

Statistica e Analisi dei Dati

June 26, 2020

1 Statistica descrittiva

1.1 Descrizione di dati

Popolazione: insieme di tutti gli elementi che ci interessano

Campione: sottoinsieme (rappresentativo) della popolazione che viene studiato - Campione casuale semplice se i membri sono i scelti in modo tale che tutte le possibili scelte dei k membri siano equiprobabili

Frequenza assoluta (f): numero di occorrenza di un dato valore in un esperimento

Frequenza relativa: $\frac{f}{n}$

Tabelle e grafici: pochi valori distinti - Grafico a bastoncini - Grafico poligonale - Grafico a barre

Istogrammi: tanti dati, li suddivido in range distinti - Istogramma delle frequenze - Istogramma delle frequenze relative

1.2 Riassumere i dati

Media campionaria: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ - Traslazione: per $i = 1, \dots, n$ abbiamo

$$y_i = x_i + c$$

$$\overline{y} = \overline{x} + c$$

- Dilatazione: per $i = 1, \dots, n$ abbiamo

$$y_i = cx_i$$

$$\overline{y} = c\overline{x}$$

Media con frequenza: abbiamo k valori x_1, x_2, \dots, x_k con le relative frequenze f_1, f_2, \dots, f_k e numero di osservazioni totali pari a $n = \sum_{i=1}^k f_i$. Con queste premesse, posso calcolare la media campionaria come:

$$\bar{x} = \frac{x_1 + \dots + x_1 + x_2 + \dots + x_2 + \dots + x_k + \dots + x_k}{n} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_k x_k}{n} = \frac{f_1}{n} x_1 + \frac{f_2}{n} x_2 + \dots + \frac{f_k}{n} x_k$$

Ma le varie divisioni $\frac{f_i}{n}$ quindi possiamo vedere il tutto come:

$$\bar{x} = w_1 x_1 + w_2 x_2 + \dots + w_k x_k$$

Scarti: $x_i - \bar{x}$

La somma totale degli scarti è sempre zero:

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = n\bar{x} - n\bar{x} = 0$$

Mediana campionaria: la media campionaria è influenzata dai valori outlier. In questi casi si usa la mediana, definita come il valore intermedio quando i dati sono disposti in ordine crescente. La mediana è un indice di centralità robusto, in quanto non subisce perturbazioni dai valori outlier. Se il numero di valori è dispari allora la mediana campionaria è il valore intermedio della lista ordinata, se è pari è la media dei due valori intermedi

Moda campionaria: valore che si verifica con maggior frequenza nell'insieme di dati. Se non c'è un valore più frequente, tutti quelli con frequenza più alta sono detti valori modali

Varianza campionaria: le statistiche precedenti misurano il centro di un insieme di dati, la varianza misura la dispersione; per questo motivo si concentra sugli scarti

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Se traslassimo di una costante c tutti i valori x_i avremmo:

$$y_i = x_i + c$$

$$y_i - \bar{y} = x_i + c - (\bar{x} + c) = x_i - \bar{x}$$

Da ciò deduciamo che la varianza non è sensibile alla traslazione. Vediamo il caso con dilatazione:

$$y_i = cx_i$$

$$s_y^2 = c^2 s_x^2$$

Deviazione standard campionaria: radice quadrata positiva della varianza

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

E nel caso di dilatazione:

$$s_y = |c| s_x$$

1.3 Note sugli insiemi di dati

Insieme di dati normale: un insieme di dati è definito normale se: - Ha il punto di massimo nell'intervallo centrale - Spostandosi verso destra o sinistra dal centro l'altezza diminuisce (forma a campana) - Istogramma simmetrico rispetto all'intervallo centrale

Se l'istogramma è approssimativamente simmetrico intorno alla mediana campionaria viene detto approssimativamente normale. Se non rientra in questi due casi, l'istogramma è detto asimmetrico. L'insieme dei dati si dice "asimmetrico a destra" o "asimmetrico a sinistra" a seconda di quale sia la coda più lunga. L'istogramma potrebbe anche avere più di un massimo. Quando, in particolare, i massimi locali sono due, parleremo di istogramma bimodale.

Regola empirica: dato un insieme di dati approssimativamente normale con media campionaria \bar{x} e deviazione standard campionaria s allora valgono:

1. Approssimamente il 68% delle osservazioni rientrano nell'intervallo $x \pm s$
2. Approssimamente il 95% delle osservazioni rientrano nell'intervallo $x \pm 2s$
3. Approssimamente il 99.7% delle osservazioni rientrano nell'intervallo $x \pm 3s$

Coefficiente di correlazione campionaria: statistica atta alla misurazione dell'associazione tra i valori di un insieme di dati a coppie. Considero la somma:

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Quando grandi valori di x tendono ad essere associati a grandi valori di y , e piccoli valori di x tendono ad essere associati a piccoli valori di y , allora i segni positivi o negativi dei vari scarti tenderanno ad essere gli stessi. La sommatoria appena vista, divisa per $n-1$, prende il nome di **covarianza campionaria**.

$$Cov = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Il problema della covarianza è l'essere un indice dimensionale. Ci serve un indice che sia adimensionale, e per ottenerlo divido la covarianza per il prodotto delle deviazioni standard di x e di y .

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

Tale indice prende il nome di **coefficiente di correlazione campionaria**. Se $r > 0$ i dati sono correlati positivamente, quando $r < 0$ i dati sono correlati negativamente.

Proprietà:

1. r è sempre compreso fra -1 e +1
 - Posso inserirlo in una scala. Tanto più è vicino a 1 e tanto più la relazione tendenziale è forte, viceversa altrimenti
2. Se r è il coefficiente di correlazione campionaria di x_i e y_i , allora, scelti a , b , c e d a scelta, r sarà il coefficiente di correlazione campionaria anche per:

$$a + bx_i$$

$$c + dy_i$$

A condizione che b e d abbiano lo stesso segno ($bd \geq 0$). Controllare le dimostrazioni dagli appunti presi a lezione.

Formula alternativa per il calcolo di r :

$$r = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)(\sum_{i=1}^n y_i^2 - n\bar{y}^2)}}$$

Quantili: supponendo di avere x_1, x_2, \dots, x_n e q appartenente all'intervallo chiuso $[0, 1]$, il quantile q -esimo è:

- \geq di almeno nq osservazioni ordinate
- \leq di almeno $n(1 - q)$ osservazioni ordinate

La mediana campionaria è un caso particolare di una statistica nota come 100p-esimo percentile campionario, dove p indica qualunque frazione compresa tra 0 e 1. Un 100p-esimo percentile campionario è un valore maggiore del 100p percento dei valori e minore del 100(1-p) di essi. Il percentile p -esimo è uguale al quantile p -esimo/100. Per calcolare il quantile q -esimo è necessario:

1. Disporre i dati in ordine crescente
2. se nq non è intero, determina il più piccolo intero maggiore di nq . Il valore dei dati in questa posizione è il q -esimo quantile campionario

3. Se nq è intero, allora la media dei valori nelle posizioni nq e $nq + 1$ è il q -esimo quantile campionario

Box Plot: grafico rappresentante i valori in una scatola determinata dal primo e terzo quantile. La mediana ricade all'interno di questa scatola e identifica, in base alla sua posizione, la distribuzione dei dati. Il box è definito dal range interquantile, dato dalla differenza fra il III e il I quantile.

