

01/10/02 2F PNI probabilità: primi elementi

1. In una classe di 20 alunni si svolge una interrogazione a sorpresa che coinvolge 3 studenti. Si sa che 15 alunni hanno studiato e 5 no. Determinare la probabilità dei seguenti eventi: A = "i 3 studenti sono tutti preparati" B = "i 3 studenti sono tutti impreparati" C = "1 studente è preparato e 2 no" D "non accade l'evento C"

L'evento corrisponde all'accadimento di 3 eventi tra loro indipendenti ma caratterizzati da una progressiva modifica dello spazio campionario (si usa la probabilità condizionale $p(E_1 \cap E_2) = p(E_1) p(E_2|E_1)$). Nel nostro caso la situazione si ripete 2 volte di fila.

$$p(A) = \frac{15}{20} \times \frac{14}{19} \times \frac{13}{18} = \frac{91}{228} \approx 0.399$$

$$\text{Analogamente } p(B) = \frac{5}{20} \times \frac{4}{19} \times \frac{3}{18} = \frac{1}{114} \approx 0.00877$$

Per determinare p(C) bisogna tener conto che lo studente preparato può essere interrogato sia come primo, sia come secondo o come terzo avendo come soci due impreparati e che la probabilità di tale evento è sempre la stessa (basta scrivere esplicitamente le altre due ed usare la proprietà commutativa; pertanto (visto che i tre casi sono tra loro incompatibili) si

$$\text{dovranno sommare 3 valori uguali e allora } p(C) = 3 \times \frac{15}{20} \times \frac{5}{19} \times \frac{4}{18} = \frac{5}{38} \approx 0.1316$$

$$\text{Infine } p(D) = 1 - p(C) = \frac{113}{114} \approx 0.868$$

Note: tranne in pochi casi non ho visto né l'analisi dell'esperimento e degli eventi né tantomeno il richiamo ai teoremi della probabilità. Qualcuno ha sbagliato anche analisi di eventi semplici; in molti hanno dimenticato di moltiplicare per 3 il calcolo di p(C). Come ho raccomandato più volte evitare di semplificare sopra il testo il che rende illeggibile l'elaborato. I valori approssimati vanno indicati con almeno 3 cifre significative.

2. In un campione di prodotti provenienti da Taiwan dove la produzione è di grande serie ma la qualità non è altrettanto elevata sono presenti 4 componenti per Pc difettosi su 10. Se ne estraggono a caso 3.

Determinare la probabilità dei seguenti eventi: A = "almeno 1 componente è difettoso"; B = "solo 1 componente è difettoso"

L'evento A è l'evento contrario di "tutti i componenti sono funzionanti" che corrisponde alla estrazione successiva caratterizzata da indipendenza ma condizionata (si veda la correzione dell'esercizio precedente); pertanto:

$$p(A) = 1 - \frac{6}{10} \times \frac{5}{9} \times \frac{4}{8} = \frac{5}{6} \approx 0.833$$

L'evento B può accadere nei seguenti casi DFF, FDF, FFD e ciascuno di questi 3 eventi (come nel problema precedente) ha la stessa probabilità; vediamo esplicitamente $p(DFF) = \frac{4}{10} \times \frac{6}{9} \times \frac{5}{8}$; $p(FDF) = \frac{6}{10} \times \frac{4}{9} \times \frac{5}{8}$; $p(FFD) = \frac{6}{10} \times \frac{5}{9} \times \frac{4}{8}$.

I tre eventi sono incompatibili e pertanto si devono sommare le probabilità ottenendo:

$$p(B) = 3 \times \frac{4}{10} \times \frac{6}{9} \times \frac{5}{8} = \frac{1}{2}$$

1 sistema 3 eq	2 sistema sostit	3 probabilità	4 probabilità	totale	voto

29/04/02 2F PNI calcolo combinatorio e calcolo probabilità

1. Calcolare quanti anagrammi si possono formare con la parola SETTEMBRE prescindendo dal senso e spiegare perché si tratta di permutazioni con ripetizione del tipo $P_9^{3,2}$

Gli anagrammi sono sempre delle permutazioni perché si utilizzano tutti i simboli forniti e si tiene conto dell'ordine. Quando però, come in questo caso alcune lettere sono ripetute k_1 o k_2 volte si possono effettuare $k_1!$ e $k_2!$ permutazioni della stessa parola ottenendo lo stesso anagramma. Per questa ragione (visto che ci sono 9 posti ma la lettera E è presente 3 volte e la lettera T lo è 2 volte) il numero di anagrammi corrisponderà alle permutazioni di 9 oggetti con due lettere ripetute 3 e 2 volte pertanto $P_9^{3,2} = \frac{9!}{3!2!} = 30 \cdot 240$

2. Risolvere l'equazione ai coefficienti binomiali prestando attenzione alla accettabilità delle soluzioni: $6 \binom{x}{5} = \binom{x+2}{5}$

Per la accettabilità delle soluzioni deve essere X intero e $x \geq 5$

Applicando la definizione di coefficiente binomiale e semplificando per 5! che è presente in entrambi i membri si ha:

$$6x(x-1)(x-2)(x-3)(x-4) = (x+2)(x+1)x(x-1)(x-2)$$

Poiché con le condizioni poste $x(x-1)(x-2) \neq 0$ si può semplificare e si ottiene l'equazione di II grado:

$$6(x-3)(x-4) = (x+2)(x+1) \Leftrightarrow 5x^2 - 45x + 70 = 0 \Leftrightarrow x^2 - 9x + 14 = 0$$

$$\Delta = 81 - 56 = 5^2 \text{ e pertanto } x_{1,2} = \frac{9 \pm 5}{2} = \begin{cases} 7 \text{ s.a.} \\ 2 \text{ s.n.a} \end{cases}$$

3. In una partita di 100 scatole di sugo ce ne sono 7 provenienti da una macchina che ne ha inquinato il contenuto. Si procede ad una indagine a campione e se ne estraggono 8

- Qual è la probabilità $p(E_1)$ che 2 scatole estratte siano inquinate?
- Qual è la probabilità $p(E_2)$ che nessuna delle scatole estratte sia inquinata?
- Qual è la probabilità $p(E_3)$ che qualcuna delle scatole estratte sia inquinata?
- Qual è la probabilità $p(E_4)$ che vengano estratte tutte le scatole inquinate?

