Densità ipergeometrica

Siano a, b, m tre numeri interi strettamente positivi tali che $m \leq a+b$. Poniamo

$$k_1 = \max(m-b, 0) \quad \text{e} \quad k_2 = \min(a, m)$$

Chiaramente $k_1 \leq k_2$. Chiamiamo densità ipergeometrica di parametri a, b, m la funzione così definita

$$q(x) = \begin{cases} \dfrac{\binom{a}{x}\binom{b}{m-x}}{\binom{a+b}{m}} & x = k_1, k_1+1, \dots, k_2 \\ 0 & \text{altrimenti} \end{cases}$$

È spesso indicata con il simbolo $\mathrm{Iper}(m, a, b)$.

Notiamo che $x \geq k_1$ implica $x \geq 0$ ed $m-x \leq b$

mentre $x \leq k_2$ implica $x \leq a$ ed $m-x \geq 0$

quindi i coefficienti binomiali hanno senso e sono non nulli. È possibile provare che $\sum_{x \in \mathbb{R}} q(x) = 1$.

Esempio Supponiamo, ancora una volta, di avere un'urna con N palline di due colori: M bianche ed $N-M$ nere. Supponiamo di eseguire m estrazioni successive senza reimmissione.

Sappiamo già che lo spazio da considerare per model=

LIZZARE QUESTA SITUAZIONE È

$$\Omega = \{ \text{COMBINAZIONE DI CLASSE } m \text{ SU } N \text{ OGGETTI} \}$$

NUMERANDO LE PALLINE ED IDENTIFICANDO OGNI POSSIBILE COMBINAZIONE CON UNA ESTRAZIONE, ED ASSEGNANDO LA PROBABILITÀ UNIFORME:

$$P(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{\binom{N}{m}} \qquad \forall \omega \in \Omega$$

CONSIDERIAMO ORA LA V.A. $X : \Omega \to \mathbb{R}$ DEFINITA DA

$$X(\omega) = \text{NUMERO DI PALLINE BIANCHE ESTRATTE}$$

OSSERVIAMO CHE X PRENDE VALORI INTERI E CHE QUESTI SONO MINORI DI $k_2 = \min\{m, M\}$ (LE PALLINE BIANCHE ESTRATTE NON POSSONO ESSERE PIÙ DELLE PALLINE BIANCHE TOTALI) E MAGGIORI DI $k_1 = \max\{0, m-(N-M)\}$ (LE PALLINE NERE ESTRATTE NON POSSONO ESSERE PIÙ DELLE PALLINE NERE TOTALI, E QUINDI QUELLE BIANCHE ESTRATTE NON POSSONO ESSERE MENO DI $m-(N-M)$.

CALCOLIAMOCI ORA LA DENSITÀ DI X. L'EVENTO $\{X = x\}$ È NON VUOTO SOLO SE x È UN INTERO $k \in [k_1, k_2]$, PER QUANTO DETTO SOPRA.

L'evento $\{X=k\}$ corrisponde alla situazione in cui k palline estratte sono bianche. Ne abbiamo calcolato in precedenza la probabilità:

$$P\{X=k\} = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

quindi la densità di X è di tipo ipergeometrico con parametri $a=M$, $b=N-M$, n.

Introduciamo la prossima densità direttamente con un esempio

<u>Esempio</u> Consideriamo uno schema di n prove indipendenti ognuna avente probabilità di successo q.

