Probabilità e statistica

Giacomo Fantoni

14 maggio 2019

Indice

Capitolo 1

Introduzione

Per dare una valutazione di incertezza è necessario definire il metodo dell'esperimento.

1.1 Esperimento casuale o esperimento aleatorio

Un esperimento determinato da un certo individuo in un certo istante. Pur avendo una serie di informazioni su di esso l'individuo non è in grado di determinarne con certezza il risultato, indipendentemente dal fatto che l'esperimento sia già avvenuto oppure no.

1.1.1 Spazio campionario

Lo spazio che contiene tutti risultati possibili dell'esperimento. Indicato con Ω. Gli elementi di Ω sono mutualmente esclusivi (l'avvenimento di uno esclude gli altri) e lo spazio è esaustivo (contiene tutti i risultati possibili). Gli elementi di Ω si chiamano eventi elementari.

1.2 Insiemistica

Si definisca:

- A^c il complementare di A.
- $A \backslash B = A \cap B^c$ la differenza tra due insiemi.
- $A \Delta B = (A \backslash B) \cup (B \backslash A)$.

1.2.1 Proprietà unione e intersezione

- Idempotenza: $A \cap A = A$ e $A \cup A = A$
- Commutatività.
- Associatività.
- Distributività: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cap \emptyset = \emptyset$, $A \cup \emptyset = A$, $A \cap \Omega = A$, $A \cap \Omega = \Omega$

1.2.2 Leggi di De Morgan

- $(A \cup B)^c = A^c \cap B^c$
- $(A \cap B)^c = A^c \cup B^c$

1.2.3 Operazioni per gruppi di insiemi

Si supponga di avere un numero finito di insiemi $A_1, A_2, \cdots, A_n$.

- L'unione viene indicata con $\bigcup\limits_{i=1}^{n} A_i$.
- l'intersezione: $\bigcap\limits_{i=1}^{n} A_i$

Regole di De Morgan

- $(\bigcup\limits_{i=1}^{n} A_i)^c = \bigcap\limits_{i=1}^{n} A_i^c$
- $(\bigcap\limits_{i=1}^{n} A_i)^c = \bigcup\limits_{i=1}^{n} A_i^c$

1.2.4 Successioni di insiemi

Le successioni di insiemi sono numerabili, contengono ovvero elementi pari al più ai numeri naturali. Ovvero esiste una funzione biettiva tra i numeri naturali e gli insiemi.Si possono allora considerare:

- Unione $\bigcup\limits_{i=1}^{\infty} A_i$
- Intersezione $\bigcap\limits_{i=1}^{\infty} A_i$

Considerando una successione più che numerabile si descrive $\bigcap\limits_{\alpha \in \Lambda} A_\alpha$, dove Λ è un insieme maggiore dei numeri naturali.

1.2.5 Considerazioni su sottoinsiemi

Siano $A, B \subseteq \Omega$, se $A \subseteq B$ $A \cap B = A$:

- $A \cup B = B$
- $B^c = \Omega \backslash B \subseteq \Omega \backslash A = A^c$
- $A \cap B^c = \emptyset$
- $A^c \cap B = \Omega$

1.3 Gli eventi

I potenziali eventi sono sottoinsiemi dell'insieme spazio campionario. Il numero di eventi è dato dalla cardinalità dell'insieme delle parti dello spazio campionario: $2^{\#\Omega}$. Questo insieme viene detto insieme potenza, indicato con $P(\Omega)$ e contiene tutti i sottoinsiemi di Ω. Gli eventi sono gli elementi dell'insieme potenza.

1.3.1 Proprietà

Una classe $\mathcal{A}$ di parti di un insieme Ω ($\mathcal{A} \subseteq P(\Omega)$) è detta tribù (o σ-algebra) se:

- $\Omega \in \mathcal{A}$
- Deve essere chiusa rispetto all'operazione complementare se $A \in \mathcal{A}$ allora $A^c \in \mathcal{A}$.
- $\{A_i\}_{i=1}^{+\infty}$ tale che $A_i \in \mathcal{A} \Rightarrow \bigcup\limits_{i=1}^{+\infty} A_i \in \mathcal{A}$ (metto in successione infinita ogni combinazione di elementi di $\mathcal{A}$).

Un algebra è una σ-algebra la cui terza proprietà cambia in: $\{A_i\}_{i=1}^{n}$ tale che $A_i \in \mathcal{A} \Rightarrow \bigcup\limits_{i=1}^{n} A_i \in \mathcal{A}$. Una σ-algebra è sempre un'algebra. La più piccola tribù è $\mathcal{A}_\infty = \{\emptyset, \Omega\}$. Una tribù generata da un'insieme di partenza è detta tribù generata da tale insieme.

1.3.2 Definizione

Un evento è un elemento della tribù che viene utilizzata.

Capitolo 2

Probabilità

2.1 Spazio probabilizzabile

Uno spazio probabilizzablie è uno spazio su cui potenzialmente si può definire una misura di probabilità. Si dice spazio probabilizzabile la coppia $(\Omega, \mathcal{A})$ dove Ω è uno spazio campionario e $\mathcal{A}$ è una tribù su Ω.

2.2 La funzione di probabilità

Dato uno spazio probabilizzabile $(\Omega, \mathcal{A})$ una funzione di probabilità (o una probabilità) è una funzione $Pr : \mathcal{A} \to [0; 1]$. La particolare funzione da associare ad un particolare spazio probabilizzato dipende ed è caratteristica di esso.

2.2.1 Assiomi di Kolmogorov

Qualunque funzione che associa ad un elmento di $\mathcal{A}$ un numero reale può essere considerata una funzione di probabilità se rispetta i seguenti assiomi:

- Non negavitività: se $A \in \mathcal{A}$ allora $Pr(A) \geq 0$.
- Normalizzazione: $Pr(\Omega) = 1$
- Di σ-additività: $\forall n \in \mathbb{R}^+ + \{\infty\}\ \ \{A_i\}_{i=1}^n : A_i \in \mathcal{A} \wedge A_i \cap A_j = \emptyset\ \forall i \neq j$ allora $Pr(\bigcup\limits_{i=1}^{n} A_i) = \sum\limits_{i=1}^{+\infty} Pr(A_i)$. In particolare se $A, B \in \mathcal{A}$ e $A \cap B = \emptyset$ allora: $Pr(A \cup B) = Pr(A) + Pr(B)$.

Conseguenze degli assiomi

- $Pr(\emptyset) = 0$.
- $Pr(A) \leq 1$.

- $Pr(A^c) = 1 - Pr(A)$.
- Se $A \subseteq B$ allora $Pr(A) \leq Pr(B)$.
- $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$. Si noti che se i due eventi sono disgiunti si riottiene il terzo assioma

2.3 Spazio probabilizzato

Uno spazio probabilizzato (o spazio di probabilità o spazio di Kolmorogov) è la terna $(\Omega, \mathcal{A}, Pr)$ che rispettivamente sono spazio campionario, tribù su Ω e una funzione di probabilità $Pr : \mathcal{A} \to \mathbb{R}^+$.

2.3.1 Regole di calcolo

Probabilità che l'evento non si verifichi

$$Pr(A^c) = 1 - Pr(A) \tag{2.1}$$

Sia $A \in \mathcal{A}$. Essendo $[A \cup A^c] = \Omega \in \mathcal{A}$ e $(A \cap A^c) = \emptyset \in \mathcal{A}$ ed essendo per il secondo assioma $Pr(\Omega) = 1$ e per il terzo $Pr(A \cup A^c) = Pr(A) + Pr(A^c)$, essendo elementi disgiunti allora $Pr(A^c) = 1 - Pr(A)$.
Osservazione: considerando A_1, A_2, A_n, con $A_i \in \mathcal{A}\ \forall i = 1, \cdots, n$, $\bigcap_{i=1}^{n} A_i$ è l'intersezione enumerabile di n eventi, che rappresenta un evento, perciò la tribù è chiusa rispetto all'intersezione.

