DIMOSTRIAMO UNA FORMULA FAMOSA

<u>TEOREMA</u> Sia $(\Omega, \mathcal{A}, P)$ uno spazio di probabilità e siano A e B due eventi per cui $P(A)>0$, $P(B)>0$. Allora

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

[FORMULA DI BAYES]

Se $P(B) \in (0,1)$, allora la formula può essere scritta come

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|B^c)P(B^c)}$$

[II VERSIONE]

Inoltre, se $\{B_k\}_{k\in K} \subset \mathcal{A}$ è una partizione (finita o numerabile) di Ω con $P(B_k)>0$ $\forall k \in K$, allora

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)}$$

[III VERSIONE]

$$= \frac{P(A|B_i)P(B_i)}{\sum_{k\in K} P(A|B_k)P(B_k)} \qquad \forall i \in K$$

PROOF

PER LA FORMULA DI BAYES BASTA OSSERVARE CHE PER LA DEFINIZIONE DI PROBABILITÀ CONDIZIONALE

$$P(B|A)P(A) = P(A\cap B) = P(A|B)P(B)$$

PER LE ALTRE DUE FORMULE, BASTA APPLICARE A DENOMINATORE LA FORMULA DELLE PROBABILITÀ TOTALI E IL SUO COROLLARIO ■

<u>OSSERVAZIONE</u> LA III VERSIONE DELLA FORMULA DI BAYES È DETTA <u>FORMULA DELLE PROBABILITÀ DELLE CAUSE</u>.
INFATTI, INTERPRETANDO GLI EVENTI DELLA PARTIZIONE $\{B_k\}_{k\in K}$ COME "CAUSE" AL VERIFICARSI DI A (E LA $P(A|B_k)$ È APPUNTO LA PROBABILITÀ CHE SI VERIFICHI A SE ACCADE B_k), LA FORMULA CI PERMETTE DI CALCOLARE LA PROBABILITÀ CHE QUANDO A SI È VERIFICATO LA CAUSA SIA STATA B_i.

<u>ESEMPIO</u> TORNIAMO ANCORA UNA VOLTA ALLE NOSTRE DUE URNE. SAPPIAMO CHE

$$P(A) = \frac{1}{2} \qquad P(N|A) = \frac{3}{4} \qquad P(N|A^c) = \frac{1}{2}$$

DOMANDA: SE LA PALLINA ESTRATTA È NERA, QUAL È LA PROBABILITÀ CHE L'URNA SCELTA SIA STATA LA α?

CIOÈ, QUANTO VALE $p(A|N)$?

DALLA FORMULA DI BAYES

$$p(A|N) = \frac{p(N|A)p(A)}{\underbrace{p(N|A)p(A) + p(N|A^c)[1-p(A)]}_{p(N)}} \qquad \frac{\frac{3}{4}\cdot\frac{1}{2}}{\frac{5}{8}} = \frac{3}{5}$$

CURIOSO CHE SAPENDO A POSTERIORI IL COLORE DELLA PALLINA, LA PROBABILITÀ SULLA SCELTA DELL'URNA α (EVENTO A) SIA PIÙ ALTA DI QUELLA NON CONDIZIONATA $p(A) = 1/2$

<u>ESEMPIO</u> È STATO CHIESTO AGLI STUDENTI DEL II ANNO QUALE SIA STATA LA CHIAVE PER PASSARE ANALISI I.

- IL 50% DEL CAMPIONE HA RISPOSTO "TANTO STUDIO"
- IL 40% " " " " "FORTUNA"
- IL 10% " " " " "AMICO DA CUI COPIARE"

È STATO POI CHIESTO IL VOTO CON CUI HANNO SUPERATO L'ESAME

- NEL PRIMO GRUPPO 2/3 HA PRESO VOTO ≥ 27
- NEL SECONDO GRUPPO 1/2 " " " ≥ 27
- NEL TERZO GRUPPO 1/4 " " " ≥ 27

VALUTARE IN CHE PERCENTUALE CHI HA PRESO ≥ 27 APPARTIENE AL PRIMO E AL TERZO GRUPPO.

