

Sommario lezioni di Probabilità

C. Franchetti

September 4, 2005

1 Lo spazio di probabilità.

1.1 Prime definizioni

I possibili risultati di un “esperimento” costituiscono lo “spazio dei campioni” o “spazio di probabilità”. Le parole chiave sono spiegate dagli esempi che seguono.

1 **Lancio di un dado** : $\{1,2,3,4,5,6\}$

2 **Lancio di una moneta** : $\{T,C\}$

3 **Due lanci successivi di una moneta** : $\{TT,TC,CT,CC\}$

In questi esempi lo spazio dei campioni ha cardinalità rispettivamente 6, 2, 4. Ma ecco un esempio con cardinalità infinita:

4 **Lanci successivi di una moneta finché venga testa**:

$$\{T, CT, CCT, CCCT, \dots, CCC\dots CT, \dots\}$$

Interessa sapere se il risultato dell’esperimento conferma (rientra in) o no un “evento” previsto. Questa nuova parola chiave sarà spiegata usando gli esempi di sopra. Sono eventi:

1 : $A = \text{“viene un numero pari”} = \{2,4,6\}$

2 : $A = \text{“non viene nè testa nè croce”} = \emptyset$

3 : $A = \text{“viene almeno una volta testa”} = \{TT,TC,CT\}$

4 : $A = \text{“viene testa in non più di 3 lanci”} = \{T,CT,CCT\}$

Come si vede gli “eventi” sono sottoinsiemi dell’insieme costituito dallo spazio dei campioni, i risultati degli esperimenti (atomi, eventi semplici) sono dunque gli elementi (punti) dello spazio dei campioni, che indicheremo con Ω . Gli eventi sono dunque una sottofamiglia $\mathcal{A}$ di $\mathcal{P}(\Omega)$ (parti di Ω).

Nell’esempio 1 se B è l’evento “viene un numero minore o uguale a 2” = $\{1,2\}$; è naturale considerare come evento “viene un numero pari oppure viene un numero minore o uguale a 2” = $\{1,2,4,6\}$ che poi è $A \cup B$ l’evento unione.

Nell’esempio 3 è naturale considerare come evento “non viene mai testa” = $\{CC\}$ che poi è $A^c = \Omega \setminus A$ l’evento complementare.

Questa ultima osservazione suggerisce di premettere la

Definizione 1.1.1. Una famiglia $\mathcal{A}$ di sottoinsiemi di un insieme Ω si dice che

è una σ -algebra di Ω se:

i) $\Omega \in \mathcal{A}$.

ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$.

iii) $A_n \in \mathcal{A}$ per $n = 1, 2, \dots \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

L'esempio 4 spiega perchè richiedere che l'unione contabile di eventi sia un evento. Si osservi che dagli assiomi segue che l'insieme vuoto è un evento così come l'intersezione contabile di eventi.

Facendo riferimento all'idea intuitiva di probabilità, nei primi tre esempi concordiamo nel dire che ogni evento semplice ha la stessa possibilità di realizzarsi: rispettivamente una su sei, una su due e una su quattro. Prendiamo ora un dado e segniamo quattro facce con T e le altre due con C, si ottiene un esempio in cui lo spazio dei campioni (come insieme) coincide con quello dell'esempio 2: $\{T, C\}$; tuttavia concorderemo nel dire che i due eventi semplici non hanno la stessa possibilità di realizzarsi, ne daremo a T quattro su sei e a C due su sei. Se A non è un evento semplice, concorderemo nell'assegnargli come possibilità di realizzarsi la somma delle possibilità degli atomi che lo compongono. Chiamiamo dunque questa possibilità $p(A)$, la probabilità che A si verifichi. Siamo indotti subito a dire che Ω = "evento certo" e $\emptyset$ = "evento impossibile" hanno probabilità rispettivamente 1 e 0. Si ha subito da qua che se Ω ha cardinalità infinita (come nell'esempio 4), i suoi atomi non possono avere tutti la stessa probabilità. (In 4 le probabilità sono rispettivamente $1/2, 1/4, \dots$ e, come deve essere, $\sum_{n=1}^{\infty} \frac{1}{2^n} = 1$. Si noti che abbiamo tacitamente ammesso questo fatto: "prima o poi deve necessariamente uscire testa" che risulta essere Ω , l'evento certo). La probabilità p dovrà essere quella che i matematici chiamano una misura (positiva) definita su una σ -algebra di un insieme Ω con la condizione che la misura di Ω sia 1, cioè

Definizione 1.1.2. Siano Ω un insieme e $\mathcal{A}$ una σ -algebra di Ω ; una *probabilità* p è una applicazione $p : \mathcal{A} \rightarrow [0, 1]$ tale che:

i) $p(\Omega) = 1$.

ii) Se per $n \in N, A_n \in \mathcal{A}$ e gli insiemi A_n sono a due a due disgiunti allora (σ -additività)

$$p\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} p(A_n)$$

Dall'esempio 4 comprendiamo la necessità di richiedere la completa additività della probabilità. Siamo ora pronti a dare la

Definizione 1.1.3. Uno *spazio di probabilità* è una terna $\{\Omega, \mathcal{A}, p\}$ dove Ω è un insieme, $\mathcal{A}$ una σ -algebra di Ω e p una probabilità su $\mathcal{A}$.

Teorema 1.1.1.

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

inoltre se $B \subset A$ si ha

$$p(B) \leq p(A); \quad p(A \setminus B) = p(A) - p(B)$$

Dimostrazione. Basta osservare che $B = (B \cap A) \cup (B \cap A^c)$, per cui $p(B) = p(A \cap B) + p(B \cap A^c)$; $A \cup B = A \cup (B \cap A^c)$, per cui $p(A \cup B) = p(A) + p(B \cap A^c)$. Nel caso che $B \subset A$ le altre formule si provano scrivendo $A = B \cup (A \setminus B)$.

Si osservi che questo risultato implica in particolare che per eventi qualsiasi A_n (non necessariamente disgiunti) si ha

$$p\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} p(A_n)$$

Per fare un esempio prendiamo il lancio di due dadi. Come spazio dei campioni si può prendere $\Omega = \{(i, j) : i, j = 1, \dots, 6\}$, dove nella coppia ordinata il primo (secondo) elemento è il numero uscito nel primo (secondo) dado. Assegniamo la probabilità uniforme, per cui se $x \in \Omega$, $p(x) = 1/36$. Calcoliamo la probabilità dell'evento $S = \text{"esce almeno una volta il 6"}$. Le coppie che non contengono il 6 sono 25, per cui $p(S) = 1 - 25/36 = 11/36$. Si poteva usare il Teorema di sopra: $S = A_6 \cup B_6$ dove $A_6(B_6)$ è l'insieme delle coppie con primo (secondo) elemento il 6. Si ha $p(A_6) = p(B_6) = 1/6$; inoltre si ha $A_6 \cap B_6 = \{(6, 6)\}$, per cui $p(A_6 \cup B_6) = p(A_6) + p(B_6) - p(A_6 \cap B_6) = 2/6 - 1/36 = 11/36$.

Teorema 1.1.2. Sia $\{\Omega, \mathcal{A}, p\}$ uno spazio di probabilità: se $\{C_n\}$ è una successione crescente di elementi di $\mathcal{A}$ allora

$$p(C_n) \rightarrow p(C), \quad C = \bigcup_{n=1}^{\infty} C_n$$

se $\{D_n\}$ è una successione decrescente di elementi di $\mathcal{A}$ allora

$$p(D_n) \rightarrow p(D), \quad D = \bigcap_{n=1}^{\infty} D_n$$

Dimostrazione. Poniamo $B_1 = C_1, B_n = C_n \setminus C_{n-1}$, allora i B_n sono in $\mathcal{A}$, sono a due a due disgiunti e inoltre $C_n = B_1 \cup \dots \cup B_n$, $C = \bigcup_{i=1}^{\infty} B_i$. Pertanto

$$p(C_n) = \sum_{i=1}^n p(B_i); \quad p(C) = \sum_{i=1}^{\infty} p(B_i)$$

e la prima formula segue dalla definizione di somma di una serie.

Poniamo ora $S_n = D_1 \setminus D_n$, allora gli S_n sono una successione crescente e $p(S_n) = p(D_1) - p(D_n)$; $D_1 \setminus D = \bigcup S_n$, applicando la formula precedente si ha

$$p(D_1) - p(D) = p(D_1 \setminus D) = \lim p(S_n) = p(D_1) - \lim p(D_n)$$

e questo prova la seconda formula.

Come applicazione di questa seconda formula dimostriamo l'utile

Teorema 1.1.3 (Lemma di Cantelli). Sia $\{E_k\}$ una successione di eventi tale che $\sum_{k=1}^{\infty} p(E_k) < \infty$; allora la probabilità che si verifichino simultaneamente infiniti eventi della successione è nulla.

Dimostrazione. Sia A l'evento "si verificano simultaneamente infiniti eventi E_k " e poniamo $B_n = \bigcup_{k=n}^{\infty} E_k$. Sia $x \in A$, se x non appartenesse a un B_n per un certo n , x apparterrebbe solo a un numero finito di E_k , pertanto x sta in ogni B_n e quindi $A \subset \bigcap_{n=1}^{\infty} B_n$. Sia ora $x \in \bigcap_{n=1}^{\infty} B_n$, Se x appartenesse solo a un numero finito di E_k , esisterebbe p tale che x non sta in E_i per $i > p$, ma allora x non starebbe in B_{p+1} : una contraddizione. Abbiamo provato che

$$A = \bigcap_{n=1}^{\infty} B_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k.$$

Siccome $\{B_n\}$ decresce si ha $p(A) = \lim_{n \rightarrow \infty} p(B_n)$; d'altra parte si ha $0 \leq p(B_n) \leq p(E_n) + p(E_{n+1}) + \dots$ e dal fatto che la serie $\sum p(E_k)$ converge segue subito che $p(B_n) \rightarrow 0$.

1.2 Esempi.

Diremo che su Ω c'è una distribuzione *uniforme* di probabilità se p è costante su gli atomi di Ω ; in questo caso si può prendere $\mathcal{A} = \mathcal{P}(\Omega)$ e per ogni $A \subset \Omega$ si ha $p(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$.

Uno schema molto comune in casi di distribuzione uniforme di probabilità è il seguente:

Collocamento a caso di r palle in n celle.

Qui r ed n sono numeri naturali positivi. Esistono n^r arrangiamenti possibili, a ciascuno dei quali attribuiremo probabilità $1/n^r$ (basta osservare che per la prima palla ci sono n scelte, altrettante per la seconda, la terza e così via). Ecco degli esempi concreti:

Totocalcio. Le " r palle" sono le 13 partite, le " n celle" sono i tre possibili risultati 1, x, 2 degli incontri. Se per ogni partita 1, x, 2 sono equiprobabili (ipotesi non realistica), allora la probabilità di "fare 13", cioè di indovinare tutti i risultati, vale $p = 3^{-13} = 6,27 \cdot 10^{-7}$.

r lanci di monete. Le " r palle" sono gli r lanci, le " n celle" sono i 2 risultati T e C. Se in ogni lancio T e C hanno probabilità $1/2$, allora le 2^r sequenze possibili di risultati sono equiprobabili, per es. la probabilità di una sequenza tutta di teste vale 2^{-r} .

Compleanni. Le " r palle" sono r persone scelte a caso, le " n celle" sono i 365 giorni dell'anno. Se si considera equiprobabile ogni data, allora ogni dis-

tribuzione di compleanni di r persone ha probabilità $p = 365^{-r}$. Consideriamo il seguente problema: quanto vale la probabilità $p_r(A)$ dell'evento $A =$ “nel gruppo almeno due persone hanno il compleanno nello stesso giorno” ? Calcoliamo la probabilità dell'evento complementare $A^c =$ “le persone del gruppo hanno il compleanno in giorni tutti distinti”. Per la prima persona si può scegliere in 365 modi, per la seconda in 364, per la terza in 363 e così via. Le scelte possibili sono $365(365 - 1) \dots (365 - r + 1)$ e quindi

$$p_r(A^c) = \frac{365(365 - 1) \dots (365 - r + 1)}{365^r}; \quad p_r(A) = 1 - p_r(A^c).$$

Per $r \geq 23$ $p_r(A)$ è maggiore di $1/2$.

Estrazione con rimpiazzamento. Le “ r palle” sono r palle estratte a caso da un'urna contenente n palle chiamate “1”, “2”, ..., “ n ” (dopo ogni estrazione la palla estratta viene reintrodotta nell'urna), le “ n celle” sono gli n nomi delle palle nell'urna. Con r estrazioni si possono ottenere n^r campioni distinti.

Consideriamo ora un problema un po' più complesso.

Distribuzione ipergeometrica. In una popolazione di n individui ce ne sono r rossi ($r \leq n$); si scelgono a caso s individui ($s \leq n$), determinare la probabilità q_k che questo gruppo contenga esattamente k individui rossi ($0 \leq k \leq \min(s, r)$). Dal gruppo dei rossi si possono scegliere k individui in $\binom{r}{k}$ modi, gli altri $(s - k)$ in $\binom{n-r}{s-k}$ modi; pertanto

$$q_k = \frac{\binom{r}{k} \binom{n-r}{s-k}}{\binom{n}{s}};$$

con semplice manipolazione dei coefficienti binomiali q_k si può anche scrivere

$$q_k = \frac{\binom{s}{k} \binom{n-s}{r-k}}{\binom{n}{r}}.$$

Diamo un semplice esempio

Gioco del lotto. Giochiamo tre numeri su una data ruota, qual'è la probabilità di vincere? Usando le notazioni di sopra si ha: $n = 90$; $r = 3$; $s = 5$; $k = 3$ per cui $q_3 = \frac{\binom{5}{3} \binom{85}{0}}{\binom{90}{3}} = \frac{1}{11748}$.

1.3 Probabilità condizionale, indipendenza.

Un gruppo Ω di n persone atte al lavoro contiene d disoccupati e f femmine. Si sceglie una persona a caso dal gruppo, sia D l'evento “disoccupato” e F l'evento “femmina”, allora $p(D) = d/n$; $p(F) = f/n$. Se consideriamo il sottogruppo delle femmine, scegliendone una a caso, la probabilità che sia disoccupata è c/f dove c è il numero di disoccupati femmine. Possiamo esprimerci così: c/f è la probabilità che si verifichi l'evento D assumendo che si sia verificato l'evento F .

Useremo la notazione $c/f = p(D|F)$ che leggeremo: “probabilità condizionale di D rispetto a F ”. Osserviamo subito che, poichè $c/f = \frac{c/n}{f/n}$, vale la formula

$$p(D|F) = \frac{p(D \cap F)}{p(F)}.$$

Si può considerare Ω come uno spazio con probabilità uniforme ($p(x) = 1/n$ su ogni atomo x) definita su tutti i sottoinsiemi. La probabilità condizionale $p(\cdot|F)$ è un'altra probabilità su Ω che assegna probabilità 0 agli eventi disgiunti da F ; oppure si può vederla come una probabilità sullo spazio F . Come si vede la probabilità condizionale non è un concetto nuovo, tuttavia risulta un modo utile di muoversi per risolvere un problema. Sia $(\Omega, \mathcal{A}, p)$ uno spazio di probabilità, diamo la formale

Definizione 1.3.1 Siano $A, B \in \mathcal{A}$, $p(A) > 0$. Si chiama *probabilità condizionale* di B rispetto ad A la quantità

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

Come esempio consideriamo le famiglie che hanno esattamente due bambini, scriviamo $m(f)$ per maschio (femmina) e la prima lettera rappresenta il primogenito. Su $\Omega = \{mm, mf, fm, ff\}$ consideriamo la probabilità uniforme ($p = 1/4$ su ogni atomo). Sapendo che una famiglia ha un maschio (evento A), qual'è la probabilità dell'evento $B =$ “figli entrambi maschi”? Poichè $A = \{mm, mf, fm\}$; $A \cap B = B = \{mm\}$, si ha che la probabilità richiesta $p(B|A)$ vale $\frac{1/4}{3/4} = 1/3$. (Si noti che un esame diretto dello spazio dei campioni porta immediatamente a questo risultato).

Consideriamo ora una partizione $\{A_i\}$ di uno spazio di probabilità Ω , per ogni altro evento $K \subset \Omega$ vale la

Formula di Bayes

$$p(A_i|K) = \frac{p(A_i)p(K|A_i)}{\sum_j p(A_j)p(K|A_j)}$$

Dimostrazione. Basta osservare che $p(A_i|K) = \frac{p(A_i \cap K)}{p(K)}$; $p(A_i \cap K) = p(A_i)p(K|A_i)$; $p(K) = \sum_j p(K \cap A_j) = \sum_j p(A_j)p(K|A_j)$.

Esempio. Supponiamo che in una famiglia la probabilità che ci siano k figli sia p_k , ($\sum_{k=0}^{\infty} p_k = 1$); supponiamo inoltre che la distribuzione dei sessi sia uniforme.

Sapendo che non ci sono femmine (evento K), qual'è la probabilità (condizionale) che nella famiglia ci siano esattamente i figli (evento A_i)? Abbiamo: $p(A_i) = p_i$; $p(K|A_i) = 2^{-i}$, per la formula di Bayes la probabilità richiesta è: $p(A_i|K) = \frac{p_i 2^{-i}}{p_0 + p_1 2^{-1} + p_2 2^{-2} + \dots}$.