Il problema ha a che fare con il concetto di combinazione perché si prescinde dall'ordine.

- a) Il numero di casi possibili in tutto l'esperimento è dato dalle combinazioni di 100 oggetti in classe 8 e cioè:

$$n_p = C_{100,8} \approx 1.861 \cdot 10^{11}$$

L'evento E_1 corrisponde ad allineamenti formati combinando 7 oggetti in classe 2 ed associando a tale allineamento tutti i casi in cui i restanti 6 posti sono occupati da scatole non inquinate (che sono 93). Il numero dei casi favorevoli all'evento E_1 sarà indicato con n_1 e si avrà: $n_1 = C_{7,2} \cdot C_{93,6} \approx 1.60 \cdot 10^{10}$

$$p_1 = \frac{n_1}{n_p} = 0.0861$$

- b) $n_2 = C_{93,8} \approx 1.0184 \cdot 10^{11}$ e pertanto $p_2 = \frac{n_2}{n_p} = 0.547$

- c) $p_3 = p(\overline{E_2}) = 1 - p_2 = 1 - 0.547 = 0.453$

- d) in questo rimane un solo posto per le scatole non inquinate e si otterrà probabilmente una probabilità molto bassa; in effetti:

$$n_4 = C_{7,7} \cdot C_{93,1} = 93$$

$$p_4 = \frac{n_4}{n_p} = 5.00 \cdot 10^{-10}$$

4. Un sistema di ponti collega mediante un unico circuito chiuso e senza intrecci un arcipelago di 10 isole garantendo i collegamenti stradali tra di esse. Per effetto di frequenti maremoti la probabilità che un ponte

si rompa è $p(R) = \frac{1}{5}$.

- a) Quanti ponti al più si possono rompere affinché sia garantito il collegamento reciproco di tutte le isole?

- b) Quanto vale la probabilità che il collegamento reciproco sia garantito?

- a) Se si rompe uno qualsiasi dei ponti il collegamento è garantito (le due isole collegate dal ponte interrotto saranno collegate facendo tutta la restante parte di anello). Se però si interrompe un secondo ponte le isole restano divise in due porzioni mutuamente irraggiungibili. Dunque è ammessa la rottura di un solo ponte.

- b) L'evento E è dato dalla unione di due eventi incompatibili e cioè $E_1 =$ "tutti i ponti funzionano" ed $E_2 =$ "si interrompe uno e un solo ponte". La probabilità sarà allora la somma delle due.

E_1 corrisponde al verificarsi simultaneo di 10 eventi tra loro indipendenti ciascuno di probabilità $1 - 1/5 = 4/5$ e pertanto

$$p(E_1) = (4/5)^{10}$$

$p(E_2)$ corrisponde al verificarsi simultaneo del caso in cui sia rotto un ponte e gli altri 9 siano funzionanti; ma il ponte rotto può essere scelto in 10 modi e si avrà pertanto:

$$p(E_2) = 10 (1/5) (4/5)^9 = 2 (4/5)^9$$

$$p(E) = p(E_1) + p(E_2) = (4/5)^{10} + 2 (4/5)^9 = 14/5 (4/5)^9 \approx 0.3758$$

1) anagrammi		
2) equazione ()		
3) probabilità		
4) probabilità		

3F PNI 30/4/2003 statistica descrittiva e primi elementi calcolo delle probabilità

Dopo aver osservato la griglia di correzione svolgere non più di un esercizio di statistica, due di c.c. e due di c.p. evitando di saltare completamente un singolo argomento. Attenzione: evitare di pasticciare un po' qui e un po' là. L'esercizio 6 che è più ampio vale per 2.

1	statistica	
2	c.c.	
3	c.c.	
4	c.c.	
5	c.p.	
6	c.p.	
7	c.p.	
Valutazione:		

- 1) Data la coppia di variabili X e Y che assume forniti a parte si svolga manualmente e in modo completo la regressione lineare con il metodo dei minimi quadrati. Nel calcolo devono essere riportati analiticamente i valori la cui somma consente il calcolo degli indicatori statistici per somma (si veda la tabella facente parte del testo). Determinare (precisione entro la quarta cifra decimale) tutti i valori della tabella e il valore estrapolato per $x = 4.5$

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	1.8	2.0972	3.24	4.398248	3.77496
2	2.2	2.5276	4.84	6.388762	5.56072
3	2.6	3.1992	6.76	10.23488	8.31792
4	3.0	5.336	9	28.4729	16.008
5	3.4	4.932	11.56	24.32462	16.7688
6	3.8	6.1692	14.44	38.05903	23.44296
somme	16.8	24.2612	49.84	111.8784	73.87336
	μ_x	μ_y	μ_{x^2}	μ_{y^2}	μ_{xy}
medie	2.8	4.043533	8.3066667	18.64641	12.31223
	cov(x,y)	$\rho(x,y)$	m	q	estrap
	0.990333	0.956685	2.122143	-1.89847	7.651176

- 2) Perché le diagonali di un poligono di n lati sono $\mathcal{C}_{n,2} - n$? Calcolare il valore corrispondente e interpretare geometricamente il risultato.
Se ci sono n lati ci sono n vertici e facendo $\mathcal{C}_{n,2}$ si contano tutte le possibili congiungenti di due vertici (cioè tutti i lati e tutte le diagonali); sottraendo il numero di lati restano le diagonali.

$$\mathcal{C}_{n,2} - n = \frac{n(n-1)}{2} - n = \frac{n^2 - n - 2n}{2} = \frac{n^2 - 3n}{2} = \frac{n(n-3)}{2}$$

L'interpretazione è la seguente: da ognuno degli n vertici si può tracciare una diagonale verso n-3 vertici (vanno esclusi quello di partenza e i due adiacenti). Il valore va diviso per 2 per evitare di contare 2 volte la stessa diagonale.

