

# DISTRIBUZIONI DI PROBABILITÀ PER DATI CATEGORIALI

L'identificazione della distribuzione di probabilità è lo step iniziale  
dell'inferenza statistica

[debora.giovannelli@gmail.com](mailto:debora.giovannelli@gmail.com)

*Debora Giovannelli*

## Sommario

<b>DISTRIBUZIONI DI PROBABILITÀ PER DATI CATEGORIALI .....</b>	<b>2</b>
IDENTIFICAZIONE DELLA DISTRIBUZIONE DI PROBABILITÀ COME PRIMO STEP DELL'ANALISI INFERENZIALE .....	2
SVILUPPO DEI METODI STATISTICI PER DATI CATEGORIALI .....	3
SCALE DI MISURA PER VARIABILI CATEGORIALI .....	3
<i>Scale nominali, ordinali e intervallari</i> .....	4
DISTRIBUZIONI DELLE PRINCIPALI VARIABILI CASUALI CATEGORIALI .....	6
<i>Distribuzione di Bernoulli</i> .....	7
<i>Distribuzione binomiale</i> .....	7
Considerazioni preliminari .....	7
Lo schema di campionamento binomiale .....	8
Distribuzione binomiale - Funzione Densità di Probabilità, valore atteso e varianza .....	9
Assunzioni per una distribuzione binomiale .....	9
Proprietà .....	10
<i>Distribuzione ipergeometrica</i> .....	11
Un esempio di applicazione – il Test esatto di Fisher .....	12
<i>Distribuzione di Poisson</i> .....	13
Assunzioni .....	14
Proprietà .....	14
Stima della media campionaria .....	15
Il fenomeno dell'Overdispersione ( <i>overdispersion</i> ) .....	16
<i>Distribuzione multinomiale</i> .....	16
Assunzioni .....	17
Proprietà .....	17
Spazio parametrico del parametro $\pi$ .....	18
Lancio di un dado in un intervallo di tempo prefissato – multinomiale o Poisson? .....	19
RIFERIMENTI .....	21
APPENDICE 1 – WORKFLOW PER L'IDENTIFICAZIONE DELLA DISTRIBUZIONE DI PROBABILITÀ PER VARIABILI CATEGORIALI IN ANALISI INFERENZIALE .....	22
APPENDICE 2 – PMF, $E(X)$ E $VAR(X)$ PER LE DISTRIBUZIONI BERNOULLI, BINOMIALE, POISSON E MULTINOMIALE .....	23

# Distribuzioni di Probabilità per Dati Categoriali

## ***Identificazione della distribuzione di probabilità come primo step dell'analisi inferenziale***

Il primo step alla base dell'analisi inferenziale è quello di identificare la distribuzione di Probabilità, assumendo che i dati campionari siano stati generati casualmente da una determinata distribuzione. Prima di effettuare qualunque analisi statistica sui dati dobbiamo fare, cioè, delle assunzioni sul meccanismo casuale o modello di campionamento che ha generato i dati.

Nel caso, ad esempio, di una variabile casuale continua, per effettuare un test delle ipotesi tipo un *1-sample t test*, dobbiamo assumere che la variabile in studio sia normo-distribuita.

Nel caso, invece, di una variabile discreta o categoriale, possiamo assumere come distribuzione di riferimento ad esempio una Bernoulli, una binomiale, una Poisson, oppure una multinomiale.

In ambito categoriale, per potersi orientare verso una distribuzione specifica, è fondamentale definire ed analizzare lo schema di campionamento o *sampling scheme*.

Lo schema di campionamento rappresenta la prospettiva attraverso la quale stiamo campionando i dati.

La definizione del metodo o schema di campionamento è molto importante perché determina il modello di riferimento (la distribuzione) per le procedure inferenziali successive.

Anche se queste distribuzioni sono accomunate dallo stesso metodo per la stima dei parametri, cioè il metodo di Massima Verosimiglianza o *Maximum Likelihood Estimation Method* (MLE), le formule di calcolo per gli stimatori dei parametri o *Maximum Likelihood Estimators* (MLEs) e per i relativi *Confidence Intervals* (CI) sono diverse.

Supponiamo di contare il numero di occorrenze di un evento. Nel caso, ad esempio, di un campionamento con un numero massimo di prove (*trials*)  $n$  prefissato (definisco un numero massimo di prove) è opportuno orientarsi verso una distribuzione binomiale (se la variabile casuale è categoriale dicotomica) o multinomiale (se la variabile casuale è categoriale politomica). Se, invece, il campionamento non definisce un numero massimo di prove (*unrestricted*) è opportuno orientarsi verso una distribuzione di Poisson.

Pertanto il primo step dell'analisi inferenziale è identificare la distribuzione di Probabilità che ha generato i dati campionari sulla base del metodo o modello di campionamento utilizzato.

Supponiamo, ad esempio, di considerare l'esperimento del lancio di un dado.

Ci sono due prospettive attraverso le quali possiamo considerare lo stesso esperimento:

1. lancio il dado tante volte in un tempo definito e conto quante volte esce ciascuna delle 6 facce
2. lancio il dado  $n$  volte e conto quante volte esce ciascuna delle 6 facce

Nel primo caso non è stato prefissato il numero massimo di lanci (numero di trials), ma è stata prefissata una *finestra temporale* entro cui contiamo le occorrenze per ciascun evento. Questo schema di campionamento corrisponde ad un campionamento di Poisson (*Poisson Sampling*) e, pertanto, la distribuzione di riferimento per la successiva procedura inferenziale sarà una distribuzione di Poisson.

Nel secondo caso è stato prefissato il numero massimo di lanci (*fixed total sample size*). Questo schema di campionamento comporta una distribuzione di riferimento binomiale o multinomiale.

### ***Sviluppo dei metodi statistici per dati categoriali***

A differenza dei metodi statistici per dati continui quelli per dati categoriali hanno avuto uno sviluppo più tardivo e frazionato.

Dopo l'importante lavoro svolto dallo statistico Britannico Karl Pearson nei primi anni del '900, i metodi statistici per l'analisi dei dati categoriali hanno avuto uno scarso sviluppo fino agli anni '60, sviluppo poi stimolato da studi di ricerca condotti nell'ambito delle scienze sociali e biomediche, ambiti nei quali le scale categoriali sono molto comuni (es. per misurare attitudini o opinioni, o per misurare l'esito di un determinato trattamento medico).

I dati di tipo categoriale sono, infatti, comuni in questi due ambiti, ma sono ampiamente diffusi anche in altri contesti, rendendo l'applicazione dei metodi statistici per dati categoriali di uso più esteso. Di seguito alcuni esempi:

- scienze sociali
- scienze biomediche
- scienze comportamentali (tipi di malattie mentali)
- epidemiologia
- salute pubblica
- genetica
- zoologia (cibo preferito di una specie)
- istruzione (esiti di questionari)
- marketing (brand selection)
- produzione industriale (esito QC)

### ***Scale di misura per variabili categoriali***

I dati categoriali sono dati di tipo discreto o qualitativo che derivano da misurazioni effettuate su una scala categoriale, assegnando le osservazioni, in termini di conteggi o proporzioni, ad una delle categorie, che possono essere o meno ordinate.