“Filosofia” della formula di Bayes. Se chiamiamo **“cause”** gli eventi A_i che

compaiono nella formula di Bayes, allora si può interpretare la formula come una regola per calcolare a posteriori la probabilità che A_i sia stato la causa dell'essersi verificato l'evento K . Per questo motivo la formula di Bayes viene molto usata in statistica; tuttavia se si usa a questo scopo la formula occorre molta cautela nell'interpretarne i risultati.

La probabilità (condizionale) $p(A|K)$ è in generale diversa dalla probabilità (assoluta) $p(A)$, intuitivamente: l'informazione che l'evento K si è verificato cambia la nostra maniera di scommettere sul verificarsi dell'evento A . Quando $p(A|K) = p(A)$ è un po' come dire che l'evento K non influenza (è indipendente dal) l'evento A ; questo caso si ha se e solo se vale l'uguaglianza $p(A \cap K) = p(A)p(K)$ e si vede quindi che se A è indipendente da K allora K è indipendente da A . Diamo dunque la fondamentale

Definizione 1.3.2 Due eventi A, B si dicono (stocasticamente) *indipendenti* se

$$p(A \cap B) = p(A)p(B),$$

Una famiglia di eventi $\{A_i\}$ si dice *indipendente* se per ogni gruppo di k indici distinti si ha

$$p(A_{n_1} \cap A_{n_2} \cap \dots \cap A_{n_k}) = p(A_{n_1})p(A_{n_2}) \dots p(A_{n_k}).$$

Se una moneta viene lanciata ripetutamente, accettiamo di ritenere che l'esito di un lancio non influenza l'esito di un altro lancio (per es. il successivo), ciò si riflette nella (stocastica) indipendenza di alcuni eventi.

1.4 Esercizi e complementi

1. Generalizzare il Teorema 1.1.1 a tre e poi a n eventi.

Abbiamo successivamente:

$$\begin{aligned} p(A_1 \cup A_2 \cup A_3) &= p(A_1) + p(A_2 \cup A_3) - p(A_1 \cap (A_2 \cup A_3)) = \\ &= p(A_1) + p(A_2) + p(A_3) - p(A_2 \cap A_3) - p[(A_1 \cap A_2) \cup (A_1 \cap A_3)] = \\ p(A_1) + p(A_2) + p(A_3) - p(A_2 \cap A_3) - p(A_1 \cap A_2) - p(A_1 \cap A_3) + p[(A_1 \cap A_2) \cap (A_1 \cap A_3)] &= \\ p(A_1) + p(A_2) + p(A_3) - p(A_2 \cap A_3) - p(A_1 \cap A_2) - p(A_1 \cap A_3) + p(A_1 \cap A_2 \cap A_3) . \end{aligned}$$

Per generalizzare a n eventi A_i poniamo:

$$p_i = p(A_i); \quad p_{ij} = p(A_i \cap A_j); \quad p_{ijk} = p(A_i \cap A_j \cap A_k); \dots$$

$$S_1 = \sum p_i; \quad S_2 = \sum p_{ij}; \quad S_3 = \sum p_{ijk}; \dots$$

Si può dimostrare che la probabilità P_1 che si verifichi uno almeno degli eventi A_i vale

$$P_1 = p(A_1 \cup \dots \cup A_n) = S_1 - S_2 + S_3 - S_4 + \dots + (-1)^{n+1} S_n .$$

2. Problema dei matches .

Un mazzo di N carte numerate $1, 2, \dots, N$ viene mescolato e steso in fila su un tavolo. Agli $N!$ ordinamenti possibili delle carte viene assegnata la probabilità $\frac{1}{N!}$. Si chiede il valore della probabilità P_1 che almeno una carta sia al posto giusto (carta denotata p al posto p -mo).

Soluzione.

Se c'è un match all' i -mo posto, le rimanenti carte possono essere disposte in $(N-1)!$ modi per cui per la corrispondente probabilità si ha $p_i = \frac{(N-1)!}{N!} = \frac{1}{N}$, similmente per la probabilità di match simultaneo al posto i e al posto (distinto) j si ha $p_{ij} = \frac{(N-2)!}{N!}$. Con le notazioni di sopra si osserva che la somma S_r contiene $\binom{N}{r}$ addendi tutti uguali a $\frac{(N-r)!}{N!}$; pertanto $S_r = \frac{1}{r!}$ e quindi

$$P_1 = 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{N+1} \frac{1}{N!} .$$

3. Esercizio.

Un'urna contiene n palline distinguibili, si estrae con rimpiazzo fino a che il campione ottenuto contenga due elementi uguali; calcolare la probabilità q_k che a questo scopo siano necessarie esattamente k estrazioni.

Soluzione.

Si osservi che $2 \leq k \leq n+1$. Siccome si rimpiazza, con k estrazioni si ottengono n^k campioni distinti; per ottenere un campione come desiderato, la prima pallina si può scegliere in n modi, la seconda in $(n-1)$ modi, ..., la $(k-1)$ -ma in $(n-k+2)$ modi e la k -ma in $(k-1)$ modi. Pertanto:

$$q_k = \frac{n(n-1)(n-2)\dots(n-k+2)(k-1)}{n^k} .$$

4. Formula di Stirling.

La formula di Stirling asserisce l'esistenza di un $\theta = \theta_n \in (0, 1)$ tale che:

$$n! = n^n e^{-n} \sqrt{2\pi n} e^{\frac{\theta}{12n}} .$$

Nella pratica è sufficiente l'approssimazione

$$n! \simeq n^n e^{-n} \sqrt{2\pi n} .$$

verifichiamo in parte la formula con i calcoli che seguono:

$$\int_{k-1}^k \log x \, dx < \log k < \int_k^{k+1} \log x \, dx ;$$

da cui sommando k da 1 a n si ottiene:

$$\int_0^n \log x \, dx < \log n! < \int_1^{n+1} \log x \, dx ;$$

$$n \log n - n < \log n! < (n+1) \log(n+1) - n ;$$

$$n^n e^{-n} < n! < (n+1)^{(n+1)} e^{-n} .$$

5. Esercizio.

Un'urna contiene 42 palline segnate con le lettere dell'alfabeto italiano: due per ogni lettera. Si estraggono a caso 21 palline. Calcolare:

- i) la probabilità che nel campione compaia una data lettera.
- ii) la probabilità che nel campione compaiano tutte le lettere.
- iii) la probabilità che si possa formare la parola "CASSA".

Soluzione.

i). Ci sono $\binom{42}{21}$ modi di scegliere un campione di 21 lettere, di questi $\binom{40}{21}$ non contengono una data lettera, pertanto la probabilità richiesta vale

$$1 - \frac{\binom{40}{21}}{\binom{42}{21}} = \frac{11}{21} .$$

ii). Per ogni lettera ci sono due scelte per cui ci sono 2^{21} modi di scegliere un campione che contenga le 21 lettere distinte, pertanto la probabilità richiesta vale

$$\frac{2^{21}}{\binom{42}{21}} \simeq 2^{-21} \sqrt{21\pi} = 3,87 \cdot 10^{-6} ,$$

dove per l'approssimazione abbiamo usato la formula di Stirling.

iii). Distinguiamo per comodità con C_1 e C_2 le due lettere "C" che sono nell'urna, ci sono $\binom{37}{16}$ campioni che contengono le lettere $C_1 A S S A$, pertanto la probabilità richiesta vale

$$2 \frac{\binom{37}{16}}{\binom{42}{21}} = 0,0478..$$

6. Esercizio.

In una metropoli scegliendo a caso un ristorante si hanno le seguenti probabilità:

$p(A_b) = 0,2$ dove A_b = "si mangia bene"

$p(A_n) = 0,4$ dove A_n = "si mangia normale"

$p(A_c) = 0,4$ dove A_c = "si mangia male"

Si hanno poi gli eventi

K = "caro", E = "equo", B = "buon mercato",

e le probabilità condizionali

$$p(K|A_b) = 0,7 \quad p(K|A_n) = 0,3 \quad p(K|A_c) = 0,2 .$$

Calcolare $p(A_b|K)$. Se poi sono date le probabilità condizionali

$$p(B|A_b) = 0,1 \quad p(B|A_n) = 0,3 \quad p(B|A_c) = 0,4$$

descrivere completamente lo spazio dei campioni.

Soluzione.

Per la formula di Bayes si ha

$$p(A_b|K) = \frac{p(A_b)p(K|A_b)}{p(A_b)p(K|A_b) + p(A_n)p(K|A_n) + p(A_c)p(K|A_c)} ,$$

dove il denominatore della frazione vale $p(K)$. Si ha

$$p(K) = 0,2 \times 0,7 + 0,4 \times 0,3 + 0,4 \times 0,2 = 0,34 ,$$

$$p(A_b|K) = \frac{p(A_b \cap K)}{p(K)} = \frac{7}{17} = 0,41..$$

Coi dati ulteriori possiamo calcolare $p(B)$:

$$p(B) = 0,2 \times 0,1 + 0,4 \times 0,3 + 0,4 \times 0,4 = 0,30$$

di conseguenza si ha pure che $p(E) = 0,36$. Consideriamo il quadro

	K	E	B	Tot
A_b		x		0,2
A_n		y		0,4
A_c		z		0,4
Tot	0,34	0,36	0,30	1

che possiamo completare coi seguenti calcoli:

$$p(K \cap A_b) = p(K|A_b)p(A_b) = 0,7 \times 0,2 = 0,14 ,$$

$$p(K \cap A_n) = p(K|A_n)p(A_n) = 0,3 \times 0,4 = 0,12 ,$$

$$p(K \cap A_c) = p(K|A_c)p(A_c) = 0,2 \times 0,4 = 0,08 ;$$

$$p(B \cap A_b) = p(B|A_b)p(A_b) = 0,1 \times 0,2 = 0,02 ,$$

$$p(B \cap A_n) = p(B|A_n)p(A_n) = 0,3 \times 0,4 = 0,12 ,$$

$$p(B \cap A_c) = p(B|A_c)p(A_c) = 0,4 \times 0,4 = 0,16 .$$

Possiamo riscrivere il quadro

	K	E	B	Tot
A_b	0,14	x	0,02	0,2
A_n	0,12	y	0,12	0,4
A_c	0,08	z	0,16	0,4
Tot	0,34	0,36	0,30	1

Otteniamo dunque i dati mancanti:

$$x = 0,04 \quad y = 0,16 \quad z = 0,16 .$$

2 Variabili aleatorie.

2.1 Variabili aleatorie discrete

Definizione 2.1.1. Una funzione $X : \Omega \rightarrow \mathbf{R}$, dove $(\Omega, \mathcal{A}, p)$ è uno spazio di probabilità, si dice che è una *variabile aleatoria* (var.al.) se per ogni $t \in \mathbf{R}$,

$$\{\omega : X(\omega) \leq t\} \in \mathcal{A}.$$

Questa denominazione è un po' fuorviante: occorrerebbe dire funzione aleatoria e non var.al.; tuttavia è ormai divenuta standard.

Tipici esempi di var.al.: il numero di successi in n prove di Bernoulli (vedi dopo), il guadagno (o perdita) di uno scommettitore in un gioco d'azzardo.

Considereremo qui solo var.al. *discrete*: supporremo cioè che la var.al. X prenda al più un insieme numerabile di valori distinti $\{x_i\}_{i \in I}$ dove l'insieme I degli indici è finito o numerabile.

Definizione 2.1.2. Data la var.al. X , la sua *densità* (discreta) è la funzione $f : \mathbf{R} \rightarrow \mathbf{R}^+$

$$f(x) = p\{\omega : X(\omega) = x\}$$

Si osservi che

- i) $f(x) = 0$ se $x \neq x_i$,
- ii) $\sum_{x \in R} f(x) = \sum_{i \in I} f(x_i) = 1$.

Ogni funzione $g : \mathbf{R} \rightarrow \mathbf{R}^+$ che soddisfi i) e ii) viene chiamata densità.

Scriveremo poi di solito $f(x) = p\{X = x\}$ anzichè più correttamente $f(x) = p\{\omega : X(\omega) = x\}$.

Consideriamo ora un esempio importante: il così detto

Schema di Bernoulli : Una prova (ad es. lancio di moneta) ha due soli possibili risultati: successo (T) con probabilità p e insuccesso (C) con probabilità $1-p$. Consideriamo il caso che si ripetano n prove di questo tipo. Le 2^n sequenze distinte che si possono formare con i simboli T e C sono uno spazio di campioni Ω . Sia $1 \leq i \leq n$; siano S_i, N_i gli eventi "T all'i-ma prova", "C all'i-ma prova" rispettivamente. Qualunque sia i si ha $p(S_i) = p$, $p(N_i) = 1-p$; inoltre se gli indici sono tutti distinti, ogni gruppo di eventi del tipo S_j, N_s costituisce una famiglia di eventi indipendenti.

Teorema 2.1.1. Se $A_k \in \Omega$ è l'evento " k successi negli n lanci", allora

$$p(A_k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Dimostrazione. Una sequenza con k successi (e quindi $n-k$ insuccessi) si ottiene come intersezione di n eventi: k di tipo S_i e $n-k$ di tipo N_i , a causa

dell'indipendenza questa sequenza ha probabilità $p^k(1-p)^{n-k}$. Si conclude notando che di tali sequenze ce ne sono $\binom{n}{k}$.

Indichiamo con S_n il numero di successi in n prove di Bernoulli. S_n è una var.al. definita su Ω : se $\omega \in \Omega$, $S_n(\omega)$ è il numero di volte che compare T nella sequenza ω ; i valori possibili di S_n sono $0, 1, \dots, n$. Per la densità f di S_n si ha $f(k) = p\{S_n = k\} = \binom{n}{k}p^k(1-p)^{n-k}$ per $k = 0, 1, \dots, n$ e 0 altrimenti; cioè f è concentrata nell'insieme $\{0, 1, \dots, n\}$ che a volte per brevità si prende come dominio di f omettendo di definire la densità dove è 0.

Vedremo che è possibile fare calcoli significativi con le probabilità anche senza considerare esplicitamente lo spazio dei campioni.

Si noti infine che

$$S_n = X_1 + \dots + X_n,$$

dove X_i è la var.al che vale 1 se all' i -ma prova si ha successo e 0 se si ha insuccesso (la sua densità vale risp. $p, (1-p)$ se $x = 1, 0$ e 0 in ogni altro caso).

Diremo che S_n segue una legge binomiale di parametri n e p che si indica con $B(n, p)$. Chiaramente ogni X_i ha legge $B(1, p)$.

Definizione 2.1.4. Data una var.al. X si chiama *funzione di ripartizione* (f.r.) la funzione $F_X : \mathbf{R} \rightarrow [0, 1]$ definita da

$$F_X(t) = p\{X \leq t\}.$$

Una f.r. è sempre non decrescente. Se f è la densità di X , si ha

$$F_X(t) = \sum_{x \leq t} f(x);$$

se X prende valori interi si ha pure

$$F_X(k) - F_X(k-1) = f(k).$$

2.2 Caso vettoriale

Spesso è necessario considerare simultaneamente più variabili aleatorie.

Definizione 2.2.1. Si chiama *vettore aleatorio* o var.al. m -dimensionale un vettore $X = (X_1, X_2, \dots, X_m)$, le cui componenti X_i siano var.al. nel senso della defin. 2.1.1.

La densità f di X si definisce allo stesso modo, cioè

$$f(x_1, x_2, \dots, x_m) = p\{X_1 = x_1, X_2 = x_2, \dots, X_m = x_m\},$$

f viene anche detta densità *congiunta* delle X_i le cui densità, dette *marginali*, possono essere indicate con f_{X_i} . Se si conosce la densità congiunta si possono calcolare le densità marginali; per semplicità trattiamo il caso $m = 2$. Sia

(X, Y) una var.al. 2-dimensionale con densità $f = f(x, y)$. Indichiamo con $\{x_i\}$, $\{y_i\}$, f_X , f_Y i valori possibili e le densità di X e Y rispettivamente. Si ha

$$f_X(z) = p\{X = z\} = p\left(\bigcup_i \{X = z, Y = y_i\}\right) = \sum_i p\{X = z, Y = y_i\} = \sum_i f(z, y_i).$$

Analogamente si ha

$$f_Y(z) = \sum_i f(x_i, z).$$

Non basta invece conoscere le densità marginali per calcolare la densità congiunta, perchè densità congiunte diverse possono avere densità marginali uguali. Del resto è chiaro che anche avendo tutti i dati relativi a una var.al. X e a un'altra var.al. Y (per esempio la conoscenza delle rispettive densità f_X e f_Y) nulla si può sapere sulle probabilità che si realizzino simultaneamente un evento concernente la var.al. X e un altro concernente la Y ; questo a meno che non si sappia a priori che X e Y sono var.al. indipendenti (vedi dopo).

2.3 Variabili aleatorie indipendenti

Definizione 2.3.1. Si dice che n var.al. $X_1, \dots, X_n$ sono *indipendenti* se e solo se per ogni scelta di sottoinsiemi $J_1, \dots, J_n$ di $\mathbf{R}$ si ha

$$p\{X_1 \in J_1, \dots, X_n \in J_n\} = p\{X_1 \in J_1\} \cdot p\{X_n \in J_n\}$$

La definizione si estende a infinite var.al. nel solito modo.

Se f_i sono le densità (marginali) delle var.al. indipendenti X_i e f la loro densità congiunta, allora si ha che $f(x_1, \dots, x_n) = f_1(x_1) \cdot \dots \cdot f_n(x_n)$. Viceversa se vale quest'ultima identità si riconosce che le var.al. X_i sono indipendenti.