PRENDIAMO COME SPAZIO CAMPIONE

$$\Omega = \{(1,n),(1,b),(2,n),(2,b),(3,n),(3,b)\}$$

DOVE

n = NERO STA PER ≥ 27

b = BIANCO STA PER < 27

QUINDI, PER $k = 1,2,3$

(k,n) SONO COLORO NEL k-ESIMO GRUPPO CON VOTO ≥ 27

(k,b) SONO COLORO NEL k-ESIMO GRUPPO CON VOTO < 27

CONSIDERIAMO GLI EVENTI

N = "IL VOTO È ≥ 27" $= \{(1,n),(2,n),(3,n)\}$

B_k = "APPARTENENZA AL GRUPPO k" $= \{(k,n),(k,b)\}$

OSSERVIAMO CHE $\{B_k\}_{k=1,2,3}$ È UNA PARTIZIONE DI Ω.

ABBIAMO

$$P(B_1) = \frac{1}{2} \qquad P(B_2) = \frac{2}{5} \qquad P(B_3) = \frac{1}{10}$$

INOLTRE

$$P(N|B_1) = \frac{2}{3} \qquad P(N|B_2) = \frac{1}{2} \qquad P(N|B_3) = \frac{1}{4}$$

ABBIAMO, PER LA FORMULA DELLE PROBABILITÀ TOTALI

$$P(N) = \sum_{k=1}^{3} P(N|B_k)P(B_k)$$

$$= \frac{2}{3}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{2}{5} + \frac{1}{4}\cdot\frac{1}{10} = \frac{40+24+3}{120} = \frac{67}{120}$$

CIOÈ LA PROBABILITÀ DI PRENDERE ≥ 27.

PER LA FORMULA DI BAYES

$$P(B_1|N) = \frac{P(N|B_1)P(B_1)}{P(N)} = \frac{\frac{2}{3}\cdot\frac{1}{2}}{\frac{67}{120}} = \frac{40}{67} \qquad \sim 60\%$$

$$P(B_3|N) = \frac{P(N|B_3)P(B_3)}{P(N)} = \frac{\frac{1}{4}\cdot\frac{1}{10}}{\frac{67}{120}} = \frac{3}{67} \qquad \sim 5\%$$

NOTARE CHE AVREMMO POTUTO USARE DIRETTAMENTE LA FORMULA DELLE CAUSE. ABBIAMO COSÌ TROVATO CHE SE UNO STUDENTE HA PRESO ≥ 27, AL 60% PROVIENE DAL PRIMO GRUPPO E AL 5% DAL TERZO. QUINDI STUDIATE!

ABBIAMO SCHERZATO, ORA UN ESERCIZIO PIÙ SERIO.

<u>ESEMPIO</u> PER DETERMINARE LA PRESENZA DEL CORONA-VIRUS È STATO ELABORATO UN TEST CLINICO CON LA SEGUENTE EFFICACIA:

- SE IL VIRUS È PRESENTE, IL TEST RISULTA POSITIVO AL 99%
- SE IL VIRUS È ASSENTE, IL TEST RISULTA POSITIVO AL 2% (I COSIDDETTI FALSI POSITIVI).

SU UN CAMPIONE DI 1000 PERSONE 2 RISULTANO AVERE IL VIRUS. SUPPONIAMO CHE UN INDIVIDUO SCELTO A CASO RISULTI POSITIVO AL TEST. CON QUALE PROBABILITÀ RISULTA REALMENTE MALATO?