1.4 Eterogeneità/Omogeneità

Massima eterogeneità (minima omogeneità): gli elementi hanno tutti frequenza 1 **Massima omogeneità (minima eterogeneità):** contiene sempre la stessa forma ripetuta **Indice di Gini:** dati m elementi v_1, v_2, \dots, v_m di frequenza f_1, f_2, \dots, f_m , l'indice di Gini è definito come:

$$I = 1 - \sum_{j=1}^m f_j^2$$

$$\forall j \rightarrow f_j \geq 0 \quad \sum f_j = 1 \quad \exists j | f_j > 0$$

$$f_j^2 > 0 \quad \sum_j f_j^2 = f_j^2 + \sum_{j \neq \bar{j}} f_j^2 > 0$$

$$I = 1 - \sum_j f_j^2 < 1$$

$$0 \leq f_j \leq 1 \quad \forall j \rightarrow f_j^2 \leq f_j$$

$$\sum_j f_j^2 \leq \sum_j f_j = 1 \rightarrow I = 1 - \sum_j f_j^2 \geq 1 - 1 = 0$$

$$0 \leq I \leq 1$$

Nel caso di eterogeneità massima l'indice di Gini è: $I = \frac{m-1}{m}$

Nel caso di omogeneità massima l'indice di Gini è: $I = 0$

Entropia: $H = \sum_j f_j \log \frac{1}{f_j} \quad H \geq 0$

Nel caso di eterogeneità massima l'entropia è: $H = \log(m)$

Nel caso di omogeneità massima l'entropia è: $H = 0$

Trasformazioni lineari: fissate due costanti a e $b \in R$, il valore x verrà trasformato nel valore x' secondo la regola:

$$x' = g(x) = ax + b$$

- *Traslazione:* $x \rightarrow x' = x \pm k$
- *Dilatazione:* $x \rightarrow x' = \frac{x}{h}k$
 - $h > 1$ applica una contrazione, $h < 1$ applica una dilatazione
 - Media, mediana, quantili, range di variazione, distanza interquantile e deviazione standard vengono scalati di $\frac{1}{h}$; la varianza viene scalata di $\frac{1}{h^2}$
- *Cambiamento di scala e di origine:* se abbiamo dei valori nell'intervallo $[a, b]$ e vogliamo adattarli in modo che appartengano all'intervallo $[c, d]$, la trasformazione da applicare sarà:

$$x \rightarrow x' = c + \frac{d-c}{b-a}(x - a)$$
- *Standardizzazione:* applico una scala il cui fattore è uguale alla deviazione standard dei valori, per poi traslare verso sinistra rispetto alla media

$$x' = \frac{x - \bar{x}}{s_x}$$

Trasformazioni logaritmiche: quando i valori di una variabile osservata sono molto grandi oppure molto distanziati, conviene pensare a tale valore come potenza di una data base, ragionando in termini del relativo esponente. Ciò corrisponde ad applicare una trasformazione logaritmica del tipo:

$$x \rightarrow x' = \log x$$

Le scelte per la base del logaritmo sono tendenzialmente 10 e 2. Nel caso in cui i valori siano molto distanti e caratterizzati da una distribuzione di frequenza unimodale fortemente asimmetrica, la trasformazione logaritmica permette di ottenere una distribuzione di frequenza più simmetrica.

1.5 Analisi della varianza

Dato un insieme di osservazione di un dato attributo, possiamo dividerlo in G gruppi diversi. Indichiamo con n_1, n_2, \dots, n_G la numerosità dei vari gruppi, con $n = n_1 + n_2 + \dots + n_G$ a indicare il numero totale di osservazioni. Fissato $g \in 1, \dots, G$ e $i \in 1, \dots, n_G$, denotiamo con x_i^g il valore della i -esima osservazione nel gruppo g .

- Media campionaria: $\bar{x} = \frac{1}{n} \sum_{g=1}^G \sum_{i=1}^{n_g} x_i^g$
- Mediana campionaria: $\bar{x}^g = \frac{1}{n_g} \sum_{i=1}^{n_g} (x_i^g - \bar{x})^2 \quad \forall g = 1, \dots, n_G$
- Somma totale degli scarti: $SS_T = \sum_{g=1}^G \sum_{i=1}^{n_g} (x_i^g - \bar{x})^2$
- Somma degli scarti entro i gruppi (within groups): $SS_W = \sum_{g=1}^G \sum_{i=1}^{n_g} (x_i^g - \bar{x}^g)^2$
- Somma degli scarti tra i gruppi (between groups): $SS_B = \sum_{g=1}^G n_g (\bar{x}^g - \bar{x})^2$
- Varianza campionaria su tutte le osservazioni: $s_T^2 = \frac{1}{n-1} SS_T$
- Varianza campionaria della media tra i gruppi: $s_B^2 = \frac{1}{G-1} SS_B$
- Varianza campionaria dei valori entro i gruppi: $s_W^2 = \frac{1}{n-G} SS_W$

Si può mostrare che:

$$SS_T = SS_W + SS_B$$

Che possiamo vedere come:

$$\frac{SS_T}{n-1} = \frac{SS_W}{n-1} + \frac{SS_B}{n-1}$$

$$\frac{SS_T}{n-1} = \frac{n-G}{n-1} \frac{SS_W}{n-G} + \frac{G-1}{n-1} \frac{SS_B}{G-1}$$

$$s_T^2 = \frac{n-G}{n-1} s_W^2 + \frac{G-1}{n-1} s_B^2$$

La dimostrazione di questa uguaglianza è possibile visionarla durante le lezioni del professor Dario Malchiodi.

1.6 Calcolo combinatorio

Permutazioni: dato un insieme di n oggetti $A = \{a_1, a_2, \dots, a_n\}$, una permutazione di tali oggetti è una qualsiasi sequenza ordinata in cui compaiono tutti gli oggetti.

Se gli n oggetti di A sono tutti distinguibili si parla di permutazioni semplici, calcolabili come:

$$P_n = n \times (n-1) \times (n-2) \dots 2 \times 1 = n!$$

Permutazioni di oggetti distinguibili a gruppi: quando gli oggetti di A non sono tutti distinguibili ma sono distinguibili a gruppi di numerosità n_1, n_2, \dots, n_k (la cui somma ovviamente vale n) allora una sequenza ordinata di tali oggetti che sia distinguibile dalle altre è detta permutazione di oggetti distinguibili a gruppi. Formalizziamo:

$$P_{n;n_1, \dots, n_k} = \frac{n!}{n_1! n_2! \dots n_k!} = \binom{n}{n_1! n_2! \dots n_k!}$$

Tale coefficiente è chiamato *coefficiente multinomiale*.

Disposizioni e combinazioni: cpmsoderoa, n oggetti distinti $A = \{a_1, a_2, \dots, a_n\}$ e selezioniamo k oggetti di questo insieme.

- Se vogliamo distinguere le configurazioni contenenti gli stessi oggetti ma estratti in ordine differente, allora parliamo di disposizioni di n oggetti su k posti, dove sono importanti sia l'oggetto selezionato che la sua posizione
- Se siamo interessati a quali oggetti sono stati estratti e non alla loro posizione nella sequenza, parliamo di combinazioni di n oggetti presi k alla volta

Parliamo di disposizioni o combinazioni senza ripetizione se gli oggetti di A possono essere usati una sola volta. Se il singolo oggetto può essere selezionato anche più di una volta, allora parliamo di disposizioni o combinazioni con ripetizione.