Probabilità che si verifichi almeno un evento tra i due

$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B) \tag{2.2}$$

Siano $A, B \in \mathcal{A}$. Essendo $A \cup B = A \cup (B \cap A^c)$, $B = (A \cap B) \cup (A^c \cap B)$, $Pr(A \cap B) - Pr(B) = Pr(A) + Pr(B \cap A^c) - Pr(A \cap B) - Pr(B \cap A^c)$, ovvero $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$.
Osservazione: una conseguenza di questa regola di calcolo è che $Pr(A \cup B) \leq Pr(A) + Pr(B)$, o, più in generale $Pr(\bigcup_{i=0}^{n} A_i) \leq \sum_{i=0}^{n} Pr(A_i)$.

2.3.2 Un evento implica un altro

Considerando i due eventi A, B tali che $A \subset B$, allora:

$$Pr(B) = Pr(A) + Pr(B \cap A^c) \geq Pr(A) \tag{2.3}$$

Perciò $Pr(B) \geq Pr(A)$ in quanto $B = A \cup (B \cap A^c)$, ed essrndo A e $B \cap A^c$ incompatibili, si ottiene che $Pr(B) = Pr(A) + Pr(B \cap A^c)$. Essendo $B \cap A^c$ un evento $Pr(B \cap A^c) \geq 0$, da cui il risultato.
Osservazione: da questa regola derivano queste considerazioni:

- $\forall A \in \mathcal{A}, Pr(A) \leq 1$ (basta porre $B = \Omega$).
- $A \subseteq B \wedge B \subseteq A \Rightarrow Pr(A) = Pr(B)$.
- $Pr(A) \leq Pr(B) \wedge Pr(A) \geq Pr(B) \Rightarrow A = B$.

2.3.3 Disuguaglianza di Bonferroni

Siano $A_1, \cdots, A_n$ eventi di una tribù,e $n \geq 1$ allora:

$$\sum_{i=1}^{n} Pr(A_i) - \sum_{1 \leq i \leq j \leq n} Pr(A_i \cap A_j) \leq Pr(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} Pr(A_i) \tag{2.4}$$

Basta verificare la disuguaglianza di sinistra, procedendo per induzione (dimostrazione incompleta): per $n = 1$ è banalmente verificata.
Per $n = 2$: $Pr(A_1) + Pr(A_2) - Pr(A_1 \cap A_2) = Pr(A_1 \cup A_2)$.
Per $n = 3$: $Pr(A_1 \cup A_2 \cup A_3) = Pr(A_1 \cup A_2) + Pr(A_3) - Pr((A_1 \cup A_2) \cap A_3)$
$= Pr(A_1) + Pr(A_2) - Pr(A_1 \cap A_2) + Pr(A_3) - Pr((A_1 \cap A_3) \cup (A_2 \cap A_3))$
$= \sum\limits_{i=1}^{3} Pr(A_i) - Pr(A_1 \cap A_2) - [Pr(A_1 \cup A_3) + Pr(A_2 \cup A_3) - Pr(A_1 \cap A_2 \cap A_3)]$
$= \sum\limits_{i=1}^{3} Pr(A_i) - \sum\limits_{1 \leq i \leq j \leq 3} Pr(A_i \cap A_j) + Pr(A_1 \cap A_2 \cap A_3)$.

2.4 Limiti e probabilità

- Se $A_1 \subset A_2 \subset \cdots \in \mathcal{A}$ è una successione di eventi non decrescente e $A = \lim A_n = \bigcup_{n=1}^{\infty} A_n$ allora:
$$Pr(\lim A) = \lim Pr(A) \tag{2.5}$$
- Se $A_1 \supset A_2 \supset \cdots \in \mathcal{A}$ è una successione di eventi non crescente e $A = \lim A_n = \bigcap_{n=1}^{\infty} A_n$ allora:
$$Pr(\lim A) = \lim Pr(A) \tag{2.6}$$

Dimostrazione

Poichè $A = \bigcup\limits_{n=1}^{\infty} A_n = \bigcup\limits_{n=1}^{\infty} A_n \backslash A_{n-1} = \bigcup\limits_{n=1}^{\infty} (A_n \cap A_{n-1}^c)$, avendo definito $A_0 = \emptyset$ ed è $A_{n-1} \subset A_n$, dunque $\forall n, (A_n \cap A_{n-1}^c)$ sono eventi incompatibili e dunque per la σ-additività $Pr(A) = Pr(\bigcup\limits_{n=1}^{\infty} A_n \backslash A_{n-1}) = \sum\limits_{n=1}^{\infty} Pr(A_n \backslash A_{n-1}) =$
$= \sum\limits_{n=1}^{\infty} Pr(A_n) - Pr(A_{n-1}) = \lim\limits_{N \to \infty} \sum\limits_{n=1}^{N} Pr(A_n) - Pr(A_{n-1}) = \lim\limits_{N \to \infty} Pr(A_N)$
Per la seconda parte si considerino i complementari in modo che $A_1^c \subset A_2^c \subset \cdots$, si ottiene: $\lim\limits_{n \to \infty} A_n^c = \bigcup\limits_{n=1}^{\infty} A_n^c = (\bigcap\limits_{n=1}^{\infty} A_n)^c = A^c$, allora $\lim\limits_{n \to \infty} Pr(A_n^c) = Pr(\lim\limits_{n \to \infty} A_n^c) = Pr(A^c)$, pertanto $\lim\limits_{n \to \infty} Pr(A_n) = \lim\limits_{n \to \infty} (1 - Pr(A_n^c)) = 1 - \lim\limits_{n \to \infty} Pr(A_n^c) = 1 - Pr(\lim\limits_{n \to \infty} A_n^c) = 1 - Pr(A^c) = Pr(A)$.

2.5 Costruzione della funzione di probabilità

Si definisca una funzione di probabilità considerando al più uno spazio Ω numerabile e come tribù $\mathcal{A} = P(\Omega)$. Si supponga $\Omega = \{w_1, w_2, \cdots\}$ e ad ogni evento $w_i \in \Omega$ si assegna un peso $p(w_i)$, con $i \in \mathbb{N}$ in modo che $p(w_i) \geq 0$ e $\sum\limits_{i=1}^{\infty} p(w_i) = 1$. Per ogni evento $A \in \mathcal{A}$ si definisce la sua probabilità come somma dei pesi di ogni w_i che appartiene ad A, ovvero $Pr(A) = \sum\limits_{w_i \in A} p(w) = \sum\limits_{w_i \in A} P(\{w\})$.

2.6 Prodotto di una famiglia di tribù

Data una qualsiasi famiglia $\{(\Omega_i, \mathcal{A}_i)\}_{i \in I}$ di spazi probabilizzabili si considerino tutte le famiglie $\{A_i\}_{i \in I}$ con $A_i \in \mathcal{A}_i$ per ogni indice i e $A_i \neq \Omega_i$ al più per un numero finito di indici. Per ogni siffatta famiglia si consideri il "rettangolo" generato $\prod\limits_{i \in I} \Omega_i$ la tribù generata su questo insieme dalla classe di questi "rettangoli" si chiama tribù prodotto della famiglia di tribù $\{\mathcal{A}_i\}_{i \in I}$ e si denota come $\otimes_{i \in I} \mathcal{A}_i$. Quando $I = \{1, 2, \cdots, n\}$ si indica come $\mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \cdots \otimes \mathcal{A}_n$.

