazione produce migliaia d'oggetti al giorno che devono avere una determinata resistenza o dimensione. In media una certa percentuale di questi oggetti non resiste correttamente (non è conforme alle specifiche) o non possiede la caratteristica dimensionale richiesta.
Per verificare ciò, ogni ora si preleva un certo numero d'oggetti per effettuare un'ispezione che, tramite una prova distruttiva o meno, ne misuri la caratteristica.
In questo caso la variabile casuale ***x*** può assumere qualsiasi valore appartenente a una distribuzione continua che costituisce l'intervallo di valori possibili che può assumere la variabile casuale (ad esempio, tutti i valori compresi tra 40 e 50).
Detta ***f***(**x**) la funzione che esprime le frequenze in termini di probabilità, si ha:

$$\mu = \int_{-\infty}^{+\infty} x \cdot f(x)\, dx$$

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x)\, dx$$

$$\gamma^3 = \frac{\int_{-\infty}^{+\infty} (x-\mu)^3 \cdot f(x)\, dx}{\sigma^3}$$

La distribuzione normale standardizzata

Come premessa all'argomento è utile richiamare il valore del seguente integrale:

$$\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}}\, dz = \sqrt{2 \cdot \pi}$$

$$\int_{-\infty}^{+\infty} z^2 \cdot e^{-\frac{z^2}{2}}\, dz = \sqrt{2 \cdot \pi}$$

Si supponga di voler descrivere l'andamento degli errori accidentali.

$$u = \frac{x-m}{\sigma}$$

Questo tipo di curva deve rispettare alcune caratteristiche:

- il valore medio deve essere nullo

$$_0\mu_1 = \int_{-\infty}^{+\infty} u \cdot f(u)\, du = 0$$

- per cui la funzione $f(u)$ deve essere positiva e simmetrica rispetto all'origine

$$f(u) = f(-u) > 0$$

- si deve annullare agli estremi, asintoticamente

$$\lim_{u \to \pm\infty} f(u) = 0$$

- deve possedere un massimo per u=0:

$$\begin{cases} f'(0) = 0 \\ f''(0) < 0 \end{cases}$$

- deve essere monotona crescente per u<0 e monotona decrescente per u>0:

$$\begin{cases} u < 0 \Rightarrow f'(u) > 0 \\ u > 0 \Rightarrow f'(u) < 0 \end{cases}$$

Il tipo di funzione che soddisfa queste esigenze è del tipo:

$$f'(u) = -u \cdot f(u) \Rightarrow \begin{cases} f(u) = k \cdot e^{-\frac{u^2}{2}} \\ k > 0 \end{cases}$$

Per determinare k basta applicare il fatto che l'integrale da -∞ a +∞ deve essere pari a 1:

$$\int_{-\infty}^{+\infty} f(u)\, du = \int_{-\infty}^{+\infty} k \cdot e^{-\frac{u^2}{2}}\, du = k \cdot \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}}\, du = k \cdot \sqrt{2 \cdot \pi} = 1$$

da cui:

$$k = \frac{1}{\sqrt{2 \cdot \pi}}$$

quindi:

$$\int f(u)\, du = \int \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{u^2}{2}}\, du$$

da cui:

$$_0\mu_2 = \int_{-\infty}^{+\infty} u^2 \cdot f(u)\, du = \int_{-\infty}^{+\infty} \frac{u^2}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{u^2}{2}}\, du = 1 \Rightarrow Var(u) = {}_0\mu_2 - {}_0\mu_1^2 = 1$$

Rammentando l'espressione di u si ha:

$$du = \frac{d}{dx}\frac{x-m}{\sigma}\,dx = \frac{dx}{\sigma}$$

da cui:

$$\int f(u)\,du = \int \phi(x)\,dx = \int \frac{1}{\sigma\sqrt{2\cdot\pi}}\cdot e^{-\frac{(x-m)^2}{2\cdot\sigma^2}}\,dx$$

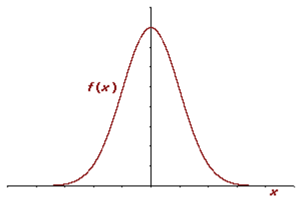
nota come distribuzione normale o di Gauss

