Mirko Calcaterra

Appunti di
Inferenza Statistica

Tratti dalle lezioni dal corso di Inferenza Statistica

fubini ⊗ tonelli
edizioni e convoluzioni

Per suggerimenti e correzioni potete contattarmi all'indirizzo:
`mirko@fubinitonelli.it`.

Revisione del 1 giugno 2023.

A Nello e Giovanna

Prefazione

Durante l'anno accademico 2019/2020, ho seguito un corso di Inferenza Statistica. Il corso quell'anno ha registrato alcuni tagli di programma dovuti al fatto che è iniziato in ritardo a causa dell'emergenza pandemica.
Proprio grazie a questi argomenti "appena accennati" è nata la curiosità di andare oltre e studiare nel dettaglio alcuni argomenti che troverete nel libro.
La svolta è arrivata quando nell'estate del 2021, insieme alla Fubini $\otimes$ Tonelli, che ringrazio con tutto il cuore per il tempo dedicato, per la pazienza avuta con me e per i suggerimenti che mi hanno dato, abbiamo deciso di scrivere questo libro, il quale mantiene gli argomenti del corso, ma che si arricchisce con alcuni argomenti e plot disegnati da me nel corso di questi 2 anni. Questo libro eredita la struttura del corso dell'omonimo corso tenuto al Politecnico di Milano. Soprattutto nei primi capitoli, ho selezionato alcuni esempi, tratti dalle lezioni di questo corso, che sono stati per me di fondamentale importanza nello studio e la comprensione della materia.

Perché questo libro? Ho deciso di scrivere questo testo perché ho notato che, quando ho iniziato a studiare questa materia, non era presente un libro di testo di riferimento in italiano. Inoltre, ho notato che tutti quelli che hanno deciso di seguire il corso negli anni successivi, hanno denunciato lo stesso problema. Ecco perché ho voluto scrivere un libro su qualcosa che mi interessa e che fosse d'aiuto a tutti quelli che decidono di approcciarsi allo studio di questa materia.

La dipendenza dal corso In origine avrei voluto scrivere questo libro cambiando l'ordine di alcuni argomenti, dando più peso ad alcuni invece che altri. Però, dato che questo libro sarà letto in particolar modo dagli studenti del omonimo corso che ho seguito, ho preferito mantenere la stessa struttura del corso dell'anno in cui l'ho seguito anni fa, e ho deciso di selezionare gli esempi più significativi tratti proprio dallo stesso corso, proprio per non disorientare lo studente che si approccia per la prima volta ad esso e ha deciso di usare questo libro. Se secondo te è più opportuno cambiare l'ordine di alcuni argomenti, di aggiungere o togliere alcuni esempi, ti invito a scrivermi una mail all'indirizzo che troverai nel prossimo paragrafo. Ogni consiglio, infatti, è molto apprezzato.

Per gli studenti del corso Ci tengo a specificare sin da subito che questo libro può essere un supporto per lo studio della materia, ma non sostituisce le lezioni e i libri consigliati dal docente del corso. Ho deciso di dare una propria identità al libro. Questo significa che possono esserci argomenti che non fanno parte degli argomenti del corso (ma magari lo saranno in futuro) come potrai, d'altra parte, non trovare argomenti che invece sono stati trattati a lezione.

Invito chiunque volesse *segnalare errori* o proporre *nuovi argomenti* per le prossime versioni del libro di scrivere un'e-mail a `mirko@fubinitonelli.it`.

L'autore Prima di lasciarti alla lettura, voglio aprire una piccola parentesi su di me, sarò veloce, non voglio tediarti.