2.6.1 Prodotto di una famiglia di uno spazio probabilizzato

Data una famiglia $\{(\Omega_i, \mathcal{A}_i, Pr_i)\}_{i \in I}$ di spazi probabilizzati si nota che sulla tribù prodotto $\otimes_{i \in I} \mathcal{A}_i$ esiste un unica funzione di probabilità tale che $Pr(\prod\limits_{i \in I} A_i) = \prod\limits_{i \in I} Pr(A_i)$ per ogni famiglia $\{A_i\}_{i \in I}$ come definita precedentemente. Questa probabilità si chiama probabilità prodotto della famiglia e si indica come $\otimes_{i \in I} Pr(A_i)$. Se $i \in \{1, 2, \cdots, n\}$ si indica come $Pr_1 \otimes Pr_2 \otimes \cdots \otimes Pr_n$.

Capitolo 3

Calcolo combinatorio

Il calcolo combinatorio si occupa di studiare l'esistenza, la costruzione, l'enumerazione e esamina le proprietà di configurazioni che soddisfano delle particolari condizioni. Si considerino $S_1, S_2, \cdots, S_r$ r insiemi di cardinalità $n_1, n_2, \cdots, n_r$ formati da elementi distinti. Si consideri il loro prodotto cartesiano $\Omega_r = S_1 \times S_2 \times \cdots \times S_r = \{(s_1, s_2, \cdots, S_3) : s_1 \in S_1, s_2 \in S_2, \cdots, s_r \in S_r\}$.

3.1 Disposizioni con ripetizione

Le disposizioni con ripetizione indicano il numero di r-uple che sono contenute nell'insieme Ω_r. Poichè da S_n si possono scegliere n_n elementi allora il numero complessivo di elementi che sono contenuti è $n_1 \cdot n_2 \cdots n_r$. In particolare se $S_1 = S_2 = \cdots = S_r$ e $n_i \;\; \forall i \leq r$ allora il numero di r-uple sarà n^r.

Definizione

Dato un insieme $S = \{a_1, a_2, \cdots, a_n\}$ formato da n oggetti distinti il numero di allineamenti che si possono formare con r elementi scelti tra gli n,ritenendo diversi due allineamenti se differiscono per almeno un elemento, per l'ordine in cui sono scelti o perchè uno stesso elemento compare un numero diverso di volte, è dato da:

$$D^{\star}_{n,r} = n^r \tag{3.1}$$

Ogni allineamento si dice disposizione con ripetizione di n oggetti in classe r.

3.2 Disposizioni senza rpetizione

Le disposizioni senza ripetizione (o semplici), indicano il numero di allineamenti che si possono formare con gli oggetti contenuti in $S = \{a_1, a_2, \cdots, a_n\}$ presi a gruppi di $1 \leq r \leq n$ ma in modo che uno stesso oggetto non appaia più di una volta.

Definizione

Dato un insieme $S = \{a_1, a_2, \cdots, a_n\}$ formato da n oggetti distinti il numero di allineamenti che si possono formare con $1 \leq r \leq n$ elementi scelti tra gli n, ritenendo diversi due allineamenti se differiscono per almeno un elemento o perchè gli stessi oggetti si susseguono in modo diverso è dato da:

$$D + n, r = n(n-1) \cdots (n-r+1) = \frac{n!}{(n-r)!} \tag{3.2}$$

3.3 Permutazioni

Le permutazioni sono un caso particolare delle disposizioni senza ripetizione in cui $r = n$ e possono differire solo per l'ordine in cui si susseguono gli oggetti.

Definizione

Dato un insieme $S = \{a_1, a_2, \cdots, a_n\}$ di n oggetti distinti, il numero degli allineamenti che si possono formare con tutti essi, ritenendo diversi due allineamenti perchè gli oggetti si susseguono in ordine diverso, è dato da $n!$ (si pone $0! = 1$).

$$P_n = D_{n,n} = n! \tag{3.3}$$

3.4 Combinazioni

Le combinazioni indicano il numero di disposizioni che differiscono unicamente se gli allineamenti contengono elementi diversi, il loro ordine risutla ininfluente.

Definizione

Dato un insieme $S = \{a_1, a_2, \cdots, a_n\}$ di n oggetti distinti, il numero degli allineamenti che si possono formare con $1 \leq r \leq n$ oggetti scelti tra gli n, ritenendo diversi due allineamenti solo perché contengono oggetti differenti, è dato da:

$$C_{n,r} = \frac{D_{n,r}}{r!} = \binom{n}{r} \tag{3.4}$$

Ogni allineamento si dice combinazione senza ripetizione di n oggetti in classe r.

3.5 Cardinalità delle parti di un insieme finito

Enunciato

Sia $S = \{a_1, a_2, \cdots, a_n\}$ un insieme di n oggetti distinti, allora $\#P(S) = 2^n$.

Dimostrazione

Si dimostri per induzione:

- Per $n = 0$, $P(\emptyset) = \{\emptyset\}$ e ha cardinalità $2^0 = 1$.
- Per $n = 1$, $P(\{a\}) = \{\emptyset, \{a\}\}$ e ha cardinalità $2^1 = 1$.
- Per $n = 2$, $P(\{a_1, a_2\}) = \{\emptyset, \{a_1\}, \{a_2\}, \{a_1, a_2\}\}$ e ha cardinalità $2^2 = 4$.

Ora si assuma che $\#P(S_k) = 2^k$ e si mostri che vale $\#P(S_{k+1}) = 2^{k+1}$. Si noti che i sottoinsiemi di S_{k+1} sono divisi in due categorie: quelli che contengono a_{k+1}, si chiamino A e quelli che non la contengono, si chiamino B. Gli insiemi A sono 2^k per ipotesi di induzione, rimane da mostrare che gli insiemi B siano 2^k. A questo scopo si noti come ogni insieme B sia esprimibile come $A \cup \{a_{k+1}\}$, ovvero esiste una funzione biettiva del tipo $A \to A \cup \{a_{k+1}\}$, che implica che gli insiemi B siano 2^k. In concluionse $2^k + 2^k = 2^{k+1}$.

3.5.1 Cardinalità dell'insieme delle parti e combinazioni

Si notino le seguenti:

- $2^n = \sum_{r=0}^{n} \binom{n}{r}$
- $\binom{m+n}{r} = \sum_{i=0}^{r} \binom{m}{i} \cdot \binom{n}{r-i}$
- $\binom{n}{r} = \frac{n!}{r!(n-r)!} = \binom{n}{n-r}$

Capitolo 4

Probabilità sui reali

4.1 Costruzione di uno spazio probabilizzato su uno spazio campionario più che numerabile

Si consideri $\Omega = \mathbb{R}$ (i seguenti risultati potranno essere estesi facendo uso degli spazi prodotto nei casi $\Omega 0 \mathbb{R}^n$). Considerare come tribù $P(\mathbb{R})$ presenta dei problemi in quanto sarebbe troppo vasta. Si considera pertanto una tribù meno fine che contenga tutti gli intervalli nella forma $]a, b]$, con $a \leq b$. Si consideri pertanto la tribù generata a partire da questa classe di insiemi, ovvero posto $\mathcal{F} = \{]a, b] : a \leq b, a, b \in \mathbb{R}\}$ la tribù più piccola che contiene tutti i suddetti intervalli è $\mathcal{A}_{\mathcal{F}} = \bigcap \{\mathcal{A} : \mathcal{A} \text{ è una tribù e } \mathcal{F} \subseteq \mathcal{A}\}$.