Disposizioni senza ripetizione: $d_{n,k} = n(n-1)(n-2)\dots(n-k+1) = n(n-1)(n-2)\dots(n-k+1) \binom{(n-k)(n-k-1)\dots 1}{(n-k)(n-k-1)\dots 1} = \frac{n!}{(n-k)!}$

Combinazioni senza ripetizione: $c_{n,k} = \left(\frac{d_{n,k}}{k}\right)! = \frac{n!}{(n-k)!k!} = \binom{n}{k}$

Disposizioni con ripetizione: $D_{n,k} = n \times n \dots \times n = n^k$

Combinazioni con ripetizione: $C_{n,k} = \binom{n+k-1}{k}$

1.7 Teoria degli Insiemi

Identifichiamo con Ω l'*insieme Universo*, ossia l'insieme che contiene tutti gli elementi possibili. Diremo che $w \in \Omega$ quando l'elemento w appartiene all'insieme Ω . Definiamo $E \subseteq \Omega$ quando E è un sottoinsieme dell'insieme Ω .

Identifichiamo infine con \emptyset l'insieme vuoto e con $\{\}$ l'insieme privo di elementi.

Unione: $E \cup F \quad x \in E \cup F \leftrightarrow x \in E \vee x \in F$

Intersezione: $E \cap F \quad x \in E \cap F \leftrightarrow x \in E \wedge x \in F$

Disgiunzione: $E \cap F = \emptyset$

Complemento di E: E^C oppure \bar{E}

Notazioni particolari:

- $\bigcap_{i=1}^n A_i = A_1 \cap A_2 \dots \cap A_n$
- $\bigcup_{i=1}^n A_i = A_1 \cup A_2 \dots \cup A_n$

Proprietà

- Commutative: $E \cap F = F \cap E \quad E \cup F = F \cup E$

- Associative: $(E \cap F) \cap G = E \cap (F \cap G)$
- Distributive: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$; $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$
- De Morgan: $\overline{E \cup F} = \overline{E} \cap \overline{F}$; $\overline{E \cap F} = \overline{E} \cup \overline{F}$

1.8 Elementi di probabilità

Interpretazione frequentista: la probabilità di un esito è considerata una proprietà dell'esito stesso. In tal caso, essa viene calcolata come rapporto fra il numero di casi favorevoli e il numero di casi possibili

Interpretazione soggettivistica: la probabilità di un esito non è una proprietà oggettiva, bensì la precisazione del livello di fiducia che lo studioso ripone nel verificarsi di un evento.

L'insieme universo Ω viene spesso chiamato anche *Spazio Campionario* oppure *Spazio degli eventi*: in entrambi i casi essi identificano l'insieme di tutti gli esiti di un esperimento casuale. Definiamo *Evento Elementare* un elemento w appartenente all'universo Ω . Definiamo banalmente *Evento* un generico sottoinsieme dello spazio degli eventi: $E \subset \Omega$. L'evento può essere interpretato come un sottoinsieme contenente eventi elementari. Gli insiemi composti da un solo elemento sono detti *insiemi singoletto*. Tra tutti gli insiemi ve ne sono due di menzione particolare:

- $\{\}$: l'insieme di partenza, si verifica sempre
- \emptyset : l'insieme vuoto, non si verifica mai

Proprietà:

- Disgiunzione logica: $E_1 \cup E_2 \dots \cup E_n$ Evento che si verifica quando si verifica almeno uno degli eventi che ho unito
- Congiunzione logica: $E_1 \cap E_2$ Evento che si verifica quando sia E che F si sono verificati
- Evento certo: $E \subseteq \Omega$ Anche l'universo Ω è un evento e contiene tutti gli esiti possibili. Ω è l'evento che si verifica sempre
- Evento impossibile: \emptyset Qualunque esito dell'esperimento io analizzi, non sarà mai nell'insieme vuoto, quindi in questo caso si parla di evento impossibile
- $\overline{E} = \Omega \setminus E$ $\overline{\Omega} = \emptyset \rightarrow \overline{\emptyset} = \Omega$
- $E \subseteq F$ E implica F
- $E \subseteq F$ e $F \subseteq E \rightarrow E = F$

Algebra degli eventi: definiamo Algebra degli eventi l'insieme:

$$A = \{E_i \subseteq \Omega\}$$

Per essere tale, è necessario che vengano soddisfatte le seguenti proprietà:

1. $\Omega \in A$
2. $\forall E$ se $E \in A$ allora $\overline{E} \in A$
3. $\forall E, F$ se $E \in A \wedge F \in A$ allora $E \cup F \in A$
4. L'algebra è chiusa rispetto all'unione infinita. Per questo motivo A è più di un'algebra e viene detta σ -algebra

Funzione di probabilità: $P : A \rightarrow [0,1]$

Dominio: generica algebra A

Codominio: tutti i valori dell'intervallo chiuso $[0,1]$

Assiomi di Kolmogorov

1. $P(\Omega) = 1$ posso calcolare tale probabilità perché $\Omega \in A$ per ipotesi
2. $\forall E \in A \quad 0 \leq P(E) \leq 1$ probabilità intesa come frequenza relativa con cui a lungo termine si verifica un determinato evento
3. $\forall E_1, E_2 \in A$ se $E_1 \cap E_2 = \emptyset$ diremo che E_1 e E_2 sono eventi disgiunti o *mutuamente esclusivi* e in tal caso vale che:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

4. $\forall E_1, \dots, E_n \in A \quad \forall i \neq j \quad E_i \cap E_j = \emptyset$

$$P(\bigcup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i)$$

Teoremi elementari di probabilità

1. $\forall E \in A \quad P(\overline{E}) = 1 - P(E)$

Dimostrazione

$$1 = P(\Omega) = P(E \cup \overline{E}) = P(E) + P(\overline{E}) \rightarrow P(\overline{E}) = 1 - P(E)$$

2. $\forall E, F \in A \quad P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Spazi equiprobabili

$(\omega A, P) =$ Spazio di probabilità

Supponiamo Ω composto da N eventi: $\Omega = \{1, 2, \dots, N\}$; se tutti gli eventi hanno la stessa probabilità, parliamo di spazi equiprobabili.

$$\forall w \in \Omega \quad P(\{w\}) = p$$

$$\Omega = \{1\} \cup \{2\} \dots \cup \{N\} \quad 1 = P(\Omega) = P(\{1\}) + \dots + P(\{N\}) = Np \rightarrow p = \frac{1}{N}$$

N.B. E' necessario che Ω sia finito, poiché altrimenti N tenderebbe a infinito e, di conseguenza, p tenderebbe a 0. Se consideriamo ora un sottoinsieme di Ω , abbiamo:

$$\forall E \subseteq \Omega \quad E = \{e_1, \dots, e_k\} \quad k \leq N$$

$$E = \{e_1\} \cup \{e_2\} \dots \cup \{e_k\} = kp = \frac{k}{N} = \frac{|E|}{N}$$

Probabilità condizionata

$$P(E|F) = \frac{P(E \cap F)}{P(F)} \quad \text{vale se } P(F) \neq 0$$

Teorema delle probabilità totali

Supponiamo di avere due eventi $E, F \in A$ con:

$$(E \cap F) \cup (E \cap \overline{F}) = E$$

$$P(E) = P(E|F)P(F) + P(E|\overline{F})P(\overline{F}) = P(E|F)P(F) + P(E|\overline{F})(1 - P(F))$$

Partizionamento

Supponiamo di avere l'insieme Ω partizionato in n sezioni di stessa grandezza: F_1, F_2, \dots, F_n . L'unione di queste partizioni costruisce lo spazio degli eventi. Inoltre: $\forall i \neq j \rightarrow F_i \cap F_j = \emptyset$