LO SPAZIO CAMPIONE È

$$\Omega = \{(+,m), (+,s), (-,m), (-,s)\}$$

DOVE

$+$	SIGNIFICA	TEST POSITIVO
$-$	"	TEST NEGATIVO
m	"	ESAMINATO MALATO (CIOÈ VIRUS PRESENTE)
s	"	ESAMINATO SANO (CIOÈ VIRUS ASSENTE)

CONSIDERIAMO GLI EVENTI

A = "IL TEST È POSITIVO" $= \{(+,m),(+,s)\}$

M = "L'ESAMINATO È MALATO" $= \{(+,m),(-,m)\}$

$P(A|M^c)$ È LA PERCENTUALE DI FALSI POSITIVI, CIOÈ PERSONE SANE CHE RISULTANO POSITIVE AL TEST.

VOGLIAMO TROVARE $P(M|A)$. SAPPIAMO CHE

$$P(M) = \frac{1}{500} \qquad P(A|M) = \frac{99}{100} \qquad P(A|M^c) = \frac{2}{100}$$

QUINDI, TRAMITE LA FORMULA DI BAYES

$$P(M|A) = \frac{P(A|M)P(M)}{P(A|M)P(M) + P(A|M^c)P(M^c)} = \frac{\frac{99}{100}\cdot\frac{1}{500}}{\frac{99}{100}\cdot\frac{1}{500} + \frac{2}{100}\cdot\frac{499}{500}}$$

$$= \frac{99}{1097} \sim 0{,}09 \quad \text{CIOÈ IL } 9\%$$

QUESTO SIGNIFICA ANCHE CHE $P(M^c|A) \sim 0{,}91$, CIOÈ TRA LE PERSONE TESTATE CHE RISULTANO POSITIVE, 91% IN VERITÀ SONO SANE!

PUÒ SEMBRARE STRANO, MA PROVIAMO A VEDERE DI NUOVO I NUMERI. TESTANDO CAMPIONE DI 100.000 PERSONE,

PRESUMIBILMENTE 200 SARANNO MALATE (1/500) E 99.800 SANE. SE EFFETTUIAMO IL TEST, AVREMO

- PER LE PERSONE MALATE UN RISULTATO POSITIVO E CORRETTO SUL 99% DI ESSE, CIOÈ

$$200 \cdot \frac{99}{100} = 198$$

- PER LE PERSONE SANE UN RISULTATO POSITIVO ED ERRATO SUL 2% DI ESSE, CIOÈ

$$99.800 \cdot \frac{2}{100} = 1996$$

ABBIAMO QUINDI $1996 + 198 = 2194$ PERSONE SU CUI IL TEST RISULTA POSITIVO, MA SOLO 198 SONO REALMENTE MALATE ($\sim 9\%$, COME CALCOLATO PRIMA).

POTREBBE SEMBRARE STRANO COME RISULTATO, DOPO TUTTO I FALSI POSITIVI SONO POCHI: $P(A|M^c) = 2/100$.

IL PUNTO È CHE

$$P(M|A) = \frac{P(A|M)}{P(A|M) + \boxed{\frac{P(A|M^c)}{P(M)}} P(M^c)}$$

SE I SANI SONO L'AMPIA MAGGIORANZA, CIOÈ $P(M^c) \sim 1$

(NEL NOSTRO CASO 499/500), AFFINCHÉ $P(M|A) \sim 1$,

CIOÈ I TESTATI POSITIVI SIANO QUASI TUTTI MALATI,

QUELLO CHE CONTA È CHE SIA PICCOLO IL RAPPORTO

FALSI POSITIVI \ MALATI $\qquad \dfrac{P(A|M^c)}{P(M)}$

CIOÈ LA PERCENTUALE DI FALSI POSITIVI DEVE ESSERE

MOLTO PIÙ BASSA DELLA PERCENTUALE DEI MALATI:

$$P(A|M^c) \ll P(M)$$

TORNIAMO ORA AL CONCETTO DI EVENTI INDIPENDENTI.