4.1.1 Tribù Boreliana

Definzione

Si chiama tribù Boreliana di $\mathbb{R}$ e si denota con $\mathcal{B}(\mathbb{R})$ la tribù generata su $\mathbb{R}$ dalla classe di tutti gli intervalli $]a, b]$ di $\mathbb{R}$. I suoi elementi si chiamano insiemi boreliani di $\mathcal{B}$ e $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è uno spazio probabilizzabilie.

Insiemi borelliani

- $]a, b[= \bigcup\limits_{n=1}^{+\infty}]a, b - \frac{1}{n}]$.

- $[a, b] = \bigcap\limits_{n=1}^{+\infty}]a - \frac{1}{n}, b]$.

- $[a, b[= \bigcup\limits_{n=1}^{+\infty} \bigcap\limits_{n=1}^{+\infty}]a - \frac{1}{n}, b - \frac{1}{n}]$.

- $] - \infty, b] = \bigcup\limits_{n=1}^{+\infty}] - n, b]$. Allo stesso modo $]a, +\infty[$.

- $a = \bigcup_{n=1}^{+\infty}]a - \frac{1}{n}, a]$.
- $A = \{a_1, a_2, \cdots, a_n\} = \bigcap_{j=1}^{n} \{a_j\}$.
- $B = \bigcup_{j=1}^{+\infty} \{a_j\}$.

Non sono borelliani tutti gli insiemi per cui il concetto di lunghezza, area, volume non ha senso come per esempio l'insieme di Vitali.

4.2 Misura di probabilità

4.2.1 Funzione di distribuzione

Per provedere all'assegnazione di una funzione di probabilità agli eventi di $\mathcal{B}(\mathbb{R})$ si fissa la probabilità da attribuire agli intervalli $]a, b]$ mediante una funzione $F(x)$ che:

- Non è decrescente.
- è continua a destra: $\forall x_0 \in \mathbb{R}, \lim_{x \to x_0^+} (x) = F(x_0)$.
- ammette limite a sinistra: $\forall x_0 \in \mathbb{R}, \exists \lim_{x \to x_0^-} (x)$.
- $\lim_{x \to +\infty} F(x) = 1$.
- $\lim_{x \to -\infty} F(x) = 0$.

Ponendo $F(]a, b]) = F(b) - F(a)$. Il calcolo effettivo di $Pr(A)$ semplice quando A è un intervallo o un'unione numerabile di intervalli disgiunti: $Pr(\bigcup_{i=1}^{\infty}]a_i, b_i]) = \sum_{i=1}^{\infty} Pr(]a_i, b_i]) = \sum_{i=1}^{\infty} (F(b_i) - f(a_i))$. Se A

Osservazione

- $Pr(]a, b]) = F(b) - F(a)$. Sia $A = \bigcup_{i=1}^{+\infty}]a_i, b_i] \quad]a_i, b_i] \cap]a_j, b_j] = \emptyset \quad \forall i \neq j$, pertanto $Pr(A) = Pr(\bigcup_{i=1}^{+\infty}]a_i, b_i]) = \sum_{i=1}^{+\infty} Pr(]a_i, b_i]) = \sum_{i=1}^{+\infty} F(b_i) - F(a_i)$.
- $Pr([a, b]) = Pr(\lim_{n \to +\infty}]a - \frac{1}{n}, b]) = \lim_{n \to +\infty} Pr(]a - \frac{1}{n}; b])$, questa operazione è possibile se e solo se la funzione di distribuzione è continua e monotona nell'intorno.
- $Pr(\{a\}) = Pr([a, b] \setminus]a, b]) = Pr([a, b]) - Pr(]a, b])$.

Capitolo 5

Probabilità condizionale

Definizione

Sia $(\Omega, \mathcal{A}, Pr)$ uno spazio probabilizzato. Fissato un elemento H di $\mathcal{A}$, con $Pr(H) \neq 0$, si chiama funzione di probabilità dedotta da Pr sotto la condizione H la funzione di probabilità Pr_H sullo spazio $(\Omega, \mathcal{A})$ probabilizzabile:

$$Pr_H(A) - \frac{Pr(A \cap H)}{Pr(H)} \tag{5.1}$$

Per ogni evento $A \in \mathcal{A}$. La probabilità $Pr_H(A)$ si chiama probabilità condizionale di A secondo Pr sotto la condizione H e si denota $Pr(A|H)$.

5.1 Spazio di probabilità condizionale

Dopo che è accaduto l'evento H per ogni evento $A \in \mathcal{A}$ tale che $A \cap H = \emptyset$ si ha che $Pr_H(A) = PR(A|H) = 0$, avremo quindi un altro modo di definire lo spazio probabilizzato $(\Omega_H, \mathcal{A}_H, Pr_h)$ dove $\Omega_H = \{A \subseteq H : A \in \Omega\}$ e poichè accadono solo gli eventi che implicano H, $\mathcal{A}_H = \{A \subseteq H : A \in \mathcal{A}\}$.

5.2 teorema delle probabilità totali

5.2.1 Enuncitato

Sia $\{A_i\}_{i=1}^{\infty}$ una famiglia di eventi che costituisce una classe completa su Ω tale che $Pr(A_i) > 0 \;\; \forall i$ e sia B un qualunque evento, allora:

$$Pr(B) = \sum_{i=1}^{\infty} Pr(A_i \cap B) = \sum_{i=1}^{\infty} Pr(A_i) Pr(B|A_i) \tag{5.2}$$

Dimostrazione

Dalla relazione $\Omega = \bigcup_{i=1}^{\infty} A_i$, intersecando con B ambo i membri e per la proprietà distributiva rispetto all'unione si ottiene: $\Omega \cap B = \bigcup_{i=1}^{\infty} (A_i) \cap B = \bigcup_{i=1}^{\infty} (A_i \cap B)$, con gli eventi $A_i \cap B$ incompatibili a due a due. Segue allora che $Pr(B) = Pr(\bigcup_{i=1}^{\infty} (A_i \cap B)) = \sum_{i=1}^{\infty} Pr(A_i \cap B) = \sum_{i=1}^{\infty} Pr(A_i)Pr(B|A_i)$.

5.3 Teorema di Bayes

Sia $\{A_i\}_{i=1}^{\infty}$ una famiglia di eventi che costituisce una classe completa su Ω tale che $Pr(A_i) > 0 \ \ \forall i$ e sia B un qualunque evento, allora:

$$Pr(A_i|B) = \frac{PR(A_i)Pr(B|A_i)}{\sum_{j=1}^{\infty} Pr(A_j)Pr(B|A_j)} \tag{5.3}$$

Dimostrazione

La dimostrazione è da fare, da ricordarsi, immediata dal teorema delle probabilità totali.

Capitolo 6

Indipendenza stocastica

6.1 Indipendenza di due eventi

Due eventi si dicono stocasticamente indipendenti quando l'avvenimento di uno non modifica la misura di incertezza dell'altro.

Definizione

In uno spazio probabilizzato $(\Omega, \mathcal{A}, Pr)$ due eventi A, B si dicono stocasticamente indipendenti se e solo se $Pr(A \cap B) = Pr(A)Pr(B)$, si noti che in particolare $Pr(A|B) = Pr(A)$ e $Pr(B|A) = Pr(B)$.