Lo spazio di probabilità discreto è

$$\Omega = \{0,1\}^n \quad \text{con probabilità} \quad P(\{\omega\}) = q^{\sum_{i=1}^{n} \omega_i} (1-q)^{n-\sum_{i=1}^{n}\omega_i}$$

Avevamo valutato la probabilità dei due eventi

$$A_k := \text{"esattamente k prove hanno successo"} = \left\{ \omega \in \Omega : \sum_{i=1}^{n} \omega_i = k \right\}$$

$$B_j := \text{"il primo successo si verifica alla j-esima prova"} = \left\{ \omega \in \Omega : \begin{array}{l} \omega_i = 0 \quad i < j \\ \omega_j = 1 \end{array} \right\}$$

Lo avevamo fatto nei due quesiti alla fine della sezione sullo schema delle prove indipendenti. Avevamo trovato che

$$P(A_k) = \binom{n}{k} q^k (1-q)^{n-k} \qquad (*)$$

$$P(B_j) = (1-q)^{j-1} q \qquad (**)$$

Introdotta la variabile aleatoria $X(\omega) = \sum_{i=1}^{n} \omega_i$, che ci dice quanti successi si verificano in un set ω di n prove, avevamo mostrato che la sua densità è binomiale usando la $(*)$ e il fatto che $\{X=k\} = A_k$.

Introduciamo una nuova variabile aleatoria

$$Y_n(\omega) = \begin{cases} \min\left\{ j \in \{1,\dots,n\} : \omega_j = 1 \right\} & \text{se } \omega \neq (0,\dots,0) \\ +\infty & \text{se } \omega = (0,\dots,0) \end{cases}$$

cioè Y_n ci dice quando si verifica in ω il primo successo. Se ω non presenta successi, cioè se $\omega = (0,\dots,0)$, allora per convenzione poniamo $Y_n(\omega) = +\infty$, che è un modo per dire "mai".

Notare che $Y_n : \Omega \to \mathbb{R} \cup \{+\infty\}$ e non in $\mathbb{R}$, ma soprassediamo su questo dettaglio tecnico.

ABBIAMO $\{Y_n = j\} = B_j$ PER $j = 1, \dots, n$

E $\{Y_n = x\} = \emptyset$ PER $x \neq 1, \dots, n$

DA (**) DEDUCIAMO CHE LA DENSITÀ DI Y_n È

$$q_n(x) = P\{Y_n = x\} = \begin{cases} (1-q)^{x-1} q & x = 1, \dots, n \\ 0 & \text{ALTRIMENTI} \end{cases}$$

OSSERVIAMO CHE $q_n(x) = q_m(x)$ $\forall x \leq \min\{n, m\}$. CIOÈ, DATO j CONSIDERATE n OPPURE m RIPETIZIONI, ENTRAMBE $\geq j$, LA PROBABILITÀ CHE IL PRIMO SUCCESSO SI VERIFICHI ALLA j-ESIMA PROVA È SEMPRE LA STESSA. PER CAPIRCI, LA PROBABILITÀ CHE ESCA 3 PER LA PRIMA VOLTA AL QUARTO LANCIO $(j=4)$ DI UN DADO A SEI FACCE È LA STESSA SE EFFETTUO 5 $(=n)$ OPPURE 100 $(=m)$ LANCI.

PONIAMOCI ORA LA SEGUENTE DOMANDA: IN UNO SCHEMA DI PROVE INDIPENDENTI, OGNUNA CON PROBABILITÀ q, QUAL È LA PROBABILITÀ CHE IL PRIMO SUCCESSO SI VERIFICHI ALLA j-ESIMA PROVA?

SIAMO PORTATI A CONSIDERARE LA VARIABILE ALEATORIA

$Y(\omega) =$ "LA PRIMA PROVA IN CUI SI HA SUCCESSO"

SAPENDO IN ANTICIPO CHE LE PROVE SARANNO IN TOTALE M, AVREMO $Y = Y_M$. SE INVECE NON ABBIAMO VINCOLI, TENENDO CONTO DI QUANTO DETTO SUL FATTO CHE $q_m = q_M$ SU $(-\infty, \min\{m, M\}]$, POSSIAMO COMUNQUE ASSEGNARE AD Y LA DENSITÀ

$$q(x) = \begin{cases} (1-q)^{x-1} q & x = 1, 2, 3, \dots \\ 0 & \text{ALTRIMENTI} \end{cases}$$

CIOÈ $P\{Y = j\} = q(1-q)^{j-1} \quad \forall j = 1, 2, 3, \dots$

QUESTA CHE ABBIAMO APPENA SCRITTO È CHIAMATA <u>DENSITÀ GEOMETRICA</u> DI PARAMETRO $q \in (0,1)$. È SPESSO INDICATA CON IL SIMBOLO $Geo(q)$.

