Formulario di probabilità e statistica

Insiemistica

Se E e F sono tali che $E \cap F = \varnothing$ allora E e F sono incompatibili o disgiunti, non si possono verificare contemporaneamente

Leggi di De Morgan:

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

Disgiunzioni: siano $F, E \in \Omega$ e $\{E, E^c\}$ partizione di Ω

- Una partizione di Ω è composta da eventi a due a due disgiunti e la cui unione forma Ω
- $F = (F \cap E) \cup (F \cap E^c)$
- $E \cup F = (E \setminus F) \cup (E \cap F) \cup (F \setminus E)$
- $E \cup F = E \cup (F \setminus E)$

Formula di inclusione-esclusione:
$P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Proprietà misura di probabilità: $P(E) = 1 - P(E^c)$

Conteggio

Disposizioni

Dati n elementi, il numero di sottinsiemi formati da k elementi, in cui due sottinsiemi differiscano tra loro per almeno un elemento o per l'ordine

- Disposizioni con ripetizione: disposizioni in cui uno stesso elemento può comparire fino a k volte nello stesso gruppo: $n \cdot n \cdot n \cdot ... \cdot n = n^k$.
- Disposizioni senza ripetizione: disposizioni in cui in ogni sottinsieme i k elementi sono tutti distinti tra loro: $n \cdot (n-1) \cdot (n-2) \cdot ... \cdot (n-k+1)$ con k fattori.
- Permutazioni: disposizioni senza ripetizione in cui $n = k$: $n!$

Combinazioni

Dati n elementi, il numero di sottinsiemi formati da k elementi, in cui due sottinsiemi differiscano tra loro per almeno un elemento e non per l'ordine, e uno stesso elemento può comparire fino a k volte nello stesso gruppo: $\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$

Probabilità condizionata

Probabilità condizionata:

- Formula di moltiplicazione: $P(E \cap F) = P(E|F) \cdot P(F)$
- $P(E^c|F) = 1 - P(E|F)$
- $P(E \cup G|F) = P(E|F) + P(G|F) - P(E \cap G|F)$

Formula delle probabilità totali: $P(E) = P(E|F) \cdot P(F) + P(E|F^c) \cdot P(F^c)$

Formula di Bayes: $P(F|E) = \frac{P(E|F) \cdot P(F)}{P(E)}$

Eventi indipendenti

Due eventi E, F si dicono indipendenti se $P(E \cap F) = P(E) \cdot P(F)$. Inoltre:

- $\{E, F\}$ indipendenti $\iff P(E|F) = P(E) \iff P(F|E) = P(F)$
- $\{E, F\}$ indipendenti $\iff$ $\{E^c, F\}$ indipendenti $\iff$ $\{E^c, F^c\}$ indipendenti $\iff$ $\{E, F^c\}$ indipendenti

Variabili aleatorie

Formula valor medio: $E(X) = \sum_{x_k \in \chi} x_k \cdot p_x(x_k)$

Proprietà valor medio:

- $E(aX) = a \cdot E(x)$
- $E(X + Y) = E(X) + E(Y)$
- $X \geq Y \Rightarrow E(X) \geq E(Y)$

Valor medio: $E(g(X)) = \sum_{x_k \in \chi} (g(x_k) \cdot p_x(x_k))$

Varianza: $Var(X) = \sum_{x_k \in \chi} ((x_k - E(X))^2 \cdot p_x(x_k))$

Varianza

Sia X v.a. con legge P_X e alfabeto composto da elementi x_k. Si definisce la varianza:

$$Var(X) = \sum_{x_k} [x_k - E(X)]^2 P_X(x_k)$$

Valgono:

- $Var(X) = E[(X - E(X))^2] = E(X^2) - E(X)^2$
- $Var(X) \geq 0$, in particolare $Var(X) = 0 \Leftrightarrow X \equiv$ *costante*
- $Var(aX) = a^2 Var(X)$, $a \in \mathbb{R}$
- $Var(X + a) = Var(X)$, $a \in \mathbb{R}$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
- $X \perp\!\!\!\perp Y \implies Var(X + Y) = Var(X) + Var(Y)$

Def Covarianza. *Siano X, Y v.a.. Si definisce la covarianza:*
$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$
Valgono:

- $Cov(X, Y) = E(XY) - E(X)E(Y)$
- $Cov(Y, X) = Cov(X, Y)$
- $Cov(aX + b, Y) = aCov(X, Y)$
- $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
- $X \perp\!\!\!\perp Y \implies Cov(X, Y) = 0$

Def Indipendenza. $X \perp\!\!\!\perp Y$ se: $P_{XY}(x_i, y_j) = P_X(x_i) P_Y(y_j) \quad \forall x_i, y_j$

- $X \perp\!\!\!\perp Y \implies Var(X + Y) = Var(X) + Var(Y)$
- $X \perp\!\!\!\perp Y \implies Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y)$