- 3) Proprietà dei fattoriali

a) Calcolare l'espressione $d_m = \frac{1}{m!} - \frac{1}{(m+1)!}$

$$d_m = \frac{1}{m!} - \frac{1}{(m+1)!} = \frac{m+1-1}{(m+1)!} = \frac{m}{(m+1)!}$$

b) Utilizzare il risultato precedente per calcolare la somma $S_n = \frac{1}{2!} + \frac{2}{3!} + \frac{3}{4!} + \dots + \frac{n}{(n+1)!}$

Usando la proprietà precedente si ha: $S_n = \frac{1}{2!} + \frac{2}{3!} + \frac{3}{4!} + \dots + \frac{(n-1)}{n!} = \left(\frac{1}{1!} - \frac{1}{2!}\right) + \left(\frac{1}{2!} - \frac{1}{3!}\right) + \dots + \left(\frac{1}{(n-1)!} - \frac{1}{n!}\right) = 1 - \frac{1}{n!}$

4) Sfruttando le proprietà dei coefficienti binomiali semplificare e poi risolvere l'equazione:

$$\binom{n}{9} - \binom{n-1}{9} + 3 \binom{n-1}{8} = k \binom{n}{8} \text{ con } k \text{ naturale diverso da } 0$$

Per la definizione di coefficiente binomiale deve essere n naturale e $n \geq 9$. Per la proprietà sulla somma dei coefficienti binomiali usata nella costruzione del triangolo di Tartaglia si ha:

$$\binom{n}{9} = \binom{n-1}{9} + \binom{n-1}{8} \text{ e pertanto l'equazione diventa:}$$

$$\binom{n-1}{8} + 3 \binom{n-1}{8} = k \binom{n}{8} \Leftrightarrow 4 \binom{n-1}{8} = k \binom{n}{8} \Leftrightarrow 4(n-1)(n-2)\dots(n-8) = kn(n-1)\dots(n-7) \Leftrightarrow 4(n-8) = kn \Leftrightarrow n(4-k) = 32 \Leftrightarrow$$

$$n = \frac{32}{4-k} \text{ si hanno soluzioni solo per } k=2 \Rightarrow n=16 \text{ e per } k=3 \Rightarrow n=32.$$

5) Data una popolazione di n individui per individuare i portatori di una malattia che presenta statisticamente una incidenza del 3.5‰ (per mille) si mette a punto un test diagnostico dotato del seguente grado di affidabilità: rivela i portatori della malattia nel 98% dei casi; dà dei falsi positivi (indica come malato un sano) nello 0.3% dei casi.

a) Determinare le 4 probabilità corrispondenti ai 4 eventi elementari possibili dopo averli enumerati

b) Determinare la probabilità che un cittadino sottoposto al test venga definito malato

Se ripartiamo la nostra popolazione in sani (s) e malati (m) con $s + m = n$ avremo i seguenti possibili eventi di interesse statistico:

E_1 = "un paziente sano viene sottoposto al test e riconosciuto sano"

E_2 = "un paziente sano viene sottoposto al test e riconosciuto malato"

E_3 = "un paziente malato viene sottoposto al test e riconosciuto sano"

E_4 = "un paziente malato viene sottoposto al test e riconosciuto malato"

$p(E_1) = (1 - 0.0035) \cdot (1 - 0.003) = 0.9935105$ perché E_1 avviene per congiunzione di due eventi indipendenti (la scelta del paziente e la affidabilità del test).

$$p(E_2) = (1 - 0.0035) \cdot 0.003 = 0.0029895$$

$$p(E_4) = 0.0035 \cdot 0.98 = 0.00343$$

$$p(E_3) = 0.0035 \cdot (1 - 0.98) = 0.000070$$

L'evento E = "un cittadino sottoposto a test viene dichiarato malato" corrisponde alla disgiunzione di E_2 ed E_4 che sono tra loro incompatibili e pertanto:

$$p(E) = p(E_2) + p(E_4) = 0.064195$$

6) Un imputato viene giudicato da una giuria che può assumere le sue decisioni a maggioranza ed è composta da 3 giurati. I giurati A e B possono decidere (in modo indipendente) per l'assoluzione con probabilità p . Il giurato C decide invece con probabilità p' . Indicare con E_A, E_B, E_C gli eventi corrispondenti al voto favorevole alla assoluzione.

a) Scrivere simbolicamente a cosa corrisponde l'evento E = "la giuria decide per la assoluzione dell'imputato" attraverso l'uso dei connettivi \neg (non) \wedge (e) \vee (o). $E = E_1 \vee E_2 \vee E_3 \vee E_4$ con $E_1 = \dots$

$$E = (E_A \wedge E_B \wedge E_C) \vee (\neg E_A \wedge E_B \wedge E_C) \vee (E_A \wedge \neg E_B \wedge E_C) \vee (E_A \wedge E_B \wedge \neg E_C)$$

b) Calcolare le probabilità $p(E_1), p(E_2), p(E_3), p(E_4)$ e $p(E)$

I 4 eventi che producono per disgiunzione E hanno probabilità

$$p(E_1) = ppp' = p^2p' \quad p(E_2) = p(E_3) = (1-p)pp' \quad p(E_4) = p^2(1-p')$$

la loro disgiunzione (trattandosi di eventi incompatibili) è pari alla somma delle probabilità