6.2 Indipendenza di n eventi

La notazione di indipendenza stocastica può essere

Definizione

Si consideri $(\Omega, \mathcal{A}, Pr)$ e $\{A_i\}_{i=1}^{n} : A_i \in \mathcal{A} \;\; \forall i$ tali eventi sno stra loro stocasticamente indipendenti se e solo se $Pr(A_{i_1} \cap A_{i_2} \cap A_{i_k}) = Pr(A_{i_1})Pr(A_{i_2})Pr(A_{i_k})$, $\forall k = 2, \cdots, n$ e per ogni allineamento $(i_1 < i_2 < \cdots < i_k)$ degli interi $1, 2, \cdots, n$, ovvero gli eventi di una famiglia si dicono mutualmente indipendenti se ogni ennupla da essa estratta risulta stocasticamente indipendente.

6.3 Tribù indipendenti

Dato uno spazio probabilizzato $(\Omega, \mathcal{A}, Pr)$ due tribù contenute in $\mathcal{A}$ si dicono tra loro stocasticamente indipendenti se ogni elemento dell'uno è indipendente da ogni elemento dell'altra.

Capitolo 7

Variabili aleatorie

In molte occasioni lo spazio campionario è composto da elementi di natura non numerica. Spesso in questo caso sono oggetto di studio le conseguenze dell'esperimento, valutabili numericamente. Per legare l'esperimento e le sue conseguenze numeriche, e pertanto gli spazi campionari è necessario introdurre le variabili aleatorie.

Definizione

Sia dato uno spazio probabilizzabile $(\Omega, \mathcal{A})$, si dice variabile aleatoria ogni funzione a valori reali definita in Ω, con $y = X(w)$ tale che $\{w \in \Omega : X(w) \leq x\} \in \mathcal{A}$. Per ogni valore reale x.

Osservazioni

- Nella definizione la probabilità non ha alcun ruolo e che quando $\mathcal{A}$ è la classe dei sottoinsiemi di Ω la condizione della definizione è sempre soddisfatta.

- Per rendersi conto della necessità di imporre alla funzione $X(w)$ la condizione sopra descritta basta considerare che intendendo assegnare una probabilità agli insiemi $\{w \in \Omega : X(w) \leq x\}$ per ogni reale x ed avendo probabilizzato la classe $\mathcal{A}$, occorre che tali insiemi appartengano a $\mathcal{A}$.

7.1 Variabili aleatorie e tribù

7.1.1 Primo teorema sul cambio di tribù

Siano $\tilde{\Omega}$ e Ω due insiemi arbitrarie sia $X : \tilde{\Omega} \to \Omega$ una funzione, se $\mathcal{A}$ è una tribù su Ω allora $\tilde{\mathcal{A}} = \{X * -1(A) : A \in \mathcal{A}\}$ è una tribù su $\tilde{\Omega}$.

Dimostrazione

- $X^{-1}(\Omega) = \tilde{\Omega}$.
- $X^{-1}(\Omega \backslash A) = \tilde{\Omega} \backslash X^{-1}(A)$.
- $X^{-1}(\bigcup_{i=1}^{\infty} A_i) = \bigcup_{i=1}^{\infty} X^{-1}(A_i)$

7.1.2 Secondo teorema sul cambio di tribù

Siano $\tilde{\Omega}$ e Ω due insiemi arbitrarie sia $X : \tilde{\Omega} \to \Omega$ una funzione, se $\tilde{\mathcal{A}}$ è una tribù su $\tilde{\Omega}$ allora $\mathcal{A} = \{A \subseteq \Omega : X^{-1}(A) \in \tilde{\mathcal{A}}\}$ è una tribù su Ω.

Dimostrazione

- Siccome $\tilde{\Omega} \in \tilde{\mathcal{A}}$ e $\tilde{\Omega} = X^{-1}(\Omega)$ allora $\Omega \in \mathcal{A}$.
- Si supponga che $A \in \mathcal{A}$ allora $X^{-1}(A) \in \tilde{\mathcal{A}}$ che implica $X^{-1}(\Omega \backslash A) = \tilde{\Omega} \backslash X^{-1}(A) \in \tilde{\mathcal{A}}$ e quindi $\Omega \backslash A \in \mathcal{A}$.
- Si suppongano $A_1, A_2, \cdots \in \mathcal{A}$. Pertanto $X^{-1}(A_1), X^{-1}(A_2), \cdots, \in \tilde{\mathcal{A}}$, da cui $X^{-1}(\bigcup_{i=1}^{\infty} A_i) = \bigcup_{i=1}^{\infty} X^{-1}(A_i) \in \tilde{\mathcal{A}}$, perciò $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

7.2 Variabili aleatorie e funzioni

Ogni funzione continua e monotona crescente oppure decrescente $f : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è una variabile aleatoria.

7.2.1 Definizione funzioni di probabilità

Il valore che assume la funzione $y = X(w) : \Omega \to \mathbb{R}$ in corrispondenza di un esperimento è aleatorio in quanto dipende dal risultato dell'esperimento, ci si chiederà con quale probabilità la funzione $X(w)$ assume un nell'intervallo $]a, b]$, ovvero la probabilità di $a < X \leq b = Pr(X \in]a, b])$, $\ -\infty \leq a < b < \infty$. A tale scopo si osservino l'intervallo $]a, b]$ e l'evento $A = \{w \in \Omega : a < X(w) \leq b\} \in \mathcal{A}$, essi sono equivalenti in quanto quando si verifica A allora $X \in]a, b]$ e viceversa. Dato che ad A è assegnata $Pr(A)$, si potrà porre per ogni $a < b$, $Pr_X(]a, b]) = Pr(X \in]a, b]) = Pr(\{w \in \Omega : a < X(w) \leq b\})$. Questa funzione di probabilità è nota come come distribuzione della variabile aleatoria X e mediante essa è possibile determinare $Pr_X(B) = Pr(X \in B)$ per ogni $B \in \mathcal{B}(\mathbb{R})$.

Definizione

Una variabile aleatoria o casuale in uno spazio campionario Ω è una qualunque funzione con valori nei numeri reali la cui controimmagine è un evento di Ω. Se a diversi eventi corrisponde lo stesso numero a, $Pr(X = a) = Pr(\{w \in \Omega | X(w) = a\})$. Più in generale, per $a < b$, $Pr(a < X \leq b) = Pr(\{w \in \Omega | a < X(w) \leq b\})$.

7.3 Variabili aleatorie discrete

Sia X una variabile discreta che assume un numero finito di valori reali $X = \{x_1, x_2, \cdots, x_n\}$, l'insieme dei valori che X può assumere è detto supporto della variabile casuale. Si assegni ad ogni valore di X una probabilità: $P(X = x_i) = \{w \in \Omega | X(w) = x_i\}$ o $p(x_i)$. La funzione $p(\cdot)$ avrà pertanto le seguenti proprietà:

- $0 \leq p(x_i) \leq 1 \;\; \forall x_i \in \{x_1, \cdots, x_n\}$.
- $\sum\limits_{i=1}^{n} p(x_i) = 1$.

Questa funzione verrà pertanto chiamata funzione di probabilità della variabile discreta X.

Variabile aletoria con infiniti valori

Una variabile aleatoria discreta può assumere una quantità al più numerabile di valori, come nell'esempio nella distribuzione geometrica, in cui $P(X = x_i) = (1-p)^{i-1}p$, ovvero, in una sequenza di prove binarie determina la probabilità che il primo risultato sia alla prova i. È immediato notare che $\sum\limits_{i=1}^{\infty} p(x_i) = 1$.