<u>ESEMPIO</u> TRE SCIMMIE GIOCANO A DADI, OGNUNA LANCIANDO UN DADO A 6 FACCE. PROVIAMO CHE GLI EVENTI

A = "LA PRIMA E LA SECONDA SCIMMIA OTTENGONO LO STESSO RISULTATO"

B = "LA SECONDA E LA TERZA SCIMMIA OTTENGONO LO STESSO RISULTATO"

SONO INDIPENDENTI. CONVIENE PRENDERE COME SPAZIO CAMPIONE $\Omega = \{1, \dots, 6\}^3$, CIOÈ LE TERNE DI NUMERI DA 1 A 6, ASSOCIANDO ALLA COORDINATA K-ESIMA IL RISULTATO DEL LANCIO DEL K-ESIMO DADO. COSÌ FACENDO POSSIAMO SCRIVERE

$$A = \{(x,y,z) \in \Omega : x = y\} \quad B = \{(x,y,z) \in \Omega : y = z\}$$

$$A \cap B = \{(x,y,z) \in \Omega : x = y = z\}$$

POICHÉ $|\Omega| = 6^3 \quad |A| = |B| = 6^2 \quad |A \cap B| = 6$, DOTANDO Ω DI PROBABILITÀ UNIFORME ABBIAMO

$$P(A) = P(B) = \frac{6^2}{6^3} = \frac{1}{6} \quad \text{E} \quad P(A \cap B) = \frac{6}{6^3} = \frac{1}{36}$$

DUNQUE $p(A \cap B) = p(A) \cdot p(B)$ E GLI EVENTI SONO INDIPENDENTI.

SE DEFINIAMO $C = \{(x,y,z) \in \Omega : x = z\}$, CIOÈ "LA PRIMA E LA TERZA SCIMMIA OTTENGONO LO STESSO RISULTATO", SI VEDE ALLO STESSO MODO CHE

$$p(A \cap C) = p(A) \cdot p(C) \qquad p(B \cap C) = p(B) \cdot p(C)$$

CIOÈ A E C SONO INDIPENDENTI, E LO SONO ANCHE B E C. QUINDI A, B, C SONO A DUE A DUE INDIPENDENTI.

USANDO LA PROBABILITÀ CONDIZIONALE POSSIAMO SCRIVERE

$$p(A) = p(A|B) = p(A|C)$$

$$p(B) = p(B|A) = p(B|C)$$

$$p(C) = p(C|A) = p(C|B)$$

IL FATTO CHE A SIA INDIPENDENTE SIA DA B CHE DA C È PIUTTOSTO INTUITIVO. È ALTRETTANTO INTUITIVO CHE PERÒ A NON È INDIPENDENTE DAL FATTO CHE B E C ACCADANO ENTRAMBI, CIOÈ DALL'EVENTO $B \cap C$.

CIOÈ $p(A) \neq p(A|B \cap C)$, INFATTI

$$p(A|B \cap C) = \frac{p(A \cap B \cap C)}{p(B \cap C)} = 1$$

IN QUANTO $A \cap B \cap C = B \cap C \quad (= A \cap B = A \cap C)$

SE A FOSSE STATO INDIPENDENTE DA $B \cap C$ AVREMMO AVUTO

$$p(A) = p(A \mid B \cap C) = \frac{p(A \cap B \cap C)}{p(B \cap C)} = \frac{p(A \cap B \cap C)}{p(B)\,p(C)}$$

ESSENDO B E C INDIPENDENTI, E QUINDI

$$p(A)\,p(B)\,p(C) = p(A \cap B \cap C) \qquad (*)$$

QUESTO SUGGERISCE LA SEGUENTE DEFINIZIONE

<u>DEFINIZIONE</u> TRE EVENTI A, B, C IN UNO SPAZIO DI PROBABILITÀ SI DICONO <u>INDIPENDENTI</u> SE SONO INDIPENDENTI A DUE A DUE E SE VALE LA $(*)$