Prima di seguire questo corso, non amavo granché questa materia e questo mondo. Era il 2020, l'anno che molti ricordano per il Covid, un anno in cui ho fatto parte del direttivo dell'Associazione degli Ingegneri Matematici (AIM) e anno in cui l'associazione è cresciuta tantissimo e io con lei.
Tra le tante cose che ho imparato durante l'anno di direttivo c'è sicuramente il linguaggio LaTeX, in cui questo libro è scritto, ma non solo. Infatti ho avuto la fortuna di confrontarmi con persone che ritengo molto ben formate nel settore della Data Science ed è probabilmente grazie a loro che è nata questa passione per questo universo.
Fatta questa lunga premessa, non mi resta che ringraziare tutti quelli che mi hanno fatto cambiare idea sul mondo della statistica (prima) e sul mondo della Data Science (dopo), quindi in primis la Professoressa del corso, in particolare per la passione con cui insegna, G.G., G.M., F.F. di AIM e tutto lo staff della F$\otimes$T.
Un ultimo consiglio prima di lasciarti alla lettura di questo libro:

> Cerca di essere curioso e mettiti sempre in discussione: non fermarti ai primi fallimenti.

Ricordati che è possibile ottenere dei risultati non troppo incoraggianti nelle prime iterazioni del training set, ma questo non significa che ci si debba fermare. Ciò che è importante è ottenere un bel risultato nel test set alla fine. Perseguita sempre i tuoi sogni!
Se non hai capito nulla di questo ultimo esempio e sei curioso di capire quello che hai appena letto, stai leggendo il libro giusto e non ti resta che

leggere tutto fino alla fine per capire cosa intendo dire. Non mi resta che augurarti buona lettura!

Mirko Calcaterra

Indice

Capitolo 0

Pillole di Probabilità e Statistica

Prima di iniziare a parlare di argomenti di Inferenza Statistica, è opportuno un ripasso di argomenti di Probabilità.

0.1 Probabilità

In questa sezione daremo le basi da conoscere per comprendere tutto il libro. Per una trattazione più completa[1], invito a consultare *Appunti di Probabilità*, sempre targato f$\otimes$t.

Il primo concetto che è fondamentale avere chiaro è quello di variabile aleatoria.

Definizione 0.1: Variabile aleatoria
Dati gli spazi misurabili $(\Omega, \mathcal{A})$ e $(F, \mathcal{F})$, una funzione $X : \Omega \to F$ si dice **funzione misurabile** o **variabile aleatoria** se:

$$(X \in B) \in \mathcal{A} \ \forall B \in \mathcal{F} \tag{1}$$

In questo capitolo considereremo variabili aleatorie reali, le quali si dividono principalmente in due classi: discrete e continue:

[1] nonché sofferta

Definizione 0.2: Variabili aleatorie reali discrete
Variabili aleatorie la cui probabilità è concentrata su singoli punti. Essi possono anche essere in quantità numerabile, quindi non finita.

Definizione 0.3: Variabili aleatorie reali continue
Una probabilità $\mathbb{P}$ su $(\mathbb{R}, \mathcal{B})$ ammette **densità (continua) di probabilità** se $\exists f : \mathbb{R} \to [0, +\infty)$ boreliana tale che:

$$F(x) = \mathbb{P}((-\infty, x]) = \int_{-\infty}^{x} f \, \mathrm{d}t = \int_{\mathbb{R}} f(t) \, \mathbb{I}_{(-\infty,x]}(t) \, \mathrm{d}t = \int_{(-\infty,x]} f \, \mathrm{d}t \tag{2}$$

Se $\mathbb{P} = P^X$ di una variabile aleatoria reale X, diciamo che X è una **variabile aleatoria reale (assolutamente) continua** e che f è densità (continua) di X.

Una volta chiarita la differenza tra le due tipologie variabili aleatorie reali, possiamo ricordare la definizione la distribuzione (o legge) di X, come:

Definizione 0.4: Distribuzione (o legge) di X
Dati $(\Omega, \mathcal{A}, \mathbb{P})$ spazio di probabilità, $(F, \mathcal{F})$ spazio misurabile e $X : \Omega \to F$ VA, la probabilità

$$P^X : \mathcal{F} \to [0, 1] \quad P^X(B) = \mathbb{P}(X \in B) \tag{3}$$

è detta **legge** o **distribuzione** di X.