Per una scala categoriale i livelli o categorie possono essere valori numerici interi oppure valori non numerici.

Misurare in una scala categoriale significa andare a "**contare**" il numero di occorrenze (numero di volte che un determinato evento si verifica) per ciascuno dei livelli della scala.

Le diverse scale categoriali si differenziano per numero e tipo di categorie, come sintetizzato nella seguente tabella.

<i>numero di categorie (k)</i>	<i>valori delle categorie</i>	<i>tipo di scala</i>	<i>tipo di misura</i>	<i>variabile categoriale (x)</i>	<i>tipo di variabile</i>
k = 2	0,1	binaria	misuro l'esito (0,1) di una singola prova	di Bernoulli	
			misuro il numero di occorrenze dell'evento di riferimento (0 o 1) su n prove	binomiale	proporzione di successi (numero di successi su n prove)
k > 2	valori qualitativi non ordinati	nominale	misuro il numero di occorrenze per ciascuna delle categorie	multinomiale nominale	proporzione di ciascuna categoria
	valori numerici o qualitativi ordinati secondo un ordine naturale	ordinale	misuro il numero di occorrenze per ciascuna delle categorie	multinomiale ordinale	proporzione di ciascuna categoria
	valori raggruppati in categorie ordinate, separati da distanze numeriche (le categorie distinte sono intervalli e differiscono in quantità)	intervallare	misuro il numero di occorrenze per ciascuna delle categorie	multinomiale intervallare	proporzione di ciascuna categoria
k > 2	valori interi non negativi (0,1,2,...)	Poisson	misuro il numero di occorrenze per ciascuna delle categorie	Poisson	conteggi

Dalla tabella di cui sopra si vede come la classificazione di una variabile categoriale dipende dal modo con cui la variabile viene misurata (dal tipo di scala di misura).

L'identificazione della scala di misura di una variabile casuale categoriale e, pertanto la sua classificazione, è fondamentale al fine di identificare i metodi statistici più appropriati per approcciare un'analisi inferenziale.

### ***Scale nominali, ordinali e intervallari***

Una variabile categoriale è una variabile discreta con una scala di misura costituita da un set di categorie, livelli o modalità. A seconda del tipo di scala le variabili di risposta si possono distinguere in ***ordinali e nominali***.

Secondo il criterio di S. S. Stevens (1946), che ne ha proposto la prima definizione, si definiscono

- Scale Nominali: scale che attribuiscono numeri alle categorie senza implicazione di ordinalità
- Scale Ordinali: scale che attribuiscono alle categorie numeri implicando un ranking su un singolo attributo (ordinabilità unimodale)

La restrizione relativa al ranking limitato ad un singolo attributo (ranking monodimensionale) è stata nel tempo contestata da diversi autori, tanto che sono state trattate come ordinali variabili con categorie ordinabili su più attributi (es. in campo di "political party affiliation").

Questa distinzione è molto importante perché è alla base della scelta dell'adeguato metodo statistico.

Relativamente ai dati di tipo ordinale è essenziale sottolineare che, nonostante l'esistenza di un ranking tra i livelli, ***la distanza tra categorie adiacenti non è nota***.

L'ordinalità è, pertanto, una caratteristica che permette di utilizzare specifici metodi statistici per questo tipo di variabili, che ne sfruttano appunto la peculiarità, ma **non giustifica l'utilizzo di modelli lineari, che invece prevedono una distanza tra i livelli nota e costante.**

Occorre specificare che la classificazione di una variabile categoriale in nominale o ordinale dipende dal modo in cui la variabile è misurata.

Se ad esempio misuriamo il livello di istruzione distinguendo tra scuola pubblica e privata, la variabile in studio è di tipo nominale, mentre se lo misuriamo in una scala che va da "nessuna istruzione" a "specializzazione post-laurea", la variabile può essere trattata come una variabile ordinale.

Oltre alle scale nominali ed ordinali esistono anche le **scale intervallari**.

Le variabili categoriali che si misurano attraverso scale intervallari sono dette variabili intervallo.

Queste variabili sono caratterizzate dal fatto che i livelli sono costituiti da *range* di valori (ad esempio fasce di età, livelli della pressione sanguigna, fasce di reddito), oppure da pochi valori (numero di volte che un individuo si è sposato, numero di anni di istruzione).

Per orientarsi verso l'approccio statistico inferenziale più adeguato occorre considerare la scala di misura della variabile categoriale in gioco.

A tal proposito esiste una gerarchia tra le tre scale di misura categoriali che condiziona i metodi statistici applicabili a ciascun tipo di variabile categoriale, che vede al livello più alto le variabili intervallari, quindi le variabili ordinali, e, al livello più basso, le variabili nominali. Secondo questa gerarchia i metodi statistici applicabili alle variabili appartenenti ad un determinato livello possono essere applicati alle variabili appartenenti a livelli di ordine superiore, ma non a quelli di ordine inferiore.

Pertanto, metodi statistici applicabili alle variabili nominali possono essere applicati alle variabili intervallari, mentre metodi per dati ordinali non possono essere applicati a dati nominali.

I metodi per dati nominali ed ordinali possono essere applicati a variabili intervallo, purché il numero di livelli sia ridotto (numero di categorie distinte limitato) o nei casi in cui i livelli siano caratterizzati da range di valori ordinati (se ad es. si considera una distinzione per età: <10, tra 10 e 30, >30 anni).

Di seguito alcuni esempi di variabili categoriali per scale nominali, ordinali ed intervallari.

<i>tipo di scala</i>	<i>esempi</i>
<b>NOMINALE</b>	<ul style="list-style-type: none"> <li>• affiliazione politica (Repubblicani, Democratici)</li> <li>• affiliazione religiosa (Cattolici, Ortodossi, Islamici, Protestanti, ecc.)</li> <li>• preferenza di cibo, musica, brand, auto, tipo di abitazione, ecc.</li> </ul>
<b>ORDINALE</b>	<ul style="list-style-type: none"> <li>• livello di classe sociale</li> <li>• pensiero politico</li> <li>• stato di salute/gravità di un paziente</li> </ul>
<b>INTERVALLARE</b>	<ul style="list-style-type: none"> <li>• livelli di pressione sanguigna</li> <li>• tempo di vita di un dispositivo</li> <li>• numero di anni di istruzione (range)</li> </ul>

Analizzando le tre scale un punto di vista quali-quantitativo possiamo fare alcune considerazioni sintetizzate nella seguente tabella.