Supponiamo di analizzare un evento E e di osservare le intersezioni fra tale evento e tutte le partizioni dello spazio degli eventi: $E \cap F_1, E \cap F_2, \dots, E \cap F_n$

$$\bigcup_{i=1}^n (E \cap F_i) = E \cap (\bigcup_{i=1}^n F_i) = E \cap \Omega = E$$

$$\forall i \neq j \rightarrow (E \cap F_i) \cap (E \cap F_j) = (E \cap E) \cap (F_i \cap F_j) = E \cap \emptyset = \emptyset$$

Dall'unione di quanto appena enunciato, posso dire che:

$$E = E \cap (\bigcup_{i=1}^n F_i) \rightarrow P(E) = \sum_{i=1}^n P(E \cap F_i) = \sum_{i=1}^n P(E|F_i)P(F_i)$$

Teorema di Bayes

Immaginiamo di avere n eventi che soddisfino le ipotesi del problema delle probabilità totali:

$$F_1, F_2, \dots, F_n$$

$$P(F_j|E)P(E) = P(F_j \cap E) = P(E \cap F_j) = P(E|F_j)P(F_j)$$

$$P(F_j|E) = \frac{P(E|F_j)P(F_j)}{P(E)} = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^n P(E|F_i)P(F_i)}$$

Eventi indipendenti

Nel caso in cui $P(E|F)$ e $P(E)$ siano uguali, diciamo che E ed F sono indipendenti. Quindi, se:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Allora, possiamo formalizzare l'indipendenza di eventi dicendo che:

$$P(E \cap F) = P(E)P(F)$$

Una proprietà da notificare è che se E ed F sono indipendenti, lo sono anche E e \bar{F}

Supponiamo ora di avere più eventi: E , F e G . Tali eventi si dicono indipendenti se:

1. $P(E \cap F) = P(E)P(F)$
2. $P(E \cap G) = P(E)P(G)$
3. $P(F \cap G) = P(F)P(G)$
4. $P(E \cap F \cap G) = P(E)P(F)P(G)$

Generalizzando, presi n eventi diversi, essi si dicono indipendenti se e solo se per ogni sottogruppo $E_{\alpha 1}, \dots, E_{\alpha n}$, vale l'equazione:

$$P(\bigcap_{i=1}^r E_{\alpha i}) = \prod_{i=1}^r P(E_{\alpha i})$$

Se tre o più eventi sono indipendenti, allora ciascuno di essi è indipendente da qualunque evento si possa costruire con gli altri due.

$$P(E \cap (F \cup G)) = P(E \cap F) \cup P(E \cap G) = P(E \cap F) + P(E \cap G) - P(E \cap F \cap G) = P(E)P(F) + P(E)P(G) - P(E)P(F \cap G) = P(E)(P(F) + P(G) - P(F \cap G)) = P(E)P(F \cup G)$$

Componenti in serie e in parallelo

Immaginiamo un sistema di n scatole poste in serie. Tali scatole conducono elettricità, quindi è necessario che siano tutte funzionanti affinché il sistema funzioni correttamente:

$$P(\text{Scatola}_i = 1) = p_i$$

Ogni scatola è indipendente dalle altre, quindi:

$$P(\text{sistemaFunziona}) = P(\bigcap_{i=1}^n \text{Scatola}_i = 1) = \prod_{i=1}^n P(\text{Scatola}_i = 1) = \prod_{i=1}^n p_i$$

Ora immaginiamo di prendere quelle scatole e di porle in parallelo. Così facendo, il sistema è più robusto ai danni, poiché anche se una componente dovesse rompersi, ne avrei ancora $n - 1$ su cui fare affidamento. Fintanto che c'è almeno un componente funzionante, il sistema funzionerà, quindi:

$$P(\text{sistemaFunziona}) = 1 - P(\text{sistemaRotto}) = 1 - P(\bigcap_{i=1}^n C_i = 0) = 1 - \prod_{i=1}^n P(C_i = 0) = 1 - \prod_{i=1}^n (1 - p_i)$$

Naive-Bayes

Supponiamo di avere una situazione ove vengono associati dei particolari valori a dei determinati attributi (e.g. occhi = attributo; nero = valore). Indichiamo:

- Attributi: X_1, X_2, \dots, X_n
- Valori: $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$

Ora, indichiamo con $Y = y_k$ il generico k -esimo elemento di un insieme di cui ci interessa studiarne gli attributi. Abbiamo:

$$P(Y = y_k | X_1 = x_1, \dots, X_n = x_n) \rightarrow \text{Applichiamo Bayes} \rightarrow \frac{P(X_1=x_1, \dots, X_n=x_n | Y=y_k) P(Y=y_k)}{P(X_1=x_1, \dots, X_n=x_n)}$$

Da cui:

$$\frac{P(X_1=x_1 | Y=y_k) \dots P(X_n=x_n | Y=y_k) P(Y=y_k)}{P(X_1=x_1, \dots, X_n=x_n)}$$

Questo rapporto è proporzionale a:

$$\prod_{i=1}^n P(X_i = x_i | Y = y_k) P(Y = y_k)$$

Chiamiamo *classificatore Naive-Bayes* il seguente valore:

$$k^* = \operatorname{argmax}_k \prod_{i=1}^n P(X_i = x_i | Y = y_k) P(Y = y_k)$$

Variabili aleatorie

Definiamo *variabile aleatoria* una qualsiasi variabile:

$$X : \Omega \rightarrow R$$

Scrivendo $X = \alpha$ indichiamo l'evento $X = \alpha$, che formalizzando sarebbe:

$$\{X = \alpha\} = \{w \in \Omega : X(w) = \alpha\}$$

Per semplicità, anziché scrivere $P(\{X = \alpha\})$ scriveremo $P(X = \alpha)$.

Funzione indicatrice

Definisco *funzione indicatrice* la funzione:

$$I = \begin{cases} 1 \\ 0 \end{cases}$$

Assume valore 1 quando si verifica un determinato evento, 0 altrimenti.

Funzione di ripartizione

Prendiamo una variabile aleatoria X e definiamo:

$$F_X : R \rightarrow [0, 1]$$

$$F_X(x) = P(X \leq x)$$

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\} \text{ per } a < b$$

Da cui deriviamo:

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b)$$

Applicando la sostituzione otteniamo:

$$F_X(b) = F_X(a) + P(a < X \leq b)$$

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

Variabili aleatorie discrete

Le variabili aleatorie discrete assumono un insieme numerabile di specificazioni. Introduciamo la *funzione di massa di probabilità* definendola come:

$$p_X : R \rightarrow [0, 1]$$

$$\forall x \in R \quad p_X(x) = P(X = x)$$

Proprietà:

1. $p_X \neq 0$ in un insieme numerabile
2. $p_X \geq 0$
3. $\sum_i p_X(x_i) = 1$

Possiamo definire la *funzione di ripartizione* in funzione di quella di massa di probabilità (ed è possibile fare anche l'opposto).

$$F_X(x) = P(X \leq x) = \sum_{a \leq x} P(X = a) = \sum_{a \leq x} p_X(a)$$

Proprietà

1. $F_X(x) \geq 0 \quad \forall x \in R$
2. $\lim_{x \rightarrow +\infty} F_X(x) = 1 \quad \lim_{x \rightarrow +\infty} P(X \leq x)$
3. F_X è continua da destra

Valore atteso

Dato la variabile aleatoria discreta X su $\{x_1, x_2, \dots, x_n\}$ e p_X come funzione di massa di probabilità, definisco valore atteso la quantità:

$$E(X) = \sum_i x_i p(x_i) = \sum_i x_i P(X = x_i)$$

Come possiamo vedere dalla formula soprastante, il valore atteso è una quantità dimensionale, che assume la stessa dimensione di x .