OSSERVIAMO CHE È UNA DENSITÀ, IN QUANTO

$$\sum_{j=1}^{\infty} q(j) = q \sum_{j=1}^{\infty} (1-q)^{j-1} = 1$$

PER QUANTO VISTO SULLA SERIE GEOMETRICA

VISTO CHE $\{Y = j\}$ SONO LE RIPETIZIONI IN CUI IL SUCCESSO AVVIENE ALLA j-ESIMA PROVA, $q(1-q)^{j-1}$ È LA RISPOSTA ALLA DOMANDA INIZIALE.

NO, UN ATTIMO! DOVE È DEFINITA Y (CIOÈ, CHI È Ω)?
COME È DEFINITA Y (CIOÈ CHI È $Y(\omega)$, $\omega \in \Omega$)?
CHE SPAZIO DI PROBABILITÀ STIAMO USANDO (CIOÈ, CHI È $(\Omega, \mathcal{A}, P)$)? BOH, NON CI È SERVITO SAPERLO...

UNA VOLTA CHE ABBIAMO APPURATO CHE $q = q(x)$ È UNA DENSITÀ, SAPPIAMO CHE ESISTONO UNO SPAZIO DI PROBABILITÀ $(\Omega, \mathcal{A}, P)$ E UNA VARIABILE ALEATORIA $\tilde{Y}: \Omega \to \mathbb{R} \cup \{+\infty\}$ AVENTE q COME DENSITÀ.
IDENTIFICHIAMO Y CON $\tilde{Y}$ E SIAMO PRONTI AD OPERARE, NEL SENSO CHE Y SEGUE LE REGOLE DI UNA V.A.
QUAL È LA PROBABILITÀ CHE IL PRIMO SUCCESSO AVVENGA TRA LA SESTA E LA DECIMA PROVA?

$$P\{Y \in [6,10]\} = \sum_{x \in [6,10]} q(x) = \sum_{k=6}^{10} q(1-q)^{k-1}$$

DOVE LA PRIMA UGUAGLIANZA È GIUSTIFICATA DALLA IDENTIFICAZIONE $Y \sim \tilde{Y}$ (ANCHE SENZA SAPERE ESPLICITAMENTE CHI SONO Y E $\tilde{Y}$) E DA COME SI CALCOLA LA DISTRIBUZIONE TRAMITE LA DENSITÀ.

VOLENDO POSSIAMO ANCHE ESPLICITARE LO SPAZIO DI PRO-BABILITÀ SOTTOSTANTE. AVENDO IN MENTE QUANTO VISTO ALLA TERZA LEZIONE, USIAMO $\Omega = \{\omega_m, m \in \mathbb{N}\} \cup \{\omega_\infty\}$

DOVE

ω_m = FAMIGLIA DELLE SUCCESSIONI CHE HANNO 0 NEI PRIMI $m-1$ ELEMENTI E 1 ALL'M-ESIMO

ω_∞ = SUCCESSIONE TUTTA COSTITUITA DA 0.