Def V.a. scorrelate. *Le v.a. X, Y sono dette scorrelate se* $Cov(X, Y) = 0$

- Indipendenza $\implies$ Scorrelazione
- Non valido il contrario

Variabili aleatorie discrete notevoli

Alfabeto finito: Variabili aleatorie di Bernoulli

- $X \sim Be(p)$ con $p \in [0, 1]$
- Alfabeto $\mathcal{X} = \{0, 1\}$
- Densità discreta $p_x(1) = p$, $p_x(0) = 1 - p$
- $E(X) = p \qquad Var(X) = p(1 - p)$

Alfabeto finito: Variabili aleatorie binomiali

Conta il numero di successi in uno schema di Bernoulli con n prove indipendenti, dove la probabilità di successo è p

- $X \sim Bin(n, p)$ con $n \in \mathbb{N} = \{1, 2, ..\}$ e $p \in [0, 1]$
- Alfabeto $\mathcal{X} = \{0, 1, .., n\}$
- Densità discreta $p_x(k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$, $k \in \mathcal{X}$
- Inoltre $E(X) = np \qquad Var(X) = n \cdot p \cdot (1 - p)$

Alfabeto infinito: Variabili aleatorie geometriche

Corrisponde al numero di prove che devo effettuare per osservare il primo successo in uno schema di Bernoulli dove il numero di prove non è necessariamente predefinito e dove la probabilità di successo è p

- $X \sim Ge(p)$ con $p \in (0, 1)$
- Alfabeto $\mathcal{X} = \mathbb{N} = \{1, 2, ..\}$
- Densità discreta $p_x(k) = (1 - p)^{k-1} \cdot p$, $k \in \mathcal{X}$
- Inoltre $E(X) = \frac{1}{p} \qquad Var(X) = \dfrac{1 - p}{p^2}$

Variante: invece di guardare alla prova che corrisponde al primo successo si guarda il numero di fallimenti prima del primo successo:

- $X' = X - 1$, $X \sim Ge(p)$ con $p \in (0, 1)$
- Densità discreta $p_x(k) = (1 - p)^k \cdot p$, $k \in \mathcal{X}$
- Inoltre $E(X) = \frac{1-p}{p} \qquad Var(X) = \dfrac{1 - p}{p^2}$

Probabilità di lunga attesa: indica la probabilità che per il primo successo si debba aspettare più di qualcosa: $P(X > k) = (1 - p)^k$

Alfabeto infinito: Variabili aleatorie di Poisson

Una v.al. Binomiale con n molto grande e p molto piccolo può essere approssimata da una v.al di Poisson di parametro $\lambda = n \cdot p$. Euristica: $n > 100$; $p < 0.01$; $np \leq 20$

- $X \sim Po(\lambda)$ con $\lambda > 0$
- Alfabeto $\mathcal{X} = \mathbb{N}_0 = \{0, 1, ..\}$
- Densità discreta $p_x(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$, $k \in \mathcal{X}$
- Inoltre $E(X) = \lambda \qquad Var(X) = \lambda$

Vettori aleatori discreti

Siano X v.al con alfabeto $\mathcal{X} = \{x_1, x_2, ..\}$ e Y v.al con alfabeto $\mathcal{Y} = \{y_1, y_2, ..\}$ allora per l'alfabeto $\mathcal{V}$ del vettore aleatorio $\underline{v}$ si ha $\mathcal{V} \subseteq \mathcal{X} \times \mathcal{Y}$

- Densità congiunta: $p_{XY}(x_i, y_j) = P(X = x_i, Y = y_j)$
- Densità marginale X: $p_X(x_i) = \sum_{y_j \in \mathcal{Y}} p_{XY}(x_i, y_j)$ per ogni $x_i \in \mathcal{X}$
- Densità marginale Y: $p_Y(y_j) = \sum_{x_i \in \mathcal{X}} p_{XY}(x_i, y_j)$ per ogni $y_j \in \mathcal{Y}$
- Valor medio: $E[g(X,Y)] = \sum_{x_i, y_j} g(x_i, y_j) P_{XY}(x_i, y_j)$
- Valor medio: $E(X,Y) = \sum_{x_i \in \mathcal{X}} \sum_{y_j \in \mathcal{Y}} x_i y_j P(X = x_i, Y = y_j)$

Siano X, Y v.al. con alfabeti composti da elementi x_i e y_j. Allora X e Y si dicono indipendenti se: $P_{XY}(x_i, y_j) = P_X(x_i) P_Y(y_j) \quad \forall x_i, y_j$

Variabili aleatorie assolutamente continue

Def Una v.a. X si dice (ass.) continua si definisce associando una densità $f_x : \mathbb{R} \to \mathbb{R}$ tale che:

- $f_X(x) \geq 0$
- $\int_{\mathbb{R}} f_X(x) dx = 1$

Valgono:

- $P(X \in I = P_X(I) = \int_I f_X(x) dx$
- $P_X(a) = 0 \quad \forall a \in \mathbb{R}$

Def Funzione di distribuzione.
$F_X : \mathbb{R} \to [0, 1]$
$x \mapsto F_X(x) = P(X \leq x)$