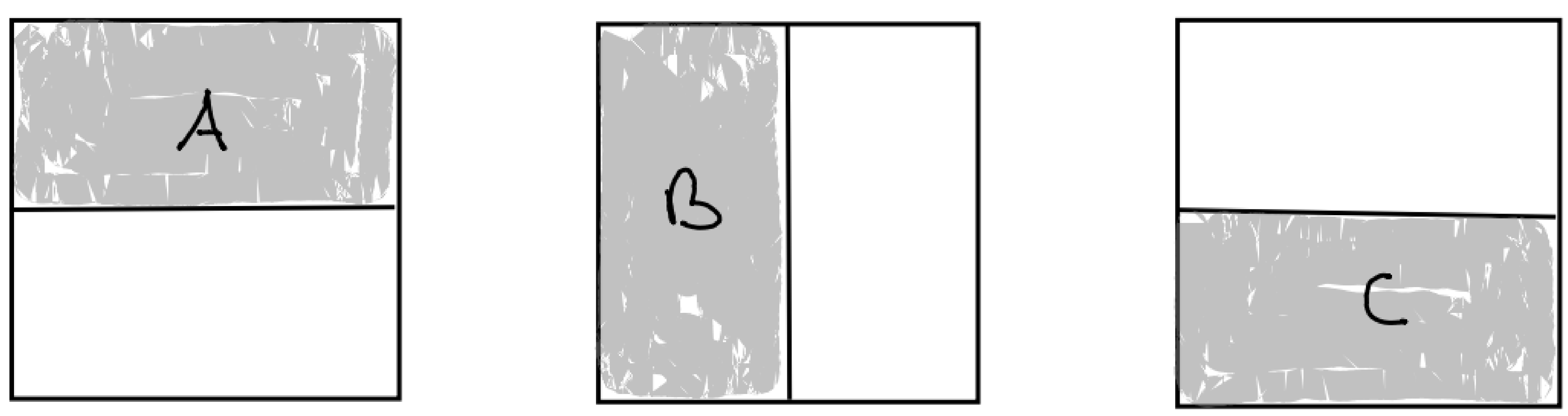
PIÙ IN GENERALE, DATA UNA FAMIGLIA $\{A_k\}_{k \in K}$ DI EVENTI, DIREMO CHE GLI EVENTI SONO INDIPENDENTI SE PER OGNI SOTTOINSIEME FINITO $J \subset K$ CON $|J| \geq 2$ SI HA

$$p\Big(\bigcap_{k \in J} A_k\Big) = \prod_{k \in J} p(A_k)$$

NEL CASO DI TRE EVENTI $K = \{1,2,3\}$ E

$$J = \{1,2\},\ \{1,3\},\ \{2,3\},\ \{1,2,3\}$$

<u>OSSERVAZIONE</u> L'INDIPENDENZA NON GODE DELLA PROPRIETÀ TRANSITIVA: SE A È INDIPENDENTE DA B E B È INDIPENDENTE DA C, NON È DETTO CHE A SIA INDIPENDENTE DA C. PER ESEMPIO, SUL QUADRATO UNITARIO CONSIDERIAMO I TRE EVENTI



ABBIAMO
$$P(A) = P(B) = P(C) = 1/2$$
$$P(A \cap B) = P(B \cap C) = 1/4$$
$$P(A \cap C) = 0$$

QUINDI $P(A \cap B) = P(A)P(B)$, $P(B \cap C) = P(B)P(C)$

MA $P(A \cap C) = 0 \neq \frac{1}{4} = P(A)P(C)$

FACCIAMO UN ULTERIORE ESEMPIO

<u>ESEMPIO</u> CONSIDERIAMO IL LANCIO DI DUE DADI A 6 FACCE. USIAMO $\Omega = \{1, \dots, 6\}^2$ DOTATO DI PROBABILITÀ UNIFORME. CONSIDERIAMO GLI EVENTI