SU QUESTO SPAZIO DISCRETO INTRODUCIAMO LA PROBABI-LITÀ

$$P(\{\omega_m\}) = (1-q)^{m-1} q \qquad m = 1, 2, 3, \ldots$$

$$P(\{\omega_\infty\}) = 0$$

E QUINDI LA VARIABILE ALEATORIA $Y : \Omega \to \mathbb{R} \cup \{+\infty\}$ DEFINITA DA

$$Y(\omega_m) = m \quad \text{SE } m = 1, 2, 3, \ldots$$

$$Y(\omega_\infty) = +\infty$$

Y FORMALIZZA IL CONCETTO DI "LA PRIMA PROVA IN CUI SI HA SUCCESSO" E SI VERIFICA SUBITO CHE LA SUA DENSITÀ È PROPRIO $Geo(q)$.

Proposizione Sia Y una variabile aleatoria avente densità geometrica di parametro q. Allora

$$P\{Y > n+m \mid Y > n\} = P\{Y > m\} \quad \forall n, m \in \mathbb{N}$$

PROOF

Osserviamo che $\{Y > n+m\} \subset \{Y > n\}$, dunque

$$\{Y > n+m\} \cap \{Y > n\} = \{Y > n+m\}.$$

Ne segue

$$P\{Y > n+m \mid Y > n\} = \frac{P\{Y > n+m,\ Y > n\}}{P\{Y > n\}} = \frac{P\{Y > n+m\}}{P\{Y > n\}}$$

Calcoliamoci ora i termini sul lato destro.

$$P\{Y > n\} = \sum_{x \in (n,\infty)} q(x) = \sum_{k=n+1}^{\infty} q(k) = \sum_{k=n+1}^{\infty} q(1-q)^{k-1}$$

tramite sostituzione $k = j+n$

$$= (1-q)^n \underbrace{\sum_{j=1}^{\infty} q(1-q)^{j-1}}_{1} = (1-q)^n$$

Similmente $P\{Y > n+m\} = (1-q)^{n+m}$.

Quindi

$$P\{Y > n+m \mid Y > n\} = \frac{(1-q)^{n+m}}{(1-q)^n} = (1-q)^m = P\{Y > m\}$$

$\blacksquare$

LA PROPRIETÀ APPENA VISTA VIENE CHIAMATA ASSENZA DI MEMORIA. IL MOTIVO È IL SEGUENTE.

IN UNO SCHEMA DI PROVE INDIPENDENTI, SUPPONIAMO DI NON AVER OTTENUTO ALCUN SUCCESSO NELLE PRIME n PROVE. QUAL È LA PROBABILITÀ DI DOVER ATTENDERE ALTRE m PROVE PER AVERE IL PRIMO SUCCESSO? CIOÈ, QUAL È LA PROBABILITÀ CONDIZIONALE $P\{Y=n+m \mid Y>n\}$?

USANDO LA PROPOSIZIONE PRECEDENTE CON n E $m-1$ ABBIAMO $P\{Y \geq n+m \mid Y>n\} = P\{Y \geq m\}$

QUINDI

$$\begin{aligned} P\{Y=n+m \mid Y>n\} &= P\{Y \geq n+m \mid Y>n\} - P\{Y>n+m \mid Y>n\} \\ &= P\{Y \geq m\} - P\{Y>m\} = P\{Y=m\} \end{aligned}$$

DOVE NELLA PRIMA UGUAGLIANZA ABBIAMO USATO CHE $P(\cdot \mid Y>m)$ È UNA PROBABILITÀ (PROPRIETÀ GENERALE DELLA PROBABILITÀ CONDIZIONALE).

QUINDI LA RISPOSTA È CHE LA PROBABILITÀ DI DOVER ATTENDERE IL PRIMO SUCCESSO PER ALTRE m PROVE È LA STESSA CHE SI AVREBBE SE LE PRIME n PROVE SENZA SUCCESSO NON FOSSERO ESISTITE.

PER CAPIRCI, SE LANCIO UNA MONETA TRE VOLTE SENZA CHE ESCA TESTA, LA PROBABILITÀ CHE ESCA TESTA PER LA PRIMA VOLTA AL SETTIMO LANCIO È LA STESSA CHE AVREI RICOMINCIANDO DA CAPO E CALCOLANDO CHE ESCA LA PRIMA VOLTA AL QUARTO LANCIO.