<i>tipo di scala</i>	<i>considerazioni sulla natura qualitativa o quantitativa delle categorie distinte</i>
<i>NOMINALE</i>	le categorie distinte differiscono in qualità (natura qualitativa)
<i>ORDINALE</i>	l'ordinalità delle categorie distinte ne conferisce una natura quantitativa, tuttavia la distanza tra le categorie, incognita, non permette di considerarle dei veri e propri numeri, limitando l'utilizzo di modelli o metodi statistici per dati continui
<i>INTERVALLARE</i>	le categorie distinte differiscono in quantità relativamente alla caratteristica misurata

### ***Distribuzioni delle principali variabili casuali categoriali***

L'analisi inferenziale, consiste in un procedimento per cui partendo dai dati campionari si inferenzia sui parametri della popolazione da cui il campione è stato estratto.

Il primo passo nell'analisi inferenziale è quello di fare assunzioni sul meccanismo casuale che generato i dati, cioè identificare la distribuzione di probabilità della popolazione di dati da cui il campione è stato estratto.

Una variabile casuale discreta  $X$  è descritta dalla sua funzione massa di probabilità  $f(x)$  o "probability mass function" (pmf), che si identifica con la sua distribuzione di probabilità  $P(X = x)$ , per cui si può scrivere:

$$f(x) = P(X = x)$$

Tenendo conto che una distribuzione dipende, oltre che dalle osservazioni della variabile casuale, anche dal parametro incognito  $\theta$ , caratteristico per una specifica distribuzione, si può scrivere  $f(\theta; x)$  (preferito dai frequentisti) o anche  $f(\theta|x)$  (preferito dai Bayesiani).

In ambito categoriale le principali distribuzioni per variabili casuali discrete sono:

- *Bernoulli*
- *binomiale*
- *Poisson*
- *multinomiale*

### ***Distribuzione di Bernoulli***

Una variabile casuale di Bernoulli è una variabile casuale discreta  $X$  che può assumere due soli valori 0 o 1, tale per cui  $X=1$  con probabilità “ $\pi$ ” e  $X=0$  con probabilità “ $1 - \pi$ ”.

Quanto sopra si può esprimere attraverso la seguente espressione:

$$f(x) = \begin{cases} \pi & x = 1 \\ 1 - \pi & x = 0 \\ 0 & \text{altrimenti} \end{cases}$$

oppure semplicemente attraverso la *pmf* della distribuzione di Bernoulli:

$$f(x) = \pi^x (1 - \pi)^{1-x} \quad x = 0,1$$

Da un punto di vista empirico questa distribuzione descrive un esperimento con due soli possibili esiti del tipo “successo” e “insuccesso”, cui corrispondono probabilità di occorrenza tra loro complementari.

Per una distribuzione di Bernoulli:

$$E(X) = 1(\pi) + 0(1 - \pi) = \pi$$

$$Var(X) = E(X^2) - E(X)^2 = 1^2(\pi) + 0^2(1 - \pi) - \pi^2 = \pi - \pi^2 = \pi(1 - \pi)$$

### ***Distribuzione binomiale***

#### **Considerazioni preliminari**

Per definire una variabile casuale discreta con distribuzione binomiale è essenziale definire il contesto sperimentale in cui la variabile viene misurata.

Si ricorre alla distribuzione binomiale quando il metodo di campionamento preveda l'esecuzione di un numero prefissato di prove, che possono avere solo due esiti (tipo successo/insuccesso), e si vada a contare il numero di occorrenze dell'evento preso come riferimento (generalmente il numero di successi) sul numero di prove (*n fixed*). A differenza della variabile di Poisson, il numero di occorrenze che possono verificarsi per l'evento di riferimento è limitato dal numero di prove (al massimo l'evento può verificarsi *n* volte).

Il tipo di campionamento alla base delle due distribuzioni è infatti diverso: nel campionamento di Poisson contiamo il numero di eventi per unità (spaziale o temporale), mentre nel campionamento binomiale, ci troviamo in condizioni in cui eseguiamo *n* prove, quindi contiamo il numero di successi nelle *n* prove.

Il campionamento binomiale prevede che il numero totale di prove (*total sample size*) sia fissato a priori e che per ogni prova si possano avere solo due esiti.

Di seguito alcuni esempi applicativi:

- *numero chiamate per nuovi contratti/ totale chiamate giornaliere (call-center)*
- *numero pezzi difettosi/ totale pezzi collaudati*



- *numero incidenti mortali dovuti all'uso del cellulare/totale incidenti mortali anno*
- *numero giornaliero di pazienti ricoverati arrivati al Pronto soccorso/totale pazienti registrati al giorno*
- *numero di studenti iscritti alla facoltà di matematica/totale iscritti*

### Lo schema di campionamento binomiale

La distribuzione binomiale è il modello di riferimento quando lo schema di campionamento è il risultato di un esperimento casuale in cui si conducono “ $n$ ” prove *bernoulliane indipendenti ed identiche* (i.i.d.) e si va a misurare il numero di volte che l'evento di riferimento si verifica nelle  $n$  prove fissate a priori (numero di successi).

Per prove bernoulliane si intendono prove che possono avere due soli possibili esiti del tipo successo/insuccesso.

Per prove **identiche** si intende che la probabilità di successo è la stessa per ciascuna prova.

Per prove **indipendenti** si intende che l'esito di ciascuna prova è una variabile casuale indipendente e che il verificarsi di un evento in una prova non deve condizionare il verificarsi dell'evento in un'altra. In queste condizioni di campionamento (*binomial sampling*) la variabile casuale  $X =$  numero di successi nelle  $n$  prove, è una variabile categoriale con distribuzione binomiale con indice  $n$  e parametro  $\pi \rightarrow X \sim \text{Bin}(n, \pi)$ .

La prospettiva di campionamento è diversa rispetto a quella del campionamento di Poisson.

In questo caso il campionamento prevede l'esecuzione di un **numero prefissato di prove** (*trials*), in ciascuna delle quali si va a misurare il numero totale di occorrenze dell'evento di riferimento (numero di successi). In queste condizioni (*fixed total sample size*) il **numero di successi è limitato dal numero di prove eseguite** ( $x \leq n$ ).

Nel campionamento di Poisson, invece, le  $n$  prove dell'esperimento casuale (*sample size*) sono fissate (prendo un campione di  $n$  unità), ma in questo caso cambiamo prospettiva di misura, in quanto andiamo a misurare il numero di occorrenze di un evento di interesse all'interno di ciascuna delle  $n$  unità osservative (intervalli spaziali o temporali costanti), senza imporre alcun limite al numero massimo di occorrenze.

Mentre nel campionamento binomiale la dimensione campionaria (*sample size*) è data dal numero totale delle prove (*trials*), nel campionamento di Poisson, la dimensione campionaria è data dal numero di unità/intervalli spaziali o temporali all'interno di ciascuno dei quali misuriamo le occorrenze dell'evento, senza porre alcun limite massimo. In quest'ultimo caso la variabile casuale categoriale che misuriamo è una variabile di Poisson.