$$p(E) = p^2p' + 2(1-p)pp' + p^2(1-p') = -2p^2p' + 2pp' + p^2$$

- c) Rappresentare in un diagramma la funzione lineare $y = p(E) = f(p')$ analizzando i seguenti elementi: come mai la retta ha coefficiente angolare positivo? Come mai per $p'=0$ (giurato colpevolista per partito preso) si ha $y = p^2$?
 Si ha $y = p(E) = f(p') = 2p'(p - p^2) + p^2$ e si tratta di una retta con coefficiente angolare positivo perché essendo p compreso tra 0 e 1 la quantità $p - p^2$ è sicuramente positiva. Se $p' = 0$ si può avere assoluzione solo se A e B sono per l'assoluzione e la probabilità è appunto p^2 .
- d) Quanto vale y quando $p' = 1$ (giurato innocentista). Interpretare il risultato trovato soffermandosi in particolare sul fatto che il valore trovato è sempre ≤ 1 . Quando si raggiunge la probabilità $y = 1$?
 Per $p' = 1$ si ha $y = 2p - p^2$. La disequazione $2p - p^2 \leq 1$ equivale a $(1-p)^2 \geq 0$ che è sempre vera e si ottiene il valore 1 (assoluzione certa) solo quando $p = 1$ (non c'è scampo per l'imputato: l'unico modo per avere la certezza della assoluzione è che tutti i giurati siano innocentisti)
- e) Come si imposta tutta l'analisi nel caso in cui il giurato B per dinamiche interne alla carica di consiglio (ha litigato con A) decide di votare sempre in modo difforme ad A?
 In questo caso dei 4 eventi non possono più accadere E_1 ed E_4 perché E_A ed E_B sono incompatibili e questa volta $p(E) = p(E_2) + p(E_3) = 2(1-p)pp'$. Al variare di p si tratta di una parabola con la concavità verso il basso che assume il suo valore massimo per $p = \frac{1}{2}$ con ordinata $y = \frac{1}{2} p'$ (il che vuol dire che fissata la disponibilità del giurato C all'imputato conviene che A e B siano sulla posizione più equa possibile).
- 7) Il dodecaedro è formato da 12 pentagoni regolari. Si consideri un dado dodecaedrico con facce numerate da 1 a 12.

- a) Trovare la probabilità $p(E)$ che il risultato non sia un multiplo di 3.

Indichiamo con $E_3 =$ "il risultato del lancio è un multiplo di 3" allora $E = \overline{E_3}$

I multipli di 3 sono {3,6,9,12} e pertanto $p(E_3) = \frac{4}{12} = \frac{1}{3}$

- b) Usando una metodologia simile a quella usata per affrontare il punto precedente trovare la probabilità $p(E)$ dell'evento $E =$ "il prodotto dei risultati di due lanci consecutivi sia un multiplo di 5".

Se il prodotto è un multiplo di 5 vuol dire che in almeno uno dei due lanci è uscito il 5 o il 10. Indichiamo con $E' =$ "nel lancio di un dado non esce né il 5 né il 10" $E' =$

{1,2,3,4,6,7,8,9,11,12} e $\text{card}(E') = 10$ pertanto $p(E') = \frac{10}{12} = \frac{5}{6}$

Sia ora $E'' =$ "nel lancio consecutivo dei due dadi non esce né il 5 né il 10". Poiché i due eventi che costituiscono E'' sono indipendenti ed hanno la stessa probabilità si ha $p(E'') = p(E') p(E') = \frac{25}{36}$

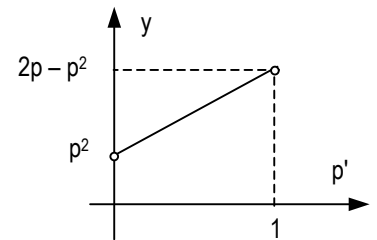
Ma $E = \overline{E''}$ e pertanto $p(E) = 1 - p(E'') = 1 - \frac{25}{36} = \frac{11}{36}$

- c) Analizzare il doppio lancio contare i casi favorevoli e quelli possibili e attraverso il rapporto determinare il risultato del punto precedente.

casi favorevoli: possono verificarsi tre tipi di eventi più semplici e cioè: I) esce 5 nel primo dado e uno qualsiasi dei dodici risultati nel secondo; ciò si può verificare in 12 modi II) esce 10 nel primo dado e uno qualsiasi dei dodici risultati nel secondo; ciò si può verificare in 12 modi III) non esce né 5 né 10 nel primo dado ed esce o 5 o 10 nel secondo; ciò si può verificare in $10 \cdot 2 = 20$ modi. In totale il numero di casi favorevoli $n_f = 12+12+20 = 44$ modi.

casi possibili $n_p = D_{12,2} = 12^2$

$$p(E) = \frac{n_f}{n_p} = \frac{44}{144} = \frac{11}{36}$$



Problema discusso in classe: Uno strano problema di calcolo combinatorio

- a) Quante coppie n si possono formare tra 9 giocatori di carte per fare in modo che essi si possano sfidare in modo contrapposto l'uno contro l'altro?
- b) A cosa corrisponde $n' = C_{9,4}$?
- c) Quante partite n'' si possono giocare considerando tutti i possibili accoppiamenti se due partite sono diverse quando almeno un giocatore è diverso o, a parità di giocatori, si scambiano gli accoppiamenti?
- d) Supponiamo ora che si voglia valutare il numero n''' di partite che si possono disputare giocando a coppie contrapposte in modo che ciascun giocatore possa incontrare tutti gli altri come avversari.
- (a) L'ordine non ha importanza e pertanto si tratta di combinazioni senza ripetizione (un giocatore non può giocare contro se stesso) si ha dunque $n = C_{9,2} = \binom{9}{2} = \frac{9 \cdot 8}{2} = 36$

(b) $n' = C_{9,4} = \binom{9}{4} = \frac{9 \cdot 8 \cdot 7 \cdot 6}{4 \cdot 3 \cdot 2 \cdot 1} = 126$ sono le combinazioni dei 9 giocatori in classe 4 e cioè le partite che si possono giocare a coppie con almeno un giocatore diverso se si prescinde dal fatto che due giocatori giochino in coppia o contrapposti.

(c) Se si vuole contare il numero di partite del tipo $(a,b;c,d)$, $(a,c;b,d)$, $(a,d;b,c)$ come se fossero diverse il valore precedente va moltiplicato per 3 (si tenga presente che scelta la prima coppia la seconda è automaticamente fissata per una data combinazione) e dunque il numero di partite che si possono giocare facendo partecipare tutte le possibili coppie di concorrenti è $n'' = 3 \cdot C_{9,4} = 378$.