Consideriamo ora:

$$g(X) = ax + b \text{ con } a, b \in R \quad X \rightarrow g(X) = aX + b$$

$$E(g(X)) = \sum_i (ax_i + b)P(X = x_i) = \sum_i ax_i P(X = x_i) + \sum_i bx_i P(X = x_i) = a \sum_i x_i P(X = x_i) + b \sum_i x_i P(X = x_i) = aE(X) + b$$

Quindi:

$$E(aX + b) = aE(X) + b$$

1. Se $a = 0 \rightarrow E(b) = b$
2. Se $b = 0 \rightarrow E(aX) = aE(X)$

Varianza

Considero la variabile aleatoria X e il suo valore atteso $E(X) = \mu$. Se considero:

$$|X - \mu|$$

difatti sto operando con una nuova variabile aleatoria, di cui posso calcolare il valore atteso:

$$E(|X - \mu|)$$

Ora, per togliere il valore assoluto ed operare esclusivamente con valori positivi, elevo al quadrato, ottenendo:

$$E((X - \mu)^2)$$

Questo nuovo valore prende il nome di *varianza* e viene indicata con σ_X^2

$$Var(X) = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2 = E(X^2) - E(X)^2$$

Proprietà:

1. $Var(aX + b) = E((aX + b - E(aX + b))^2) = E((aX + b - a\mu - b)^2) = E(a^2(X - \mu)^2) = a^2 E((X - \mu)^2) = a^2 Var(X)$
2. $Var(aX + b) \rightarrow \text{se } a = 0 \rightarrow Var(b) = 0$

La varianza ha la stessa unità di misura della variabile aleatoria elevata al quadrato. Se vogliamo una misura adimensionale utilizziamo la *deviazione standard*, calcolata come:

$$\sigma_x = \sqrt{Var(X)}$$

Variabile aleatoria multivariata

Ragioniamo in termini di coppie di variabili aleatorie: X, Y .

- *Funzione di ripartizione congiunta:* $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$
 - Nel caso in cui uno dei due valori tendesse a infinito, avremo: $\lim_{y \rightarrow +\infty} F_{X,Y}(x, y) = \lim_{y \rightarrow +\infty} P(X \leq x, Y \leq y) = P(X \leq x) = F_X(x)$
 - * Tale valore è detto *funzione di ripartizione marginale*
- *Funzione di massa di probabilità congiunta:* $p_{X,Y}(x_i, y_j) = P(X = x_i, Y = y_j)$
 - Se fisso una delle due variabili e muovo l'altra: $\sum_j P(X = x_i, Y = y_j) = P(X = x_i) \rightarrow \sum_j p_{X,Y}(x_i, y_j) = p_X(x_i)$
 - * Tale valore è detto *funzione di massa di probabilità marginale*

Diciamo che due variabili aleatorie X, Y sono indipendenti se e solo se:

$$\forall A, B \subseteq R \rightarrow P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Questa formula vale anche nei seguenti casi:

1. $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$
2. $F_{X,Y}(x, y) = F_X(x)F_Y(y)$
3. $p_{X,Y}(x, y) = p_X(x)p_Y(y)$

La dimostrazione di qui enunciato è possibile visionarla durante le lezioni del professor Dario Malchiodi.

Valore atteso di funzione di più variabili

Supponiamo di avere due variabili aleatorie X, Y e una funzione $g : R^2 \rightarrow R$

Avremo:

$$E(g(X, Y)) = \sum_{x,y} g(x, y)P(X = x, Y = y) = \sum_{x,y} g(x, y)p_{X,Y}(x, y)$$

Se prendessi $g(X, Y) = X + Y$ avrei $E(X + Y) = E(X) + E(Y)$

La dimostrazione di qui enunciato è possibile visionarla durante le lezioni del professor Dario Malchiodi.

Per un generico n numero di variabili aleatorie avrei:

$$E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i)$$

Covarianza

Date su variabili aleatorie X, Y e ponendo $\mu_x = E(X)$ e $\mu_y = E(Y)$, definisco *covarianza* la quantità:

$$Cov(X, Y) = E((X - \mu_x)(Y - \mu_y))$$

Proprietà:

1. $Cov(X, Y) = Cov(Y, X)$
2. $Cov(X, Y) = E(XY - \mu_x Y - \mu_y X + \mu_x \mu_y) = E(XY) - E(X)E(Y)$
3. $Cov(aX, Y) = E(aXY) - E(aX)E(Y) = aCov(X, Y)$
4. $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z) \rightarrow Cov(\sum_i X_i, Z) = \sum_i Cov(X_i, Z)$
5. $Cov(\sum_i \alpha_i X_i, Z) = \sum_i \alpha_i Cov(X_i, Z)$
6. $Cov(\sum_i \alpha_i X_i, \sum_j \beta_j Y_j) = \sum_i \sum_j \alpha_i \beta_j Cov(X_i, Y_j)$
7. $Cov(X, X) = Var(X)$
8. $Cov(X + b, Y) = Cov(X, Y)$
9. $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
10. $Var(X + X) = Var(X) + Var(X) + 2Cov(X, X) = 4Var(X)$
11. $Var(\sum_i X_i) = \sum_{i,j} Cov(X_i, Y_j)$

Nel caso in cui le variabili aleatorie X, Y fossero indipendenti, avremmo che il valore atteso del loro prodotto è il prodotto dei valori attesi: $E(XY) = E(X)E(Y)$. Ma allora, la covarianza di due variabili aleatorie indipendenti è 0.

$$Cov(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$$

E la varianza, di conseguenza, diventa:

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y) = Var(X) + Var(Y)$$

Generalizzando:

$$Var(\sum_i X_i) = \sum_i Cov(X_i)$$

Correlazione

Consideriamo $A, B \subseteq \Omega$ con $X = I_A, E(X) = P(X = 1)$ e $Y = I_B, E(Y) = P(Y = 1)$. Allora:

$XY = 1$ se $X = 1 \wedge Y = 1$ $XY = 0$ altrimenti

Il valore atteso delle due variabili sarà:

$$E(XY) = P(X = 1, Y = 1)$$

Mentre la covarianza:

$$Cov(X, Y) = E(XY) - E(X)E(Y) = P(X = 1, Y = 1) - P(X = 1)P(Y = 1)$$

$$Cov(X, Y) > 0 \leftrightarrow P(X = 1, Y = 1) > P(X = 1)P(Y = 1)$$

$$\frac{P(X=1, Y=1)}{P(Y=1)} > P(X = 1) \leftrightarrow P(X = 1|Y = 1) > P(X = 1)$$

Da ciò arriviamo all'*indice di correlazione lineare*:

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

Proprietà:

1. $\rho_{X,Y} \in [-1, 1]$
2. Se X, Y sono variabili aleatorie indipendenti, allora $Cov = 0 \rightarrow \rho_{X,Y} = 0$
3. Se ρ è -1 oppure 1 significa che tra X e Y c'è una dipendenza di tipo deterministico e, quindi, lineare
4. $\rho_{2X, 2Y} = \frac{Cov(2X, 2Y)}{\sigma_{2X} \sigma_{2Y}} = \rho_{X,Y}$

Variabili aleatorie continue

Una variabile aleatoria X è detta continua se $f_X : R \rightarrow R^+$ e se $\forall B \subseteq R$ e $P(X \in B) = \int_B f_X(x)dx$. Definiamo infine *densità di probabilità* la f_X .