La distribuzione normale o di Gauss

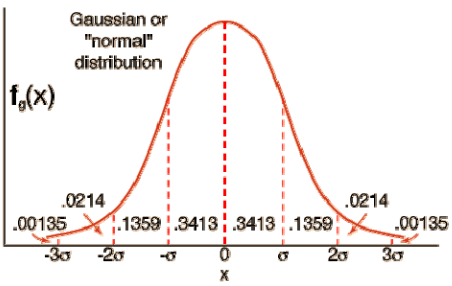
In un processo caratterizzato da una variabile casuale continua che è soggetta solo a variabilità di tipo naturale la distribuzione delle probabilità è rappresentabile tramite l'equazione:

$$\Pr(x) = \int \frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot e^{-\frac{(x-m)^2}{2\cdot\sigma^2}}\,dx$$

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:

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

Parametrizzando la variabile casuale:

$$\begin{cases} z = \dfrac{a-\mu}{\sigma} \\ \sigma = 1 \end{cases}$$

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

$$\Pr(x) = \frac{1}{\sqrt{2\cdot\pi}} \cdot e^{-\frac{z^2}{2}}$$

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

La distribuzione lognormale è:

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

$$\mu = e^{\vartheta+\frac{\omega^2}{2}}$$

$$\sigma^2 = \left(e^{\omega^2}-1\right)\cdot e^{2\cdot\vartheta+\omega^2}$$

con

$$x = \ln(w)$$

dove ***w*** è una variabile casuale avente distribuzione normale.

La distribuzione esponenziale

La **distribuzione esponenziale** è:

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

$$\mu = \frac{1}{\lambda}$$

$$\sigma^2 = \frac{1}{\lambda^2}$$

dove λ>0 è una costante.
Conseguentemente la distribuzione cumulata risulta essere:

$$F(a) = \int_0^a \lambda \cdot e^{-\lambda \cdot x}\, dx$$

La funzione esponenziale è molto utilizzata nel campo dell'affidabilità dei sistemi. In questa applicazione il parametro $\boldsymbol{\lambda}$ assume il significato di **tasso di guasto** del sistema, mentre la media rappresenta il valore del **tempo medio al guasto**.
È evidente che si può individuare un legame tra tale funzione e la distribuzione di Poisson:

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

per $\boldsymbol{x}$=0:

$$\Pr(0) = e^{-\lambda \cdot t}$$

Che è la probabilità che nell'intervallo [0; $\boldsymbol{t}$] non si realizzi l'evento. Ora:

$$F(t) = 1 - e^{-\lambda \cdot t} \Rightarrow f(t) = \frac{dF(t)}{dt} = \lambda \cdot e^{-\lambda \cdot t}$$

La distribuzione gamma

La **distribuzione gamma** è una distribuzione di probabilità continua (essa descrive anche la distribuzione esponenziale e la distribuzione del chi quadrato).
Tale distribuzione di probabilità è definita sui numeri reali non negativi ($\boldsymbol{x}\geq 0$), e solitamente è parametrizzata in due modi diversi tramite una coppia di numeri positivi: ($\boldsymbol{k}$, $\boldsymbol{\theta}$) o ($\boldsymbol{\alpha}$, $\boldsymbol{\beta}$)=($\boldsymbol{k}$, 1/$\boldsymbol{\theta}$).
La sua funzione di densità di probabilità è:

$$\Pr(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\beta \cdot x} \Leftrightarrow \begin{cases} \alpha > 0 \\ \beta > 0 \\ x \geq 0 \end{cases}$$

$$\mu = \frac{\alpha}{\beta}$$

$$\sigma^2 = \frac{\alpha}{\beta^2}$$

dove, posto $\boldsymbol{\alpha}$>0, si ha:

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} \cdot e^{-x}\, dx$$

detta **funzione gamma** o integrale euleriano[3] di seconda specie.
Usando l'integrazione per parti è possibile dimostrare che:

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} \cdot e^{-x}\, dx = \left[-x^{\alpha-1} e^{-x}\right]_0^{+\infty} + (\alpha-1) \cdot \int_0^{+\infty} x^{\alpha-2} \cdot e^{-x}\, dx = (\alpha-1) \cdot \Gamma(\alpha-1)$$

Da tale definizione derivano le seguenti proprietà per $\boldsymbol{\alpha} \in \aleph^+$ (cioè intero):

[3] Dal grande matematico svizzero Leonardo Eulero (1707-1783) che introdusse questa funzione.

$$\Gamma(\alpha) = (\alpha - 1)! \Rightarrow \Gamma(1) = 1$$

I valori assunti dalla funzione gamma per $\boldsymbol{\alpha} \in [1, 2]$
È utile ricordare che:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Viene utilizzata come modello per i tempi di attesa nella teoria delle code, mentre nella statistica bayesiana è comune come distribuzione a priori e a posteriori.
Posto $\boldsymbol{\alpha}$=1 e $\boldsymbol{\beta}$=$\boldsymbol{\lambda}$ si ha:

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

$$\mu = \frac{1}{\lambda}$$

$$\sigma^2 = \frac{1}{\lambda^2}$$

da cui emerge che la distribuzione esponenziale non è altro che un caso particolare della distribuzione gamma.
Sia i valori della funzione gamma, sia quelli della distribuzione gamma sono riportate in apposite tabelle.

La distribuzione chi-quadrato

La **distribuzione chi-quadrato** descrive la somma dei quadrati di alcune variabili aleatorie indipendenti aventi distribuzione normale standard. In statistica viene particolarmente utilizzata per l'omonimo test di verifica d'ipotesi per verificare l'accostamento di distribuzioni di frequenze osservate (reali) con il modello teorico di riferimento (teoriche).
Posto

$$x = \sum_{i=1}^{n} u_i^2$$

allora:

$$\Pr(x) = \frac{1}{2^{\frac{\nu}{2}} \cdot \Gamma\left(\frac{\nu}{2}\right)} \cdot x^{\frac{\nu}{2}-1} \cdot e^{-\frac{x}{2}} \Leftrightarrow \begin{cases} x \geq 0 \\ \nu \geq 1 \end{cases}$$

$$\mu = \nu$$

$$\sigma^2 = 2 \cdot \nu$$

Si nota che posto $\boldsymbol{\alpha}$= $\boldsymbol{\nu}$/2 e $\boldsymbol{\beta}$=1/2 la distribuzione chi-quadro non è altro che un caso particolare della distribuzione gamma.
I valori $\boldsymbol{\alpha}$ che individuano la coda destra della distribuzione chi-quadro sono riportate in apposite tabelle in funzione del valore $\boldsymbol{\nu}$ che rappresenta i gradi di libertà per quanto concerne gli addendi $\boldsymbol{u}_i$ che possono essere scelti indipendentemente.
In statistica la distribuzione chi-quadro viene utilizzata anche per verificare, tramite test d'ipotesi, la stima di una varianza.

La distribuzione t di Student

La **distribuzione *t* di Student**[4] descrive il comportamento tra due variabili aleatorie, la prima proveniente da una popolazione avente distribuzione normale e varianza ignota, la seconda il cui quadrato ha distribuzione chi quadrato.
Rammentando che:

$$\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$$

allora:

$$\Pr(x)=\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{\nu}\cdot\Gamma\left(\frac{1}{2}\right)\cdot\Gamma\left(\frac{\nu}{2}\right)}\cdot\left(1-\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}=\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{\nu\cdot\pi}\cdot\Gamma\left(\frac{\nu}{2}\right)}\cdot\left(1-\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

$$\mu=0$$

$$\sigma^2=\frac{\nu}{\nu-2}\Leftrightarrow\nu>2$$

Si nota che:

- se i gradi di libertà $\boldsymbol{\nu}$ sono finiti, risulta sempre $\boldsymbol{\sigma}^2>1$, per cui la variabile ***t*** di Student risulta essere più dispersa della normale standardizzata;
- la varianza converge all'unità per $\boldsymbol{\nu}\rightarrow+\infty$, manifestando la stessa variabilità della curva normale standardizzata.