(d) L'ultima domanda appare più complessa da analizzare perché ciò che conta è la scelta dell'avversario. Indichiamo con S_n il numero di partite per n giocatori ed operiamo poi per arricchimenti successivi per trovare il valore ma anche per scoprire la eventuale legge di costruzione del risultato (per comodità indicheremo i giocatori con le lettere minuscole dell'alfabeto)

$n = 4$ $S_4 = 2$

$ab \leftrightarrow cd$ è necessaria uno scambio tra a e c per garantire che a si scontri con b e c con d
scambio

$ac \leftrightarrow bd$

$n = 5$ $S_5 = 4$

Si aggiunge un nuovo giocatore che crea una ridondanza perché bisognerà necessariamente accoppiarlo con un giocatore che ha già fatto la partita contro gli stessi avversari (indicheremo la libertà di scelta del giocatore con \square)

$ab \leftrightarrow cd$ $cd \leftrightarrow \square e$

$ab \leftrightarrow \square e$

scambio

$ac \leftrightarrow bd$

$n = 6$ $S_6 = 5$

Quando il numero di giocatori è pari si colmano alcune ridondanze ma se ne possono creare altre

$ab \leftrightarrow cd$ $cd \leftrightarrow ef$

$ab \leftrightarrow ef$

scambio

$ab \leftrightarrow cd$ $e \square \leftrightarrow f \square$

$n = 7$ $S_7 = 8$

$ab \leftrightarrow cd$ $cd \leftrightarrow ef$ $ef \leftrightarrow g \square$

$ab \leftrightarrow ef$ $cd \leftrightarrow g \square$

$ab \leftrightarrow g \square$

scambio

$ab \leftrightarrow cd$ $eg \leftrightarrow f \square$

$n = 8$ $S_8 = 8$

$ab \leftrightarrow cd$ $cd \leftrightarrow ef$ $ef \leftrightarrow gh$

$ab \leftrightarrow ef$ $cd \leftrightarrow gh$

$ab \leftrightarrow gh$

scambio

$ab \leftrightarrow cd$ $eg \leftrightarrow fh$

le ridondanze si eliminano e pertanto $S_7 = S_8$

$n = 9$ $S_9 = 12$

$ab \leftrightarrow cd$ $cd \leftrightarrow ef$ $ef \leftrightarrow gh$ $gh \leftrightarrow i \square$

$ab \leftrightarrow ef$ $cd \leftrightarrow gh$ $ef \leftrightarrow i \square$

$ab \leftrightarrow gh$ $cd \leftrightarrow i \square$

ab↔i□
 scambio
 ab↔cd

eg↔fh

n = 10 S₁₀ = 13

ab↔cd cd↔ef ef↔gh gh↔il
 ab↔ef cd↔gh ef↔il
 ab↔gh cd↔il
 ab↔il
 scambio

ab↔cd eg↔fh i□↔l□

n = 11 S₁₁ = 18

ab↔cd cd↔ef ef↔gh gh↔il il↔m□
 ab↔ef cd↔gh ef↔il gh↔m□
 ab↔gh cd↔il ef↔m□
 ab↔il cd↔m□
 ab↔m□
 scambio
 ab↔cd eg↔fh im↔l□

La legge che consente di costruire il numero di partite diventa evidente da n = 5 in poi. Ogni volta che si aggiunge un giocatore per n dispari si aggiunge una colonna con un elemento in più rispetto al caso precedente. Se invece n è pari si saturano le ridondanze lasciate dal caso precedente e la parte superiore non muta.

Nelle configurazioni denominate scambio si ha invece un incremento di una partita ogni 4 casi. Così se indichiamo con s₁ il numero di partite della prima colonna cioè le partite giocate dalla prima coppia contro tutti gli altri e con s₂ il numero di partite necessarie a garantire che chi ha giocato in coppia giochi ora in contrapposizione si ha:

n	s _{1n}	s _{2n}	Calcolo	S _n
4	1	1	diretto	2
5	2	1	2 + 2·1 + 0	4
6	2	2	4+2·0+1	5
7	3	2	5+3·1+0	8
8	3	2	8+3·0+0	8
9	4	2	8+4·1+0	12
10	4	3	12+4·0+1	13
11	5	3	13+5·1+0	18
12	5	3	18+5·0+0	18
13	6	3	18+6·1+0	24
14	6	4	24+6·0+1	25

La legge ricorsiva di costruzione è $S_n = S_{n-1} + s_{1n}\Delta s_1 + \Delta s_2$ mentre $s_{1n} = (\text{int}(n:2) + r) - 1$ e $s_{2n} = \text{int}[(n+2):4]$

3F PNI statistica descrittiva 25/02/04

1) E' data la seguente tabella a doppio ingresso relativa a due caratteri $X = \{x_1, x_2, x_3, x_4\}$ e $Y = \{y_1, y_2, y_3\}$ con i seguenti valori di frequenza:

X\Y	y ₁	y ₂	y ₃		X\Y	y ₁	y ₂	y ₃	
x ₁	80	20	60		x ₁				
x ₂	162	41	121		x ₂				
x ₃	126	32	95		x ₃				
x ₄	34	9	26		x ₄				

- a) completare la tabella calcolando le frequenze marginali
- b) costruire a fianco una seconda tabella con le medesime frequenze marginali e caratterizzata da indipendenza dei diversi caratteri (approssimare le frequenze teoriche alla prima cifra decimale e prestare attenzione al fatto che i totali devono dare la frequenza marginale).

2) Con riferimento alle frequenze dell'esercizio precedente supponiamo che $X = \{2.45, 2.54, 2.67, 2.81\}$ e che $Y = \{3.87, 4.22, 5.81\}$

- a) determinare μ_x, σ_x , nel caso che $y = y_2$
- b) nel caso precedente quanto vale la media armonica?

3) Un insieme di dati sperimentali ha portato alle seguenti misurazioni:

x	10	20	30	40	50	60	70
y	7.80	7.60	7.00	6.95	6.80	6.00	5.90

- a) rappresentare i dati (non è indispensabile costruire un diagramma enorme ma si tenga presente che y varia tra ...)
- b) costruire e completare la tabella che fornisce tutti gli elementi necessari al calcolo dei parametri statistici
- c) determinare il baricentro e il coefficiente angolare della retta di regressione lineare da x a y
- d) tracciare la retta di regressione sul diagramma dei dati sperimentali
- e) determinare il valore del coefficiente di correlazione
- f) calcolare i valori previsto per x = 5

1a	1b	2a	2b	3a	3b	3c	3d	3e	3f	1	2	3	tot
2	3	3	2	3	6	2	2	2	1	5	5	16	26

Nome e cognome : _____ 5F PNI 3/12/2004 distribuzioni di variabili aleatorie

Svolgere 3 esercizi scegliendone uno solo tra 1/2, 3/4, 5/6

1. In un gioco d'azzardo si vincono i seguenti importi con le probabilità indicate tra parentesi 20€(0.20), 30€(0.10) e 70€(0.05).