$A =$ "IL RISULTATO DEL SECONDO DADO È 1, 2, OPPURE 5" $= \{(x,y) \in \Omega : y = 1,2,5\}$

$B =$ "IL RISULTATO DEL SECONDO DADO È 4, 5, OPPURE 6" $= \{(x,y) \in \Omega : y = 4,5,6\}$

$C =$ "LA SOMMA DEI RISULTATI DEI DUE DADI VALE 9" $= \{(x,y) \in \Omega : x+y = 9\}$

ABBIAMO $C = \{(3,6), (4,5), (5,4), (6,3)\}$

$$A \cap B = \{(x,y) \in \Omega : y = 5\} \qquad A \cap C = \{(4,5)\}$$

$$B \cap C = \{(3,6), (4,5), (5,4)\} \qquad A \cap B \cap C = \{(4,5)\}$$

QUINDI, TENUTO CONTO CHE $|\Omega| = 36$

$$P(A) = P(B) = \frac{18}{36} = \frac{1}{2} \qquad P(C) = \frac{4}{36} = \frac{1}{9} \qquad P(A \cap B) = \frac{6}{36} = \frac{1}{6}$$

$$P(B \cap C) = \frac{3}{36} = \frac{1}{12} \qquad P(A \cap C) = P(A \cap B \cap C) = \frac{1}{36}$$

DA CUI

$$\begin{cases} P(A \cap B \cap C) = \frac{1}{36} = P(A)P(B)P(C) \\ P(A \cap B) = \frac{1}{6} \neq \frac{1}{4} = P(A)P(B) \\ P(A \cap C) = \frac{1}{36} \neq \frac{1}{18} = P(A)P(C) \\ P(B \cap C) = \frac{1}{12} \neq \frac{1}{18} = P(B)P(C) \end{cases}$$

GLI EVENTI NON SONO INDIPENDENTI

OSSERVAZIONE IL FATTO CHE $P(A \cap B \cap C) = P(A)P(B)P(C)$ NULLA DICE SUL FATTO CHE GLI EVENTI A, B, C SIANO A DUE A DUE INDIPENDENTI.

OSSERVAZIONE SE A E B SONO EVENTI INDIPENDENTI, ALLORA ANCHE A^c E B LO SONO, IN QUANTO

$$P(A^c \cap B) = P(B \setminus (A \cap B)) \underset{\text{PROPOSIZIONE 1}}{=} P(B) - P(A \cap B)$$

$$\underset{\substack{\text{PER INDIPENDENZA} \\ \text{TRA } A \text{ E } B}}{=} P(B) - P(A)P(B) = P(B)\left[1 - P(A)\right] = P(B)P(A^c)$$

OVVIAMENTE, VISTO CHE $(A^c)^c = A$, VALE ANCHE IL VICEVERSA. PIÙ ESTENSIVAMENTE, ABBIAMO MOSTRATO CHE LE SEGUENTI AFFERMAZIONI SONO EQUIVALENTI

- A E B SONO INDIPENDENTI
- A E B^c " "
- A^c E B " "
- A^c E B^c " "

TUTTO QUESTO PUÒ ESSERE GENERALIZZATO A FAMIGLIE CON PIÙ DI DUE EVENTI

Proposizione Sia $\{A_k\}_{k\in K}$ una famiglia di eventi indipendenti. Sia $J \subset K$ un sottoinsieme di indici. Definiamo

$$B_k := \begin{cases} A_k & \text{se } k \in J \\ A_k^c & \text{se } k \in K \setminus J \end{cases}$$

Allora $\{B_k\}_{k\in K}$ è una famiglia di eventi indipendenti.

Proof omessa.

Esempio Consideriamo la famiglia $\{A_1, A_2, A_3, A_4\}$, quindi $K = \{1,2,3,4\}$. Prendiamo $J = \{2,3\}$. Allora $B_1 = A_1^c$, $B_2 = A_2$, $B_3 = A_3$, $B_4 = A_4^c$.