Funzione di ripartizione

Si introduce la funzione di ripartizione come la funzione nella forma $F(x) = Pr(X \leq x) = \sum\limits_{x_i < x} p(x_i)$. Questa funzione è tale che:

- non è decrescente (dimostrazione banale).
- $F(x) = 0$ se $x = \min(x_i)$ e $F(x) = 1$ se $x = \max(x_i)$
- $F(x)$ è continua a destra, ovvero la funzione presenta dei salti nei punti x in cui $Pr(X = x) > 0$.

7.4 Distribuzioni particolari

7.4.1 Distribuzione binomiale

Si consideri la classe di problemi in cui un esperimento elementare con due possibili risultati, successo e insuccesso, viene ripetuto indipendentemente m volte e si chiede la probabilità di ottenere k successi sapendo che la probabilità di successo in ogni prova è p. Si consideri lo spazio campionario Ω costituieto dalle 2^m sequenze di f e s di lunghezza m, viene definita la variabile casuale B come $B(\Omega) = \{0, 1, \cdots, m\}$, e si vuole determinare la probabilità di B. SI deve calcolare la probabilità dell'evento $A \subset \Omega$ costituito da tutte le seguenze con k successi e $m - k$ insuccessi, A possiede $\binom{m}{k}$ sequenze e la probabilità di ognuna è data da $p^k(1 - p =^{m-k}$. Se B è una variabile aleatoria che si distribuisce in accordo alla distribuzione binomiale allora si può dimostrare che $\mathbb{E}(B) = mp$, mentre $\sigma(B) = mp(1-p)$. Si indicherà con $B \sim Bin(m,p)$

7.4.2 Distribuzione bernoulliana

Nel caso particolare di una variabile binomiale con $m = 1$, tale distribuzione viene chiamata bernoulliana $B \sim Ber(p)$. $\mathbb{E} = p$, mentre la varianza $\sigma = p(1-p)$. Data $B_i \sim Ber(p)\ \ i = 1, \cdots, m$, $\sum\limits_{i=1}^{m} B_i \sim Bin(m,p)$.

7.4.3 Distribuzione di Poisson

Questa distribuzione studia cosa succede alla variabile aleatoria quando il numero di prove m tende all'infinito e la probabilità θ_m di successo decresce in modo che il prodotto $m\theta_m$ tenda ad una costante, ovvero $\theta_m = \frac{\lambda}{m}$. La risposta per un m fissato viene fornita dalla distribuzione binomiale: $Pr(P = k) = \binom{m}{k}\frac{\lambda}{m}^k(1 - \frac{\lambda}{m})^{-1}$. Facendo tendere m all'infinito si ottiene $\lim\limits_{m\to\infty} \frac{m!}{m^k(m-k)!}\frac{\lambda}{k!}^k(1-\frac{\lambda}{m})^m(1-\frac{\lambda}{m})^{-k} = \frac{\lambda^k e^{-\lambda}}{k!}$. L'espressione trovata permette di trovare la probabilità del numero di successi indicati nel parametro k che avvengono in un numero infinito di tentativi. Viene chiamata variabile di Poisson e viene indicata con $P \sim Pois(\lambda)$, $Pr(P = k) = \frac{\lambda^k e^{-\lambda}}{k!}$.

Definire Poisson ricorsivamente

La distribuzione di Poisson può essere definita ricorsivamente nel seguente modo:
$$\begin{cases} Pr(X = k+1) = \frac{\lambda}{k+1}Pr(X = k) \\ Pr(X = 0) = e^{-\lambda} \end{cases}.$$

7.4.4 Distribuzione geometrica

Questa distribuzione studia la probabilità che il primo successo avvenga all'i-esimo tentativo: si può mostrare come data la variabile aleatoria $Y \sim Geo(p)$, si nota come $Pr(Y = y) = (1-p)^{y-1}p$ e che $F_Y(y) = Pr(Y \leq y) = \sum\limits_{i=1}^{[y]}(1-p)^{i-1}p = p\sum\limits_{i=0}^{[y]-1}(1-p)^i$.

7.4.5 Distribuzione binomiale negativa

Questa distribuzione studia la probabilità che in x tentativi avvengano r successi. $X \sim NBin(r,p)$. Si costruisce come sommatoria di distribuzioni geometriche. La sua funzione di probabilità è:
$$Pr(X = x) = \begin{cases} \binom{x-1}{r-1}p^r(1-p)^{x-r} & x \geq r \in \mathbb{Z} \\ 0 & altrimenti \end{cases}.$$

Distribuzioni binomiali e binomiali negative

Siano $X \sim Bin(n, p)$ e $Z \sim (r, p)$, allora $Pr(Z \geq r) = Pr(X \leq n)$.

7.5 Variabili aleatorie continue

Si consideri ora Ω come uno spazio campionario infinito e la variabile ad esso associatà $Y(\Omega)$ è un sottoinsieme non numerabile dei numeri reali. In questo caso si è in presenza di una variabile aleatoria continua.

7.5.1 Funzione di densità

Disponendo i dati in un grafo ad istogrammi separato per classi è naturale notare come aumentando indefinitamente il numero di queste ultime si formi una funzione $f(y)$ chiamata funzione di densità. Essendo l'area sotto tale curva la somma dei rettangolini contenuti in $[a.b]$ pari alla frequenza relativa dei campioni tra a e b, e notando inoltre che l'area sottesa alla curva vale 1, si ottiene che $Pr(a < Y \leq b) = \int_a^b f(y)dy$.

Proprietà

Del tutto analoghe al caso discreto:

- $f(y) \geq 0 \forall y \in \mathbb{R}$.
- $\int_{-\infty}^{+\infty} f(y)dy = 1$.

7.5.2 Funzione di ripartizione

Si può pertanto definire la funzione di ripartizione per una variabile aleatoria continua come $F(y) = \int_{-\infty}^{y} f(t)dt$, da cui si ottiene che $Pr(a < Y \leq b) = F(b) - F(a) = \int_{-\infty}^{b} f(t)dt - \int_{-\infty}^{a} f(t)dt$. Se la funzione di ripartizione è continua allora ogni singoletto ha proprietà nulla, che implica che $Pr(a < Y \leq b) = Pr(a < Y < b) = Pr(a \leq Y \leq b) = Pr(a \leq Y < b)$. Se la funzione di ripartizione è derivabile, la sua derivata è la funzione di densità, pertanto $f(y) = \frac{dF(y}{dy}$, viceversa $F(y) = \int_{-\infty}^{y} f(t)dt$.

7.6 Distribuzioni continue notevoli

7.6.1 Distribuzione uniforme

La funzione di densità di questa distribuzione è $f(y) = \begin{cases} \frac{1}{b-a} & a \leq y \leq b \\ 0 & altrimenti \end{cases}$.

7.6.2 Distribuzione normale o gaussiana

La funzione di densità di questa distribuzione è $f(y;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}e^{\frac{1}{2}(\frac{y-\mu}{\sigma})^2}$, dove μ indica il valore atteso e σ^2 la varianza. Si indicherà con $Y \sim \mathcal{N}(\mu,\sigma^2)$ tale distribuzione.

Standardizzazione

Indicata $Z \sim \mathcal{N}(0,1)$ come la distribuzione normale standard, qualsiasi altra distribuzione normale Y di parametri μ e σ^2 è esprimibile come $Z = \frac{Y-\mu}{\sigma}$.

Capitolo 8

Valori notevoli di variabili aleatorie

8.1 Valore atteso

Un valore atteso è una sintesi della variabile aleatoria in un numero: $\mathbb{E}$: $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_x) \to \mathbb{R}$. Il valore atteso rappresenta la media pesata di tale variabile aleatoria.