- a) Calcolare la puntata che corrisponde ad un gioco equo
 b) In media si hanno 250 puntate al giorno e i costi fissi di gestione della casa da gioco sono di 200€; a quanto va posta la puntata per garantire un utile netto giornaliero di 300€?

Basta calcolare il valor medio delle vincite $\mu = 20 \cdot 0.2 + 30 \cdot 0.1 + 70 \cdot 0.05 = 10.5\text{€}$

Indicata con x la posta dovrà essere $250 \cdot x = (200 + 300) + 250 \cdot 10.5$ da cui si ha $x = 12.5\text{€}$

Nota di correzione: in molti non hanno tenuto conto, nel caso b, che le vincite sono un costo.

2. Per avere l'idoneità ad un concorso bisogna superare tre prove. Dai risultati dei concorsi precedenti si sa che statisticamente il 90% dei candidati viene eliminato alla prima prova, l'80% dei rimanenti alla seconda e il 70% dei rimanenti alla II prova. Raggiunta l'idoneità il concorso viene superato dal 5% degli idonei.

- a) Se si vogliono assumere 250 persone quanti saranno i candidati da esaminare?
 b) Quale sarà lo scarto quadratico medio?

Le probabilità di superare le varie fasi sono date dai complementi a 1 e si tratta di eventi indipendenti; pertanto

$x \cdot 0.1 \cdot 0.2 \cdot 0.3 \cdot 0.05 \cdot 250$ e dunque $x = 833'333$

$p = 0.1 \cdot 0.2 \cdot 0.3 \cdot 0.05 = 3 \cdot 10^{-4}$ e $\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{250 \cdot (1-3 \cdot 10^{-4})} = 16$

3. Una ditta produce Cd vergini che vengono confezionati in pacchi da 50. Da indagini statistiche si sa che il 2% della produzione presenta difetti.

- a) Si prende un pacco a caso e si vuol sapere la probabilità che ne contenga al più uno difettoso. Eseguire il calcolo con la distribuzione di Poisson
 b) Come cambia l'analisi se si chiede che dalla apertura di 5 pacchi ce ne siano al più 5 difettosi e si esegue il calcolo con la distribuzione di Bernoulli?

Possiamo utilizzare la distribuzione di Poisson con $\lambda = 50 \cdot p = 1$

$p(X \leq 1) = p(X=0) + p(X=1) = e^{-1} (\lambda^0 + \lambda^1) = 0.736$

Se si utilizza la bernoulliana

$$p(X \leq 5) = \sum_{k=0}^5 \binom{250}{k} 0.02^k 0.98^{250-k}$$

ciò ci porta alla tabella

K	0	1	2	3	4	5	Somma
$\binom{250}{k} 0.02^k 0.98^{250-k}$	0.006405	0.032679	0.083030	0.140078	0.176527	0.177248	0.616967

Nota di correzione: le differenze tra le due distribuzioni incidono oltre la seconda cifra significativa; nel testo dato durante il compito non veniva precisato cosa usare e sarebbe stato opportuno usare la bernoulliana per il caso a e la poissoniana per il caso b

4. In un processo produttivo la percentuale di pezzi esenti da difetti è del 74%.
 a) Su una produzione di 10'000 pezzi qual è il limite inferiore di probabilità per il quale il numero di pezzi difettosi si discosta per meno del 5% dal suo valore medio?
 b) Il valore numerico precedentemente determinato è detto *grado di fiducia*. Se lo si volesse portare al 95% a quale quantitativo si dovrebbe fare riferimento?

Useremo la disuguaglianza di Cebyshev. Indicata con X la variabile casuale numero di pezzi difettosi 0,1, 2, ..., 10⁴ sarà $p = 0.26$ e $\mu_x = 0.26 \cdot 10^4$ mentre $\varepsilon = 0.05 \cdot 2600 = 130$ e $\sigma = \sqrt{n p (1-p)} = 439$

Indicata con p' il valore richiesto sarà

$p' = p(|X - \mu| < \varepsilon) = 1 - p(|X - \mu| \geq \varepsilon)$ con $p(|X - \mu| \geq \varepsilon) \leq (\sigma/\varepsilon)^2 = 0.0766$ pertanto

$p' \geq 1 - 0.0766 = 0.9234$

Se si vuole portare il grado di fiducia da 0.9234 a 0.95 sarà:

$0.95 = 1 - (\sigma/\varepsilon)^2 = 1 - \frac{n p (1-p)}{(0.05 n)^2} = 1 - \frac{1-p}{25 \cdot 10^{-4} \cdot n \cdot p}$ da cui $n = 22'770$ pezzi

Nota di correzione: in diversi hanno usato la disuguaglianza di Bernoulli o la legge dei grandi numeri. Consiglio di rileggersi a quale variabile casuale esse si riferiscono.

5. Di una variabile gaussiana non standardizzata si conoscono $f(\mu) = 1.140$ e $f(2.67) = 1.013$

Nome e cognome: _____ 5F PNI 18 gennaio 2006 – probabilità e statistica

1) Relazioni di base sulla probabilità

- a) Siano A e B due eventi casuali con $p(A) = 0.2$ e $p(B) = 0.4$. Spiegare perché non può essere $p(A \cup B) = 0.7$?

In generale dati due eventi A e B si ha sempre (teorema della probabilità totale) $p(A \cup B) = p(A) + p(B) - p(A \cap B) \leq p(A) + p(B) = 0.6$

- b) Siano A e B due eventi casuali con $p(A) = 0.6$ e $p(B) = 0.5$. Perché A e B non possono essere incompatibili? Se $p(A \cap B) = 0.3$ quanto valgono $p(A|B)$ e $p(A \cup B)$?