Concludiamo la sezione di probabilità con le densità di probabilità. Esse possono essere continue e discrete. Daremo solo una definizione per entrambe, lasciando tutti gli esempi notevoli nell'appendice.

Definizione 0.5: Densità di probabilità continua
Si dice **densità continua di probabilità** una $f : \mathbb{R} \to [0, +\infty)$, non negativa, Riemann-integrabile e tale che:

$$\int_{-\infty}^{+\infty} f(x) \, \mathrm{d}x = 1 \tag{4}$$

Si dice inoltre che F *ammette densità* f se vale:

$$F(x) = \int_{-\infty}^{x} f(t) \, \mathrm{d}t \tag{5}$$

Definizione 0.6: Densità di probabilità discreta
La funzione positiva $p : S \to [0, 1]$, con S al più numerabile, è detta **densità discreta di probabilità** se:

$$\sum_{x \in S} p(x) = 1 \quad \text{e} \quad P^X(B) = \sum_{x \in S \cap B} p(x) \tag{6}$$

0.2 Statistica inferenziale

La statistica inferenziale è fortemente legata alla teoria della probabilità, ma bisogna fare attenzione a definire alcuni concetti come media e varianza campionaria. Questi infatti differiscono dalle definizioni di media e varianza che vengono visti nei corsi di probabilità, in quanto sono definiti direttamente sul campione di riferimento e non sulla distribuzione di probabilità.

NB: Esiste una differenza tra campione e popolazione. Il primo infatti è in genere più piccolo di quest'ultima.

Lavorare con un campione è sicuramente meno costoso dal punto di vista computazionale rispetto ad una popolazione (che presenta quindi più elementi). Però sorge spontanea una domanda:

Come faccio ad essere certo che il campione sia rappresentativo nei confronti della popolazione?

Per dare una una risposta a questa domanda, definiamo in primis dei termini che ricorreranno nel prossimo capitolo:

Definizione 0.7: Media Campionaria
Definiamo la **media campionaria** per un campione di n elementi la seguente funzione:

$$\bar{X}_n \coloneqq \frac{1}{n} \sum X_i$$

Definizione 0.8: Varianza Campionaria
Definiamo la **varianza campionaria** per un campione di n elementi la seguente funzione di dispersione statistica:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

A partire dalla varianza campionaria, è possibile definire la **deviazione standard campionaria**, come la radice quadrata della varianza campionaria, in formule:

$$S \coloneqq \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$$

Risulta ora facile rispondere alla domanda posta prima.

Definizione 0.9: Campione Rappresentativo
Diciamo che un campione è **rappresentativo** per una popolazione se

$$\mu \simeq \bar{X}_n \quad \text{e} \quad \sigma \simeq S^2$$

Dove il simbolo $\simeq$ indica il concetto di "circa uguale".

Il significato di questo concetto è delicato ed è strettamente legato al tema dei test statistici, argomento che viene introdotto e ben sviluppato nel capitolo 3. Per adesso possiamo anticipare che il concetto di "quasi uguale", essendo arbitrario, si misura con una soglia, e α è questa soglia e ha natura probabilistica.
Per esempio se vogliamo testare che $\mu \simeq \bar{X}_n$, dovremo prima definire il test (che scriveremo attraverso la definizione di un'ipotesi nulla, $H_0: \ \mu = \bar{X}_n$ e un'ipotesi alternativa che può accadere $H_1: \ \mu \neq \bar{X}_n$) e poi il livello α (un candidato può essere $\alpha = 5\%$). Successivamente, se il test dà esito positivo, accetteremo l'ipotesi nulla, che si traduce nell'affermare che, a patto di commettere un errore α, i due termini possiamo considerarli uguali.

Esempio. Per capire meglio l'importanza di avere un campione rappresentativo, supponiamo di voler valutare l'altezza degli spettatori che utilizzano i seggiolini in un palazzetto in cui si giocano partite di pallavolo e basket affinché possano essere più comodi possibile durante le partite. Per ottenere un campione rappresentativo, non possiamo considerare come campione *solo* i componenti delle squadre che giocano in quel palazzetto, poiché probabilmente saranno più alti in media rispetto agli spettatori.