Ne deriva che questa distribuzione, pur essendo campanulare e simmetrica, differisce dalla normale standardizzata perché ha una forma più allungata e le code più ingrossate (iponormale), in termini pratici ciò implica una minor concentrazione dei valori attorno alla media e, di conseguenza, una minor precisione. Tali differenze si attenuano al crescere di ***n*** e, conseguentemente, dei gradi di libertà $\boldsymbol{\nu}$.
I valori $\boldsymbol{\alpha}$ che individuano la coda destra della distribuzione ***t*** di Student sono riportate in apposite tabelle in funzione del valore $\boldsymbol{\nu}$ che rappresenta i gradi di libertà.

La distribuzione F di Fisher-Snedecor

La distribuzione ***F*** di Fisher-Snedecor è:

$$\Pr(x)=$$

$$\mu=$$

$$\sigma^2=$$

La distribuzione di Fréchet

La distribuzione di Fréchet è una distribuzione di probabilità continua definita sui numeri reali positivi ed è e descritta tramite il parametri (reale, non nullo e positivi) $\boldsymbol{\beta}$:

[4] La distribuzione venne descritta nel 1908 da William Sealy Gosset, che pubblicò il suo risultato sotto lo pseudonimo "*Student*" perché la birreria presso la quale era impiegato vietava ai propri dipendenti di pubblicare articoli affinché questi non divulgassero segreti di produzione.

$$\Pr(x) = \beta \cdot x^{-\beta-1} \cdot e^{-x^{-\beta}}$$
$$\mu = \Gamma\left(1+\frac{1}{\beta}\right)$$
$$\sigma^2 = \Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2\left(1+\frac{1}{\beta}\right)$$

La distribuzione di Gumbel

La distribuzione di Gumbel

$$\Pr(x) =$$
$$\mu =$$
$$\sigma^2 =$$

La distribuzione di Rayleigh

La distribuzione di Rayleigh

$$\Pr(x) =$$
$$\mu =$$
$$\sigma^2 =$$

La distribuzione di Weibull

La distribuzione di Weibull[5] è una distribuzione di probabilità continua definita sui numeri reali positivi e descritta tramite due parametri (reali, non nulli e positivi):

- $\boldsymbol{\beta}$, detto **parametro di forma**;
- $\boldsymbol{\vartheta}$, detto **parametro di posizione**.

Tale distribuzione, di tipo esponenziale, è derivabile direttamente dalla distribuzione di Gumbel tramite una trasformazione della variabile aleatoria di Gumbel nella forma utilizzata per i valori minimi.
La caratteristica peculiare che ha reso particolarmente nota questo tipo di distribuzione è l'ampia gamma di conformazioni (quindi l'ampia adattabilità a svariate situazioni sperimentali) che può assumere grazie al parametro di forma.

$$\Pr(x) = \frac{\beta}{\vartheta} \cdot \left(\frac{x}{\vartheta}\right)^{\beta-1} \cdot e^{-\left(\frac{x}{\vartheta}\right)^{\beta}} = \frac{\beta}{\vartheta^{\beta}} \cdot x^{\beta-1} \cdot e^{-\left(\frac{x}{\vartheta}\right)^{\beta}}$$
$$\mu = \vartheta \cdot \Gamma\left(1+\frac{1}{\beta}\right)$$
$$\sigma^2 = \vartheta^2 \cdot \left[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2\left(1+\frac{1}{\beta}\right)\right]$$

[5] Prende il nome dal matematico svedese Waloddi Weibull che la descrisse nel 1951, la distribuzione era comunque stata già trattata dal matematico francese Maurice Fréchet nel 1927.

Questa distribuzione è particolarmente utile per descrivere sistemi con tasso di guasto variabile nel tempo, ovvero come estensione della distribuzione esponenziale che, invece, 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 du**rata di vita per un fenomeno la cui probabilità di morire può variare nel tempo**, in funzione di β, in particolare:

- $\beta<0$, il tasso di guasto diminuisce nel tempo (fenomeni ad alta "*mortalità infantile*");
- $\beta=0$, il tasso di guasto è invariante nel tempo (fenomeni con "*mancanza di memoria*");
- $\beta>0$, il tasso di gusto aumenta con il tempo (fenomeni con "*morte per invecchiamento*").