Siano X e Y var.al. indipendenti e siano $\phi, \psi : \mathbf{R} \rightarrow \mathbf{R}$ due funzioni. E' intuitivo che anche le var.al. $\phi(X), \psi(Y)$ sono indipendenti. Più in generale si ha il

Teorema 2.3.1. Siano $X_1, \dots, X_m; Y_1, \dots, Y_n$ var.al. indipendenti e siano $\phi : \mathbf{R}^m \rightarrow \mathbf{R}$, $\psi : \mathbf{R}^n \rightarrow \mathbf{R}$ delle funzioni. Allora le var.al. $\phi(X_1, \dots, X_m)$, $\psi(Y_1, \dots, Y_n)$ sono indipendenti.

2.4 Densità di una somma

Consideriamo il problema di calcolare la densità della somma di due var.al. X e Y . Si noti che l'evento $\{X + Y = z\}$ uguaglia l'unione (di eventi disgiunti) $\bigcup_i \{X = x_i, Y = z - x_i\}$; risulta quindi chiaro che in generale per calcolare la densità di una somma non servono le densità marginali, ma occorre conoscere la densità congiunta del vettore (X, Y) .

Teorema 2.4.1. Siano U e V var.al. di densità congiunta f . Allora la densità g di $U + V$ vale

$$g(z) = \sum_{t \in \mathbf{R}} f(t, z - t);$$

se poi U e V sono indipendenti con densità f_1, f_2 , si ha

$$g(z) = \sum_{t \in \mathbf{R}} f_1(t) f_2(z - t).$$

Dimostrazione. Premettiamo in generale che per ogni var.al. m -dimensionale X di densità f e per ogni $A \subset \mathbf{R}^m$ si ha:

$$p\{X \in A\} = \sum_{x \in A} f(x);$$

e se $\phi : \mathbf{R}^m \rightarrow \mathbf{R}$

$$p\{\phi(X) = z\} = p\{X \in \phi^{-1}(z)\} = \sum_{x \in \phi^{-1}(z)} f(x).$$

Si usi ora l'ultima formula con $X = (U, V)$ e $\phi(u, v) = u + v$, basta osservare che $\phi^{-1}(z)$ è l'insieme dei vettori (u, v) con $u + v = z$.

Teorema 2.4.2. Siano $V_1, \dots, V_s$ var.al. indipendenti con leggi binomiali risp. $B(n_1, p), \dots, B(n_s, p)$. Allora $(V_1 + \dots + V_s)$ ha legge $B(n_1 + \dots + n_s, p)$.

Dimostrazione. Basta rappresentare $(V_1 + \dots + V_s)$ come somma di $(n_1 + \dots + n_s)$ var.al. indipendenti di legge $B(1, p)$.

Esempio 1. Si lanciano 2 dadi a n facce, si chiede la probabilità $g(k)$ che la somma dei risultati sia esattamente k ($k = 2, 3, \dots, 2n$). Si ha

$$g(k) = \sum_{i=1}^n f(i, k-i);$$

dove $f(i, k-i) = \frac{1}{n^2}$ se $1 \leq (k-i) \leq n$, $f(i, k-i) = 0$ altrimenti. Pertanto nella sommatoria ci sono $(k-1)$ termini non nulli se $k \leq n+1$ e $n - (k-n) + 1 = 2n - k + 1$ termini non nulli se $k > n+1$; si vede quindi che $g(k) = g(2n - k + 2)$ e che per $k = 1, 2, \dots, n+1$ $g(k) = \frac{k-1}{n^2}$.

2.5 Speranza matematica

Sia X una var.al.(discreta) che prende valori x_i ed ha densità f .

Definizione 2.5.1. Si dice che X ha *speranza matematica* finita se

$$\sum_i |x_i| f(x_i) < +\infty;$$

in tal caso si chiama *speranza matematica* (o *media* o *valore atteso*) di X il numero

$$E[X] = \sum_i x_i f(x_i) = \sum_i x_i p\{X(\omega) = x_i\};$$

se $E[X] = 0$ si dice che la var.al. X è centrata.

Supponiamo che le var.al. X e Y aventi media finita prendano i valori $\{x_i\}$ e $\{y_j\}$ e sia $g(x, y) = p\{X = x; Y = y\}$ la loro densità congiunta. Mostriamo che anche la var.al. $(X + Y)$ ha media finita. Si ha

$$\begin{aligned}\sum_{ij} |x_i + y_j| g(x_i, y_j) &\leq \sum_{ij} |x_i| g(x_i, y_j) + \sum_{ij} |y_j| g(x_i, y_j) = \\ &= \sum_i |x_i| \sum_j g(x_i, y_j) + \sum_j |y_j| \sum_i g(x_i, y_j) = \\ &= \sum_i |x_i| f_X(x_i) + \sum_j |y_j| f_Y(y_j)\end{aligned}$$

dove f_X, f_Y sono le densità marginali di X, Y . Con calcoli simili senza il valore assoluto si prova poi la formula fondamentale

$$E[X + Y] = E[X] + E[Y].$$

Diamo ora un'importante generalizzazione di questo risultato. Sia $X = (X_1, \dots, X_m)$ un vettore aleatorio di densità f e i cui valori denoteremo con $x^{(1)}, x^{(2)}, \dots$; $h: \mathbf{R}^m \rightarrow \mathbf{R}$, ponendo $Z = h(X)$ si definisce una nuova var.al. i cui valori denoteremo con $z_1, z_2, \dots$:

Teorema 2.5.1. Z ha speranza matematica finita se e solo se

$$\sum_i |h(x^{(i)})| f(x^{(i)}) < \infty$$

e in questo caso

$$E[Z] = \sum_i h(x^{(i)}) f(x^{(i)})$$

Dimostrazione. Se $A_j = h^{-1}(z_j)$ allora $p\{Z = z_j\} = \sum_{x^{(i)} \in A_j} f(x^{(i)})$. Abbiamo

allora:

$$\sum_j |z_j| p\{Z = z_j\} = \sum_j |z_j| \sum_{x^{(i)} \in A_j} f(x^{(i)}) = \sum_j \sum_{x^{(i)} \in A_j} |z_j| f(x^{(i)}) =$$

tenendo conto che se $x^{(i)} \in A_j$ allora $h(x^{(i)}) = z_j$

$$= \sum_j \sum_{x^{(i)} \in A_j} |h(x^{(i)})| f(x^{(i)}) = \sum_i |h(x^{(i)})| f(x^{(i)}).$$

Dunque se la serie $\sum_i h(x^{(i)}) f(x^{(i)})$ converge assolutamente Z ha speranza matematica finita. La formula per $E[Z]$ si dimostra ripetendo gli stessi calcoli senza i valori assoluti; si osservi soltanto che il passaggio

$$\sum_j \sum_{x^{(i)} \in A_j} h(x^{(i)}) f(x^{(i)}) = \sum_i h(x^{(i)}) f(x^{(i)}),$$

che consiste nel sommare riordinando i termini, è lecito appunto perchè sappiamo che la serie è assolutamente convergente.

Dal teorema precedente, specializzando la funzione h , si deducono come corollari importanti risultati. Ad esempio

Corollario 2.5.1. Siano X e Y var.al. con media finita. Allora:

i) $(aX + bY)$ ha media finita e $E[aX + bY] = aE[X] + bE[Y]$;

ii) $|X|$ ha media finita e $|E[X]| \leq E[|X|]$;

se inoltre X e Y sono indipendenti

iii) XY ha media finita e $E[XY] = E[X]E[Y]$

Dimostrazione. Limitiamoci a provare iii). Siano f_X, f_Y, f risp. le densità di X, Y e del vettore (X, Y) ; per la supposta indipendenza si ha $f(x_i, y_j) = f_X(x_i)f_Y(y_j)$. Applicando il teorema con $h(x, y) = xy$ si ottiene

$$\begin{aligned} \sum_{i,j} |x_i| |y_j| f(x_i, y_j) &= \sum_{i,j} |x_i| |y_j| f_X(x_i) f_Y(y_j) = \\ &= \sum_i |x_i| f_X(x_i) \sum_j |y_j| f_Y(y_j) = E[|X|] E[|Y|] < \infty; \end{aligned}$$

ripetendo gli stessi calcoli senza i valori assoluti si prova che $E[XY] = E[X]E[Y]$.

2.6 Momenti, varianza

Definizione 2.6.1. Sia X una var.al.; diremo che X ha *momento* di ordine k finito ($k = 1, 2, \dots$) se X^k ha media finita. In questo caso $E[X^k]$ si chiama momento di ordine k della var.al. X . Se $(X - E[X])^k$ ha media finita diremo che X ha *momento centrato* di ordine k finito e chiameremo $E[(X - E[X])^k]$ momento centrato di ordine k di X .

Applicando il Teorema 2.5.1, indicando con f, μ risp. la densità e la media di X otteniamo:

$$E[X^k] = \sum_i x_i^k f(x_i); \quad E[(X - \mu)^k] = \sum_i (x_i - \mu)^k f(x_i).$$

Teorema 2.6.1. i) Se X ha momento finito di ordine k , allora ha anche momento finito di ordine r per ogni $r \leq k$.

ii) Se X e Y hanno momento finito di ordine k , allora anche $(X + Y)$ ha momento finito di ordine k . In particolare se X ha momento finito di ordine k , allora ha anche momento centrato finito di ordine k .

Dimostrazione . Siano $\{x_i\}$ i valori assunti da X , per ogni i si ha $|x_i|^r \leq 1 + |x_i|^k$ e di conseguenza

$$|x_i|^r f(x_i) \leq f(x_i) + |x_i|^k f(x_i),$$

pertanto la serie che definisce il momento di ordine k converge assolutamente per il criterio del confronto. Questo prova i). Per quanto riguarda la ii) basta

osservare che, per la convessità della funzione t^k , si ha la disuguaglianza:

$$(x + y)^k \leq 2^{k-1}(x^k + y^k).$$

Di particolare importanza è il momento centrato di ordine 2 che viene chiamato *varianza*:

$$\text{Var}(X) = E[(X - E[X])^2] = E[(X - \mu)^2].$$

Per quanto elementare il teorema che segue si rivela estremamente utile in molte circostanze:

Teorema 2.6.2 (disuguaglianza di Chebyshev). Se per la var.al. X esiste il momento $E[X^2]$ (e di conseguenza la varianza) per ogni $c > 0$ si ha:

$$p\{|X| > c\} \leq \frac{E[X^2]}{c^2};$$

in particolare, posto $\mu = E[X]$,

$$p\{|X - \mu| > c\} \leq \frac{\text{Var}(X)}{c^2}.$$

Dimostrazione. Applicando la prima disuguaglianza alla var.al. $(X - \mu)$, si deduce la seconda. Si noti ora che

$$p\{|X| > c\} = \sum_{|x_i| > c} f(x_i),$$

si ha poi

$$\sum_{|x_i| > c} f(x_i) = \sum_{|x_i| > c} f(x_i) \frac{x_i^2}{x_i^2} \leq \sum_i \frac{f(x_i)x_i^2}{c^2} = \frac{E[X^2]}{c^2}.$$

Fissata la var.al. X , di solito media e varianza si denotano risp. con μ e σ^2 ; $\sigma = \sqrt{\text{Var}(X)}$ si chiama *deviazione standard* di X . Se X e Y hanno varianza finita, è finita anche la quantità $E[(X - \mu_X)(Y - \mu_Y)]$ che si chiama *covarianza* di X e Y e si indica con $\text{Cov}(X, Y)$.

Si provano facilmente le seguenti proprietà:

- i) $\text{Var}(X) = E[X^2] - E[X]^2$;
- ii) $\text{Var}(aX) = a^2\text{Var}(X)$, $\text{Var}(a + X) = \text{Var}(X)$;
- iii) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$;
- iv) $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$.

Se X e Y sono indipendenti, applicando il Corollario 2.5.1 iii) si ha

$$\text{Cov}(X, Y) = 0; \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Per quanto visto, la covarianza misura in un certo senso l'indipendenza delle var.al.; tuttavia può capitare che $\text{Cov}(X, Y) = 0$ senza che X e Y siano indipendenti. Quando $\text{Cov}(X, Y) = 0$ si dice che le var.al. X e Y non sono *correlate*.

Si definisce pertanto il *coefficiente di correlazione* $\rho_{X,Y}$ mediante

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Vogliamo provare che $-1 \leq \rho_{X,Y} \leq 1$. Per questo notiamo che si ha sempre $E[XY]^2 \leq E[X^2]E[Y^2]$ (si vede col segno di un trinomio); basta poi applicare la disuguaglianza alle var.al. $(X - \mu_X), (Y - \mu_Y)$.

2.7 Esempi

1. Schema di Bernoulli. Siano $X_i, i = 1, \dots, n$ le var.al. di legge $B(1, p)$ relative alle prove successive in uno schema di Bernoulli. Le X_i sono indipendenti. Per ogni i si ha $E[X_i] = 1p + 0(1-p) = p$ e quindi $E[S_n] = np$ dove $S_n = X_1 + \dots + X_n$.

Si noti che abbiamo calcolato la media di S_n senza bisogno di conoscere la sua densità f_n . Questa come sappiamo (Teorema 2.1.1) è

$$f_n(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{se } k = 0, 1, \dots, n \\ 0 & \text{altrimenti} \end{cases}$$

Si ha poi $\text{Var}(X_i) = E[X_i^2] - E[X_i]^2 = p - p^2 = p(1-p)$ e di conseguenza

$$\text{Var}(S_n) = np(1-p).$$

La disuguaglianza di Chebyshev ci dice

$$p\{|S_n - np| > h\} \leq \frac{np(1-p)}{h^2};$$

ovvero, considerando la frequenza $\frac{S_n}{n}$ dei successi nelle n prove,

$$p\left\{\left|\frac{S_n}{n} - p\right| > \frac{h}{n}\right\} \leq \frac{np(1-p)}{h^2}.$$

Se scegliamo $h = n\epsilon$ in questa disuguaglianza si ottiene

$$p\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \leq \frac{p(1-p)}{n\epsilon^2};$$

in particolare se il numero n delle prove tende all'infinito, tende a zero la probabilità di una deviazione della frequenza $\frac{S_n}{n}$ dei successi dalla probabilità p di successo in ogni singola prova. Questo è un caso particolare della legge (debole) dei grandi numeri. Questo giustifica il metodo di osservare la frequenza di un risultato per stabilirne la probabilità.

2. Densità di Poisson. Si chiama distribuzione di Poisson di parametro λ , ($\lambda > 0$), la densità

$$f(k) = \begin{cases} e^{-\lambda} \frac{\lambda^k}{k!} & \text{se } k = 0, 1, 2, \dots \\ 0 & \text{altrimenti} \end{cases}$$

La densità di Poisson si può vedere come un'approssimazione di una densità binomiale $B(n, p)$ con n grande e p piccolo. Siano dunque S_n var.al. con densità binomiale $B(n, \lambda/n)$, si ha:

$$p\{S_n = k\} = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k},$$

se si fa tendere n all' ∞ si vede, con facili calcoli, che

$$p\{S_n = k\} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda};$$

inoltre la convergenza è uniforme rispetto a k .

La distribuzione di Poisson è detta anche legge degli eventi rari, infatti si presenta come legge di una var.al. S che rappresenta il numero di successi su un numero molto grande di prove ripetute indipendenti, in ciascuna delle quali la probabilità di successo è molto piccola.

Siano X e Y due var.al. indipendenti di Poisson di parametri λ, μ e di densità f_1, f_2 resp.; vogliamo trovare la densità g della somma $X + Y$. Si ha, ricordando che $f_2(z) = 0$ se z è negativo,

$$\begin{aligned} g(k) &= \sum_{i=0}^{\infty} f_1(i) f_2(k-i) = \sum_{i=0}^k e^{-\lambda} \frac{\lambda^i}{i!} e^{-\mu} \frac{\mu^{k-i}}{(k-i)!} = \\ &= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{i=0}^k \binom{k}{i} \lambda^i \mu^{k-i} = e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^k}{k!}. \end{aligned}$$

Dunque $X + Y$ è di Poisson con parametro $(\lambda + \mu)$.

Sia X una var.al. con densità di Poisson di parametro λ , si ha

$$E[X] = \lambda; \text{ Var}(X) = \lambda$$

3. Distribuzione ipergeometrica. Un'urna contiene n palline di cui r rosse ($r \leq n$); se ne estraggono (senza rimpiazzo) s ($s \leq n$); la var.al. X che dà il numero di palline rosse estratte ha la densità ipergeometrica

$$f(k) = \begin{cases} \frac{\binom{r}{k} \binom{n-r}{s-k}}{\binom{n}{s}} & \text{se } k = 0, 1, 2, \dots, \min(r, s) \\ 0 & \text{altrimenti} \end{cases}$$

Anche in questo caso per calcolare $E[X]$ non abbiamo bisogno di conoscere la densità. Possiamo ragionare così: sia X_i la var.al. che vale 1 se nell' i -esima estrazione viene una pallina rossa e 0 altrimenti. Sono state estratte s palline, essendoci simmetria (non ci sono palline privilegiate nel gruppo estratto) ogni pallina ha la stessa probabilità di essere rossa per esempio quella della prima estratta. Al primo colpo la probabilità di successo di X_1 vale r/n , per cui per

ogni i si ha $E[X_i] = (r/n)1 + (1-r/n)0$. Pertanto $E[X] = \sum E[X_i] = s(r/n)$. Si osservi che essendo s il numero di estrazioni ed avendo ogni estrazione la stessa probabilità $p = r/n$ di successo, la var.al X si comporta per quanto riguarda il calcolo della media come se fosse binomiale e questo nonostante che le var.al X_i non siano indipendenti.