Potremmo dire che nel campionamento di Poisson lo spazio delle prove è delimitato in senso verticale (numero definito di unità analizzate), ma non in senso orizzontale (numero di occorrenze dell'evento o numero di conteggi con range  $0 \rightarrow \infty$ ).

Quindi potremmo dire che quello che distingue i due tipi di campionamento, non è propriamente il numero di prove eseguite (in entrambi i casi è fisso), ma la prospettiva con cui misuriamo l'evento di interesse, che genera, appunto, due variabili casuali categoriali con due diverse distribuzioni di probabilità.

### Distribuzione binomiale - Funzione Densità di Probabilità, valore atteso e varianza

Supponiamo che un esperimento consista in un numero fisso di ***n* prove Bernoulliane** (*n* osservazioni binarie), per cui ciascuna delle *n* prove può avere come unici esiti l'esito "successo" ( $X_i = 1$ ) con probabilità  $P(X_i = 1) = \pi$  o l'esito "insuccesso" ( $X_i = 0$ ) con probabilità  $P(X_i = 0) = 1 - \pi$ , e che le prove siano **indipendenti** (le  $X_i$  sono variabili casuali indipendenti, per cui la probabilità di successo di ogni prova non condiziona la probabilità di successo delle altre) ed **identiche** (tutte le prove hanno la *stessa probabilità di successo*), la variabile che misura il numero totale di successi nelle *n* prove:

$$X = \sum_{i=1}^n X_i$$

è una variabile casuale con distribuzione binomiale la cui *pmf* è data da:

$$f(x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} \quad x = 0, 1, 2, \dots, n$$

dove il termine  $\binom{n}{x}$  è il coefficiente binomiale:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Dalla *pmf* di cui sopra, si vede come la funzione che descrive la distribuzione di probabilità dipenda da due parametri, *n* e  $\pi$ . La distribuzione binomiale è, infatti, definita da due parametri, l'indice "*n*", inteso come numero prefissato di prove (*fixed total sample size*), ed il parametro  $\pi$ , che rappresenta la proporzione dell'evento di riferimento o, più generalmente, la proporzione di successi o probabilità di successo.

Quanto sopra è sintetizzato dalla seguente espressione:

$$X \sim \text{Bin}(n, \pi)$$

che indica che la variabile *X* si distribuisce secondo una distribuzione binomiale con indice *n* e parametro  $\pi$ .

Da quanto sopra emerge che la distribuzione Bernoulliana è un caso particolare della distribuzione binomiale quando il numero di prove è ridotto ad una singola prova, per cui si può scrivere:

$$X \sim \text{Bin}(1, \pi)$$

### Assunzioni per una distribuzione binomiale

Per una variabile casuale con distribuzione binomiale il numero di prove ***n* è fisso**.

Una variabile casuale *X* segue una distribuzione binomiale soltanto se le *n* prove Bernoulliane (le *n* osservazioni binarie) sono **identiche** ed **indipendenti**.

Se tali condizioni non sono garantite si deve ricorrere a distribuzioni alternative.

Nel caso, ad esempio, di un campionamento senza rimessa da una popolazione finita, o quando la dimensione campionaria ( $n$ ) è grande rispetto alla dimensione della popolazione ( $N$ ), è più appropriato ricorrere alla distribuzione ipergeometrica.

### Proprietà

Tenendo conto che:

$$X = \sum_{i=1}^n X_i$$

dove  $X_1, X_2, \dots, X_n$  sono variabili casuali di Bernoulli indipendenti e identiche con probabilità di successo  $\pi$ , per le proprietà additive del valore atteso e della varianza:

$$E(X) = \sum_{i=1}^n E_i = n\pi$$

$$Var(X) = \sum_{i=1}^n Var(X_i) = n\pi(1 - \pi)$$

La varianza per una distribuzione binomiale dipende da  $n$  e da  $\pi$ :

- a parità di  $n$  è massima per  $\pi = 0,5$
- a parità di  $\pi$  aumenta linearmente all'aumentare di  $n$

Se:

$$X_1 \sim Bin(n_1, \pi)$$

$$X_2 \sim Bin(n_2, \pi)$$

Allora:

$$X_1 + X_2 \sim Bin(n_1 + n_2, \pi)$$

La *skewness* è descritta da:

$$\frac{E(X - \mu)^3}{\sigma^3} = \frac{(1 - 2\pi)}{\sqrt{n\pi(1 - \pi)}}$$

da cui si vede che la distribuzione è simmetrica quando  $\pi = 0,50$  e l'asimmetria aumenta come  $\pi$  si avvicina ai valori estremi.

Per  $n$  fisso la forma della curva è più simmetrica (*bell-shaped*) quanto più  $\pi$  si avvicina a 0,50.

Quando  $n$  è "grande" la distribuzione binomiale si approssima ad una distribuzione normale con  $\mu = n\pi$  e varianza  $Var = n\pi(1 - \pi)$ .

### Approssimazione normale

All'aumentare di  $n$ , per  $\pi$  fissa, la distribuzione binomiale tende sempre di più ad assumere una forma "bell-shaped", fino a convergere asintoticamente verso una distribuzione normale con valore atteso  $E(X) = n\pi$  e varianza  $Var(X) = n\pi(1 - \pi)$  [ $X \sim N(n\pi, n\pi(1 - \pi))$ ]. Di regola, perché tale approssimazione possa considerarsi valida, il numero di occorrenze per ciascuno dei due eventi (successo e insuccesso) dovrebbe essere almeno 5. Questo significa che per  $\pi = \pi(1 - \pi) = 0,50 \rightarrow n \geq 10$  (perché l'approssimazione sia valida il sample size deve essere almeno 10) mentre per valori di  $\pi$  vicini agli estremi ( $\pi = 0,1$  o  $\pi = 0,9$ )  $\rightarrow n \geq 50$ . Questo a causa della maggiore asimmetria quando la probabilità di successo è vicina ai valori estremi (occorre  $n$  più grande per bilanciare la maggiore asimmetria. Pertanto quando  $\pi \rightarrow 0$  o  $\pi \rightarrow 1$  perché la curva assuma una forma simmetrica del tipo "bell-shaped", occorrono campioni più grandi.

### Approssimazione a Poisson

Per  $n \rightarrow \infty$ , e  $\pi \rightarrow 0$  la distribuzione binomiale si approssima ad una Poisson, per cui nelle condizioni in cui  $n$  è grande e  $\pi$  è piccolo:

$$\frac{n!}{x!(n-x)!} \pi^x (1 - \pi)^{n-x} \approx \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{dove } \lambda = n\pi$$

L'approssimazione asintotica alla Poisson risulta utile in tali condizioni, in quanto la formula di calcolo è generalmente più facile da applicare.