<u>ESERCIZIO</u> UN DADO VIENE LANCIATO PIÙ VOLTE FINO A CHE NON SI OTTIENE 3. QUAL È LA PROBABILITÀ CHE OCCORRANO ESATTAMENTE 5 LANCI?

LA PROBABILITÀ SUL SINGOLO LANCIO È $q = 1/6$ (SUPPONENDO IL DADO EQUILIBRATO). QUINDI LA PROBABILITÀ È

$$\frac{1}{6}\left(1-\frac{1}{6}\right)^4 = \frac{1}{6}\left(\frac{5}{6}\right)^4 \sim 8\%$$

<u>OSSERVAZIONE</u> NEL DIMOSTRARE LA PROPRIETÀ DI ASSENZA DI MEMORIA, ABBIAMO PROVATO CHE

$$P\{Y > n\} = (1-q)^n$$

<u>ESERCIZIO</u> DA UN'URNA CON 10 PALLINE BIANCHE E 15 NERE SI ESEGUONO ESTRAZIONI CON REINSERIMENTO FINO ALL'ESTRAZIONE DI UNA PALLINA NERA. CALCOLARE LA PROBABILITÀ CHE SERVANO ALMENO 10 ESTRAZIONI.

VISTO CHE C'È REINSERIMENTO, SIAMO IN UNO SCHEMA DI PROVE INDIPENDENTI, CON PROBABILITÀ DI SUCCESSO (CIOÈ ESTRAZIONE DEL NERO) SULLA SINGOLA PROVA PARI A $q = 15/25 = 3/5$. SIA Y LA VARIABILE ALEATORIA CHE INDIVIDUA QUANDO AVVIENE IL PRIMO SUCCESSO.

DOBBIAMO CALCOLARE LA PROBABILITÀ DELL'EVENTO $\{Y \geq 10\}$.

ABBIAMO, PER L'OSSERVAZIONE PRECEDENTE

$$P\{Y \geq 10\} = P\{Y > 9\} = \left(1 - \frac{3}{5}\right)^9 = \left(\frac{2}{5}\right)^9$$

DENSITÀ DI POISSON

SIA $\lambda > 0$ UN NUMERO REALE FISSATO. CHIAMIAMO DENSITÀ DI POISSON DI PARAMETRO λ (SPESSO INDICATA CON IL SIMBOLO $\wp_\lambda$) LA FUNZIONE COSÌ DEFINITA

$$q(x) = \begin{cases} \dfrac{\lambda^x}{x!} e^{-\lambda} & x = 0, 1, 2, 3, \ldots \\ 0 & \text{ALTRIMENTI} \end{cases}$$

È EFFETTIVAMENTE UNA DENSITÀ IN QUANTO $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^\lambda$.

QUESTA È UNA SERIE NOTEVOLE CHIAMATA SERIE ESPONENZIALE.

LA DENSITÀ DI POISSON PUÒ ESSERE OTTENUTA COME LIMITE DELLE DENSITÀ BINOMIALI $B(n, \lambda/n)$ PER $n \to \infty$ NEL MODO CHE SEGUE. RICORDO CHE PER DEFINIZIONE

$$B(n, \lambda/n)(k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \qquad k = 0,1,2,\dots,n$$

FISSATO k, SIA n GRANDE ABBASTANZA COSÌ CHE $n \geq k$.