4. Densità geometrica. Consideriamo sempre prove ripetute con probabilità p di successo. Problema: determinare la probabilità che ci vogliano esattamente $k + 1$ ($k = 0, 1, 2, \dots$) lanci per ottenere il (primo) successo. Si ha successo esattamente al $(k+1)$ -mo lancio se e solo se si è verificata la sequenza $\underbrace{\text{CCC}\dots\text{CT}}_k$;

questa ha probabilità $(1-p)^k p$. Si chiama densità geometrica di parametro p la densità

$$f(k) = \begin{cases} p(1-p)^k & \text{se } k = 0, 1, 2, \dots \\ 0 & \text{altrimenti} \end{cases}$$

Sia T la var.al. “numero di lanci necessari per il primo successo”, allora f è la densità di $X = (T - 1)$. Si ha:

$$E[X] = \sum_{k=0}^{\infty} k p (1-p)^k = \frac{1-p}{p},$$

$$\text{Var}(X) = E[X^2] - E[X]^2 = \frac{1-p}{p^2}.$$

Mostriamo ora la “mancanza di memoria” della legge geometrica. Cominciamo ad osservare che

$$p\{X \geq k\} = \sum_{j=k}^{\infty} p(1-p)^j = (1-p)^k.$$

Calcoliamo adesso una probabilità condizionale:

$$\begin{aligned} p(\{X = k+m\} | \{X \geq k\}) &= \frac{p(\{X = k+m\} \cap \{X \geq k\})}{p\{X \geq k\}} = \frac{p\{X = k+m\}}{p\{X \geq k\}} = \\ &= \frac{p(1-p)^{k+m}}{(1-p)^k} = p(1-p)^m = p\{X = m\}. \end{aligned}$$

Applichiamo questi risultati alla var.al. T : supponiamo di non aver avuto alcun successo nelle prime k prove: qual’è la probabilità di dover attendere ancora m prove? Risposta: la stessa che si avrebbe ricominciando da capo le prove. Questo spiega l’espressione “mancanza di memoria”.

Se diamo credito a questo modello probabilistico, allora la pratica diffusa di giocare al lotto i numeri “ritardatari” non è giustificata.

5. Distribuzione binomiale negativa. Si generalizza l’esempio precedente nel seguente modo: consideriamo una sequenza di prove di Bernoulli con probabilità di successo p e di insuccesso $q = (1-p)$, ci si chiede quanto si deve

aspettare per l' r -mo successo. Per il numero ν di prove necessarie sarà $\nu \geq r$ scriviamo $\nu = r + k$. Indicheremo con $f(k; r, p)$ la probabilità che l' r -mo successo avvenga alla prova $r + k$, ($k = 0, 1, 2, \dots$); ovvero la probabilità che esattamente k insuccessi precedano l' r -mo successo. Ciò avviene se e solo se tra le prime $(r+k-1)$ prove ci sono k insuccessi e la seguente è un successo; pertanto avremo $f(k; r, p) = \binom{r+k-1}{k} p^r q^k = \binom{-r}{k} p^r (-q)^k$. (Si ricordi che $(-1)^k \binom{r+k-1}{k} = \binom{-r}{k}$). Chiameremo distribuzione binomiale negativa di parametri r e p la densità

$$f_r(k) = \begin{cases} \binom{-r}{k} p^r (-q)^k & \text{se } k = 0, 1, 2, \dots \\ 0 & \text{altrimenti} \end{cases}$$

Ricordando lo sviluppo in serie binomiale di $(1 - q)^{-r}$ si verifica, come deve essere, che $\sum_{k=0}^{\infty} f_r(k) = 1$. Si osservi che per $r = 1$ la densità binomiale negativa si riduce alla densità geometrica.

Non è difficile calcolare la media di una binomiale negativa usando la densità in base alla definizione; ma è più istruttivo procedere nel modo seguente: sia X_i la var.al. che indica il numero di insuccessi che ci sono tra l' $(i-1)$ -mo e l' i -mo successo. Qualunque sia i , X_i ha densità geometrica $\{pq^k\}$ e media $E[X_i] = q/p$. La var.al. $Y_r = X_1 + \dots + X_r$ rappresenta il numero di insuccessi che precedo l' r -mo successo e quindi ha densità f_r . Siccome $E[Y_r] = \sum E[X_i]$, si ottiene $E[Y_r] = rq/p$.

2.8 Funzioni generatrici

Sia X una var.al. che prende valori interi positivi, per ogni $z \in \mathbf{R}$ poniamo

$$\psi_X(z) = E[z^X].$$

Definizione 2.8.1. La funzione ψ_X si chiama *funzione generatrice delle probabilità*.

Se f è la densità di X , per il Teorema 2.5.1, si ha

$$\psi_X(z) = \sum_{n=0}^{\infty} z^n f(n).$$

Questa serie di potenze ha raggio di convergenza $R \geq 1$ (perchè $\sum f(n) = 1$), R può anche essere $+\infty$, come nel caso che X prenda solo un insieme finito di valori e ψ_X si riduce a un polinomio.

Si vede facilmente che due var.al. a valori interi positivi hanno la stessa densità se e solo se hanno la stessa funzione generatrice. Se la funzione generatrice di X è data in forma esplicita (non come serie), la densità di X si calcola con derivazioni successive nell'origine. Diamo subito alcuni esempi

Binomiale. Sia X una var.al. $B(n, p)$, si ha

$$\psi_X(z) = \sum_{k=0}^n z^k \binom{n}{k} p^k (1-p)^{n-k} = (1-p+pz)^n$$

Poisson . Sia X una var.al. che segue la legge di Poisson di parametro λ , si ha

$$\psi_X(z) = e^{-\lambda} \sum_{k=0}^{\infty} z^k \frac{\lambda^k}{k!} = e^{\lambda(z-1)}$$

Risulta molto utile il

Teorema 2.8.1. Siano X e Y var.al. indipendenti. Allora

$$\psi_{X+Y} = \psi_X \psi_Y.$$

Dimostrazione. Per il Teorema 2.3.1 anche z^X, z^Y sono var.al. indipendenti, pertanto si ha

$$\psi_{X+Y}(z) = E[z^{X+Y}] = E[z^X]E[z^Y] = \psi_X(z)\psi_Y(z).$$

Questo teorema può essere usato per calcolare la densità della somma di due var.al. come si vede nel seguente esempio.

Somma di geometriche. Siano X e Y due var.al. indipendenti di densità geometrica $\{pq^k\}(q = 1 - p)$. La loro funzione generatrice vale

$$\psi(z) = \sum_{k=0}^{\infty} pq^k z^k = \frac{p}{1 - qz}.$$

Se $Q = X + Y$ si ha $\psi_Q = \psi_X \psi_Y$ ossia $\psi_Q(z) = \frac{p^2}{(1 - qz)^2}$, con facili calcoli si ha poi

$$\frac{p^2}{(1 - qz)^2} = p^2(1 + 2qz + 3q^2z^2 + \dots)$$

e quindi per la densità di Q si ottiene

$$f_Q(k) = p^2(k + 1)q^k.$$

Mostriamo anche il

Teorema 2.8.2. Se la serie che definisce la funzione generatrice ψ_X ha raggio di convergenza $R > 1$, allora la var.al. X ha momenti di ogni ordine; in particolare

$$E[X] = \psi'_X(1); \quad \text{Var}(X) = \psi''_X(1) + \psi'_X(1) - \psi'_X(1)^2.$$

Dimostrazione. Derivando ψ_X e valutando in 1 si ottiene

$$\psi'_X(1) = \sum_{k=1}^{\infty} kf(k) = E[X];$$

derivando una seconda volta e valutando in 1 si ottiene

$$\psi''_X(1) = \sum_{k=1}^{\infty} (k^2 - k)f(k) = E[X^2] - E[X]^2$$

da cui segue subito la formula per la varianza. Allo stesso modo (con derivate successive) si possono determinare i momenti di ordine maggiore.

2.9 Esercizi e complementi

1. Il problema del collezionista.

Una popolazione contiene n elementi distinti, tutti in ugual proporzione, (ad es. figurine per una collezione). Si fanno estrazioni con rimpiazzo. Un campione di dimensione d in generale conterrà meno di d elementi distinti a causa delle ripetizioni (i doppi). Sia V_t la var.al. “numero di estrazioni necessarie per avere t elementi distinti”; i valori possibili per V_t sono $t, t+1, \dots$. Si vuole calcolare la media di V_t .

Soluzione.

Diciamo che un'estrazione ha successo se risulta nell'aggiunta di un nuovo elemento nel campione. Poniamo poi $X_k = V_{k+1} - V_k$; allora $(X_k - 1)$ è il numero di estrazioni con insuccesso tra il k -mo successo e il $(k+1)$ -mo successo. Durante tutte queste estrazioni la popolazione contiene $(n-k)$ elementi che non sono entrati nel campione, per cui $(X_k - 1)$ è il numero di insuccessi che precedono il primo successo in prove ripetute di Bernoulli con $p = \frac{n-k}{n}$. X_k prende il valore $s+1$ con probabilità $(1-p)^s p$ per cui, facendo i calcoli, si ottiene $E[X_k] = \frac{1}{p} = \frac{n}{n-k}$. Si noti ora che $V_t = 1 + X_1 + \dots + X_{t-1}$, per cui

$$E[V_t] = n \left(\frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-t+1} \right) \simeq n \int_{n-t+1/2}^{n+1/2} \frac{1}{x} dx = n \log \frac{n+1/2}{n-t+1/2}.$$

In particolare $E[V_n] \simeq n \log n$ e (assumendo n pari) $E[V_{n/2}] \simeq n \log 2$. Si osservi come il numero atteso di prove per completare la collezione sia in generale assai più grande del doppio del numero atteso di prove per completare metà collezione.

2. Varianza di somme.

Supponiamo che n var.al. X_i abbiano varianza finita, allora per la varianza della loro somma si ha la formula generale

$$(*) \quad \text{Var}(X_1 + X_2 + \dots + X_n) = \sum_i \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j).$$

3. Problema dei matches

Facciamo riferimento a 2 di 1.4. Sia $S = S_N$ la var.al. “numero di matches”, si chiedono la media e la varianza di S .

Soluzione.

Sia X_i la var.al. che vale 1 (con probab. $1/N$) se all' i -mo posto c'è match e 0 altrimenti. Abbiamo

$$E[X_i] = 1/N, \quad S = \sum_{i=1}^N X_i, \quad E[S] = 1, \quad \text{Var}(X_i) = (N-1)/N^2.$$

Calcoliamo le covarianze. Per $i \neq j$ la var.al. $X_i X_j$ prende solo valori 0 e 1, quest'ultimo con prob. $\frac{1}{N(N-1)}$. Pertanto si ha

$$\text{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j] = \frac{1}{N(N-1)} - \frac{1}{N^2}.$$

Applicando la formula (*) si ottiene

$$\text{Var}(S) = N \frac{N-1}{N^2} + N(N-1) \frac{1}{N^2(N-1)} = 1.$$

4. Polinomi di Bernstein

Diamo ora una interessante applicazione dei metodi probabilistici alla teoria dell'approssimazione.

Sia f una funzione reale continua su $[0, 1]$, per $n \in \mathbf{N}$, $x \in [0, 1]$; l' n -mo polinomio di Bernstein della f è definito da

$$B_n^f(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}.$$

Faremo vedere che B_n^f converge uniformemente a f ; dando così una dimostrazione del teorema di Weierstrass: "I polinomi sono densi nello spazio delle funzioni continue". Si ha

$$\begin{aligned} |B_n^f(x) - f(x)| &= \left| \sum_{k=0}^n \left[f\left(\frac{k}{n}\right) - f(x) \right] \binom{n}{k} x^k (1-x)^{n-k} \right| \leq \\ &\leq \sum_{k=0}^n \left| f\left(\frac{k}{n}\right) - f(x) \right| \binom{n}{k} x^k (1-x)^{n-k} \end{aligned}$$

Poichè f è uniformemente continua, fissato $\epsilon > 0$ esiste un $\delta > 0$ tale che: $|u - v| \leq \delta \Rightarrow |f(u) - f(v)| < \epsilon/2$; sia poi $\|f\| = M$. Si ha

$$|B_n^f(x) - f(x)| \leq \sum_{|\frac{k}{n} - x| \leq \delta} \dots + \sum_{|\frac{k}{n} - x| > \delta} \dots < \epsilon/2 + 2M \sum_{|k - nx| > n\delta} \binom{n}{k} x^k (1-x)^{n-k}.$$

Se si interpreta x come la probabilità parametro di una var.al. S_n binomiale (avente media $\mu = nx$), allora l'ultima somma eguaglia

$$p\{|S_n - \mu| > n\delta\}.$$

Questa quantità, usando la disuguaglianza di Chebyshev, si maggiora con

$$\frac{\text{Var}(S_n)}{n^2\delta^2} = \frac{nx(1-x)}{n^2\delta^2}.$$

Otteniamo dunque

$$|B_n^f(x) - f(x)| < \epsilon/2 + \frac{M}{2n\delta^2},$$

e quindi $\|B_n^f - f\| < \epsilon$ per n sufficientemente grande.

3 La legge dei grandi numeri.

3.1 La legge debole

Teorema 3.1.1 (Legge debole dei grandi numeri). Sia $\{X_k\}$ una successione di var.al. indipendenti aventi la stessa densità. Se esiste la media $\mu = E[X_k]$, allora qualunque sia $\eta > 0$ si ha

$$\lim_{n \rightarrow \infty} p\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| > \eta\right\} = 0.$$

(Per la dimostrazione di questo teorema vedi appendice del capitolo). Conviene qui premettere qualche osservazione. Poniamo $\frac{X_1 + \dots + X_n}{n} = M_n$: il teorema afferma che $M_n \rightarrow \mu$ secondo un certo tipo di convergenza per var.al. (che poi sono funzioni). Nel linguaggio della teoria delle funzioni reali questa convergenza viene chiamata *convergenza in misura* o anche *convergenza in probabilità*. In realtà le ipotesi di questo teorema implicano una convergenza più forte, la *convergenza quasi ovunque* (per funzioni): vale cioè, come vedremo, una “legge forte dei grandi numeri”.

Se aggiungiamo un’ipotesi speciale, cioè che esistano le varianze s_k^2 delle X_k , il teorema ammette una dimostrazione molto semplice e per altri versi più generale (non occorre supporre che le X_k abbiano la stessa densità) che qui riportiamo.

Teorema (Legge debole dei grandi numeri in ipotesi di varianza). Sia $\{X_k\}$ una successione di var.al. indipendenti di media μ_k e di varianza σ_k^2 . Poniamo

$$S_n = X_1 + \dots + X_n ;$$

$$m_n = \sum_{k=1}^n \mu_k , \quad s_n^2 = \sum_{k=1}^n \sigma_k^2$$

allora, se $\frac{s_n^2}{n} \rightarrow 0$, per ogni $\eta > 0$ si ha

$$p\left\{\left|\frac{S_n}{n} - \frac{m_n}{n}\right| > \eta\right\} \rightarrow 0$$

Dimostrazione. Applicando la disuguaglianza di Chebyshev si ha

$$p\{|S_n - m_n| > n\eta\} \leq \frac{\text{Var}(S_n)}{n^2\eta^2} = \frac{s_n^2}{n^2\eta^2}$$

e l’ultimo termine è infinitesimo.

3.2 La legge forte

Cominciamo coll’ enunciare una disuguaglianza che è più fine di quella di Chebyshev. (Dimostrazione in appendice).

Teorema 3.2.1 (Disuguaglianza di Kolmogorov). Siano $X_1, X_2, \dots, X_n$ var.al.

indipendenti aventi media μ_k e varianza σ_k^2 e sia $S_k = X_1 + \dots + X_k$, $m_k = \mu_1 + \dots + \mu_k$. Dato $t > 0$ consideriamo le n disuguaglianze

$$(*) \quad |S_k - m_k| \geq t\sqrt{\text{Var}(S_n)} \quad (k = 1, 2, \dots, n).$$

Sia C_n l'evento "si verifica almeno una delle disuguaglianze (*)"; allora $p(C_n) \leq \frac{1}{t^2}$.

Si osservi che per $n = 1$ la disuguaglianza di Kolmogorov si riduce alla disuguaglianza di Chebyshev. In effetti basta prendere $t = \frac{c}{\sqrt{\text{Var}(S_n)}}$

Se le var.al. hanno la stessa media μ e la stessa varianza σ^2 , allora con la scelta di $t = \frac{c}{\sqrt{n}\sigma}$ in luogo delle n disuguaglianze (*) si hanno le n disuguaglianze

$$(**) \quad \left| \frac{S_k}{k} - \mu \right| \geq \frac{c}{k} \quad (k = 1, 2, \dots, n).$$

Per illustrare il significato (e i limiti) della legge debole e per introdurre la legge forte consideriamo per semplicità lo schema di Bernoulli $B(n, p)$. Indichiamo con S_n il numero di successi nelle n prove. La nozione intuitiva di probabilità vorrebbe che per la frequenza di successi $\frac{S_n}{n}$ si avesse $\frac{S_n}{n} \rightarrow p$: in effetti questo evento si verifica con probabilità 1, ma la questione andrebbe trattata con la probabilità nel continuo. Comunque questo risultato non è contenuto nella legge debole (ma in quella forte). La legge debole afferma che

$$\lim_{n \rightarrow \infty} p\left\{ \left| \frac{S_n}{n} - p \right| > \eta \right\} = 0.$$

Questa formula ci dice che, per n fissato sufficientemente grande, la frequenza di successi $\frac{S_n}{n}$ ha una buona probabilità di essere vicina a p ; ma non ci dice che tale frequenza sia vincolata a rimanere vicino a p . Per esempio viene lasciata aperta la possibilità che per infiniti j si verifichi l'evento $A_j = \left\{ \left| \frac{S_j}{j} - p \right| \geq \eta \right\}$. In effetti non è così: faremo ora vedere che con probabilità 1, $\left| \frac{S_n}{n} - p \right|$ diventa e resta piccolo.