Valore atteso per variabili aleatorie discrete

Si supponga X una variabile aleatoria discreta con funzione di probabilità $p_x(x)$, si chiamerà $\mathbb{E}(X)$ la quantità $\mathbb{E}(X) = \sum\limits_{x \in R_x} x p_x(x)$ se $\sum\limits_{x \in R_x} |x| p_x(x) < \infty$.

Valore atteso per variabili aleatorie continue

Si supponga X una variabile aleatoria continua con funzione di densità $p_x(x)$, si chiamerà $\mathbb{E}(X)$ la quantità $\mathbb{E}(X) = \int_{-\infty}^{+\infty} x p_x(x)$ se $\int_{-\infty}^{+\infty} |x| p_x(x) < \infty$.

8.2 Trasformazioni di variabili aleatorie

Sia X una variabile aleatoria e si definisca $Y = a + bX$, $\mathbb{E}(Y) = \mathbb{E}(a + bX) = \sum\limits_{x \in R_X} (a + bx) p_x(x) = \sum\limits_{x \in R_X} (a p_x(x) + b x p_x(x)) = \sum\limits_{x \in R_X} a p_x(x) + \sum\limits_{x \in R_x} b x p_x(x) = a \sum\limits_{x \in R_X} p_x(x) + b \sum\limits_{x \in R_x} x p_x(x) = a + b\mathbb{E}(X)$. Il valore atteso di una trasformazione lineare di una variabile aleatoria è la trasformazione lineare del valore atteso. Più in generale il valore atteso di $h(Y)$, dove Y è una variabile aleatoria discreta può essere espresso come $\mathbb{E}(h(Y)) = \sum\limits_{i=1}^{k} h(y_i) p(y_i)$. Analogamente con l'integrale per le variabili aleatorie continue.

8.3 Momenti

8.3.1 Momenti non centrati

Sono valori attesi $\mathbb{E}(X^r)$ dove r (deciso arbitrariamente, fissato) è un intero positivo, chiamato ordine del momento. $\mathbb{E}(X^r) = \sum\limits_{x^r \in R_x} x^r p_x(x)$.

8.3.2 Momenti centrati

Si definisce il momento centrato come $\mathbb{E}((X - \mathbb{E}(X))^r)$.

8.3.3 Varianza

$\sigma^2 = \mathbb{E}((X - E(X))^2)$ è la varianza della variabile aleatoria. Il valore atteso è nella stessa unità di misura della variabile aleatoria. È una misura di dispersione, dice quanto si discosta il valore dal valore atteso.

Deviazione standart

Si definisce la deviazione standard come la radice della varianza $\sigma = \sqrt{\mathbb{E}((X - E(X))^2)}$.

8.3.4 Coefficiente di variazione

In modo analogo alla varianza indica la dispersione ma indipendentemente dall'unità di misura. $CV = \frac{\sqrt{\mathbb{E}((X-E(X))^2)}}{|\mathbb{E}(X)|}$.

8.3.5 Momento terzo centrato

$\rho_3 = \frac{\mathbb{E}((X-E(X))^3)}{\sigma^3}$. Se la variabile è simmetrica rispetto al suo valore atteso il numeratore è zero, che è il valore minore del momento terzo, pertanto valori piccoli quando la variabile tende verso la simmetria. L'implicazione al contrario non è vera.

8.3.6 Momento quarto centrato

$\rho_4 = \frac{\mathbb{E}((X-E(X))^4)}{\sigma^4}$, indice di Curtosi, ovvero la misura di quanto la massa è concentrata la media e di quanto velocemente vanno a zero le code, indici grandi code che scendono a zero lentamente e appuntimento in centro. La curtosi per la normale è tre.

8.4 Variabile aleatoria degenere

La variabile aleatoria degenere assume valore di probabilità 1 in un punto c. Il valore atteso di una costante è la costante stessa. Si consideri la variabile aleatoria Y costruita come $Y = cg(X)$, allora se $c = 1$, $Y = g(X)$, R_Y, il supporto della nuova variabile aleatoria, e si calcoli $Pr(Y_j)$, $\mathbb{E}(Y) = \sum_{i \in R_Y} yPr(y) = \sum_{j=1}^{J} y_j Pr(y_j) = \sum_{j=1}^{J} y_j P_x(g^{-1}(y_j))$. $Y = g(x)$, $\mathbb{E}(X) = \mathbb{E}(g(X))$. Se $c \neq 1$, si definisca $h(x) = cg(x)$, $\mathbb{E}(Y) = \sum_{x \in R_X} h(x)p_x(x) = \sum_{x \in R_X} cg(x)p_x(x) = c \sum_{x \in R_X} g(x)p_x(x) = x\mathbb{E}(g(X))$.

Capitolo 9

Variabili aleatorie multidimensionali

Si consideri $(\Omega, \mathcal{A}, P) \to (\mathbb{R}^P, \mathcal{B}(\mathbb{R}^P), P)$, ovvero l'idea è che $X(w)$ variable aleatoria $w \to (X_1(w), X_2(w), \cdots, X_P(w))$, dove $X(w)$ è una funzione $\Omega \to \mathbb{R}^P$.

9.1 Boreliani

Si considerano gli intervalli della stessa forma (chiuso a destra, aperto a sinistra) e si fa il prodotto cartesiano: $]x_{11}, x_{12}] \times]x_{21}, x_{22}] \times]x_{P1}, x_{P2}]$, ovvero $\mathcal{A} = \{$tutti i rettangoli ottenuti come prodotto cartesiano di tutti i rettangoli aperti a sinistra e chiusi a destra $\}$. $\mathcal{B}(\mathbb{R}^P) = \sigma(\mathcal{A})$.

9.2 Variabili aleatorie discrete bidimensionali

Si indicherà con $P_{x,y}(x, y) = Pr(\{X(w) = x\} \cap \{Y(w) = y\})$. La funzione di probabilità per la variabile aleatoria bivariata viene chiamata funzione di probabilità congiunta. $p_x(x) = Pr(\{X(w) = x\} \cap \Omega\}$, le antiimmagini di una variabile aleatoria formano partizioni di Ω, pertanto $Pr(\{X(w) = x\} \cap \{\bigcup_{y_i \in R_y} Y(w) = y_i\})$, utilizzando la proprietà distributiva si ottiene $\sum_{y_i \in R_y} Pr(X(w) = x \cap Y(w) = y_i)$, ovvero $\sum_{y_i \in R_y} P_{x,y_i}$. Pertanto si può esprimere la funzione di probabilità di una variabile aleatoria a partire dalla funzione di probabilità della variabile congiunta. Le funzioni delle singole variabili aleatorie vengono chiamate funzioni di probabilità marginali. Date le funzioni di probabilità marginali non posso determinare la funzione congiunta. Pertanto $p_x(x) = \sum_{y \in R_y} P_{X,Y}(x, y)$ e $p_y(y) = \sum_{x \in R_x} P_{X,Y}(x, y)$.

9.3 Funzione di ripartizione congiunta

$F_{x,y}(x, y) = Pr(\{w \in \Omega|X(w) \leq x\} \cap \{w \in \Omega|Y(w) \leq y\}) = Pr(\{w \in \Omega|(X(w), Y(w)) \in]-\infty, x] \times]-\infty, y]\})$. $F_{X,Y}(x, y) = \sum p_{X,Y}(x, y)$.