Per ottenere un campione rappresentativo, dovremmo considerare un gruppo di spettatori scelti in modo *totalmente casuale* tra coloro che frequentano il palazzetto.

Teorema 0.10: di Slutsky

Siano X_n, X, Y_n variabili aleatorie reali tali che $X_n \xrightarrow{\mathcal{L}} X$ e $Y_n \xrightarrow{\mathcal{L}} c$ dove c è una costante reale. Allora:

- $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$.
- $X_n Y_n \xrightarrow{\mathcal{L}} cX$.
- $Y_n, c \neq 0 \implies \frac{X_n}{Y_n} \xrightarrow{\mathcal{L}} \frac{1}{c} X$.

Teorema 0.11: Legge Forte dei Grandi Numeri

Siano $X_n \in L^1$ una successione di variabili aleatorie indipendenti e identicamente distribuite, e $\mu \in \mathbb{R}$ tale che $\mathbb{E}[X_n] = \mu$. Allora:

$$\bar{X}_n \xrightarrow{\text{qc}} \mu$$

Si noti che le variabili devono essere distribuite allo stesso modo, ma non per forza devono essere Gaussiane. Inoltre si noti che è la variabile aleatoria *media campionaria* che converge quasi certamente al valore atteso, non la singola osservazione.

Una premessa prima del prossimo enunciato è doverosa. Il nome del teorema è *Teorema Centrale del Limite*, nel senso che il suo ruolo nella statistica è *centrale*, di fondamentale importanza. Non è il *Teorema del Limite Centrale*[2].

Teorema 0.12: Centrale del Limite

Siano $X_n \in L^2$ una successione di variabili aleatorie indipendenti e identicamente distribuite, con $\mu = \mathbb{E}[X_n]$ e $\sigma^2 = \mathrm{Var}(X_n)$. Allora:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Si noti che non è corretto dire che *il campione diventa Gaussiano*, ma che la variabile aleatoria *media campionaria*, opportunamente trasformata[3] converge in legge a una variabile aleatoria distribuita come una Normale standard.

[2] Altrimenti ci dovrebbe essere anche un Teorema del Limite Destro e uno del Limite Sinistro – Piercesare Secchi.

[3] standardizzata.

Teorema 0.13: Distribuzione di alcune statistiche importanti

Sia $X_1, \ldots, X_n$ un campione casuale estratto da una popolazione $\mathcal{N}(\mu, \sigma^2)$, $\sigma^2 > 0$, $n \geq 2$. Allora:

1. $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.
2. $S^2 \sim \dfrac{\sigma^2}{n-1}\chi^2(n-1)$, ovvero $\dfrac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$.

   Di conseguenza si può affermare che:

   $$\mathbb{E}[S^2] = \sigma^2 \quad \text{e} \quad \mathrm{Var}\left(S^2\right) = \frac{2\sigma^4}{n-1}$$

   In particolare, la varianza della varianza campionaria tende a 0 al crescere di n.
3. $\bar{X} \perp\!\!\!\perp S^2$.
4. $\frac{\bar{X}-\mu}{\sqrt{\frac{S^2}{n}}} \sim t(n-1)$.

Capitolo 1

Il modello statistico

1.1 Introduzione

Come abbiamo già avuto di vedere già nello scorso capitolo, lo studio della statistica ruota attorno al concetto di campione,più nello specifico di campione casuale.

Definizione 1.1: Campione casuale

Un **campione casuale** è una successione $\overrightarrow{X} = \{X_1, \ldots, X_n\}$ di n variabili aleatorie **indipendenti e identicamente distribuite (iid)**.
In particolare, se le variabili del campione casuale hanno distribuzione $\mathcal{L}$, scriviamo:

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{L}$$

oppure, laddove è inteso che $i = 1, \ldots, n$:

$$X_i \overset{\text{iid}}{\sim} \mathcal{L}$$

n è detta **dimensione del campione** o **sample size**.