### Distribuzione ipergeometrica

Si ricorre a questa distribuzione in caso di campioni estratti da popolazioni relativamente piccole ( $N$  piccolo), ed in condizioni di estrazione senza rimessa.

Questa distribuzione è la distribuzione nulla (distribuzione di riferimento sotto l'ipotesi nulla) nel test esatto di Fisher per tabelle di contingenza 2x2 e nel test di indipendenza di *Cochran-Mantel-Haenszel* per tabelle di contingenza 2x2xk.

Prevede le seguenti condizioni di applicabilità:

- $N$  piccolo
- $n \geq 20\%N$

Se, ad esempio, si volesse calcolare la probabilità di trovare  $x$  pezzi difettosi campionando  $n$  pezzi da un lotto di  $N$  unità totali, per cui:

- $N$  = numero totale di pezzi
- $n$  = numero di pezzi campionati
- $M$  = numero totale di pezzi difettosi
- $x$  = numero di pezzi difettosi nel campione

La *pmf* (*Probability Mass Function*) che descrive la probabilità di trovare  $x$  pezzi difettosi nel campione estratto di  $n$  pezzi,  $P(X=x)$ , è la seguente:

$$f(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

dove:

$$\binom{M}{x} = C_x^M = \text{possibili combinazioni contenenti difetti}$$

$$\binom{N-M}{n-x} = C_{n-x}^{N-M} = \text{possibili combinazioni che non contengono difetti}$$

$$\binom{N}{n} = C_n^N = \text{tutte le possibili combinazioni}$$

Di seguito il valore atteso e la varianza della variabile casuale  $X$  (numero di pezzi difettosi in un campione di  $n$  pezzi) per la distribuzione ipergeometrica:

$$E(X) = E(M) = \frac{M \cdot n}{N}$$

$$\text{var}(X) = \text{var}(M) = \frac{M(N-M)n(N-n)}{N^2(N-1)}$$

### Un esempio di applicazione – il Test esatto di Fisher

Quando  $n$  (la dimensione campionaria) è piccolo si ricorre a distribuzioni esatte per campioni piccoli (*exact small sample distributions*).

Questo test è stato proposto da R.A Fisher (1935) per testare l'indipendenza tra due variabili categoriali a due livelli (è un test per tabelle di contingenza  $2 \times 2$ ), quando  $n$  è piccolo e non si può inferenziare con metodi inferenziali basati su distribuzioni approssimate per campioni grandi (*large-samples approximation*).

Il test assume sotto l'ipotesi nulla di indipendenza una distribuzione ipergeometrica.

Fissare entrambi i marginali rende la distribuzione campionaria altamente discreta, limitando le combinazioni possibili ad un numero definito.

Dati i totali marginali,  $n_{11}$  determina i valori dei restanti 3 conteggi di cella ( $n_{12}$ ,  $n_{21}$ ,  $n_{22}$ ).

	pezzi nel campione	pezzi restanti	
difetti	$x$ (difetti campionati)	$M-x$ (difetti non campionati)	$M$ (difetti totali)
non difetti	$n-x$ (non difetti campionati)	$(N-M) - (n-x)$ (non difetti non campionati)	$N-M$ (non difetti totali)
	$n$	$N-n$	$N$

Data la seguente tabella di contingenza 2X2:

$n_{11}$	$n_{1+}$	
	$n_{2+}$	
$n_{+1}$	$n_{+2}$	$n$

fissando i marginali riga e colonna possiamo esprimere la *pmf* in funzione di  $n_{11}$ :

$$f(n_{11}) = \frac{\binom{n_{1+}}{n_{11}} \binom{n_{2+}}{n_{+1}-n_{11}}}{\binom{n}{n_{+1}}}$$

### ***Distribuzione di Poisson***

Si ricorre alla distribuzione di Poisson quando si abbia a che fare con variabili casuali rappresentate da conteggi, intesi come numero di occorrenze di un evento in un intervallo spaziale o temporale **finito** (il termine “frequenze” o “rates” è legato al fatto che i conteggi sono effettuati in un intervallo di tempo o di spazio prefissato), cioè si misura o conta il numero di volte che un evento si verifica in questo intervallo di spazio o di tempo.

Il modello si adatta bene nei casi in cui la probabilità di occorrenza degli eventi per i vari conteggi è molto bassa mentre il numero di conteggi è molto grande (distribuzione degli eventi rari).

Di seguito alcuni esempi applicativi:

- *numero chiamate/giorno (call-center)*
- *numero difetti/unità collaudata*
- *numero reclami/trimestre*
- *numero ordini/settimana*
- *numero fatture emesse/mese*
- *numero Non Conformità/reparto (o mese)*
- *numero Non Conformità registrate per Audit*
- *numero incidenti mortali/anno*
- *numero pazienti in arrivo al Pronto soccorso accettati e dimessi in giornata /giorno*
- *numero pratiche aperte e chiuse nello stesso giorno/operatore*
- *numero attività/interventi giornalieri chiusi/operatore*
- *numero mail ricevute tra le 16,00 e le 17,00 del Venerdì*
- *numero visitatori di un museo/giorno*
- *numero visitatori sito web/giorno*
- *numero di goal fatti dalle diverse squadre nel primo tempo di un campionato*

Una variabile casuale di Poisson è una variabile casuale discreta  $X$  che può assumere valori interi non negativi ( $X = 0, 1, \dots$ ).

Quanto sopra si può esprimere attraverso la seguente espressione che rappresenta l'equazione della *pmf* per una Poisson:

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0, 1, \dots,$$

### Assunzioni

Una variabile casuale  $X$  segue una distribuzione di Poisson se sono rispettate tutte le seguenti condizioni:

- **Not Fixed sample size:** i dati sono conteggi di eventi in un intervallo di osservazione prefissato ( $\rightarrow \Delta t$  o  $\Delta s$  definito) ma il numero totale di conteggi non è prefissato (non è predefinito un limite massimo di conteggi)  $\rightarrow$  *overall n not fixed* (differenza rispetto alla binomiale)
- **Indipendenza**  $\rightarrow$  si assume che gli eventi siano indipendenti tra loro (il numero di difetti riscontrati in una unità non deve condizionare il numero di difetti riscontrati in un'altra unità)
- **Omogeneità**  $\rightarrow$  si assume la stessa frequenza media " $\lambda$ " per ciascun  $\Delta t$  o  $\Delta s$  (il numero medio di difetti si assume costante per tutte le unità). Si assume che tutte le realizzazioni della variabile casuale  $X$  siano variabili indipendenti di Poisson, tutte con la stessa frequenza  $\lambda$ , o differenti frequenze ma tali che per cui la loro somma sia uguale a  $\lambda$ .

Quando alcune assunzioni sono violate, in particolare in presenza del fenomeno dell'**overdispersione** (varianza osservata maggiore di quella assunta dal modello), si può ricorrere alla distribuzione binomiale negativa.