Quindi abbiamo:

1. $P(a \leq X \leq b) = P(X \in [a, b]) = \int_a^b f_X(x)dx$
2. $P(X \in R) = \int_{-\infty}^{+\infty} f_X(x)dx$
3. $P(X = \alpha) = \int_{\alpha}^{\alpha} f_X(x)dx = 0$
4. $F_X(a) = P(X \leq a) = \int_{-\infty}^a f_X(x)dx$
5. $P(a - \frac{\epsilon}{2} \leq X \leq a + \frac{\epsilon}{2}) = \int_{a-\frac{\epsilon}{2}}^{a+\frac{\epsilon}{2}} f_X(x)dx \approx \epsilon f_X(a)$

Valore atteso

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)dx$$

$$Var(X) = E((X - \mu_X)^2)$$

La funzione di ripartizione di due variabili aleatorie X, Y è uguale al caso discreto, da cui, se $X \geq 0$:

$$E(X) = \int_0^{+\infty} 1 - F_X(x)dx$$

Disuguaglianza di Markov

Supponendo $X \geq 0$, allora $\forall a > 0$ abbiamo:

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Questa disuguaglianza vale sia nel caso discreto che continuo. La dimostrazione nel continuo è possibile studiarla attraverso le lezioni del professor Dario Malchiodi. Cosa accade nel caso in cui abbiamo $P(X < a)$:

$$P(X < a) = 1 - P(X \geq a) \geq 1 - \frac{E(X)}{a}$$

Disuguaglianza di Chebyshev

Presa una variabile aleatoria X di valore atteso $E(X) = \mu$ e $Var(X) = \sigma^2$, allora $\forall r > 0$:

$$P(|X - \mu| \geq r) \leq \frac{\sigma^2}{r^2}$$

Vale sia per variabili aleatorie discrete che continue. Cosa accade nel caso in cui $P(|X - \mu| < r)$?

$$P(|X - \mu| < r) = 1 - P(|X - \mu| \geq r) \geq 1 - \frac{\sigma^2}{r^2}$$

L'importanza di queste due disequazioni risiede nel fatto che ci consentono di avere stime della probabilità a partire dal valore atteso o dal valore atteso e dalla varianza.

1.9 Modellizzazione discreta

Abbiamo bisogno di astrarre esperimenti casuali simili che si presentano in spoglie diverse. Partiamo analizzando gli esperimenti bernoulliani.

Modello di Bernoulli

Un esperimento bernoulliano è caratterizzato da una variabile aleatoria di bernoulli, parametrizzata rispetto alla probabilità di successo. Tale variabile aleatoria assume quindi due valori:

$X = 1$ se l'esito è positivo $X = 0$ se l'esito è negativo

Definiamo distribuzione bernoulliana, la distribuzione identificata da:

$$X \sim B(p)$$

da leggersi “la variabile aleatoria X è distribuita come $B(p)$ ”.

Funzione di massa di probabilità

$$p_X(0) = P(X = 0) = 1 - p \quad p_X(1) = P(X = 1) = p$$

$$p_X(x) = 0 \quad \forall x \neq 0 \wedge x \neq 1$$

Generalizzando:

$$p_X(x) = p^x(1 - p)^{1-x} I_{\{0,1\}}(x)$$

Funzione di ripartizione

$$F_X(x) = P(X \leq x)$$

Considerando che X può assumere solo valori pari a 0 o 1, abbiamo:

$$F_X(0) = P(X \leq 0) = P(X = 0) = 1 - p \quad F_X(1) = P(X \leq 1) = 1$$

Generalizzando:

$$F_X(x) = 0 I_{R^-}(x) + (1 - p) I_{[0,1)}(x) + 1 I_{[1,+\infty)}(x) = (1 - p) I_{[0,1)}(x) + 1 I_{[1,+\infty)}(x)$$

Valore atteso

$$E(X) = \sum_x x P(X = x) = 0 P(X = 0) + 1 P(X = 1) = P(X = 1) = p$$

Varianza

$$\begin{aligned} \text{Var}(X) &= E((X - \mu)^2) = E((X - p)^2) = \sum_x (x - p)^2 P(X = x) = (0 - p)^2 P(X = 0) + (1 - p)^2 P(X = 1) \\ &= p^2(1 - p) + (1 - p)^2 p = p(1 - p)(p + 1 - p) = p(1 - p) \end{aligned}$$

La varianzasi massimizza per $p = \frac{1}{2}$

Vediamo alcuni casi estremi:

$$p = 1 \rightarrow E(X) = 1 \quad \text{Var}(X) = 0$$

$$p = 0 \rightarrow E(X) = 0 \quad \text{Var}(X) = 0$$

Per valori intermedi abbiamo che il valore medio e p crescono proporzionalmente.

Modello binomiale

Supponiamo di eseguire un esperimento bernoulliano per n volte. Otterremo una nuova variabile aleatoria chiamata *variabile aleatoria binomiale*. Essa ha due parametri:

- La variabile aleatoria su cui si basa
- Il numero di ripetizioni dell'esperimento bernoulliano

Formalizziamo:

$$X \sim B(n, p)$$

Con $n \in \mathbb{N}$ e $p \in [0, 1]$. Il dominio di X sarà $D_X = \{0, \dots, n\}$

Funzione di massa di probabilità

$$p_X(i) = P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i} I_{\{0, \dots, n\}}(i)$$

Funzione di ripartizione

$$\begin{aligned} F_X(x) &= P(X \leq x) = \sum_{i \leq x} P(X = i) = \sum_{i=0}^{\lfloor x \rfloor} P(X = i) = \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1 - p)^{n-i} = \\ &= \left(\sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1 - p)^{n-i} \right) I_{[0, n]}(x) + I_{[n, +\infty)}(x) \end{aligned}$$

Valore atteso

Possiamo vedere la nostra distribuzione $X \sim B(n, p)$ come $X_1, \dots, X_n \sim B(p)$, da cui:

$$X = \sum_{i=1}^n X_i$$

Quindi:

$$E(X) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p = np$$

Quindi il valore atteso cresce linearmente al numero di prove effettuate.

Varianza

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^n X_i\right) \rightarrow \text{assumendo } X_i \text{ indipendenti} \rightarrow \sum_{i=1}^n \text{Var}(X_i) = \sum_{i=1}^n p(1 - p) = np(1 - p)$$

Più aumenta n e più aumenta l'intervallo possibile di valori su cui può variare.

Cosa accade se cerchiamo di sommare due variabili aleatorie X_1 e X_2 che seguono il modello binomiale?

$$X_1 \sim B(n, p)$$

$$X_2 \sim B(m, p)$$

Abbiamo che:

$$X_1 = \sum_{i=1}^n X_{1,i}$$

$$X_2 = \sum_{i=1}^m X_{2,i}$$

Quindi:

$$X_1 + X_2 = \sum_{i=1}^{n+m} X_i \sim B(n + m, p)$$

Modello uniforme discreto

Supponiamo ora di non essere più nel caso bernoulliano. Abbiamo uno spazio degli eventi equiprobabili: cerchiamo di modellarlo con le variabili aleatorie.