In generale, la legge $\mathcal{L}(X_i)$ di una variabile aleatoria del campione casuale dipende anche da parametri incogniti $\overrightarrow{\vartheta} \in \Theta \subseteq \mathbb{R}^n$, che vivono nello spazio parametrico Θ.

Esempio. Esempi di parametri della distribuzione di un campione casuale sono:

- $X_i \sim \mathsf{Be}(p)$, dove $\vartheta = p \in \Theta = [0,1]$ è incognita.
- $X_i \sim \mathcal{N}(\mu, \sigma^2)$, dove $\overrightarrow{\vartheta} = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}^+$ sono incogniti.

Definizione 1.2: Modello statistico
Il **modello statistico** è la terna data da:

$$\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_{\overrightarrow{\vartheta}}\right)$$

Dove:

- $\mathbb{R}^n$ è lo spazio dei reali in n dimensioni, dove n è la dimensione del campione casuale.
- $\mathcal{B}(\mathbb{R}^n)$ indica lo spazio dei boreliani[1] definiti su $\mathbb{R}^n$.
- $P_{\overrightarrow{\vartheta}}$ indica la probabilità che $\overrightarrow{\vartheta} \in \Theta \subseteq \mathbb{R}^n$.

Definizione 1.3
Una **statistica** T è una qualsiasi funzione $T = T(X_1, \dots, X_n)$ di un campione casuale. La sua distribuzione probabilistica è detta **distribuzione campionaria**.

NB: T dev'essere funzione del campione, quindi calcolabile a partire dai dati, e non deve invece dipendere dal parametro incognito.

Esempio. Supponiamo di voler studiare quante volte sia uscita testa a partire da due monete $X_1, X_2 \overset{\text{iid}}{\sim} \mathsf{Be}(p)$.[2] Sia quindi $T(\overrightarrow{X}) = X_1 + X_2$, ovvero la Statistica che risponde alla domanda "Quante volte è uscita testa". Verifichiamo che legge congiunta dipenda *solo* da p. Anzitutto notiamo che $T(\overrightarrow{X}) \in \{0, 1, 2\}$. È quindi possibile calcolare singolarmente tutti i valori della sua distribuzione:

- *Media campionaria*: $T(X_1, \dots, X_n) = \bar{X}_n = \frac{1}{n} \sum X_i$.
- *Varianza campionaria*: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$.
- *Deviazione standard campionaria*: $S = \sqrt{S^2}$, definita a partire dalla varianza campionaria.

[1] Se non avete idea di cosa sia il magico mondo dei Boreliani, vi invito ancora una volta a consultare *Appunti di Probabilità*. Fubini-Tonelli, 2019. ISBN: 9788894431308. URL: `https://www.fubinitonelli.it/probabilita/`.

[2] Questo significa che esce testa con probabilità p in entrambi i casi.

- *Minimo*: $X_{(1)} = \min\{X_1, \ldots, X_n\}$.
- *Massimo*: $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

Se $X_1, ..., X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, allora

$$\frac{\bar{X}_n - \mu}{\sigma\sqrt{n}}$$

non è una statistica perché dipende anche da μ e ϑ che sono parametri incogniti.
Come facciamo a conoscere i parametri incogniti? Usando la statistica[3]!

Definizione 1.4: Ancillarità
Una statistica $T(\vec{x})$ la cui legge non dipende da $\vec{\vartheta}$ è detta **ancillare**.

1.2 Sufficienza, minimalità e completezza

Definizione 1.5: Principio di sufficienza
Dato un campione casuale $X_1, \ldots, X_n$ la cui legge $\mathcal{L}(X_i)$ dipende da $\vec{\vartheta}$, la statistica $T(\vec{X})$ è detta **statistica sufficiente** per $\vec{\vartheta}$ quando *ogni* inferenza su $\vec{\vartheta}$ dipende da $\vec{X}$ solo tramite il valore di $T(\vec{X})$.