Schema di Bernoulli: (legge forte dei grandi numeri).

Per ogni $\eta > 0$ è nulla la probabilità che si verifichino simultaneamente infiniti eventi

$$A_i = \left\{ \left| \frac{S_i}{i} - p \right| \geq \eta \right\}$$

Dimostrazione. Per il lemma di Cantelli (Teorema 1.1.3) basta dimostrare che la serie $\sum p(A_i)$ converge. Indichiamo con B_ν l'evento: "per almeno un n , con $2^{\nu-1} < n \leq 2^\nu$, vale la disuguaglianza $\left| \frac{S_n}{n} - p \right| \geq \eta$ ". Basterà provare che la serie $\sum p(B_\nu)$ converge. Il verificarsi di B_ν implica che $|S_n - np| \geq \eta 2^{\nu-1}$. Appliciamo ora la disuguaglianza di Kolmogorov: (abbiamo qui $\mu = p; \sigma^2 = p(1-p); t = \frac{\eta 2^{\nu-1}}{\sqrt{n}\sigma}$) si avrà

$$p(B_\nu) \leq t^{-2} = \frac{4n\sigma^2}{\eta^2 2^{2\nu}} \leq \frac{4\sigma^2}{\eta^2 2^\nu}$$

e questo prova che la serie converge.
Enunciamo dunque il

Teorema 3.2.2 (Legge forte dei grandi numeri). Sia $\{X_n\}$ una successione di var.al. indipendenti aventi la stessa densità. Se esiste la media $\mu = E[X_k]$, allora vale la legge forte dei grandi numeri, cioè per ogni $\eta > 0$ è nulla la probabilità che si verifichino simultaneamente infinite disuguaglianze del tipo

$$|\frac{S_i}{i} - \mu| \geq \eta.$$

Si osservi che non si richiede che le var.al. X_n abbiano varianza finita. Quando esiste la varianza delle X_n vale un risultato molto più forte (il Teorema Limite Centrale).

Definizione. Si chiama *distribuzione normale* la funzione Φ definita da

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy.$$

Ricordiamo che come si sa dall'analisi $\Phi(+\infty) = 1$ e naturalmente $\Phi(-\infty) = 0$; inoltre $1 - \Phi(x) = \Phi(-x)$. Possiamo ora enunciare il

Teorema 3.2.3 (Teorema Limite Centrale). Sia $\{X_n\}$ una successione di var.al. indipendenti aventi la stessa densità. Supponiamo che esistano la media $\mu = E[X_k]$, e la varianza $\sigma^2 = \text{Var}(X_k)$. Allora, posto $S_n = X_1 + \dots + X_n$, qualunque sia β si ha

$$(1) \quad \lim_{n \rightarrow \infty} p\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} < \beta\right\} = \Phi(\beta).$$

Tratteremo la dimostrazione di questo teorema nel capitolo sulla probabilità nel continuo.

Osserviamo intanto che per $c > 0$ si ha

$$p\left\{\left|\frac{S_n}{n} - \mu\right| > \frac{c}{\sqrt{n}}\right\} = p\left\{\frac{|S_n - n\mu|}{\sigma\sqrt{n}} > \frac{c}{\sigma}\right\} =$$

$$1 - p\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{c}{\sigma}\right\} + p\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} < -\frac{c}{\sigma}\right\}.$$

Pertanto, applicando la (1), si ottiene

$$\lim_{n \rightarrow \infty} p\left\{\left|\frac{S_n}{n} - \mu\right| > \frac{c}{\sqrt{n}}\right\} = 2\Phi\left(-\frac{c}{\sigma}\right)$$

Mostriamo ora che, se $\eta > 0$, per ogni funzione g tale che $\frac{\sqrt{n}}{g(n)} \rightarrow \infty$ per $n \rightarrow \infty$, si ha

$$(2) \quad \lim_{n \rightarrow \infty} p\{g(n) \left| \frac{S_n}{n} - \mu \right| > \eta\} = 0$$

che è un risultato ben più forte della legge debole. Si ha infatti

$$p\{g(n) \left| \frac{S_n}{n} - \mu \right| > \eta\} = p\left\{ \left| \frac{S_n}{n} - \mu \right| > \frac{\eta}{g(n)} \right\}$$

quest'ultima probabilità è definitivamente minore o uguale di $p\left\{ \left| \frac{S_n}{n} - \mu \right| > \frac{c}{\sqrt{n}} \right\}$. Si ha quindi

$$0 \leq \liminf p\{g(n) \left| \frac{S_n}{n} - \mu \right| > \eta\} \leq \limsup p\{g(n) \left| \frac{S_n}{n} - \mu \right| > \eta\} \leq$$

$$\lim_{n \rightarrow \infty} p\left\{ \left| \frac{S_n}{n} - \mu \right| > \frac{c}{\sqrt{n}} \right\} = 2\Phi\left(-\frac{c}{\sigma}\right)$$

da cui, facendo tendere c a $+\infty$ e ricordando che $\Phi(-\infty) = 0$, segue la (2).

3.3 Esercizi e complementi

1. Dimostrazione della legge debole.

Sia f la densità delle X_k . Non è restrittivo supporre $\mu = 0$. Fissato $\delta > 0$, per ogni n definiamo n coppie di var.al.

$$U_k^{(n)} = X_k, \quad V_k^{(n)} = 0 \quad \text{se } |X_k| \leq n\delta$$

$$U_k^{(n)} = 0, \quad V_k^{(n)} = X_k \quad \text{se } |X_k| > n\delta.$$

Si ha evidentemente che $X_k = U_k^{(n)} + V_k^{(n)}$.

Per provare la (*) basterà mostrare che, dato $\epsilon > 0$, δ può essere scelto in modo che per $n \rightarrow \infty$

$$(**) \quad p\{|U_1^{(n)} + \dots + U_n^{(n)}| > \frac{1}{2}\epsilon n\} \rightarrow 0, \quad p\{|V_1^{(n)} + \dots + V_n^{(n)}| > \frac{1}{2}\epsilon n\} \rightarrow 0.$$

Sia $a = E[|X_k|]$, allora (omettendo l'indice n in alto)

$$E[U_1^2] = \sum_{|x_k| \leq n\delta} x_k^2 f(x_k) \leq \delta n \sum_j |x_j| f(x_j) = a\delta n.$$

Le U_k sono var.al. indipendenti e con la stessa distribuzione, si ha quindi

$$\text{Var}(U_1 + \dots + U_n) = n\text{Var}(U_1) \leq nE[U_1^2] \leq a\delta n^2.$$

Si osservi che per $n \rightarrow \infty$, $E[U_1] \rightarrow E[X_1] = 0$ per cui per n abbastanza grande:

$$E[(U_1 + \dots + U_n)^2] = \text{Var}(U_1 + \dots + U_n) + E[U_1 + \dots + U_n]^2 \leq 2a\delta n^2.$$

Se ora applichiamo la disuguaglianza di Chebyshev otteniamo:

$$p\{|U_1^{(n)} + \dots + U_n^{(n)}| > \frac{1}{2}\epsilon n\} \leq \frac{8a\delta}{\epsilon^2};$$

siccome $\delta > 0$ è arbitrario, questa implica la prima delle (**). Per quanto riguarda la seconda delle (**) osserviamo che si ha

$$\{(V_1 + \dots + V_n) \neq 0\} \subset \bigcup_k \{V_k \neq 0\},$$

$$p\{(V_1 + \dots + V_n) \neq 0\} \leq \sum_k p\{V_k \neq 0\} = np\{V_1 \neq 0\};$$

$$p\{V_1 \neq 0\} = p\{|X_1| > \delta n\} = \sum_{|x_j| > \delta n} f(x_j) \leq \frac{1}{\delta n} \sum_{|x_j| > \delta n} |x_j| f(x_j);$$

l'ultima somma tende a zero se n tende all'infinito e quindi la dimostrazione è completa.

2. Dimostrazione della disuguaglianza di Kolmogorov.

Poniamo $x = p(C_n)$. Per $j = 1, \dots, n$ definiamo var.al. Y_j nel seguente modo: Y_j vale 1 nei punti in cui la j -ma è la prima delle disuguaglianze (*) ad essere soddisfatta, vale 0 altrimenti. Dunque Y_j vale 1 se

$$|S_j - m_j| \geq t\sqrt{\text{Var}(S_n)}, \quad \text{ma} \quad |S_r - m_r| < t\sqrt{\text{Var}(S_n)} \quad \text{per } r = 1, \dots, j-1;$$

Y_j vale 0 in ogni altro caso. Pertanto in ogni punto al più una delle Y_k vale 1 e quindi la somma $Y_1 + \dots + Y_n$ prende solo i valori 0 e 1; vale 1 se e solo se una almeno delle disuguaglianze (*) è soddisfatta; questo significa che $x = p\{(Y_1 + \dots + Y_n) = 1\}$. Dalla disuguaglianza $\sum Y_k \leq 1$ moltiplicando per $(S_n - m_n)^2$ e prendendo la media E si ottiene

$$\sum_{k=1}^n E[Y_k(S_n - m_n)^2] \leq E[(S_n - m_n)^2] = \text{Var}(S_n).$$

Stimeremo ora il primo membro di questa disuguaglianza. Ponendo

$$U_k = (S_n - m_n) - (S_k - m_k) = \sum_{j=k+1}^n (X_j - \mu_j)$$

si ottiene

$$E[Y_k(S_n - m_n)^2] = E[Y_k(S_k - m_k)^2] + E[Y_k U_k^2] + 2E[Y_k U_k(S_k - m_k)].$$

Notiamo che U_k dipende solo da $X_{k+1}, \dots, X_n$ mentre Y_k e S_k solo da $X_1, \dots, X_k$; pertanto U_k è indipendente da $Y_k(S_k - m_k)$. Tenendo conto che $E[U_k] = 0$, otteniamo

$$E[Y_k U_k(S_k - m_k)] = E[Y_k(S_k - m_k)]E[U_k] = 0.$$

Questo implica che $E[Y_k(S_n - m_n)^2] \geq E[Y_k(S_k - m_k)^2]$. D'altra parte $Y_k \neq 0$ solo se $|S_k - m_k| \geq t\sqrt{\text{Var}(S_n)}$ per cui $Y_k(S_k - m_k)^2 \geq t^2\text{Var}(S_n)Y_k$. Possiamo quindi concludere che

$$\text{Var}(S_n) \geq t^2\text{Var}(S_n)E[Y_1 + \dots + Y_n] = t^2\text{Var}(S_n)x$$

ossia $x \leq t^{-2}$.

4 La probabilità nel continuo

4.1 Prime definizioni

Definizione 4.1.1. Una $F : \mathbf{R} \rightarrow \mathbf{R}$ è detta *funzione di ripartizione* (f.r.) se gode delle seguenti proprietà:

- 1) F è non decrescente e continua a destra;
- 2) $\lim_{t \rightarrow -\infty} F(t) = 0$; $\lim_{t \rightarrow +\infty} F(t) = 1$.

Ricordiamo che per una var.al. X abbiamo già definito la sua funzione di ripartizione ponendo: $F_X(t) = p\{X \leq t\}$. Le definizioni sono consistenti: infatti F_X verifica le proprietà 1) e 2). Proviamo la 1). E' ovvio che F_X sia non decrescente. Sia $\{t_n\}$ una successione che tende a x decrescendo, dobbiamo mostrare che $F_X(t_n) \rightarrow F_X(x)$. Sia $A = \bigcap_{k=1}^{\infty} \{X \leq t_n\}$, avendosi una successione decrescente di eventi, per il Teorema 1.1.2 $F_X(t_n) \rightarrow p(A)$. Resta da provare che $A = \{X \leq x\}$. E' ovvio che $\{X \leq x\} \subset A$; se $y \in A$, allora $X(y) \leq t_n$ per ogni n e pertanto anche $X(y) \leq x$, cioè $A \subset \{X \leq x\}$. In modo simile si prova la 2).

Ogni f.r. F , come si vede subito, soddisfa pure:

- 3) $0 \leq F(t) \leq 1$;
- 4) $p\{a < X \leq b\} = p\{X \leq b\} - p\{X \leq a\} = F(b) - F(a)$.

Come ogni funzione monotona la F ammette in ogni punto x anche il limite sinistro che si denoterà con $F(x^-)$. Avremo dunque sempre

$$F(x^-) \leq F(x) = F(x^+);$$

al solito modo si verifica poi che

$$F(x) - F(x^-) = p\{X = x\};$$

in particolare $p\{X = x\} = 0$ se F è continua in x .

E' importante sottolineare il fatto che le proprietà 1) e 2) caratterizzano le f.r.

Definizione 4.1.2. Una $f : \mathbf{R} \rightarrow \mathbf{R}$ è una *densità* se è misurabile (secondo Lebesgue) e

$$f \geq 0; \quad \int_{-\infty}^{+\infty} f(x) dx = 1.$$

Convieni osservare subito che se $\mathcal{L}$ è la σ -algebra dei sottoinsiemi di $\mathbf{R}$ misurabili secondo Lebesgue, avendo a disposizione una densità f , si ottiene uno spazio di probabilità $(\mathbf{R}, \mathcal{L}, p)$ ponendo, per $A \in \mathcal{L}$, $p(A) = \int_A f$.

Sempre essendo f una densità, se poniamo

$$G(t) = \int_{-\infty}^t f(x) dx,$$

G risulta una f.r. con la proprietà ulteriore di essere *assolutamente continua* (vedi appendice), inoltre $G'(t) = f(t)$ quasi ovunque (q.o.) cioè tranne al più

un insieme di misura (di Lebesgue) nulla.

L'assoluta continuità è una proprietà più forte della continuità. In generale se H è una funzione assolutamente continua (a.c.) H è (derivabile q.o. ed è) primitiva della sua derivata, cioè

$$H(t) = H(a) + \int_a^t H'(x) dx.$$

Se F è una f.r. e f una densità, diremo che F ammette f come densità se $F(t) = \int_{-\infty}^t f(x) dx$. Se una f.r. G è a.c. allora ammette come densità la sua derivata G' . Esistono tuttavia f.r. F continue che non ammettono densità. (Questo nonostante che ogni f.r., essendo monotona, sia derivabile q.o.).

Definizione 4.1.3. Diremo che una var.al. X è a.c. se la sua f.r. F_X ammette densità (ed è quindi a.c.).

Sia X a.c. con f.r. F e densità f , allora si ha $p\{a \leq X \leq b\} = F(b) - F(a)$. Più in generale se A è misurabile in $\mathbf{R}$ si ha

$$p\{X \in A\} = \int_A f(x) dx.$$

4.2 Esempi

1. **Densità esponenziale.** Per ogni $\lambda > 0$ si definisce una f.r. ponendo

$$F(t) = 1 - e^{-\lambda t} \text{ se } t > 0, \quad F(t) = 0 \text{ altrimenti.}$$

F è a.c. ; derivando si ottiene la sua densità

$$f(t) = \lambda e^{-\lambda t} \text{ se } t > 0, \quad f(t) = 0 \text{ altrimenti.}$$

2. **Densità uniforme.** Definiamo una f.r. ponendo

$$F(t) = 0 \text{ se } t < 0, \quad F(t) = t \text{ se } 0 \leq t \leq 1; \quad F(t) = 1 \text{ se } t > 1.$$

F è a.c. ; derivando si ottiene la sua densità

$$f(t) = 0 \text{ se } t < 0, \quad f(t) = 1 \text{ se } 0 \leq t \leq 1; \quad f(t) = 0 \text{ se } t > 1.$$

Se X è una var.al. con questa densità uniforme e se $A \subset [0, 1]$ è misurabile, allora $p\{X \in A\} = \int_A 1 dx$, la misura di Lebesgue di A .

3. **Densità normale.** La f.r. (distribuzione) normale Φ , che abbiamo introdotto nel capitolo 3, è definita da:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy$$

e la sua densità è

$$f(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

La f.r. del primo esempio non è derivabile nell'origine, quella del secondo non lo è nell'origine e nel punto 1.

4.3 Densità congiunte

Considereremo, per semplicità solo vettori aleatori bidimensionali. Sia $Z = (X, Y)$ un tale vettore; la f.r. F_Z congiunta di Z è definita da:

$$F_Z(x, y) = p\{X \leq x, Y \leq y\} = p\{Z \in A_{x,y}\},$$

dove $A_{x,y} = \{(u, v) : u \leq x, v \leq y\}$.