### Proprietà

La distribuzione di Poisson è il modello di riferimento quando si abbia una variabile casuale  $X =$  conteggio del numero di eventi che si verificano in un intervallo spaziale o temporale **finito**  $\rightarrow$  si parla di frequenze (*rates*).

Il numero totale dei conteggi (*total sample size*) non è fissato a priori (*total count not fixed*) come nella distribuzione binomiale (numero di *trials* fisso), ma è fissato l'intervallo spaziale o temporale all'interno del quale registriamo i conteggi (l'intervallo osservazionale è fissato a priori).

Se  $X = (X_1, X_2, \dots)$  è una variabile casuale di Poisson, allora la sua distribuzione dipende da un unico parametro,  $\lambda$  (*rate*), che descrive la frequenza media della distribuzione (numero medio di occorrenze dell'evento nell'intervallo definito):

$$X \sim \text{Poisson}(\lambda)$$

la funzione massa di probabilità è:

$$pmf = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0, 1, 2, \dots$$

$\lambda > 0$  (solo valori positivi)

e il valore atteso (media delle occorrenze) è uguale alla varianza:

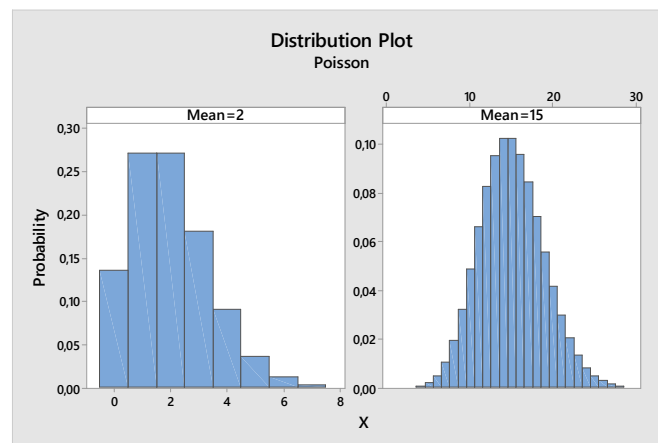
$$E(X) = var(X) = \lambda$$

La distribuzione è unimodale con moda = al valore intero della media.

La *skewness* è descritta da:

$$\frac{E(X - \mu)^3}{\sigma^3} = \frac{1}{\sqrt{\lambda}}$$

da cui si vede che all'aumentare di  $\lambda$  la Poisson si approssima ad una Normale (l'approssimazione è buona quando  $\lambda$  è almeno =10).



La Poisson rappresenta una buona approssimazione per la distribuzione binomiale quando  $n$  è grande e  $\pi$

è piccolo. Per  $n \rightarrow \infty$ , e  $\pi \rightarrow 0$ ,  $\lambda \rightarrow n\pi$ :

$$\frac{n!}{x!(n-x)!} \pi^x (1-\pi)^{n-x} \approx \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{dove } \lambda = n\pi$$

In tali condizioni l'approssimazione asintotica della distribuzione binomiale ad una Poisson risulta utile, in quanto la formula di calcolo per una Poisson è generalmente più facile da applicare.

### Stima della media campionaria

Sia  $X$  è una variabile casuale di Poisson e " $n$ " è il numero totale di unità spaziali o temporali all'interno delle quali si contano le occorrenze. La media campionaria dei conteggi per unità spaziale o temporale è data dalla seguente formula:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

dove il numeratore  $\sum_{i=1}^n x_i$  è il numero totale di occorrenze (somma di tutti gli eventi), mentre il denominatore  $n$  è il numero di osservazioni o realizzazioni della variabile casuale di Poisson (**sample size**), cioè il numero totale delle unità spaziali o temporali all'interno delle quali si contano le occorrenze.



## Il fenomeno dell'Overdispersione (*overdispersion*)

Questo fenomeno si verifica quando si ha a che fare con dati di tipo categoriale.

Il fenomeno deve il suo nome al fatto che la dispersione effettiva dei dati, cioè la varianza osservata sui dati campionari, è maggiore rispetto a quella teorica predetta dal modello.

Nel caso specifico della distribuzione di Poisson abbiamo che:

$$\text{Var}(x) > E(x)$$

Analisi statistiche che assumono distribuzioni di tipo Poisson, binomiale e multinomiale, sono spesso invalidate a causa dell'overdispersione.

Il fenomeno può essere spiegato considerando che il modello assume la stessa probabilità di occorrenza per tutti gli eventi, tuttavia, nella pratica esistono dei fattori che possono condizionare tale probabilità.

Se, ad esempio, consideriamo come variabile di Poisson il numero dei laureati in corso ogni un anno, la probabilità associata a ciascun anno di corso di laurea non è la stessa, ma dipende dalle competenze dei singoli studenti appartenenti allo stesso anno, dai professori che hanno tenuto i corsi, da eventi esterni che potrebbero aver inciso sul rispetto degli appelli di laurea.

Quando alcune assunzioni sono violate, in particolare in presenza del fenomeno dell'overdispersione (varianza osservata maggiore di quella assunta dal modello), si può ricorrere alla distribuzione binomiale negativa.

### ***Distribuzione multinomiale***

La distribuzione multinomiale nasce come estensione di un esperimento binomiale a situazioni in cui il numero di possibili esiti di ciascuna prova è maggiore di 2.

Supponiamo che un esperimento consista in un numero fisso di ***n prove identiche ed indipendenti*** e che, per ciascuna delle prove, si possano avere "*k*" possibili esiti con  $k > 2$ .

La variabile casuale categoriale che misura su *n* prove il numero di occorrenze per ciascuna delle *k* categorie è una variabile categoriale politomica o multinomiale.

***La variabile casuale multinomiale è un vettore le cui componenti sono variabili casuali con distribuzione binomiale.***

Questi eventi sono ***mutualmente esclusivi*** (per ogni prova si può verificare uno solo dei *k* esiti possibili) e ***collettivamente esaustivi*** (la loro somma è vincolata dal numero totale di prove *n*).

### **Eventi mutualmente esclusivi**

Per ogni *i*-esima prova, dati *k* possibili esiti o eventi, considerando l'evento generico "*j*", avremo  $x_{ij} = 1$  se l'*i*-esima prova ha avuto esito "*j*", altrimenti  $x_{ij} = 0$  (ciascuna  $x_{ij}$  è una realizzazione di una variabile casuale binomiale  $x_j$ )

Dato che per ogni prova dei *k* possibili esiti solo uno si verifica, la sommatoria dei componenti del vettore  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ik})$  è unitaria.

Se, ad esempio,  $k = 5$  e nella  $i$ -esima prova è uscito 3, scriveremo  $x_i = (0, 0, 1, 0, 0)$ .