Se $\vec{x}, \vec{y}$ sono due relazioni diverse del campione, ma tali che $T(\vec{x}) = T(\vec{y})$ allora l'inferenza su ϑ sarà la stessa sia se si osservi $\vec{x}$, sia che si osservi $\vec{y}$.

Esempio. Siano $X_1, X_2 \overset{\text{iid}}{\sim} \mathsf{Be}(p)$ e sia $T(\vec{X}) = X_1 + X_2$. Verifichiamo che legge congiunta dipenda *solo* da p. Anzitutto notiamo che $T(\vec{X}) \in \{0, 1, 2\}$. È quindi possibile calcolare singolarmente tutti i valori della sua distribuzione:

- $T = 0$:
$$\mathbb{P}(X_1 = 0, X_2 = 0 | T = 0) = 1$$
- $T = 2$:
$$\mathbb{P}(X_1 = 1, X_2 = 1 | T = 2) = 1$$

[3] qui intesa come nome della disciplina astratta, e non come l'oggetto matematico definito in 1.3.

- $T = 1$ (svolgendo i calcoli per uno dei due casi possibili):

$$\begin{aligned}\mathbb{P}(X_1 = 1, X_2 = 0 | T = 1) &= \frac{\mathbb{P}(X_1 = 1, X_2 = 0, T = 1)}{\mathbb{P}(T = 1)} \\ &= \frac{p(1-p)}{\binom{2}{1} p(1-p)} = 0.5\end{aligned}$$

Poiché non c'è alcuna dipendenza da p in nessun valore della sua distribuzione, concludiamo che T è statistica sufficiente per p.

Teorema 1.6: Teorema! di fattorizzazione
Sia $f(\overrightarrow{X}, \vartheta)$ la densità congiunta (discreta o continua) del campione $\overrightarrow{X}$. Una statistica $T(\overrightarrow{X})$ è sufficiente per ϑ se e solo se esiste $g(t, \vartheta), h(\overrightarrow{X})$ tali che $\forall \overrightarrow{X}, \vartheta$ la densità congiunta è tale che:

$$f(\overrightarrow{x}, \vartheta) = g(T(\overrightarrow{x}), \vartheta) \cdot h(\overrightarrow{x})$$

Prima di dimostrare questo teorema introduciamo ora una convenzione notazionale. Con la lettera maiuscola X intendiamo una o più variabili aleatorie o un campione casuale, quindi un oggetto matematico che possiede una distribuzione. Al contrario, con la lettera minuscola x indichiamo un generico numero o vettore fissato, per esempio l'argomento di una densità di probabilità di $f(x)$, che è invece una funzione deterministica. Notiamo infine che le due notazioni sono spesso usate interscambiabilmente, poiché è facile intuire dal contesto se il simbolo indica una variabile aleatoria o una quantità fissata.

Dimostrazione
La dimostrazione è divisa nelle due direzioni:

$\Longrightarrow$ Sia T una statistica sufficiente. Possiamo scrivere che:

$$\begin{aligned}f(\overrightarrow{X}, \vartheta) &= \mathbb{P}_\vartheta(\overrightarrow{X} = \overrightarrow{x}) = \\ &= \underbrace{\mathbb{P}(\overrightarrow{X} = \overrightarrow{x} | T(\overrightarrow{X}) = T(\overrightarrow{x}))}_{h(\overrightarrow{x})} \cdot \underbrace{\mathbb{P}_\vartheta(\mathbb{T}(\overrightarrow{X}) = T(\overrightarrow{x}))}_{g(T(\overrightarrow{x}), \vartheta)}\end{aligned}$$

Dove è evidente che il primo termine **non** dipende da ϑ per ipotesi. Notiamo infine che il secondo termine rappresenta, a meno di costanti, la *legge della statistica*.