9.4 Variabili aleatorie bivariate continue

9.4.1 Densità

Definizione

Una densità per una variabile aleatoria bivariata è una funzione $f(x, y)$ su $\mathbb{R}^2$ se e solo se $f(x, y) \geq 0 \ \ \forall (x, y) \in \mathbb{R}^2$ e $\int_{\mathbb{R}^2} f(x, y)dxdy = 1$ ($= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y)dxdy$, ovvero si integra prima per una variabile e poi per l'altra, considerandone una costante). La probabilità è data dall'area sottesa alla curva, pertanto $Pr((X, Y) \in C) = \int_C f(x, y)dxdy$, in particolare $Pr(\{X \leq x\} \cap \{Y \leq y\}) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u, v)dudv = F(x, y)$. Quest'ultima è la relazione tra la densità e la funzione di ripartizione bivariata continua. Si considerino ora due intervalli, uno sulla X e uno sulla Y e di voler andare a calcolare la probabilità del rettangolo costruito come prodotto cartesiano dei due intervalli, $Pr(\{a < X \leq b\} \cap \{c < Y \leq d\}) = \int_c^d \int_a^b f(x, y)dxdy$.

9.4.2 Relazione tra densità congiunta e marginali

$X \sim f_X(x)$ e $Y \sim f_Y(y)$ le due densità marginali e $(X, Y) \sim f_{X,Y}(x, y)$ la densità congiunta. Il passaggio dalla densità congiuta a quella marginale è: $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dy$ e $f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dx$.

Capitolo 10

Dipendenza tra variabili aleatorie

10.1 Discrete

10.1.1 Funzione di probabilità condizionata

$Pr(\{Y = y\}|\{X = x\}) = \frac{Pr(\{Y=y\}\cap\{X=x\})}{P(\{X=x\})}$, vera se $Pr(\{X = x\}) > 0$.

Definizione

$Pr(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)}$. Per ogni valore di x fissato questa descrive una funzione di probabilità in quanto non è negativa perchè rapporto tra probabilità, inoltre $\sum\limits_{y\in\mathbb{R}_Y} P_{Y|X}(y|x) = \sum\limits_{y\in\mathbb{R}_Y} \frac{P_{X,Y}(x,y)}{P_X(x)} = \frac{1}{P_X(x)} \sum\limits_{y\in\mathbb{R}_Y} P_{X,Y}(x,y) = \frac{P_X(x)}{P_X(x)} = 1$. Per "tornare indietro" necessito della funzione condizionata e la marginale condizionante.

Notazione

$(X, Y) \sim P_{X,Y}(x, y)$, si consideri $Y|X = x \sim P_{Y|X}(y|x)$ è una variabile aleatoria.

10.1.2 Indipendenza stocastica

Si deve considerare $\frac{P_{x,Y}(X,Y)}{P_x(x)} = P_y(y)$ se e solo se $P_{X,Y}(x, y) = P_X(x)P_Y(y) \ \ \forall(x, y)$.

10.1.3 Valore atteso condizionato

Si supponga la variabile aleatoria $Y \sim P_Y(y)$, il suo valore atteso è $\mathbb{E}(Y) = \sum\limits_{\mathbb{R}_Y} yPr(y)$, il valore atteso $\mathbb{E}(Y|X = x) = \sum\limits_{\mathbb{R}_Y} yP_{Y|X}(y|x)$.

10.2 Si ridica

$X, X \sim p_{X,Y}(x,y)$, si può definire $Y|X = x$, indicizzata su X, che ha la distribuzione $(Y|X = x) \sim p_{Y,X}(y|x)$, $\mathbb{E}(X|X = x) = \sum_{\mathbb{R}_Y} y p_{Y|X=x}(y|x)$, questa pertanto si può considerare come una trasformazione di una variabile aleatoria. $\mathbb{E}(g(x)) = \mathbb{E}_X(\mathbb{E}(Y|X))) = \mathbb{E}(Y)$, dove $g(x)$ è la trasformazione data dalla condizionata.

10.3 Varianza condizionata

$\mathbb{V}ar(Y|X = x) = h(x)$, pertanto $\mathbb{E}(h(X)) = \mathbb{E}_X(\mathbb{V}ar(Y|X))$.

10.3.1 Formula della decomposizione della varianza

$\mathbb{V}ar(Y) = \mathbb{E}_X(\mathbb{V}ar(Y|X) + \mathbb{V}ar_X(\mathbb{E}(Y|X))$.

Dimostrazione

La varianza di Y è $\mathbb{V}ar(Y) = \sum_{j=1}^{J}(y_j - \mathbb{E}(Y))^2 p_y(y_j) = \sum_{j=1}^{J}(y_j - \mathbb{E}(Y))^2 \sum_{i=1}^{I} p_{X,Y}(x_i, y_j) = \sum_{j=1}^{J}\sum_{i=1}^{I}(y_j - \mathbb{E}(Y))^2 p_{X,Y}(x_i, y_j) = \sum_{j=1}^{J}\sum_{i=1}^{I}(y_j - \mathbb{E}(Y|X = x_i) + \mathbb{E}(Y|X = x_i) - \mathbb{E}(Y))^2 p_{X,Y}(x_i, y_j) = \sum_{j=1}^{J}\sum_{i=1}^{I}(y_j - \mathbb{E}(Y|X = x_i))^2 p_{X,Y}(x_i, y_j) + \sum_{j=1}^{J}\sum_{i=1}^{I}(\mathbb{E}(Y|X = x_i) - \mathbb{E}(Y))^2 p_{X,Y}(x_i, y_j) + \sum_{j=1}^{J}\sum_{i=1}^{I} 2(y_j - \mathbb{E}(Y|X = x_i))(\mathbb{E}(Y|X = x_i) - \mathbb{E}(Y))p_{X,Y}(x_i, y_j)$.

Si consideri ora il terzo elemento $\sum_{i=1}^{I}(\mathbb{E}(Y|X = x_i) - \mathbb{E}(Y))[\sum_{j=1}^{J}(y_i - \mathbb{E}(Y|X = x_i))p_{Y|X}(y_j|x_i)]p_x(x_i) = 0$ in quanto momento primo centrato, si può riscrivere $p_{X,Y}(x,y) = p_{Y|X}(y|x)p_x(x)$. Ora pertanto si consideri il primo elemento $\sum_{i=1}^{I}[\sum_{j=1}^{J}(y_j - \mathbb{E}(Y|X = x_i))^2 p_{Y|X}(y_j, x_i)]p_x(x_i) = \sum_{i=1}^{I} \mathbb{V}ar(Y|X = x_i)p_x(x_i) = \mathbb{E}(\mathbb{V}ar(Y|X))$. Si consideri ora il secondo elemento $\sum_{i=1}^{I}[\sum_{j=1}(\mathbb{E}(Y|X = x_i) - \mathbb{E}(Y))^2 p_{Y|X}(y_j, x_i)]p_x(x_i) = \sum_{i=1}^{I}(\mathbb{E}(Y|X = x_i) - \mathbb{E}(Y))^2[\sum_{j=1} p_{Y|X}(y_j, x_i)]p_x(x_i) = \sum_{i=1}^{I}(\mathbb{E}(Y|X = x_i) - \mathbb{E}(Y))^2 p_x(x_i) = \mathbb{V}ar(\mathbb{E}(Y|X))$.

Si può definire un indice, chiamato indice rapporto di correlazione $0 \leq \eta^2_{Y|X} = \frac{\mathbb{V}ar(\mathbb{E}(Y|X))}{\mathbb{V}ar(Y)} \leq 1$, se l'indice vale zero è il caso peggiore di previsione.

La varianza tra i gruppi è la varianza del valore atteso della condizionata, ovvero quanto le medie condizionate cambiano al variare della X.