Def Valor medio v.a.c.. $E(X) = \int_{\mathbb{R}} x f_X(x) dx$

Def Varianza v.a.c.. $Var(X) = \int_{\mathbb{R}} [x - E(X)]^2 f_X(x) dx$

Def Valor medio di una funzione $g(X)$, con X avente densità f_X:.. $E(g(X)) = \int_{\mathbb{R}} g(x) f_X(x) dx$

Variabili aleatorie continue notevoli

Variabili aleatorie uniformi

$X \sim U(a, b)$ con $a < b$ se ha:

- densità $f_x(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b \\ 0 & \text{altrimenti} \end{cases}$
- distribuzione $F_x(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{b-a} & \text{se } a \leq x \leq b \\ 1 & \text{se } x > b \end{cases}$
- $E(X) = \frac{b+a}{2} \qquad Var(X) = \frac{(b-a)^2}{12}$

Variabili aleatorie esponenziali

Descrive la durata della vita di un fenomeno privo di memoria.

Valgono: $X \sim Exp(\lambda)$ con densità: $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{altrimenti} \end{cases}$

Funzione di distribuzione: $F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{altrimenti} \end{cases}$

- $E(X) = \frac{1}{\lambda}$
- $Var(X) = \frac{1}{\lambda^2}$
- $P(X > T + t \mid X > T) = P(X > t)$

Variabili aleatorie gaussiane(o normali)

$X \sim N(\mu, \sigma^2)$ e densità: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$
σ^2=Varianza
σ=Deviazione standard

- $E(X) = \mu$
- $Var(X) = \sigma^2$
- $P(Z \leq x) = \Phi(x)$
- $P(|Z| \leq x) = 2\Phi(x) - 1$
- $P(Z \geq x) = 1 - \Phi(x) = \Phi(-x)$
- $P(|Z| \geq x) = 2(1 - \Phi(x)) = 2\Phi(-x)$
- $P(x \leq Z \leq y) = \Phi(y) - \Phi(x)$
- $P(\mu - 4\sigma \leq X \leq \mu + 4\sigma) \approx 1$

Def Gaussiana standard. $X \sim N(0, 1)$.
Prop Trasformazioni affini di v.a. normali. *Sia* $X \sim N(\mu, \sigma^2)$ *e* $Y = aX + b$ *allora* $Y \sim N(a\mu + b, a^2\sigma^2)$
quindi se $X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$
Se due v.al. X, Y sono indipendenti e $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma$
$X + Y$ allora:
$Z \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Teoremi Limite

Def Legge dei grandi numeri (LLN). *Siano* $X_1, \ldots, X_n$ *v.a. i.i.d. con media finita. Allora* $\overline{X_n} \to E(X_1)$

Def Metodo Monte Carlo. $\mathcal{I} = \int_a^b f(x) dx \quad b > a$
$\mathcal{I} = (b-a) \int_a^b \frac{f(x)}{b-a} dx = (b-a) E(f(x))$ *con* $X = U(a, b)$

Def Teorema centrale del limite (CLT). *Siano* $X_1, \ldots, X_n$ *v.a. i.i.d. con* $E(X_1) = \mu$, $Var(X_1) = \sigma^2$. *Allora:*
$\sqrt{n}\left(\frac{\overline{X_n} - \mu}{\sigma}\right) \to N(0, 1)$
$\frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma\sqrt{n}} \to N(0, 1)$
$\sqrt{n}(\overline{X_n} - \mu) \to N(0, \sigma^2)$

Statistica Descrittiva

Indici di posizione

Media campionaria: $\overline{x} = \frac{1}{n} \sum_{i=1}^n x_i$

Mediana: $M = \begin{cases} x_{(n+1)/2} & \text{se } n \text{ dispari} \\ \frac{1}{2}\left(x_{n/2} + x_{(n/2)+1}\right) & \text{se } n \text{ pari} \end{cases}$

Moda: dato a cui corrisponde la frequenza assoluta massima

Indici di dispersione

Varianza campionaria: $s^2 = \dfrac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x})^2$

p-esimo quantile $(0 < p < 1)$:

$q_p = \begin{cases} x_{\lfloor np \rfloor + 1} & \text{se } np \text{ non è intero} \\ \frac{1}{2}(x_{np} + x_{np+1}) & \text{se } np \text{ è intero} \end{cases}$

con n ampiezza del campione

- 25-esimo percentile=Q_1=primo quartile = $q_{0,25}$
- 50-esimo percentile=Q_2=secondo quartile (mediana) = $q_{0,5}$
- 75-esimo percentile=Q_3=terzo quartile = $q_{0,75}$

Differenza interquantile: $IQR = Q_3 - Q_1$
Boxplot:

- IQR
- Limite del baffo inferiore = $L = Q_1 - 1.5 \cdot IQR$
- Limite del baffo superiore = $U = Q_3 + 1.5 \cdot IQR$
- Outliers: dati che rimangono fuori dai baffi

Statistica Inferenziale

Stimatori

Uno stimatore T si dice corretto/non distorto se $E_{\underline{\theta}}(T) = \tau(\underline{\theta})$

Uno stimatore T si dice consistente se $\text{Var}_{\underline{\theta}}(T) \xrightarrow{n \to +\infty} 0$