ABBIAMO

$$B(n, \lambda/n)(k) = \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}$$

$$= \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k} \frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{\underbrace{n\cdot n\cdot\ldots\cdot n}_{k \text{ FATTORI}}}$$

TENUTO CONTO CHE

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda} \qquad \text{(LIMITE NOTEVOLE)}$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^k = 1 \qquad (k \text{ È FISSATO})$$

$$\lim_{n\to\infty} \frac{n}{n}\cdot\frac{(n-1)}{n}\cdots\frac{(n-k+1)}{n} = 1$$

ABBIAMO CHE

$$\boxed{\lim_{n\to\infty} B(n, \lambda/n)(k) = \frac{\lambda^k}{k!}e^{-\lambda} = \mathcal{P}_\lambda(k)} \qquad \forall k \in \mathbb{N}$$

COSA SIGNIFICA IN TERMINI DI V.A.? SUPPONIAMO DI AVERE PER OGNI INTERO n UNA V.A. X_n AVENTE DENSITÀ $B(n, \lambda/n)$.
DUNQUE

$$P\{X_n = k\} = B(n, \lambda/n)(k)$$

IL LIMITE VISTO PRIMA CI PERMETTE DI SAPERE A QUALI VALORI TENDE LA PROBABILITÀ DI $P\{X_n = k\}$ QUANDO IL NUMERO DELLE PROVE RIPETUTE AUMENTA MENTRE LA PROBABILITÀ DI SUCCESSO $q_n = \lambda/n$ SULLA SINGOLA PROVA DIMINUISCE. VA NOTATO CHE q_n È DELL'ORDINE $1/n$, CIOÈ DECRESCE INVERSAMENTE AL NUMERO DI PROVE.
IL RISULTATO CHE ABBIAMO PROVATO RIMANE VALIDO SE INVECE DI $q_n = \lambda/n$ ABBIAMO SOLO UNA SUCCESSIONE q_n TALE CHE $\lim_{n \to \infty} n q_n = \lambda$

POSTO $\lambda_n = n q_n$, ABBIAMO PER CONTINUITÀ CHE

$$\lim_{n \to \infty} \mathcal{P}_{\lambda_n}(k) = \mathcal{P}_\lambda(k) \quad \forall k \in \mathbb{N}$$

E QUINDI IL RISULTATO DI APPROSSIMAZIONE SI PUÒ SCRIVERE COME

$$\lim_{n \to \infty} |B(n, q_n)(k) - \mathcal{P}_{\lambda_n}(k)| = 0 \quad \forall k \in \mathbb{N}.$$

COME SI USA IN TERMINI PRATICI IL RISULTATO OTTENUTO?
VISTO IL COEFFICIENTE BINOMIALE $\binom{n}{k}$, È PIÙ FACILE FARE CONTI USANDO $\mathcal{P}_\lambda$ INVECE CHE $\mathcal{B}(n,q)$.
PER QUESTO, NEI CASI IN CUI IL NUMERO DI PROVE n SIA GRANDE E q PICCOLA, INVECE DELLA DENSITÀ BINOMIALE SI USA QUELLA DI POISSON CHE APPROSSIMA IL RISULTATO A CUI SIAMO INTERESSATI. LA SCELTA PER λ È $\lambda = nq$.

<u>ESEMPIO</u> IL NUMERO MEDIO DI ERRORI DI BATTITURA PER PAGINA IN UN LIBRO È 1,2. QUAL È LA PROBABILITÀ DI TROVARE IN UNA DATA PAGINA DI 2000 CARATTERI ZERO ERRORI? ED ALMENO TRE ERRORI?

POSSIAMO VEDERE LA BATTITURA DEL SINGOLO CARATTERE COME EVENTO INDIPENDENTE, CON PROBABILITÀ DI SUCCESSO (CIOÈ DI FARE ERRORE) PARI A $q = 1{,}2/_{2000} = 0{,}0006$.
USIAMO LA DISTRIBUZIONE DI POISSON CON $\lambda = nq = 1{,}2$, ESSENDO $n = 2000$ LE PROVE. QUINDI

$$P\{X=0\} = e^{-\lambda} \sim 0{,}3$$

$$P\{X \geq 3\} = 1 - P\{X=0\} - P\{X=1\} - P\{X=2\} \sim 0{,}12$$