Diremo che Z ha densità congiunta $f_Z = f$ se $f \geq 0$, $\int_{R^2} f(u, v) du dv = 1$, (cioè f è una densità) e inoltre

$$F_Z(x, y) = \int_{-\infty}^x du \int_{-\infty}^y f(u, v) dv = \int_{A_{x,y}} f(u, v) du dv .$$

Per sottoinsiemi “ragionevoli” A di R^2 avremo in generale la formula

$$p\{Z \in A\} = \int_A f(u, v) du dv .$$

Conoscendo f.r. e densità congiunte di un vettore al. $Z = (X, Y)$, calcoliamone f.r. e densità marginali. Si osservi che

$$\{X \leq x\} = \bigcup_{n=1}^{\infty} \{X \leq x, Y \leq n\} ,$$

per cui

$$F_X(x) = \lim_{y \rightarrow \infty} F_Z(x, y)$$

e, allo stesso modo

$$F_Y(y) = \lim_{x \rightarrow \infty} F_Z(x, y) .$$

Per calcolare le densità marginali osserviamo che, detta A la striscia $A = \{(u, v) : a \leq u \leq b, v \in R\}$, si ha

$$p\{a \leq X \leq b\} = \int_A f(u, v) du dv = \int_a^b du \int_{-\infty}^{+\infty} f(u, v) dv .$$

Posto

$$f_X(u) = \int_{-\infty}^{+\infty} f(u, v) dv ,$$

otteniamo $p\{a \leq X \leq b\} = \int_a^b f_X(u) du$. Questo mostra che $f_X(u)$ è una densità per X . Analogamente si vede che è una densità per Y :

$$f_Y(v) = \int_{-\infty}^{+\infty} f(u, v) du .$$

4.4 Indipendenza

Due var.al. X e Y si dicono *indipendenti* (la definizione si estende a un numero qualsiasi) se per ogni scelta dei numeri a, b, c, d si ha:

$$p\{a \leq X \leq b, c \leq Y \leq d\} = p\{a \leq X \leq b\}p\{c \leq Y \leq d\}.$$

Se esistono le densità marginali f_X, f_Y e la densità congiunta f , allora la condizione di indipendenza si traduce nella formula

$$\int_a^b du \int_c^d f(u, v) dv = \int_a^b f_X(u) du \int_c^d f_Y(v) dv;$$

questa uguaglianza è certo verificata se per ogni u, v si ha

$$f(u, v) = f_X(u)f_Y(v).$$

Si può provare che quest'ultima condizione è anche necessaria per l'indipendenza.

Esempio 1. *Ago di Buffon.* Il piano Oxy è diviso con strisce parallele all'asse y di larghezza unitaria. Un ago di lunghezza unitaria viene gettato a caso sul piano. Si domanda quanto vale la probabilità che l'ago stia in due strisce. Consideriamo il centro dell'ago: la sua posizione è determinata da due coordinate, ma possiamo ignorare la y e ridurre la x modulo 1 (considerare cioè come se la x prendesse solo valori in $[0, 1]$). La direzione dell'ago sia l'angolo che fa l'ago con l'asse x ; una rotazione dell'ago (attorno al centro) di π riproduce la sua posizione per cui l'angolo può essere preso variabile tra $-\pi/2$ e $\pi/2$. Indicheremo con X la var.al. "centro dell'ago" e con πY la var.al. "direzione". Si suppone che X e Y siano indipendenti con densità uniforme negli intervalli $[0, 1]$ e $[-1/2, 1/2]$. Si ha che l'ago taglia un confine se e solo se $1/2 \cos \pi Y > X$ oppure $1/2 \cos \pi Y > 1 - X$. Nel quadrato $[0, 1] \times [-1/2, 1/2]$ si disegnino le curve $1/2 \cos \pi y = x$ e $1/2 \cos \pi y = 1 - x$. La probabilità P richiesta è misurata dall'area di quella parte del quadrato limitata dalle curve e dalle parallele all'asse y . Tenendo conto delle simmetrie questa area vale $4 \int_0^{1/2} 1/2 \cos \pi y dy$ per cui $P = 2/\pi$.

Esempio 2. *Triangoli.* Si prendono tre punti a caso nel piano, qual'è la probabilità P che il triangolo ottenuto sia ottusangolo? Possiamo ritenere i tre punti distribuiti uniformemente sulla circonferenza unitaria; due di essi saranno agli estremi di una corda intercettata da un angolo t , in tal caso il terzo punto determina un triangolo ottusangolo quando si trovi nella parte di semicirconferenza determinata da angolo minore o uguale di $\pi - t$. Tenuto conto delle simmetrie si trova $P = \frac{2}{\pi} \int_0^{\pi/2} \frac{\pi-t}{\pi} dt = 3/4$. Senza calcoli si perviene allo stesso risultato nel modo seguente. Siano α e β due degli angoli del triangolo: $0 \leq \alpha, \beta \leq \pi$; $\alpha + \beta \leq \pi$; si consideri poi nel piano cartesiano la densità

uniforme sul triangolo di vertici $(0, 0), (\pi, 0), (0, \pi)$.

Esempio 3. *Densità uniforme sul cerchio.* Nel cerchio C_R con centro nell'origine e raggio R sia data la densità uniforme $f(x, y) = \frac{1}{\pi R^2}$ se $(x, y) \in C_R$ e 0 altrimenti. Vogliamo calcolare le densità marginali f_X, f_Y . Sappiamo che $f_X(u) = \int_{-\infty}^{+\infty} f(u, v) dv$. Nel nostro caso: se $|u| > R$, $f_X(u) = 0$; se $|u| \leq R$ si ha $f_X(u) = \frac{1}{\pi R^2} \int_{-\sqrt{R^2-u^2}}^{\sqrt{R^2-u^2}} dv = \frac{2\sqrt{R^2-u^2}}{\pi R^2}$. Analogo risultato si avrà per f_Y . Si osservi che le var.al. X e Y non sono indipendenti. Infatti se Q_R è il quadrato di lato $2R$ centrato nell'origine, allora se

$$(x, y) \in Q_R \setminus C_R : f(x, y) = 0 \text{ mentre } f_X(x)f_Y(y) > 0.$$

Come nel caso discreto vale il seguente risultato:

Se X_i sono vettori al. indipendenti e h_i funzioni a valori reali “ragionevoli”, allora anche $h_i(X_i)$ sono indipendenti.

4.5 Densità condizionali

Due var.al. X e Y abbiano densità congiunta f e marginali f_X, f_Y . Si chiama *densità condizionale* di X dato $Y = y$ la quantità

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

se $f_Y(y) > 0$, se $f_Y(y) = 0$ porremo $f_{X|Y}(x|y) = 0$. Si osservi che per ogni y con $f_Y(y) > 0$ la funzione $x \rightarrow f_{X|Y}(x|y)$ è una densità, dal momento che $\int_{-\infty}^{+\infty} f_{X|Y}(x|y) dx = 1$. Nell'esempio 3 del paragrafo precedente si ha che, se $|y| < R$, allora

$$f_{X|Y}(x|y) = \begin{cases} \frac{1}{2\sqrt{R^2-y^2}} & \text{se } -\sqrt{R^2-y^2} \leq x \leq \sqrt{R^2-y^2} \\ 0 & \text{altrimenti} \end{cases}$$

quindi la densità condizionale di X dato $Y = y$ è la densità uniforme sull'intervallo $[-\sqrt{R^2-y^2}, \sqrt{R^2-y^2}]$.

Osserviamo che si possono definire densità condizionali anche per vettori al. *misti* cioè aventi una componente a.c. e una discreta.

4.6 Calcolo di leggi

Un problema comune è il seguente: sia X un vettore al. n -dimensionale di cui si conosce la densità (congiunta) f e sia $h : \mathbf{R}^n \rightarrow \mathbf{R}$ una funzione “ragionevole”: calcolare la densità della var.al. $h(X)$.

Una soluzione possibile consiste nel calcolare la f.r. di $h(X)$ e poi derivarla. In effetti si ha: $F_{h(X)}(t) = p\{h(X) \leq t\} = p\{X \in h^{-1}(-\infty, t] = A_t\} = \int_{A_t} f$.

Esempio. Sia X una var.al. di densità f , calcolare le densità g di $aX + b$ ($a \neq 0$) e di X^2 .

Nel primo caso, se $a > 0$, $p\{aX + b \leq t\} = p\{X \leq \frac{t-b}{a}\} = F_X(\frac{t-b}{a})$, pertanto derivando si ottiene $g(t) = \frac{1}{a}f(\frac{t-b}{a})$. Se $a < 0$, $p\{aX + b \leq t\} = p\{X \geq \frac{t-b}{a}\} = 1 - F_X(\frac{t-b}{a})$, derivando si ottiene ora $g(t) = -\frac{1}{a}f(\frac{t-b}{a})$. In definitiva si ha per ogni $a \neq 0$: $g(t) = \frac{1}{|a|}f(\frac{t-b}{a})$.

Indichiamo con g la densità di X^2 . Si ha, prendendo $t \geq 0$,

$$p\{X^2 \leq t\} = p\{-\sqrt{t} \leq X \leq \sqrt{t}\} = F_X(\sqrt{t}) - F_X(-\sqrt{t})$$

e derivando si ottiene $g(t) = \frac{f(\sqrt{t}) + f(-\sqrt{t})}{2\sqrt{t}}$.

Teorema. Siano X e Y var.al. di densità congiunta f , allora la densità g di $X + Y$ vale

$$g(z) = \int_{-\infty}^{+\infty} f(x, z-x) dx.$$

Dimostrazione. Sia $S_t = \{(x, y) : x + y \leq t\}$, allora la f.r. G della somma vale $G(t) = \int_{S_t} f(x, y) dx dy = \int_{-\infty}^{+\infty} dx (\int_{-\infty}^{t-x} f(x, y) dy)$, col cambiamento di variabile $y = z - x$ si ottiene $G(t) = \int_{-\infty}^{+\infty} dx (\int_{-\infty}^t f(x, z-x) dz)$ da cui cambiando l'ordine di integrazione si ha infine

$$G(t) = \int_{-\infty}^t dz (\int_{-\infty}^{+\infty} f(x, z-x) dx)$$

ossia $G(t) = \int_{-\infty}^t g(z) dz$.

A volte la densità di una somma si calcola in modo diretto come nel seguente

Esempio. Siano X e Y var.al. indipendenti e uniformi in $[0,1]$. La var.al. $X + Y = Z$ prende valori in $[0,2]$. Sul quadrato unitario del piano Oxy si disegni la retta di equazione $x + y = t$. $F_Z(t)$ vale l'area della parte di quadrato a sinistra di questa retta. Conviene distinguere i casi $t > 1$ e $t < 1$. Avremo

$$F_Z(t) = t^2/2 \text{ se } 0 \leq t \leq 1, \quad F_Z(t) = 1 - (2-t)^2/2 \text{ se } 1 < t \leq 2.$$

Di conseguenza la densità vale

$$f_Z(t) = t \text{ se } 0 \leq t \leq 1, \quad f_Z(t) = 2 - t \text{ se } 1 < t \leq 2.$$

Nel caso vettoriale per trovare la densità di $h(X)$ bisogna ricorrere alla formula del cambiamento di variabili negli integrali multipli.

Sia $X = (X_1, X_2)$ un vettore al. di densità congiunta f e sia $h : R^2 \rightarrow R^2$ una funzione per cui valga la formula del cambiamento di variabili negli integrali doppi. Si cerca la densità g del vettore al. $Y = h(X)$. Abbiamo intanto per un "ragionevole" $A \subset R^2$ che

$$\int_{h^{-1}(A)} f(x_1, x_2) dx_1 dx_2 = p\{X \in h^{-1}(A)\} = p\{h(X) \in A\} = \int_A g(y_1, y_2) dy_1 dy_2.$$

Applicando il cambiamento di variabile $\underline{x} = h^{-1}(\underline{y})$ si ha

$$\int_{h^{-1}(A)} f(\underline{x}) d\underline{x} = \int_A f(h^{-1}(\underline{y})) |Dh^{-1}(\underline{y})| d\underline{y}$$

dove Dh^{-1} è il determinante jacobiano della trasformazione h^{-1} . Siccome per “quasi tutti” gli A vale l’uguaglianza

$$\int_A g(\underline{y}) d\underline{y} = \int_A f(h^{-1}(\underline{y})) |Dh^{-1}(\underline{y})| d\underline{y},$$

possiamo concludere che

$$g(\underline{y}) = f(h^{-1}(\underline{y})) |Dh^{-1}(\underline{y})|.$$

A volte per studiare un problema conviene cambiare coordinate. Trattiamo il caso delle coordinate polari sul piano.

Sia $V = (X, Y)$ un vettore al. di densità congiunta f ; passiamo alle coordinate polari ponendo

$$X = R \cos \Theta, \quad Y = R \sin \Theta$$

dove $W = (R, \Theta)$ è ristretto a $R \geq 0, -\pi < \Theta \leq \pi$. Se $A \subset R^2$ avremo

$$p(A) = \int_A f(x, y) dx dy = \int_{A'} f(r \cos \theta, r \sin \theta) r dr d\theta$$

dove A è l’immagine di A' nella trasformazione. Ma gli eventi $V \in A$ e $W \in A'$ sono identici; questo significa che la densità congiunta g di W vale $g(r, \theta) = f(r \cos \theta, r \sin \theta) r$.

Applichiamo questo risultato alla densità uniforme sul cerchio già considerata (vedi Esempio 3 in 4.4.). Avremo

$$g(r, \theta) = \frac{r}{\pi R^2} \quad 0 \leq r \leq R, \quad -\pi < \theta \leq \pi,$$

da cui si ottengono con semplici calcoli le densità marginali

$$g_R(r) = \frac{2r}{R^2}, \quad g_\Theta(\theta) = \frac{1}{2\pi}.$$

Siccome $g(r, \theta) = g_R(r)g_\Theta(\theta)$ si vede che le var.al. R e Θ sono indipendenti (mentre X e Y non lo sono).

Esempio 4.6.1. Sia $X = (X_1, X_2)$ una var.al. uniforme sul cerchio unitario, si cerca la densità della var.al. $Z = \frac{2}{\|X\|} \sqrt{-\log \|X\|} X$.

Passando alle coordinate polari sia $Y = (R, \Theta)$, per quanto visto Y ha densità $\frac{r}{\pi}$ sul rettangolo (del piano $r \theta$) $[0, 1] \times [-\pi, \pi]$. Si ha $W = Z(R \cos \Theta, R \sin \Theta) = (2\sqrt{\log 1/R}, \Theta) = h(Y)$. Indicando con s, ψ le variabili della inversa, si vede

che $h^{-1}(s, \psi) = (e^{-\frac{s^2}{4}}, \psi)$. Il modulo del determinante Jacobiano della h^{-1} vale $\frac{s}{2}e^{-\frac{s^2}{4}}$. La densità di W vale pertanto $\frac{e^{-\frac{s^2}{4}}}{\pi} \frac{s}{2}e^{-\frac{s^2}{4}} = \frac{s}{2\pi}e^{-\frac{s^2}{2}}$. Dividendo per s e rimettendo le coordinate cartesiane si ottiene la densità richiesta di Z espressa appunto con le coordinate cartesiane:

$$g(y_1, y_2) = \frac{1}{2\pi} e^{-\frac{y_1^2 + y_2^2}{2}}.$$

Poichè g è una densità, deve avere integrale su R^2 uguale a 1, da questa osservazione si deduce che

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

4.7 Leggi normali

Riprendiamo qui la densità normale f :

$$f(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Sia X una var.al. che abbia per densità questa f e siano σ e μ numeri reali con $\sigma > 0$. La var.al. $Y = \sigma X + \mu$ ha, come abbiamo visto, densità

$$g(y) = \frac{1}{|\sigma|} f\left(\frac{y - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y - \mu)^2}{2\sigma^2}}.$$

Una densità g come sopra si dice *normale* o *gaussiana* di parametri μ e σ^2 , o anche una $N(\mu, \sigma^2)$, per la var.al. Y si usa anche scrivere $Y \sim N(\mu, \sigma^2)$. La f considerata è quindi una $N(0, 1)$; una $N(\mu, \sigma^2)$ si può riguardare come densità di una var.al. $\sigma X + \mu$ dove X ha legge $N(0, 1)$. La f.r. Φ_{μ, σ^2} di una $N(\mu, \sigma^2)$ si deduce, come segue, dalla f.r. di una $N(0, 1)$:

$$\Phi_{\mu, \sigma^2}(t) = p\{\sigma X + \mu \leq t\} = p\left\{X \leq \frac{t - \mu}{\sigma}\right\} = \Phi\left(\frac{t - \mu}{\sigma}\right).$$

Faremo vedere, con il metodo delle funzioni caratteristiche, che la somma di n var.al. indipendenti $X_i \sim N(\mu_i, \sigma_i^2)$ è una $G_n \sim N(\mu, \sigma^2)$ con $\mu = \sum \mu_i$, $\sigma^2 = \sum \sigma_i^2$.

4.8 Leggi gamma

Si chiama Γ la funzione da $\mathbf{R}^+$ a $\mathbf{R}^+$:

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

Con integrazioni per parti si vede che

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha).$$

Se n è un intero positivo allora

$$\Gamma(n) = (n-1)!.$$

Si dice che una var.al. X segue una legge gamma di parametri $\alpha > 0$, $\lambda > 0$ oppure che $X \sim \Gamma(\alpha, \lambda)$ se la sua densità è concentrata su $\mathbf{R}^+$ e vale

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}.$$

(Non è difficile verificare con un'integrazione che f_X è effettivamente una densità).