Pertanto alla luce di quanto sopra **ogni  $i$ -esima realizzazione** è un **vettore**  $x_i = (x_{i1}, x_{i2}, \dots, x_{ik})$ , tale per cui la **somma dei suoi componenti è unitaria**,  $\sum_{j=1}^k x_{ij} = 1$ .

Questo vincolo rende la componente  $x_{ik}$  **ridondante**, in quanto linearmente dipendente dalle altre ( $x_{ik} = 1 - \sum_{j=1}^{k-1} x_{ij}$ ).

### **Eventi collettivamente esaustivi**

La somma delle occorrenze di tutti gli eventi è fissa:

$$\sum_{j=1}^k x_j = n$$

pertanto anche se le  $n$  prove sono indipendenti, le occorrenze per ciascuna  $j$ -esima categoria ( $x_j$ ), pur essendo singolarmente casuali, sono tra loro negativamente correlate.

Questo vincolo rende  $x_k$ , componente del vettore  $X = (X_1, X_2, \dots, X_k)$ , **ridondante**, in quanto linearmente dipendente dalle altre componenti del vettore ( $x_k = n - \sum_{j=1}^{k-1} x_j$ ).

Per questo motivo la *pmf* della distribuzione multinomiale ha “ $k - 1$ ” dimensioni.

### **Assunzioni**

Una variabile casuale  $X = (X_1, X_2, \dots, X_k)$  segue una distribuzione multinomiale se sono rispettate tutte le seguenti condizioni:

- **Fixed sample size:** l'esperimento casuale prevede  $n$  prove indipendenti multinomiali ( $n$  è fisso), ciascuna delle quali può avere  $k$  possibili esiti (categorie). A differenza del contesto sperimentale di Poisson e analogamente allo schema di campionamento binomiale, il numero di categorie della scala categoriale all'interno delle quali si vanno a contare le occorrenze è limitato, fissato a priori  $(0, 1, \dots, k)$ .
- **Indipendenza ed equiprobabilità** → si assume che le  $n$  prove siano indipendenti tra loro ed identiche (stessa probabilità di occorrenza di ciascuno dei  $k$  esiti all'interno di ogni prova).

### **Proprietà**

Se  $X$  è una variabile casuale  $X = (X_1, X_2, \dots, X_k)$  dove  $X_j$  rappresenta il numero di occorrenze per la categoria  $j$ -esima,  $X$  è una variabile casuale multinomiale con indice  $n$  e come parametro un vettore di parametri  $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_k)$ , tale per cui  $\sum_{j=1}^k \pi_j = 1$ . Quanto espresso può essere sintetizzato come segue:

$$X \sim \text{Mult}(n, \boldsymbol{\pi})$$

Inoltre ciascuna componente individuale del vettore casuale  $X = (X_1, X_2, \dots, X_k)$ , è, a sua volta, una variabile casuale binomiale:

$$X_j \sim \text{Bin}(n, \pi_j)$$

con valore atteso, varianza e covarianza:

$$E(x_j) = n\pi_j$$

$$\text{var}(x_j) = n\pi_j(1 - \pi_j)$$

$$\text{cov}(x_j, x_{j'}) = -n\pi_j\pi_{j'}$$

Se con  $x_j$  (dove  $x_j = \sum_{i=1}^n x_{ij}$ ), si indica il numero di occorrenze della categoria  $j$ -esima e con  $\pi_j = P(x_{ij} = 1)$  la probabilità di occorrenza della  $j$ -esima categoria, l'equazione che permette di calcolare la probabilità che il vettore  $X = (X_1, X_2, \dots, X_k)$  assuma un determinato valore  $x = (x_1, x_2, \dots, x_k)$  è data dalla la *pmf* della distribuzione multinomiale:

$$f(x_1, x_2, \dots, x_k) = \frac{n!}{x_1! x_2! \dots x_k!} \prod_{j=1}^k \pi_j^{x_j} \quad x_j = 0, 1, \dots, n$$

dove  $\sum_j \pi_j = 1$

Dato che  $\sum_{j=1}^k x_j = n$  e che, quindi  $x_k = n - (x_1 + \dots + x_{k-1})$ , la *pmf* di una variabile casuale multinomiale è multidimensionale con  $(k - 1)$  dimensioni, e la **distribuzione marginale di ciascun  $x_j$  è binomiale**.

Da notare che la distribuzione binomiale ( $k = 2$ ) rappresenta un caso speciale della distribuzione multinomiale.

### Spazio parametrico del parametro $\pi$

Il parametro  $\pi$  di una distribuzione multinomiale è un vettore e, come tale, il suo dominio sarà delimitato da uno spazio parametrico di dimensioni dipendenti dal numero di elementi che lo costituiscono.

Lo spazio parametrico di  $\pi$  è il set di vettori che soddisfano le seguenti condizioni vincolanti (*constraints*):

- $\pi_j \in [0, 1], j = 1, \dots, k$
- $\sum_j \pi_j = 1$

Questo set di vettori è chiamato "*simplex*" ed è una porzione di iperpiano a " $k-1$ " dimensioni in uno spazio parametrico a  $k$  dimensioni.

Data la restrizione per cui  $\sum_j \pi_j = 1$ , l'elemento  $\pi_k$  non è un parametro libero, ma può essere ottenuto da tutti gli altri elementi:

$$\pi_k = 1 - (\pi_1 + \dots + \pi_j + \dots + \pi_{k-1})$$

Pertanto lo spazio parametrico si riduce di una dimensione, cioè da  $k$ -dimensionale a  $(k-1)$  dimensionale.

### Lancio di un dado in un intervallo di tempo prefissato – multinomiale o Poisson?

Per poter stabilire quale delle due distribuzioni meglio si adatti a descrivere la distribuzione di probabilità della variabile casuale in gioco nell'esperimento in oggetto, è essenziale definire nel dettaglio il contesto sperimentale.

L'esperimento casuale prevede il lancio ripetuto in un tempo prefissato, pertanto possiamo fare le seguenti considerazioni:

- ogni volta che lancio il dado, cioè eseguo una prova, ho 6 possibili esiti (può uscire una delle sei facce)
- le prove sono identiche ed indipendenti
- lancio il dado più volte fino al raggiungimento del tempo prefissato (pertanto non conosco a priori il numero totale di lanci  $n$ )
- osservo e registro le occorrenze  $X_i$  per ciascuna delle 6 facce (quante volte si verifica ciascuno dei 6 esiti possibili dell'esperimento)

Assumendo che ogni prova (lancio) sia indipendente ed identica alle altre, teoricamente, se non specificassi che le osservazioni sono effettuate in un intervallo di tempo prefissato, mi troverei nelle condizioni di campionamento tipiche di una distribuzione multinomiale (*multinomial sampling scheme*).

Tuttavia nell'intervallo temporale definito, il numero di lanci  $n$  non è prefissato, ma è casuale.