$\Longleftarrow$ Supponiamo ora che f sia fattorizzabile. Sia $q(t,\vartheta)$ la densità di una statistica $t = T(\overrightarrow{X})$ e consideriamo l'insieme:

$$\mathcal{A}_{T(\overrightarrow{X})} \coloneqq \{\overrightarrow{y} : T(\overrightarrow{y}) = T(\overrightarrow{x})\}$$

Consideriamo la legge condizionata:

$$\begin{aligned}
\mathcal{L}(\overrightarrow{X}|T(\overrightarrow{X})) = \frac{f(\overrightarrow{X},\vartheta)}{q(T(\overrightarrow{X}),\vartheta)} &\stackrel{*}{=} \frac{g(T(\overrightarrow{X}),\vartheta)h(\overrightarrow{X})}{\sum_{\overrightarrow{y}\in\mathcal{A}_{T(\overrightarrow{X})}} f(\overrightarrow{y},\vartheta)} = \\
&= \frac{g(T(\overrightarrow{X}),\vartheta)h(\overrightarrow{X})}{\sum_{\overrightarrow{y}\in\mathcal{A}_{T(\overrightarrow{X})}}[g(T(\overrightarrow{y}),\vartheta)h(\overrightarrow{y})]} = \\
&= \frac{g(T(\overrightarrow{X}),\vartheta)h(\overrightarrow{x})}{g(T(\overrightarrow{x},\vartheta))\sum_{\overrightarrow{y}\in\mathcal{A}} h(\overrightarrow{y})} = \mathrm{cost}(\vartheta) \qquad (1.1)
\end{aligned}$$

Nel passaggio $*$ possiamo notare che al numeratore abbiamo sostituito ciò che sappiamo per ipotesi, ovvero che la funzione è fattorizzabile, e al denominatore abbiamo applicato il teorema delle probabilità totali.
Avendo dimostrato 1.1, abbiamo dimostrato che T è sufficiente per ϑ. $\square$

Osservazione: È possibile trovare $f(\overrightarrow{x},\vartheta)$ scritto come $f(\overrightarrow{X},\vartheta)$, poiché l'approccio bayesiano immagina ϑ come una variabile aleatoria. Questo però esula dagli obiettivi di questo libro.

Esempio. Prendiamo il **campione uniforme** $X_1,\ldots,X_n \stackrel{\text{iid}}{\sim} U_{[0,\vartheta]}$. Dimostriamo che il massimo $X_{(n)}$ è sufficiente per ϑ. Scriviamo la legge congiunta:

$$\begin{aligned}
f(\overrightarrow{x},\vartheta) = \prod_{i=1}^{n} f(x_i;\vartheta) &= \prod_{i=1}^{n}\frac{1}{\vartheta}\mathbb{I}_{[0,\vartheta]}(x_i) \\
&= \frac{1}{\vartheta^n}\prod_{i=1}^{n}\mathbb{I}_{[0,\vartheta]}(x_i) \\
&= \left\{\begin{array}{ll} 1 & \text{se } 0 \le x_i \le \vartheta \ \forall i \\ 0 & \text{altrimenti} \end{array}\right\} \\
&= \frac{1}{\vartheta^n}\mathbb{I}_{[0,\vartheta]}x_{(n)} \\
&= \frac{1}{\vartheta^n}\mathbb{I}_{[0,\vartheta]}x_{(n)} \cdot 1 \\
&= g(T(\overrightarrow{x}),\overrightarrow{\vartheta})\cdot h(\overrightarrow{x})
\end{aligned}$$

La scomposizione è possibile, quindi $T(\overrightarrow{X}) = X_{(n)}$ è statistica sufficiente per ϑ.

Esempio. Prendiamo il **campione normale** $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ con $\overrightarrow{\vartheta} = (\mu, \sigma^2)$. La sua distribuzione è:

$$\begin{aligned} f(\overrightarrow{x}, \overrightarrow{\vartheta}) &= \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x}_n) + n(\bar{x}_n-\mu)^2 - \right. \\ &\qquad \left. -\sum_{i=1}^{n}(\bar{x}_n-\mu)(x_i-\bar{x}_n)\right\} \\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2 - 2\mu\sum x_i + \mu^2\right\} \end{aligned}$$

Quindi il vettore $T(\overrightarrow{X}) = (\sum_{i=1}^n X_i^2, \sum X_i)$ è statistica sufficiente per (μ, σ^2).