Esempio. Sia $X \sim N(0, \sigma^2)$ e quindi di densità $\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}}$. La densità g di X^2 sappiamo calcolarla e vale

$$g(y) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{2\sqrt{y}} [e^{-\frac{y}{2\sigma^2}} + e^{-\frac{y}{2\sigma^2}}] = \frac{1}{\sqrt{2\pi}\sigma} y^{1/2-1} e^{-\frac{y}{2\sigma^2}}.$$

Poichè g è una densità, si vede che X^2 deve essere una $\Gamma(\frac{1}{2}, \frac{1}{2\sigma^2})$; si deduce quindi che per $\alpha = \frac{1}{2}$ e per $\lambda = \frac{1}{2\sigma^2}$ si ha

$$\frac{\lambda^\alpha}{\Gamma(\alpha)} = \frac{1}{\sqrt{2\pi}\sigma};$$

cioè l'importante uguaglianza

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

Teorema 4.8.. Siano X e Y due var.al. indipendenti di legge risp. $\Gamma(\alpha, \lambda)$, $\Gamma(\beta, \lambda)$, allora $(X + Y) \sim \Gamma(\alpha + \beta, \lambda)$.

Dimostrazione. Ricordiamo che la formula per la densità della somma di var.al. indipendenti è

$$g(z) = \int_{-\infty}^{+\infty} f_1(x) f_2(z-x) dx.$$

Siccome le f_i sono nulle per valori negativi della variabile, avremo in effetti un integrale esteso all'intervallo $[0, z]$:

$$\begin{aligned} g(z) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\lambda^\beta}{\Gamma(\beta)} \int_0^z x^{\alpha-1} e^{-\lambda x} (z-x)^{\beta-1} e^{-\lambda(z-x)} dx = \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \int_0^z x^{\alpha-1} (z-x)^{\beta-1} dx; \end{aligned}$$

facendo ora nell'integrale il cambiamento di variabile $x = tz$ si ottiene

$$g(z) = \left[\frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt \right] z^{\alpha+\beta-1} e^{-\lambda z}.$$

Dunque $g \sim \Gamma(\alpha + \beta, \lambda)$ e il termine in parentesi quadra vale $\frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)}$; si ha quindi l'importante uguaglianza

$$\int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Notiamo due casi particolari:

La legge $\Gamma(1, \lambda)$ è la legge esponenziale di densità $f(x) = \lambda e^{-\lambda x}$ già considerata. La legge $\Gamma(\frac{n}{2}, \frac{1}{2})$ è detta legge del *chi quadro* a n gradi di libertà e si denota con $\chi^2(n)$. Nell'esempio precedente abbiamo visto che se $X \sim N(0, 1)$ allora $X^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$, pertanto se $X_1, \dots, X_n$ sono var.al. indipendenti $\sim N(0, 1)$, avremo che $(X_1^2 + \dots + X_n^2) \sim \chi^2(n)$.

4.9 Media, momenti e varianza

Definizioni e risultati sono del tutto simili al caso discreto (2.5, 2.6).

Sia X una var.al. continua con densità f , si dice che X ha media finita se

$$\int_{-\infty}^{+\infty} |x|f(x) dx < +\infty,$$

in tal caso la sua media $E[X]$ vale

$$E[X] = \int_{-\infty}^{+\infty} xf(x) dx < +\infty.$$

Teorema 4.9.1. Siano $X_1, \dots, X_n$ var.al. di densità congiunta f ; sia $h : R^n \rightarrow R$ una funzione "ragionevole". Allora la var.al. $Z = h(X_1, \dots, X_n)$ ha media finita se e solo se

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} |h(x_1, \dots, x_n)| f(x_1, \dots, x_n) dx_1 \dots dx_n < +\infty,$$

in tal caso la media $E[Z]$ vale

$$E[Z] = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} h(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n < +\infty.$$

Ecco alcune proprietà della media.

Teorema 4.9.2. Siano X e Y var.al. con media finita. Allora:

i) $(aX + bY)$ ha media finita e $E[aX + bY] = aE[X] + bE[Y]$;

ii) $|X|$ ha media finita e $|E[X]| \leq E[|X|]$;

se inoltre X e Y sono indipendenti

iii) XY ha media finita e $E[XY] = E[X]E[Y]$.

Sia X una var.al. di densità f ; diremo che X ha momento di ordine k finito

($k = 1, 2, \dots$) se X^k ha media finita. In questo caso $E[X^k]$ si chiama momento di ordine k della var.al. X :

$$E[X^k] = \int x^k f(x) dx.$$

Se $(X - E[X])^k$ ha media finita diremo che X ha momento centrato di ordine k finito e chiameremo $E[(X - E[X])^k]$ momento centrato di ordine k di X .

Teorema 4.9.3. i) Se X ha momento finito di ordine k , allora ha anche momento finito di ordine r per ogni $r \leq k$.

ii) Se X e Y hanno momento finito di ordine k , allora anche $(X + Y)$ ha momento finito di ordine k . In particolare se X ha momento finito di ordine k , allora ha anche momento centrato finito di ordine k .

Ricordiamo che il momento centrato di ordine 2 viene chiamato varianza:

$$\text{Var}(X) = E[(X - E[X])^2].$$

Vale la disuguaglianza di Chebyshev. Se per la var.al. X esiste il momento $E[X^2]$ (e di conseguenza la varianza) si ha:

$$p\{|X| > c\} \leq \frac{E[X^2]}{c^2};$$

in particolare, posto $\mu = E[X]$,

$$p\{|X - \mu| > c\} \leq \frac{\text{Var}(X)}{c^2}.$$

Fissata la var.al. X , denotiamo al solito media e varianza con μ e σ^2 .

Se X e Y hanno varianza finita, è finita anche la covarianza di X e Y :

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

Teorema 4.9.3. Si ha:

i) $\text{Var}(X) = E[X^2] - E[X]^2$;

ii) $\text{Var}(aX) = a^2 \text{Var}(X)$, $\text{Var}(a + X) = \text{Var}(X)$;

iii) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$;

iv) $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$.

Se X e Y sono indipendenti si ha

$$\text{Cov}(X, Y) = 0; \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Quando $\text{Cov}(X, Y) = 0$ le var.al. X e Y non sono correlate. Per il coefficiente di correlazione $\rho_{X,Y}$ si ha

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Ricordiamo che $-1 \leq \rho_{X,Y} \leq 1$.

4.10 Esempi

1 Legge uniforme. Sia la var.al. X uniforme su $[0,1]$. Allora

$$E[X] = \int_0^1 x \, dx = \frac{1}{2}, \quad E[X^2] = \int_0^1 x^2 \, dx = \frac{1}{3}, \quad \text{Var}(X) = \frac{1}{12}.$$

2 Leggi normali. Sia $X \sim N(0, 1)$. La sua densità vale $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Siccome $xf(x)$ è una funzione dispari su $\mathbf{R}$, avremo $E[X] = 0$. In questo caso $\text{Var}(X) = E[X^2]$; l'integrale si calcola per parti e vale 1. Riassumendo

$$E[X] = 0, \quad \text{Var}(X) = 1.$$

Se $Y \sim N(\mu, \sigma^2)$ ($f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$), allora si può considerare che $Y = \sigma X + \mu$. Si ottiene subito che

$$E[Y] = \mu, \quad \text{Var}(Y) = \sigma^2.$$

3 Leggi gamma. Se $X \sim \Gamma(\alpha, \lambda)$ e $\beta > 0$, allora con facili calcoli si vede che

$$E[X^\beta] = \frac{\Gamma(\alpha + \beta)}{\lambda^\beta \Gamma(\alpha)}.$$

Otteniamo quindi subito

$$E[X] = \frac{\alpha}{\lambda}, \quad \text{Var}(X) = \frac{\alpha}{\lambda^2}.$$

Per le leggi esponenziali ($\alpha = 1$) si ha

$$E[X] = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}.$$

Se $X \sim \chi^2(n)$ ($\alpha = n/2, \lambda = 1/2$) si ha

$$E[X] = n, \quad \text{Var}(X) = 2n.$$

4.11 Numeri a caso, simulazione

Consideriamo il seguente problema: è data una var.al. X con legge uniforme su $[0,1]$ (per esempio “numeri a caso” generati da un calcolatore) ed è assegnata una densità g ; si cerca una funzione h tale che la var.al. $h(X)$ abbia come densità proprio la g .

Supporremo $g(x)$ nulla per $x < a$, $x > b$ e strettamente positiva nell'intervallo (a, b) . In queste ipotesi la sua f.r. G è strettamente crescente e quindi invertibile nell'intervallo (a, b) . Sia dunque G^{-1} l'inversa di G in (a, b) e sia $Y = G^{-1}(X)$; avremo (ricordando che la f.r. di X è l'identità in $[0,1]$)

$$F_Y(t) = p\{G^{-1}(X) \leq t\} = p\{X \leq G(t)\} = G(t).$$

Dunque $F_Y = G$ e $f_Y = g$. Facciamo ora degli esempi di simulazioni.

1 Legge esponenziale. La f.r. F , che è concentrata sulla semiretta positiva, vale $F(t) = 1 - e^{-\lambda t}$. Si ha $F^{-1}(x) = -\frac{1}{\lambda} \log(1-x)$. Se la var.al X è uniforme su $[0,1]$, allora la var.al. $Y = -\frac{1}{\lambda} \log(1-X)$ è esponenziale di parametro λ .

2 Simulazione prove di Bernoulli. Sia assegnato un numero p con $0 < p < 1$ e, per $i = 1, \dots, n$, sia K_i un numero a caso in $[0,1]$. Definiamo le var.al. X_i ponendo $X_i = 1$ se $K_i \leq p$ e $X_i = 0$ altrimenti. La somma $(X_1 + \dots + X_n)$ è una var.al. $\sim B(n, p)$.

3 Legge normale. Prendiamo una var.al. X uniforme sul cerchio unitario e consideriamo, vedi l'esempio 4.6.1, la var.al $Z = \frac{2\sqrt{-\log \|X\|}}{\|X\|} X$ che ha densità

$$f(y_1, y_2) = \frac{1}{2\pi} e^{-\frac{y_1^2 + y_2^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y_1^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y_2^2}{2}}.$$

Questo significa che le componenti Z_1, Z_2 di Z sono indipendenti e $\sim N(0, 1)$.

4.12 Esercizi e complementi

1. Esercizio.

Siano X_i var.al. indipendenti con funzioni di ripartizione F_i , $i = 1, \dots, n$. Determinare le funzioni di ripartizione di $M = \max(X_i)$ e di $m = \min(X_i)$.

Soluzione.

$$F_M(t) = p\{M \leq t\} = p\{X_1 \leq t, \dots, X_n \leq t\} = F_1(t)F_2(t) \dots F_n(t).$$

$$p\{m > t\} = p\{X_1 > t, \dots, X_n > t\} = (1 - F_1(t))(1 - F_2(t)) \dots (1 - F_n(t)),$$

$$F_m(t) = p\{m \leq t\} = 1 - (1 - F_1(t))(1 - F_2(t)) \dots (1 - F_n(t)).$$

2. Esercizio.

Siano X_1, X_2, X_3 tre var.al. indipendenti uniformemente distribuite in $[0,1]$, calcolare $E[Z]$ dove Z è la var.al. “valore intermedio delle X_i ”.

Soluzione.

Siano $M = \max(X_i)$, $m = \min(X_i)$, avremo la rappresentazione

$$Z = X_1 + X_2 + X_3 - M - m$$

e pertanto

$$E[Z] = 3/2 - E[M] - E[m].$$

Per le funzioni di ripartizione del massimo e del minimo (tenendo presente l'indipendenza) si ha:

$$F_M(t) = t^3, \quad F_m(t) = 1 - p\{m > t\} = 1 - (1-t)^3,$$

da cui si ha facilmente $E[M] = 3/4$, $E[m] = 1/4$. Si trova quindi $E[Z] = 1/2$. Naturalmente si può risolvere l'esercizio, con calcoli un po' più lunghi, determinando la densità di Z . Per fare questo poniamo:

$$A_1 = \{X_2 \leq t, X_3 \leq t\}, A_2 = \{X_3 \leq t, X_1 \leq t\}, A_3 = \{X_1 \leq t, X_2 \leq t\}.$$

Si osservi ora che

$$\begin{aligned} F_Z(t) &= p(A_1 \cup A_2 \cup A_3) = \\ &= \sum_i p(A_i) - \sum_{i \neq j} p(A_i \cap A_j) + p(A_1 \cap A_2 \cap A_3) = \\ &= 3t^2 - 3t^3 + t^3 = 3t^2 - 2t^3. \end{aligned}$$

Pertanto la densità vale

$$f_Z(t) = 6t - 6t^2$$

e si ritrova che $E[Z] = 1/2$.

3. Esercizio.

Siano X_i $i = 1, \dots, n$ var.al. indipendenti uniformemente distribuite in $[0,1]$. Dimostrare che per la funzione di ripartizione del loro prodotto W_n si ha:

$$W_n(t) = t \sum_{k=1}^n (-1)^{k-1} \frac{(\log t)^{k-1}}{(k-1)!}$$

Soluzione.

Verifichiamo la formula per $n = 2$ e poi useremo l'induzione. Si ha

$$\{X_1 X_2 \leq t\} = \cup_{a \in [t,1]} \{X_1 \leq t/a, X_2 = a\} \cup_{a \in [0,t]} \{X_2 = a\}$$

di conseguenza si ha

$$W_2(t) = \int_t^1 t/a \, da + \int_0^t da = t - t \log t.$$

Osserviamo ora che

$$\{W_n \leq t\} = \{W_{n-1} X_n \leq t\} = \cup_{a \in [t,1]} \{X_n \leq t/a, W_{n-1} = a\} \cup_{a \in [0,t]} \{W_{n-1} = a\}$$

di conseguenza si ha

$$W_n(t) = \int_t^1 \frac{t}{a} W'_{n-1}(a) \, da + \int_0^t W'_{n-1}(a) \, da = W_{n-1}(t) + t \int_t^1 \frac{W'_{n-1}(a)}{a} \, da.$$

Si verifica ora facilmente che

$$W'_{n-1}(t) = (-1)^{n-2} \frac{(\log t)^{n-2}}{(n-2)!}$$

da cui si ricava, come si voleva, che

$$t \int_t^1 \frac{W'_{n-1}(a)}{a} da = (-1)^{n-1} t \frac{(\log t)^{n-1}}{(n-1)!}.$$

4. Esercizio.

Si prendano tre numeri a caso (in $[0,1]$), detti m e M il minimo e il massimo dei tre, per ogni $a \in [0,1]$ si calcoli $p\{m \leq a \leq M\}$

Soluzione.

Risolveremo l'esercizio usando la densità condizionale:

sia $0 < x < y$ e $M = y$; calcoliamo la probabilità condizionale

$$A = p\{m \leq x \text{ dato } M = y\}.$$

Si ha $A = 1 - p\{m > x \text{ dato } M = y\} = 1 - \frac{(y-x)^2}{y^2}$. La derivata (rispetto a x) di questa ultima è la densità condizionale, cioè

$$2 \frac{(y-x)}{y^2} = f_{m|y}(x|y).$$

Per la densità congiunta f del vettore aleatorio (m, M) si ha la formula $f(x, y) = f_M(y)f_{m|y}(x|y)$, siccome $f_M(y) = \frac{d}{dy}y^3 = 3y^2$, si ottiene $f(x, y) = 3y^2 2 \frac{(y-x)}{y^2} = 6(y-x)$. Se S_a è il rettangolo $0 \leq x \leq a$; $a \leq y \leq 1$, allora per la probabilità richiesta P si ha

$$P = \int \int_{S_a} f(x, y) dx dy = 3a(1-a)$$

Questo esercizio si risolve in modo più semplice come segue:

sia Q è l'evento $\{m \leq a \leq M\}$, allora l'evento complementare

$$Q^c = \{m > a\} \cup \{M < a\}$$

è unione di due eventi disgiunti, per cui

$$p(Q^c) = a^3 + (1-a)^3, \quad p(Q) = 3a(1-a).$$

4.13 Funzioni caratteristiche

Date la var.al. U e V , diremo che $Z = U + iV$ è una var.al. complessa. Diremo che Z ha media finita se e solo se U e V hanno media finita e porremo $E[Z] = E[U] + iE[V]$. Si verifica che $|E[Z]| \leq E[|Z|]$.

Sia X una var.al. (reale), si chiama *funzione caratteristica* di X la funzione a valori complessi ϕ_X definita su $\mathbf{R}$ da:

$$\phi_X(\theta) = E[e^{i\theta X}].$$

Conviene osservare che

$$|\phi_X(\theta)| = |E[e^{i\theta X}]| \leq E[|e^{i\theta X}|] = 1; \quad \phi_X(0) = 1.$$

Se X ha densità f avremo

$$\phi_X(\theta) = \int_R e^{i\theta x} f(x) dx.$$

Vale la pena notare che, a meno di una costante, ϕ_X è la trasformata di Fourier della densità f .

Valgono i seguenti semplici risultati:

(i) X, Y indipendenti $\Rightarrow \phi_{X+Y}(\theta) = \phi_X(\theta)\phi_Y(\theta)$;

(ii) $\phi_{-X}(\theta) = \overline{\phi_X(\theta)}$ (il coniugato);

(iii) Se $Y = aX + b$, allora $\phi_Y(\theta) = \phi_X(a\theta)e^{i\theta b}$.