Da tutto ciò deriva che la distribuzione di Weibull risulta molto utile negli ambienti in cui si effettua l'analisi dei guasti, l'analisi di sopravvivenza, l'affidabilità e il controllo della qualità.
Infine, questa distribuzione di probabilità viene anche adoperata nella meteorologia per le previsioni del tempo, in tal caso come generalizzazione della distribuzione di Rayleigh.

Campioni si nasce

Seppure la scelta della tecnica e l'utilizzo delle formule appaiano spesso difficoltose, una volta acquisita la pratica ci s'accorge che il vero problema di un controllo statistico è quello di determinare le dimensioni del campione, poiché tale dimensione influisce sull'affidabilità o meno del risultato estratto.

Analisi statistica per attributi

Poniamoci quindi di fronte ad un'**analisi statistica per attributi**. In essa l'attributo cercato può presentarsi con una probabilità p e non presentarsi con una probabilità $q=(1-p)$. Definiamo p' il risultato che invece emerge da un campionamento, ne consegue che il valore $\varepsilon(p')=|p'-p|$ rappresenta l'errore da cui è affetta la stima effettuata.

$$p'-E(p') < p < p'+E(p')$$

È ovvio e opportuno che in un campionamento sia possibile scegliere una dimensione n del campione tale da assicurare che $\Pr(|p'-p|<\varepsilon)=1$.
Ora, la distribuzione di p' per campioni limitati può essere definita tramite un semplice schema:

		Modalità di campionamento	
		con reinserimento	**senza reinserimento**
Popolazione	**finita**	binomiale	ipergeometrica[6]
	infinita	binomiale	binomiale

Da tale tabella nasce la possibilità di definire una prima condizione che permetta di svincolare il campione dalle caratteristiche di **limitata numerosità**, per condurlo alle condizioni di normalità tipica dei **grandi campioni**. Tale obiettivo si può ottenere facilmente ponendo la condizione $n \cdot p \cdot q>5$.
Introdotta questa premessa, è possibile asserire che nei grandi campioni provenienti da popolazione infinita la frequenza relativa campionaria si distribuisce con media $M(p')$ e varianza $V(p')$:

[6] Detta anche **estrazione in blocco**.

$$M(p') = p$$
$$E^2(p') = V(p') = \frac{p \cdot q}{n}$$

nei grandi campioni provenienti da popolazione finita la frequenza relativa campionaria si distribuisce invece con media $M(\boldsymbol{p}')$ e varianza V(p'):

$$M(p') = p$$
$$E^2(p') = V(p') = \frac{p \cdot q}{n} \cdot \frac{N-n}{N-1}$$

dove il rapporto

$$\kappa = \sqrt{\frac{N-n}{N-1}}$$

è definito **fattore di correlazione per popolazioni finite**.
Rammentando che $\boldsymbol{q}=(1-\boldsymbol{p})$ e posto che:

$$\lim_{N \to \infty} \kappa = 1$$

possiamo generalizzare le precedenti formule scrivendo più semplicemente:

$$E^2(p') = V(p') = \frac{p - p^2}{n} \cdot \kappa^2$$

Tale funzione è una parabola che nell'intervallo d'esistenza $\boldsymbol{p} \in [0;\ 1]$ è positiva. In tale intervallo essa presenta un massimo:

$$\frac{d}{dp} V(p') = \frac{1 - 2 \cdot p}{n} \cdot \kappa^2 = 0 \Rightarrow p = \frac{1}{2} = 50\%$$

Ciò significa che la condizione più sfavorevole per la scelta della dimensione del campione ***n*** è collegata a ***p***=50%. Tale valore risulta plausibile se ***p*** è incognita, tuttavia se ***p*** è nota (o si ha un'idea della sua dimensione), per ragioni di economicità si sceglie la dimensione del campione ipotizzando un valore pessimistico arrotondato per eccesso o difetto in modo che sia orientato verso il valore 50%.