$n \in N \setminus \{0\}$ è il numero di esiti equiprobabili

$$X \sim U(n)$$

$$P(X = 1) = P(X = 2) = \dots = P(X = n) = \frac{1}{n}$$

Funzione di massa di probabilità

$$P(X = x) = \frac{1}{n} I_{\{1, \dots, n\}}(x) = p_X(x)$$

Funzione di ripartizione

$$F_X(x) = P(X \leq x) = \sum_{i \leq x} p_X(i) = \sum_1^{\lfloor x \rfloor} \frac{1}{n} = \frac{\lfloor x \rfloor}{n} I_{[1, n]}(x) + I_{(n, +\infty)}(x)$$

Valore atteso

$$E(X) = \sum_{x=1}^n x P(X = x) = \sum_{x=1}^n x \frac{1}{n} = \frac{1}{n} \sum_{x=1}^n x = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}$$

Varianza

$$E(X^2) = \sum_{x=1}^n x^2 \frac{1}{n} = \frac{1}{n} \sum_{x=1}^n x^2 = \frac{1}{n} \frac{(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{n^2-1}{12}$$

La dispersione intorno al valore centrale aumenta all'aumentare del numero di eventi equiprobabili.

Modello geometrico

La *distribuzione geometrica* descrive il numero di insuccessi necessari affinché si verifichi il primo successo in una successione di esperimenti bernoulliani indipendenti e identicamente distribuiti. Questa distribuzione, che ha quindi come supporto l'insieme dei numeri naturali (zero incluso), è quindi completamente descritta specificando il parametro p del corrispondente esperimento bernoulliano. Più precisamente, $p \in (0, 1]$: il caso $p = 0$ va infatti escluso a priori, altrimenti sarebbe impossibile avere come esito un successo (e dunque sarebbe impossibile contare il numero di insuccessi prima che si verifichi un successo). Il caso $p = 1$ può essere incluso nell'insieme dei valori validi per il parametro, sebbene questa scelta identifichi un esperimento bernoulliano che ha sempre successo: in tal caso, la variabile aleatoria che conteggia il numero di insuccessi prima del primo successo degenera nel valore costante pari a zero.

Abbiamo:

$X = x \rightarrow x$ insuccessi prima del primo successo durante le ripetizioni indipendenti di un esperimento di Bernoulli

$X \sim G(p) \rightarrow p$ indica la probabilità di successo dell'esperimento bernoulliano

Funzione di massa di probabilità

$$p_X(i) = p(1-p)^i I_{N \cup \{0\}}(i)$$

Vediamo che il supporto è infinito, in quanto non so mai quando si verifica il primo successo. La variabile i indica gli insuccessi.

$$\sum_{i=0}^{+\infty} p_X(i) = \sum_{i=0}^{+\infty} p(1-p)^i = p \sum_{i=0}^{+\infty} (1-p)^i = p \frac{1}{1-(1-p)} = 1$$

Vediamo ora un breve lemma che ci concederà di calcolare più agevolmente il valore atteso.

Per ogni $\alpha \in (-1, 1)$

$$\sum_{i=0}^{+\infty} i\alpha^i = \frac{\alpha}{(1-\alpha)^2}$$

Dimostrazione.

$$\sum_{i=0}^{+\infty} i\alpha^i = \alpha \sum_{i=0}^{+\infty} i\alpha^{i-1} = \alpha \sum_{i=0}^{+\infty} \frac{d}{d\alpha} \alpha^i = \alpha \frac{d}{d\alpha} \sum_{i=0}^{+\infty} \alpha^i = \alpha \frac{d}{d\alpha} \frac{1}{1-\alpha} = \alpha \frac{1}{(1-\alpha)^2}$$

Valore atteso

$$E(X) = \sum_{i=0}^{+\infty} ip_X(i) = \sum_{i=0}^{+\infty} ip(1-p)^i = p \sum_{i=0}^{+\infty} i(1-p)^i$$

e quindi per il lemma appena dimostrato si ha, ponendo $\alpha = 1-p$,

$$E(X) = p \frac{1-p}{p^2} = \frac{1-p}{p}$$

Varianza

$$E(X^2) = \sum_{i=0}^{+\infty} i^2 p_X(i) = \sum_{i=0}^{+\infty} i^2 p(1-p)^i = p(1-p) \sum_{i=0}^{+\infty} i^2 (1-p)^{i-1} = p(1-p) \sum_{i=0}^{+\infty} \frac{d}{dp} (-i(1-p)^i) = -p(1-p) \frac{d}{dp} \sum_{i=0}^{+\infty} i(1-p)^i$$

Applicando il lemma visto precedentemente, abbiamo:

$$E(X^2) = -p(1-p) \frac{d}{dp} \frac{1-p}{p^2} = p(1-p) \frac{p^2 + 1p(1-p)}{p^4} = \frac{(1-p)(2-p)}{p^2}$$

Possiamo ora calcolare la varianza come:

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

Funzione di ripartizione

Concentriamoci ora su un evento $X > n$ con $n \in N$ generico. Qual è la formalizzazione di tale probabilità?

$$P(X > n) = \sum_{i=n+1}^{+\infty} p_X(i) = \sum_{i=n+1}^{+\infty} (1-p)^i p = p(1-p)^{n+1} \sum_{i=n+1}^{+\infty} (1-p)^{i-(n+1)} = p(1-p)^{n+1} \sum_{j=0}^{+\infty} (1-p)^j = p(1-p)^{n+1} \frac{1}{1-(1-p)} = (1-p)^{n+1}$$

Pertanto, fissato $n \in N$ si avrà

$$F_X(n) = P(X \leq n) = 1 - P(X > n) = 1 - (1-p)^{n+1}$$

Fissato invece un generico $x \in R^+$ e indicato con $[x]$ l'intero ottenuto troncando x (o, equivalentemente, arrotondandolo per difetto), l'evento $X \leq x$ equivarrà a $X \leq [x]$. Otteniamo quindi:

$$F_X(x) = (1 - (1 - p)^{\lfloor x \rfloor + 1}) I_{[0, +\infty]}(x)$$

Si noti infine che $P(X \geq x) = P(X \geq \lfloor x \rfloor) = P(X > \lfloor x \rfloor - 1) = (1 - p)^{\lfloor x \rfloor}$, e quindi

$$P(X \geq x + y | X \geq x) = \frac{P(X \geq x + y, X \geq x)}{P(X \geq x)} = \frac{P(X \geq x + y)}{P(X \geq x)} = \frac{(1 - p)^{\lfloor x \rfloor + \lfloor y \rfloor}}{(1 - p)^{\lfloor x \rfloor}} = (1 - p)^{\lfloor y \rfloor} = P(X \geq y)$$

Questa proprietà prende il nome di *assenza di memoria*. Essa indica che durante la ripetizione dell'esperimento bernoulliano, il fatto che sia avvenuto un numero n (anche elevato) di insuccessi consecutivi non permette di dire alcunché sul numero di successivi insuccessi prima che si verifichi il primo successo. In altre parole, non c'è nessuna differenza, da un punto di vista probabilistico, dalla ripetizione degli esperimenti che vanno dal $n + 1$ -esimo in poi e dal ricominciare da capo la ripetizione.