Se A e B sono incompatibili $p(A \cap B) = 0$ e allora $p(A \cup B) = p(A) + p(B) = 1.1 > 1$ dunque i due eventi non possono essere incompatibili. Se $p(A \cap B) = 0.3$ $p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{0.3}{0.5} = 0.6$ e $p(A \cup B) = p(A) + p(B) - p(A \cap B) = 0.8$

- c) Supponiamo che due eventi A e B siano indipendenti cioè che sia $p(A \cap B) = p(A) \cdot p(B)$. Ricordando la legge di De Morgan dimostrare che $p(\overline{A \cap B}) = p(\overline{A}) \cdot p(\overline{B})$ ovvero che anche gli eventi contrari sono indipendenti.

La dimostrazione è di tipo diretto: poiché $\overline{A \cap B} = \overline{A \cup B}$

$$p(\overline{A \cap B}) = p(\overline{A \cup B}) = 1 - p(A \cup B) = 1 - [p(A) + p(B) - p(A \cap B)] = 1 - p(A) - p(B) + p(A)p(B) = 1 - p(A) - p(B) + p(A)p(B) = (1 - p(A))(1 - p(B)) = p(\overline{A}) \cdot p(\overline{B})$$

Nota di correzione: scarsa conoscenza dei teoremi, delle definizioni, della terminologia. Il teorema della probabilità totale non richiede di fare due casi a seconda che sia $p(A \cap B) = 0$ oppure $p(A \cap B) \neq 0$

2) Variabili casuali continue. Sia X una variabile casuale continua diversa da 0 solo in $[2,5]$ e caratterizzata da una densità $f(x) = kx^2$. N.B. Svolgere le determinazioni scrivendo i risultati in forma frazionaria semplificata e al termine del calcolo dare il valore approssimato con 4 cifre decimali significative.

- a) Determinare k

Si tratta di normalizzare la funzione densità:

$$1 = \int_2^5 kx^2 dx = \frac{k}{3} x^3 \Big|_2^5 = \frac{k}{3} (125 - 8) = \frac{117}{3} k = 39 k = 1 \text{ da cui } k = \frac{1}{39}$$

- b) Calcolare il valor medio μ_x

$$\mu_x = \int_2^5 kx^3 dx = \frac{k}{4} x^4 \Big|_2^5 = \frac{k}{4} (625 - 16) = \frac{609}{4} k = \frac{203}{52} \approx 3.904$$

- c) Calcolare la varianza σ_x^2

$$\mu_{x^2} = \int_2^5 kx^4 dx = \frac{k}{5} x^5 \Big|_2^5 = \frac{k}{5} (3125 - 32) = \frac{k}{5} 3093 = \frac{3093}{195} = \frac{1031}{65}$$

$$\sigma^2 = \mu_{x^2} - \mu_x^2 = \frac{1031}{65} - \frac{203^2}{52^2} = \frac{1031 \cdot 13 \cdot 16 - 203^2 \cdot 5}{13^2 \cdot 16 \cdot 5} = \frac{8403}{13520} = 0.6215$$

$$\sigma \approx 0.7884$$

Nota di correzione: assolutamente sconveniente usare la definizione; sostituire i valori di k solo alla fine

3) Calcolo combinatorio applicato alla probabilità

Calcolare la probabilità di servire 5 carte dello stesso seme estraendole casualmente da un mazzo di 52.

Calcolare separatamente i casi favorevoli ed i casi possibili.

Casi possibili: sono le combinazioni (si prescinde dall'ordine) di 52 oggetti in classe 5. Dunque $n = \binom{52}{5} = 2'598'960$

Casi favorevoli sono $m = 4 \cdot \binom{13}{5} = 4 \cdot 1287$. Si moltiplica per 4 perché i 4 semi sono egualmente probabili e corrispondono al caso favorevole.

$$p = \frac{m}{n} = 0.00198$$

Nota di correzione: soluzione alternativa: l'evento può essere visto come intersezione di eventi l'uno condizionato dall'altro. La prima carta è qualsiasi ($p = 1$) ma determina il seme. La seconda ha probabilità di $12/51$ e così via.

4) Formula di Bayes. In un gruppo di laureati il 40% ha una laurea in ambito umanistico giuridico (LU), il 25% in ambito economico (LE), e il 35% in ambito tecnico scientifico (LT). Nessuno ha la doppia laurea. E' noto che la probabilità di disoccupazione è del 10% per LU, del 7% per LE e del 3% per LT.

a) Trovare la probabilità che un soggetto scelto a caso sia disoccupato.

Gli eventi LU, LE, LT sono incompatibili e creano una partizione dello spazio degli eventi.

Pertanto indicato con D l'evento "il laureato è disoccupato" si ha:

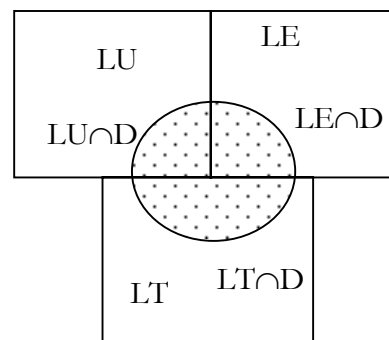
$$p(D) = p(D \cap LU) + p(D \cap LE) + p(D \cap LT) = p(D|LU) p(LU) + p(D|LE) p(LE) + p(D|LT) p(LT) = 0.10 \cdot 0.40 + 0.07 \cdot 0.25 + 0.03 \cdot 0.35 = 0.068$$

Nota di correzione: simboli a caso, quasi mai la giustificazione del calcolo

b) Nella ipotesi che il soggetto scelto sia disoccupato determinare la probabilità che esso sia del gruppo LT.

$$p(LT|D) = \frac{p(LT \cap D)}{p(D)} = \frac{p(D|LT) p(LT)}{p(D)} = \frac{0.03 \cdot 0.35}{0.068} = 0.154$$

Nota di correzione: ho visto chili di formule sparate a memoria senza ragionamento (se si usa la formula di Bayes bisogna spiegare a quali eventi la si applica e perché). Eppure ricordo con precisione di avervi messo in guardia a suo tempo e invitato a ragionare sulle partizioni negli spazi degli eventi. Per chi non ha capito aggiungo una figura:



5) Distribuzione binomiale. Il Preside di una scuola ha verificato che il 5% degli iscritti alla prima classe (iscrizione a gennaio quando si frequenta la III media) non confermano la loro iscrizione e pertanto trovandosi con $n = 320$ iscritti decide di formare 12 classi di 25 alunni per classe.

a) Determinare μ e σ per la variabile che descrive il numero di alunni che confermano l'iscrizione alla prima.