In questo contesto sperimentale se mi riconduco ad uno schema di campionamento di Poisson (*Poisson sampling scheme*), considero le 6 variabili come variabili casuali indipendenti di Poisson ( $X_1, X_2, X_3, X_4, X_5, X_6$ ), con parametro  $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6)$ , e posso calcolare la *pmf* congiunta (*joint probability mass function*) per  $\{X_i\}$ , come prodotto delle *pmf* delle singole variabili:

$$P = \prod_i \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!}$$

Inoltre il numero totale di tutte le osservazioni  $n = \sum_{i=1}^6 X_i$ , essendo casuale, è, a sua volta, una variabile casuale con distribuzione di Poisson con parametro  $\sum_{i=1}^6 \lambda_i$ :

$$n \sim \text{Poisson}(\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6)$$

Se, invece, fisso un numero di lanci  $n$ , le  $\{X_i\}$  non hanno più una distribuzione di Poisson in quanto nessuna  $X_i$  può eccedere  $n$  e le  $\{X_i\}$  non sono più indipendenti tra loro a causa del vincolo su  $n$  (il valore di una delle variabili condiziona il range di valori delle altre).

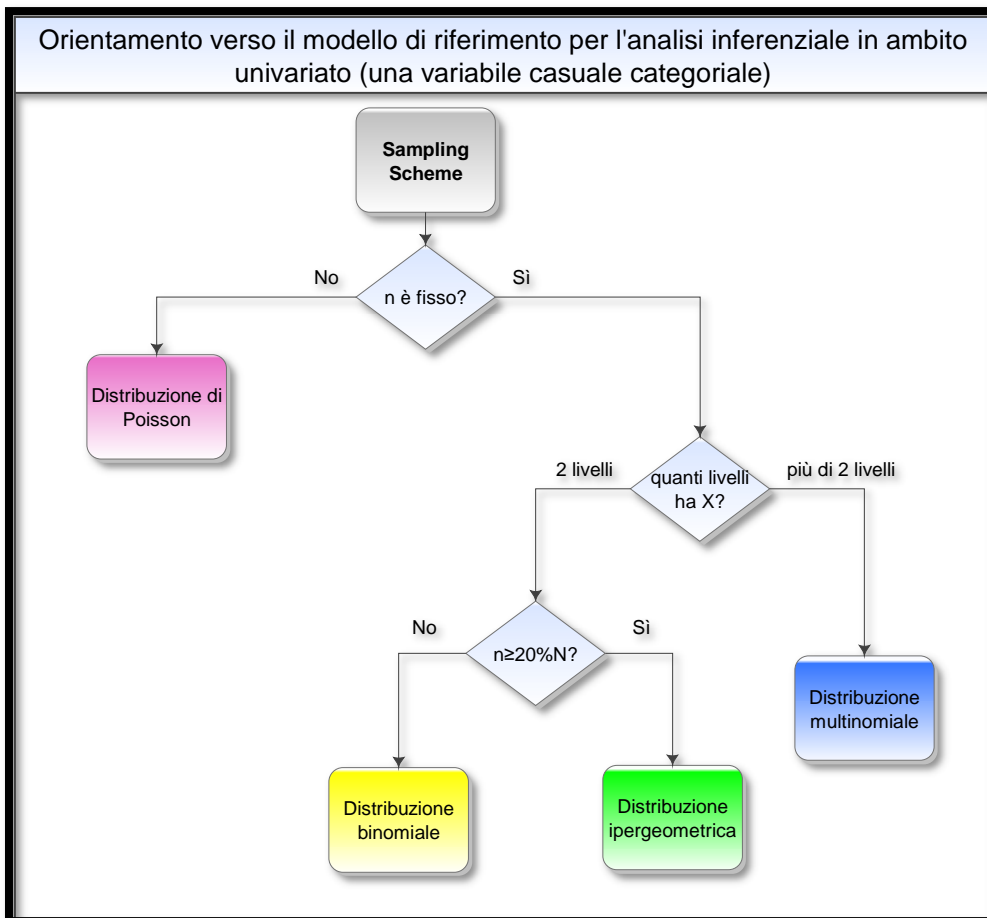
Condizionando su  $n$  il contesto sperimentale è cambiato e la *pmf* congiunta condizionata ( $P$  congiunta dato  $n$ ) è data da:

$$P\left((X_1, X_2, X_3, X_4, X_5, X_6) \mid \sum_j X_j = n\right) = \frac{P(X_1, X_2, X_3, X_4, X_5, X_6)}{P(\sum_j X_j = n)} =$$
$$= \frac{\prod_i \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!}}{\frac{e^{-\sum_j \lambda_j} (\sum_j \lambda_j)^n}{n!}} = \frac{n!}{\prod_i x_i!} \prod_i \pi_i^{x_i}$$

Da cui si vede che condizionando su  $n$  il modello Poisson genera un modello multinomiale con indice  $n$  e parametro  $\{\pi_i\}$ .

**Riferimenti**

1. Agresti A. 2015. *Foundations of Linear and Generalized Linear Models*. John Wiley & Sons, Inc.
2. Agresti A. 2013. *Categorical Data Analysis*. 3<sup>rd</sup> ed. John Wiley & Sons, Inc.
3. Agresti A. 2018. *An Introduction to Categorical Data Analysis*, 3rd ed., John Wiley & Sons, Inc.
4. Agresti A. 2018. *Statistical Methods for the Social Sciences*, 5th edition, Pearson.

**Appendice 1 – workflow per l'identificazione della distribuzione di probabilità per variabili categoriali in analisi inferenziale**

## Appendice 2 – pmf, $E(X)$ e $Var(X)$ per le distribuzioni Bernoulli, binomiale, Poisson e multinomiale

distribuzione	variabile casuale	pdf	$x$	$E(X)$	$Var(X)$
Bernoulli	$X \sim \text{Bernoulli}(\pi)$	$P(x) = \pi^x (1 - \pi)^{1-x}$	$x: 0, 1$	$E(X) = \pi$	$Var(X) = \pi(1 - \pi)$
Binomiale	$X \sim \text{Bin}(n, \pi)$	$P(x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}$	$x: 0, 1, \dots, n$	$E(X) = n\pi$	$Var(X) = n\pi(1 - \pi)$
Ipergeometrica		$P(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$			
Poisson	$X \sim \text{Poisson}(\lambda)$	$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$	$x: 0, 1, \dots$	$E(X) = \lambda$	$Var(X) = \lambda$
Multinomiale	$X \sim \text{Mult}(n, \boldsymbol{\pi})$	$P(x_1, x_2, \dots, x_k) = \frac{n!}{x_1! x_2! \dots x_k!} \prod_{j=1}^k \pi_j^{x_j}$	$x_j: 0, 1, \dots, n$ $\sum_{j=1}^k x_j = n$	$E(X_j) = n\pi_j$	$Var(X_j) = n\pi_j(1 - \pi_j)$