Proviamo per esempio (i) (si tenga conto che dall'ipotesi segue anche l'indipendenza delle var.al. $e^{i\theta X}, e^{i\theta Y}$):

$$\phi_{X+Y}(\theta) = E[e^{i\theta(X+Y)}] = E[e^{i\theta X}e^{i\theta Y}] = E[e^{i\theta X}]E[e^{i\theta Y}] = \phi_X(\theta)\phi_Y(\theta).$$

Ci domandiamo se una funzione caratteristica ϕ sia derivabile. La funzione $\theta \rightarrow e^{i\theta X}$ è analitica qualunque valore prenda X ; se l'operazione di derivazione e quella di prendere la media E (che è una integrazione) si possono scambiare, allora ϕ risulta derivabile e le sue derivate si calcolano subito. Vale il seguente

Teorema. Qualunque sia X ϕ_X è continua. Se X ha momento di ordine k , allora ϕ_X è k volte derivabile e si ha

$$\frac{d^k \phi_X}{d\theta^k}(\theta) = E[(iX)^k e^{i\theta X}].$$

Viceversa se ϕ_X è k volte derivabile e k è pari, allora X ha momento di ordine k .

Si noti in particolare che

$$\frac{d^k \phi_X}{d\theta^k}(0) = (i)^k E[X^k],$$

questa formula permette di calcolare i momenti.

E' importante sapere che se X e Y hanno la stessa funzione caratteristica, allora hanno la stessa densità. Non solo questo è vero, ma si ha pure una formula esplicita per la densità di X in termini della sua funzione caratteristica:

Teorema (Formula di inversione). Se ϕ_X è sommabile, allora X è a.c. e per la sua densità f_X si ha

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-i\theta x} \phi_X(\theta) d\theta.$$

Diamo ora degli esempi.

1. Densità esponenziale.

$$\phi(\theta) = \lambda \int_0^{+\infty} e^{i\theta x} e^{-\lambda x} dx = \frac{\lambda}{\lambda - i\theta}.$$

2. Densità normale.

$$\phi(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{i\theta x} e^{-\frac{x^2}{2}} dx;$$

$$\phi'(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} i x e^{i\theta x} e^{-\frac{x^2}{2}} dx.$$

Con una integrazione per parti si trova

$$\phi'(\theta) = -\theta\phi(\theta)$$

da cui, tenendo conto che $\phi(0) = 1$, si ricava

$$\phi(\theta) = e^{-\frac{\theta^2}{2}}.$$

Se $Y \sim N(\mu, \sigma^2)$ sappiamo che si ha $Y = (\sigma X + \mu)$ con $X \sim N(0, 1)$, pertanto applicando la (iii) si ottiene

$$\phi_Y(\theta) = e^{-\frac{\sigma^2 \theta^2}{2}} e^{i\theta \mu}.$$

Applichiamo questo risultato e (i) alla somma di var.al. normali.

Per $i = 1, 2$ siano $X_i \sim N(\mu_i, \sigma_i^2)$ e indipendenti; avremo

$$\phi_{X_1+X_2}(\theta) = e^{-\frac{(\sigma_1^2 + \sigma_2^2)\theta^2}{2}} e^{i\theta(\mu_1 + \mu_2)} = \phi_X(\theta)$$

dove $X \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. L'uguaglianza delle funzioni caratteristiche ci porta a concludere che la somma $(X_1 + X_2)$ è una var.al. $\sim N(\mu_i, \sigma_i^2)$, come avevamo già enunciato.

1. **Densità discrete.** Le funzioni caratteristiche si definiscono anche per va.al. discrete: se X prende i valori $\{x_k\}$ con densità f si pone

$$\phi_X(\theta) = \sum_k f(x_k) e^{i\theta x_k}.$$

Diamo in particolare tre esempi.

1. $X \sim B(n, p)$:

$$\phi_X(\theta) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} e^{i\theta k} = (1-p + p e^{i\theta})^n.$$

2. X geometrica:

$$\phi_X(\theta) = \sum_k p(1-p)^k e^{i\theta k} = \frac{p}{1 - (1-p)e^{i\theta}}.$$

3. X Poisson di parametro λ :

$$\phi_X(\theta) = e^{-\lambda} \sum_k \frac{\lambda^k}{k!} e^{i\theta k} = e^{-\lambda} e^{\lambda e^{i\theta}}.$$

4.14 Convergenza in legge e Teorema Limite Centrale

Definizione. Siano X_n , X var.al e siano F_n , F le loro f.r. Diremo che X_n converge *in legge* a X se per ogni punto x di continuità di F vale

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

Esempio. Sia $\{X_i\}$ una successione di var.al. indipendenti aventi la stessa densità $f(x) = \frac{2}{\pi(1+x^2)}$ concentrata in $(0, +\infty)$ e sia $G_n = \frac{\max(X_1, \dots, X_n)}{n}$; mostriamo che le G_n convergono in legge:

Sia F la f.r. delle X_i e siano F_n le f.r. delle G_n , si ha

$$F_n(t) = p\{\max(X_1, \dots, X_n) \leq nt\} = (F(nt))^n = a_n.$$

Dobbiamo calcolare il limite (per $n \rightarrow \infty$) di a_n che è una forma indeterminata del tipo 1^∞ . Si ha $\log a_n = \frac{\log F(nt)}{1/n}$; usando la regola di de l'Hôpital si trova:

$$\lim_n a_n = \exp(-t \lim_n n^2 f(nt)) = \exp(-\frac{2}{\pi t}).$$

Siano ϕ_n , ϕ le funzioni caratteristiche di X_n , X . Vale il seguente importante risultato:

Teorema (P. Lévy). X_n converge in legge a X se e solo se $\phi_n(\theta)$ converge a $\phi(\theta)$ per ogni $\theta \in \mathbf{R}$.

Teorema Limite Centrale Sia $\{X_n\}$ una successione di var.al. indipendenti aventi la stessa densità. Supponiamo che esistono la media $\mu = E[X_k]$, e la varianza $\sigma^2 = \text{Var}(X_k)$. Allora, posto $S_n^* = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$, si ha che le var.al. S_n^* convergono in legge a una var.al X che è $N(0, 1)$ (cioè normale di media 0 e varianza 1).

Dimostrazione. Poniamo $Y_k = \frac{X_k - \mu}{\sigma}$, le var.al. Y_k hanno media nulla, varianza 1 e $S_n^* = \frac{Y_1 + \dots + Y_n}{\sqrt{n}}$. Indichiamo con ϕ la funzione caratteristica comune alle Y_k : abbiamo che $\phi_{S_n^*}(\theta) = \phi(\frac{\theta}{\sqrt{n}})^n$. Per il teorema di P. Lévy basterà mostrare che $\phi_{S_n^*}(\theta) \rightarrow e^{-\frac{\theta^2}{2}}$, che è la funzione caratteristica di una $N(0, 1)$. Intanto, visto che $\phi'(0) = iE[Y_k] = 0$; $\phi''(0) = -\text{Var}(Y_k) = -1$, possiamo scrivere

$$\phi(x) = 1 - \frac{x^2}{2} + o(x^2),$$

e quindi, per ogni fissato θ ,

$$\phi(\frac{\theta}{\sqrt{n}}) - 1 = -\frac{\theta^2}{2n} + o(\frac{1}{n}).$$

Abbiamo dunque

$$\phi_{S_n^*}(\theta) = (1 + (\phi(\frac{\theta}{\sqrt{n}}) - 1))^n =$$

$$= \exp[n \log(1 + (\phi(\frac{\theta}{\sqrt{n}}) - 1))] =$$

(ricordando che $\log(1+x) \sim x$)

$$= \exp[-\frac{\theta^2}{2} + o(1)].$$

L'ultima espressione per $n \rightarrow \infty$ tende a $e^{-\frac{\theta^2}{2}}$.

4.15 Approssimazione normale

Spesso la tesi del Teorema Limite Centrale viene usata come una formula di approssimazione: si dice allora che si fa uso dell'approssimazione normale.

Il Teorema Limite Centrale afferma che

$$\lim_n p\left\{\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq t\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{y^2}{2}} dy.$$

Facendo uso dell'approssimazione normale porremo

$$p\left\{\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq t\right\} \simeq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{y^2}{2}} dy = \Phi(t).$$

Pertanto se per un numero reale α si deve stimare

$$A = p\{X_1 + \dots + X_n \leq \alpha\},$$

tenuto conto che

$$\{X_1 + \dots + X_n \leq \alpha\} = \left\{\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq \frac{\alpha - n\mu}{\sigma\sqrt{n}}\right\},$$

avremo

$$A \simeq \Phi\left(\frac{\alpha - n\mu}{\sigma\sqrt{n}}\right).$$

Se per esempio si vuole stimare una deviazione dalla media

$$D = p\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| > c\right\}$$

come compare nella disuguaglianza di Chebyshev; si opera come segue:

$$\begin{aligned} D &= 1 - p\{X_1 + \dots + X_n \leq n\mu + nc\} + p\{X_1 + \dots + X_n < n\mu - nc\} \simeq \\ &\simeq 1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right) + \Phi\left(-\frac{c\sqrt{n}}{\sigma}\right) = 2\Phi\left(-\frac{c\sqrt{n}}{\sigma}\right). \end{aligned}$$

Questa stima è in generale molto migliore di quella fornita dalla disuguaglianza di Chebyshev.

Tradizionalmente l'approssimazione normale si considera buona per gli n maggiori di 50.

5 Appendice

5.1 Misura e integrazione astratta

1. Uno spazio di misura è una terna $\{\Omega, \mathcal{A}, \mu\}$ dove Ω è un insieme, $\mathcal{A}$ una σ -algebra di Ω e μ una misura positiva su $\mathcal{A}$ (μ soddisfa gli stessi assiomi di una probabilità p salvo che non si chiede che $\mu(\Omega) = 1$, inoltre è permesso il valore $+\infty$). Gli elementi della σ -algebra $\mathcal{A}$ sono i sottoinsiemi misurabili di Ω .
2. Una $f : \Omega \rightarrow \mathbf{R}$ si dice misurabile se $f^{-1}(V)$ è misurabile (cioè appartiene ad $\mathcal{A}$) per ogni aperto V di $\mathbf{R}$.
3. Indicheremo con $\mathcal{B}$ la σ -algebra generata dagli aperti di $\mathbf{R}^n$, i suoi elementi vengono chiamati boreliani. Una funzione $f : \mathbf{R}^n \rightarrow \mathbf{R}$ si dice Borel misurabile se è misurabile quando in $\mathbf{R}^n$ ci sia la σ -algebra dei boreliani. Ogni f continua risulta dunque Borel misurabile.
4. In generale se $f_n : \Omega \rightarrow \mathbf{R}$ sono misurabili, tali risultano anche le funzioni $\sup_n f_n$, $\limsup_n f_n$, in particolare, se esiste, è misurabile $\lim_n f_n$.
5. Sia $f : \Omega \rightarrow \mathbf{R}$, definiamo $f^+(x) = \max(f(x), 0)$; $f^-(x) = -\min(f(x), 0)$, allora $f = f^+ - f^-$ porge una decomposizione canonica di una funzione nella differenza di due funzioni positive. Per quanto detto se f è misurabile, tali sono anche f^+ e f^- .
6. Una funzione s su Ω che prende solo un numero finito di valori (positivi) è detta semplice. Se $a_1, \dots, a_n$ sono i valori assunti da s , allora $s = \sum_i a_i \chi_{A_i}$, dove $A_i = s^{-1}(a_i)$ e χ_A indica la funzione caratteristica di A . Chiaramente s è misurabile se e solo se è misurabile ogni insieme A_i .
7. Sia $f : \Omega \rightarrow [0, +\infty]$ misurabile. Esiste una successione di funzioni semplici $\{s_n\}$ tale che $0 \leq s_1 \leq \dots \leq s_n \leq \dots \leq f$; $s_n(x) \rightarrow f(x)$ per ogni $x \in \Omega$. Inoltre se f è limitata la convergenza è uniforme.
8. Sia s una funzione semplice (misurabile), $s = \sum_i a_i \chi_{A_i}$, $E \subset \Omega$ misurabile, l'integrale (di Lebesgue) di s su E è definito da

$$\int_E s d\mu = \sum_i a_i \mu(A_i \cap E).$$

Sia f una funzione positiva misurabile, l'integrale di f su E è definito da

$$\int_E f d\mu = \sup_{0 \leq s \leq f} \int_E s d\mu.$$

9. Teorema della convergenza monotona: sia $0 \leq f_1 \leq \dots \leq f_n \leq \dots$, siano le f_n misurabili e $f_n(x) \rightarrow f(x)$ per ogni $x \in \Omega$, allora (f è misurabile e) $\int_E f_n \rightarrow \int_E f$.
10. $L_1(\mu)$ è lo spazio delle funzioni f μ -misurabili su Ω e tali che $\int_\Omega |f| < \infty$. Se $f \in L_1(\mu)$, l'integrale di f su E è definito da $\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu$; poichè $|f| = f^+ + f^-$ si ha $|\int_E f d\mu| \leq \int_E |f| d\mu$.
11. Teorema della convergenza dominata: siano f_n misurabili e $f_n(x) \rightarrow f(x)$

per ogni $x \in \Omega$. Supponiamo che esista $g \in L_1(\mu)$ tale che per ogni n ed x si abbia $|f_n(x)| \leq g(x)$; allora $f \in L_1(\mu)$, $\int_E |f_n - f| d\mu \rightarrow 0$ e inoltre $\int_E f_n d\mu \rightarrow \int_E f d\mu$.

12. Sia P una proprietà che un punto x può avere o non avere. Si dice che P vale quasi ovunque (q.o.) in un insieme E se il sottoinsieme di E in cui P non vale ha misura nulla. Gli insiemi di misura nulla sono trascurabili per l'integrazione nel senso che se $f = g$ q.o. in E allora $\int_E f d\mu = \int_E g d\mu$. Pertanto in questa teoria si parla spesso di funzioni definite q.o.

13. Una misura μ su una σ -algebra si dice completa se i sottoinsiemi degli insiemi (misurabili) di misura nulla sono misurabili (e di misura nulla). E' sempre possibile completare una misura ampliando in modo naturale la sua σ -algebra.

14. Sia μ una misura di Borel. Un boreliano A si dice regolare se

$$\mu(A) = \inf\{\mu(V) : A \subset V, V \text{ aperto}\}$$

e se, nel caso che $\mu(A) < \infty$,

$$\mu(A) = \sup\{\mu(K) : K \subset A, K \text{ compatto}\}.$$

Se ogni boreliano è regolare, μ si dice regolare.

15. Ogni misura di Borel in $\mathbf{R}^n$ è regolare.

5.2 Misura di Lebesgue

1. Teorema di esistenza della misura di Lebesgue in $\mathbf{R}^n$. Esiste una misura positiva completa $m = m_n$ definita su una σ -algebra $\mathcal{M}$ in $\mathbf{R}^n$, che ha le seguenti proprietà:

i) $m(W) = \text{Vol}(W)$ per ogni iperrettangolo di $\mathbf{R}^n$.

ii) $\mathcal{M}$ contiene i boreliani e m è regolare.

iii) m è invariante per traslazione, cioè $m(E + x) = m(E)$ per ogni $E \in \mathcal{M}$ e per ogni $x \in \mathbf{R}^n$.

iv) (Unicità di m). Se μ è una qualsiasi misura di Borel in $\mathbf{R}^n$ invariante per traslazione e a valori finiti sui compatti, allora esiste una costante c tale che $\mu(E) = c m(E)$ per ogni boreliano $E \in \mathbf{R}^n$.

2. Si dice che $f_n \rightarrow f$ in misura se, per ogni fissato $\eta > 0$,

$$\lim_{n \rightarrow \infty} m\{x : |f_n(x) - f(x)| > \eta\} = 0.$$

Si ha che: $f_n \rightarrow f$ q.o. $\Rightarrow f_n \rightarrow f$ in misura.

3. Le funzioni monotone (e quindi anche le differenze di monotone) sono derivabili q.o.

4. Una funzione f si dice assolutamente continua (a.c.) se: per ogni $\epsilon > 0$ esiste $\delta > 0$ tale che, comunque si scelga un numero finito di intervalli (a_i, b_i) non coprentisi e con $\sum_i (b_i - a_i) < \delta$, si abbia $\sum_i |f(b_i) - f(a_i)| < \epsilon$. Le funzioni

lipschitziane sono a.c. Le funzioni a.c. sono derivabili q.o.

5. Se $g \in L_1(m)$, posto, per ogni $x \in \mathbf{R}$, $f(x) = \int_{-\infty}^x g(t)dt$, allora f è a.c. e inoltre $f'(x) = g(x)$ q.o.

6. Per le funzioni a.c. vale la formula fondamentale del calcolo integrale (per l'integrale di Lebesgue). Precisamente vale il seguente fatto. Condizione necessaria e sufficiente a che la formula

$$f(x) - f(a) = \int_a^x f'(t)dt$$

valga per tutti gli x di un intervallo $[a, b]$, è che f sia a.c. su $[a, b]$.

7. Teorema di Fubini in $\mathbf{R}^2$. Sia $f \in L_1(m_2)$. Allora

i) per quasi tutti gli $x \in \mathbf{R}$ la funzione $y \rightarrow f(x, y)$ appartiene a $L_1(m_1)$;

ii) la funzione $g(x) = \int_{\mathbf{R}} f(x, y)dy$ appartiene a $L_1(m_1)$;

iii) risulta

$$\int_{\mathbf{R}^2} f(x, y)dxdy = \int_{\mathbf{R}} g(x)dx = \int_{\mathbf{R}} \left(\int_{\mathbf{R}} f(x, y)dy \right) dx.$$

Analogamente per quasi tutti gli $y \in \mathbf{R}$, esiste l'integrale $h(y) = \int_{\mathbf{R}} f(x, y)dx$, e si ha

$$\int_{\mathbf{R}^2} f(x, y)dxdy = \int_{\mathbf{R}} h(y)dy = \int_{\mathbf{R}} \left(\int_{\mathbf{R}} f(x, y)dx \right) dy.$$