Uno qualsiasi degli alunni può confermare o non confermare con probabilità $p = 0.95$ e rappresenta una variabile che può prendere valori 0 o 1. Noi siamo interessati alla variabile somma che, astrattamente può prendere tutti i valori da 0 a 320. Nella distribuzione binomiale si ha $\mu = np = 320 \cdot 0.95 = 304$ mentre $\sigma = \sqrt{np(1-p)} = \sqrt{320 \cdot 0.95 \cdot 0.05} = 3.90$

b) Calcolare la probabilità che si debba formare anche una sola classe di 26 alunni.

La nostra variabile binomiale può prendere tutti i valori da 0 a 320 e noi siamo interessati al caso in cui si formino 11 classi da 25 e 1 classe da 26 cioè $X = 301$

$$p(X = 301) = \binom{320}{301} 0.95^{301} 0.05^{19} = 0.071$$

c) Calcolare la probabilità che tutte le classi abbiano più di 25 alunni

Se tutte le classi hanno più di 25 alunni ne hanno almeno 26 e dunque siamo interessati a calcolare $p(X \geq 12 \cdot 26) = p(X \geq 312)$ Convienne ovviamente utilizzare la possibilità di approssimare la bernoulliana con la gaussiana.

A questo scopo si tenga presente che poiché la gaussiana opera nel continuo e la bernoulliana nel discreto dovremo valutare la probabilità che $311.5 \leq X \leq 320.5$

$$\text{Standardizziamo la variabile tramite la trasformazione } z = \frac{x - \mu}{\sigma} \text{ e ciò ci porta a } z_1 = \frac{311.5 - 304}{3.90} = 1.923$$

$$\text{e } z_2 = \frac{320.5 - 304}{3.90} = 4.231$$

$$p(311.5 \leq X \leq 320.5) = F(z_2) - F(z_1) = 0.999988 - 0.972760 = 0.027228 \approx 0.272$$

Nota di correzione: errore medio dimenticarsi delle classi quando si passa da variabili discrete a variabili continue; errore grave dimenticarsi che gli studenti sono 320 e non infiniti.

6) Stima per intervallo di una media. Da una partita di mele si estrae un campione bernoulliano con $n = 350$. Per tale campione si valutano la media campionaria e lo scarto quadratico campionario corretto ottenendo $\bar{x} = 225$ g e $s^* = 32$ g.

a) Determinare con livello di confidenza del 95% l'intervallo di confidenza della massa media μ ($z_c = 1.96$)

Poiché per grandi campioni \bar{X} è una variabile gaussiana normale con media μ e varianza σ^2/n e poiché s^* è un buon stimatore di σ consideriamo la variabile standardizzata

$$z = \frac{(\bar{x} - \mu)\sqrt{n}}{s^*} \text{ a cui corrisponde lo } z \text{ critico } 1.96. \text{ Si avrà dunque la semidispersione } z_c \frac{s^*}{\sqrt{n}} = 1.96 \frac{32}{\sqrt{350}} = 3.35 \text{ g}$$

Dunque l'intervallo di confidenza va da $225 - 3.35 = 222$ a 228

Nota di correzione: l'intervallo ha lo stesso numero di cifre della media.

- b) Se si vuole ridurre a 1 g la semiampiezza dell'intervallo di confidenza quanto deve essere la dimensione del campione?

La semiampiezza dell'intervallo è $z_c \frac{s^*}{\sqrt{n}}$ e dunque dovrà essere $z_c \frac{s^*}{\sqrt{n}} \leq 1 \Rightarrow n \geq (z_c s^*)^2 = 3934$

Nota di correzione: è una disequazione non una equazione.

- 7) Distribuzione di Poisson. Un ospedale è dotato di 3 posti di rianimazione che vengono utilizzati per gestire nelle 24 ore la fase acuta degli accessi di Pronto Soccorso. Il numero di accessi che richiedessero l'uso della rianimazione degli ultimi 5 anni è stato di 682, 698, 710, 630, 723.

- a) Determinare il parametro λ del processo poissoniano.

Per determinare λ basta calcolare il valor medio del processo (si ricordi che $\mu = \lambda$) e a questo scopo una statistica relativa a ben 5 anni ci fornisce quanto serve. In effetti $\lambda = \frac{682 + 698 + 710 + 630 + 723}{365 \cdot 5} = 1.89$ accessi/giorno. La presenza di 1 o 2 anni bisestili non porta a significative variazioni.

- b) Stimare la probabilità che i 3 posti disponibili non siano sufficienti.

$$p(x > 3) = 1 - [p(x=0) + p(x=1) + p(x=2) + p(x=3)] = 1 - e^{-\lambda} \left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} \right) = 1 - e^{-1.89} \cdot 5.80 = 0.124$$

- c) Se si vuole portare il livello di rischio (cioè la probabilità che il numero di posti non riesca a coprire la richiesta) a 0.01 quanto deve essere il numero di posti disponibili?

Basta eseguire il calcolo precedente finché

$$p(x > n) = 1 - e^{-\lambda} \left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots + \frac{\lambda^n}{n!} \right) < 0.01 \text{ e cioè } \left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots + \frac{\lambda^n}{n!} \right) > 6.55$$

Per $n = 4$ si ha 6.33

Per $n = 5$ si ha 6.53

Per $n = 6$ si ha 6.59

Dunque con 6 posti in rianimazione il livello di rischio (cioè la probabilità di dover mandare presso un altro ospedale un paziente bisognoso della rianimazione perché tutti i posti sono occupati) è inferiore all'1%.

Nota di correzione: curare gli aspetti formali.