Modello di Poisson

Consideriamo:

$$\lambda \in \mathbb{R}^+$$

$$X \sim P(\lambda)$$

Funzione di massa di probabilità

$$p_X(i) = P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!} I_{\mathbb{N} \cup \{0\}}(i)$$

Vediamo di confermarne la correttezza:

1. Non negatività: è non negativa in quanto operante con tutti valori maggiori di 0
2. La somma dei valori assoluti della funzione è uguale a 1. Dimostriamolo:

$$\sum_{i=0}^{+\infty} p_X(i) = \sum_{i=0}^{+\infty} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \sum_{i=0}^{+\infty} \frac{\lambda^i}{i!} = \rightarrow \text{utilizzando Taylor} \rightarrow e^{-\lambda} e^{\lambda} = 1$$

Valore atteso

$$E(X) = \sum_{i=0}^{+\infty} i p_X(i) = \sum_{i=0}^{+\infty} i e^{-\lambda} \frac{\lambda^i}{i!} = \sum_{i=1}^{+\infty} i e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \sum_{i=1}^{+\infty} \frac{\lambda^i}{(i-1)!} = e^{-\lambda} \lambda \sum_{i=1}^{+\infty} \frac{\lambda^{i-1}}{(i-1)!} = e^{-\lambda} \lambda \sum_{i=0}^{+\infty} \frac{\lambda^i}{i!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Varianza

$$E(X^2) = \lambda + \lambda^2 \quad \text{Dimostrazione lasciata al lettore}$$

$$\text{Var}(X) = E(X^2) - E(X)^2 = \lambda + \lambda^2 - \lambda^2 = \lambda = E(X)$$

Quindi valore atteso e varianza coincidono. Consideriamo:

$$X \sim B(n, p)$$

Supponiamo di considerare tante variabili aleatorie binomiali, dove il parametro n è di volta in volta più grande ($n \rightarrow +\infty$). Vogliamo che il parametro p diminuisca di conseguenza, in modo da mantenere costante il prodotto np . Ossia, fissato un certo λ deve valere $np = \lambda \rightarrow p = \frac{\lambda}{n}$

Quindi possiamo interpretare la distribuzione di Poisson come la distribuzione limite che si ottiene considerando la distribuzione binomiale e facendo tendere a $+\infty$ il numero di prove (diminuendo p in modo che $np = \lambda$ fissata).

Quindi, una variabile aleatoria con distribuzione di Poisson può essere usata come approssimazione di una variabile aleatoria binomiale, purché siano n abbastanza grande e p abbastanza piccola. Dunque, formalizzando, dato:

$X \sim B(n, p)$ con n grande e p piccolo

Indichiamo con:

$$X \sim P(\lambda)$$

l'approssimazione di X con distribuzione di Poisson, dove $\lambda = np$

Vediamo ora la proprietà di riproducibilità della distribuzione di Poisson. Dati:

1. $X_1 \sim P(\lambda_1)$
2. $X_2 \sim P(\lambda_2)$

Abbiamo:

$$X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$$

Quindi, la distribuzione della somma fa parte della stessa famiglia delle distribuzioni originarie.

Modello ipergeometrico

Sappiamo che la distribuzione binomiale è la descrizione formale della ripetizione di un esperimento bernoulliano in cui vengono calcolati il numero di successi. Immaginiamo ora di avere una ripetizione di un esperimento dove vengono fatte una serie di estrazioni: in questo caso il numero di successi è ancora descrivibile tramite una trasformazione binomiale? Dipende dalla reimmissione: se è presente posso rappresentare l'esperimento con una binomiale, altrimenti devo utilizzare la distribuzione ipergeometrica. Consideriamo il caso in cui abbiamo:

- N oggetti corretti
- M oggetti errati
- $N + M$ oggetti totali

X = numero di oggetti corretti estratti senza reimmissione

Allora avremo:

$$P(X = i) = \frac{\binom{N}{i} \binom{M}{n-i}}{\binom{N+M}{n}}$$

Definisco la variabile aleatoria bernoulliana X_i che può assumere valori:

$$X_i = 1 \quad \text{se l}'i\text{-esima estrazione è un successo} \quad X_i = 0 \quad \text{altrimenti}$$

Valore atteso

$$E(X_i) = \frac{N}{N+M} := p$$

$$X = \sum_{i=1}^n X_i \rightarrow \text{numero di successi su } n \text{ estrazioni}$$

Quindi:

$$E(X) = E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i) = n \frac{N}{N+M} := np$$

Se interpreto $\frac{N}{N+M}$ come p sembrerebbe non esserci differenza fra distribuzione binomiale e ipergeometrica. Tuttavia, se col valore atteso sembrano non esserci differenze, il discorso non è il medesimo con la varianza.

Varianza

$$Var(X_i) = \frac{N}{N+M} \left(1 - \frac{N}{N+M}\right) = \frac{NM}{(N+M)^2}$$

Covarianza

$$Cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$$

Il prodotto di due variabili bernoulliane è anch'esso una variabile bernoulliane. La covarianza, nel caso della binomiale, sarebbe stata pari a 0 in quanto, in quel caso, vi è indipendenza fra X_i e X_j . Ora non posso dire più la stessa cosa.

$$Cov(X_i, X_j) = \frac{-NM}{(N+M)^2(N+M-1)}$$

La dimostrazione è lasciata al lettore.

Quindi posso vedere la varianza come:

$$Var(X) = Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j)$$

Da cui, attraverso calcoli, ottengo:

$$Var(X) = np(1-p) \left(1 - \frac{n-1}{N+M-1}\right)$$

Il termine $\left(1 - \frac{n-1}{N+M-1}\right)$ era mancante nel caso binomiale. Esso è introdotto a causa della non indipendenza. Una cosa interessante da osservare è che all'aumentare del numero N di oggetti da estrarre e mantenendo fissato il numero di estrazioni da poter fare, $N+M \rightarrow \infty$. Ma se ciò accade allora:

$$1 - \frac{n-1}{N+M-1} \rightarrow 1$$

Quindi sembrerebbe che all'aumentare di $N+M$ la distribuzione ipergeometrica possa essere approssimata alla distribuzione binomiale.

1.10 Modellizzazione continua

Distribuzione uniforme continua

Consideriamo:

$$a, b \in R \quad a < b$$

$$X \sim U([a, b])$$

Funzione di densità

Un generico valore appartenente all'intervallo $[a, b]$ ha la stessa densità. Definisco:

$$f_U(x) = k I_{[a,b]}(x)$$

Sappiamo che:

$$\int_a^b f_U(x) dx = 1 \rightarrow f_U(x) = \frac{1}{b-a} I_{[a,b]}(x)$$

Quindi, sostituendo all'integrale:

$$\int_a^b \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b dx = \frac{1}{b-a} [x]_a^b = \frac{b-a}{b-a} = 1$$

Funzione di ripartizione

$$F_U(x) = P(X \leq x) = \int_{-\infty}^x f_U(n) dn$$

Tale funzione può assumere tre diversi valori:

- 0 se $x < a$
- $\int_a^x \frac{1}{b-a} dn$ se $x \in [a, b]$
- 1 se $x > b$

$$\int_a^x \frac{1}{b-a} dn = \frac{1}{b-a} [n]_a^x = \frac{x-a}{b-a}$$

Quindi i valori assumibili sono:

- 0 se $x < a$
- $\frac{x-a}{b-a}$ se $x \in [a, b]$
- 1 se $x > b$

Quindi possiamo formalizzare dicendo:

$$F_U(x) = \frac{x-a}{b-a} I_{[a,b]}(x) + I_{(b,+\infty)}(x)$$

Valore atteso

$$E(X) = \int_{-\infty}^{+\infty} x f_U(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}$$

La dimostrazione dei passi intermedi è lasciata al lettore.

Varianza

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_U(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{a^2+ab+b^2}{3}$$

La dimostrazione dei passi intermedi è lasciata al lettore.

$$Var(X) = E(X^2) - E(X)^2 = \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}$$

La dimostrazione dei passi intermedi è lasciata al lettore.

Quindi la varianza di una variabile aleatoria continua uniforme dipende dal quadrato della lunghezza di un intervallo.