Prendiamo il **campione poissoniano** $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{P}(\lambda)$, dove:

$$h(x) = \frac{1}{\prod_{i=1}^n x_i!}\prod_{i=1}^{n}\mathbb{I}_{[\mathbb{N}]}(x_i)$$

Si ha che:

$$\begin{aligned} f(\overrightarrow{x}, \lambda) &= \frac{e^{-\lambda n}\prod_{i=1}^n \lambda^{x_i}}{\prod_{i=1}^n x_i!}\prod_{i=1}^{n}\mathbb{I}_{[\mathbb{N}]}(x_i) \\ &= \frac{1}{\prod_{i=1}^n x_i!}e^{-\lambda n}\lambda^{\sum_{i=1}^n x_i}\prod_{i=1}^{n}\mathbb{I}_{[\mathbb{N}]}(x_i) \end{aligned}$$

Le distribuzioni sopracitate appartengono tutte alla famiglia esponenziale:

Definizione 1.7: Famiglia esponenziale

Una distribuzione $f(x, \vartheta)$ appartiene alla **famiglia esponenziale** se può essere scritta nel seguente modo:

$$f(x, \vartheta) = h(x)c(\vartheta)\exp\left\{\sum_{j=1}^{k} w_j(\overrightarrow{\vartheta}) \cdot t_j(x)\right\} \tag{1.2}$$

	$h(x)$	$c(\vartheta)$	$w_j(\vec{\vartheta})$	$t_j(x)$	$f(\vec{x},\vartheta)$
Bernoulli	$\mathbb{I}_{\{0,1\}}(x)$	$(1-p)$	$(\frac{p}{1-p})$	$\exp(x)$	$(1-p)\mathbb{I}_{\{0,1\}}(x)\exp\left\{x\log\left(\frac{p}{1-p}\right)\right\}$
Poisson	$\frac{1}{x!}$	$\exp\{-\lambda\}$	λ	$\exp(x)$	$\frac{1}{x!}\exp\{-\lambda\}\exp\{x\log(\lambda)\}$
Normale	1	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{\mu^2}{2\sigma^2}\right\}$	$\begin{pmatrix} -\frac{1}{2\sigma^2} \\ \frac{\mu}{\sigma^2} \end{pmatrix}$	$\begin{pmatrix} x^2 \\ x \end{pmatrix}$	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{\mu^2}{2\sigma^2}\right\}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}$

Tabella 1.1: coefficienti funzionali per esprimere alcune distribuzioni notevoli nella forma 1.2

Nella tabella 1.1 si possono trovare i coefficienti funzionali per esprimere le distribuzioni viste finora nella forma enunciata.

Definizione 1.8: Minimalità

Una statistica $T(\vec{x})$ sufficiente è anche **minimale**, se per ogni altra statistica sufficiente $T'(\vec{x})$, $T(\vec{x})$ è una funzione di $T'(\vec{x})$. Equivalentemente, si dice che $T(\vec{x})$ sufficiente è minimale se, per ogni $\vec{x}$, $\vec{y}$, è verificata la seguente implicazione:

$$T'(\vec{x}) = T(\vec{y}) \implies T(\vec{x}) = T(\vec{y})$$

Enunciamo ora un criterio di sufficienza e minimalità per statistiche.

Teorema 1.9: di Lehmann–Scheffé

Sia $f(\vec{x},\vartheta)$ la densità di probabilità congiunta di $\vec{x}$. Si supponga che esista una funzione $T(\vec{x})$ tale che, per ogni $\vec{x}$, $\vec{y}$, è verificata la seguente co-implicazione:

$$\frac{f(\vec{x},\vartheta)}{f(\vec{y},\vartheta)} \text{ è costante in } \vartheta \iff T(\vec{x}) = T(\vec{y})$$

allora, $T(\vec{x})$ è statistica sufficiente e minimale per ϑ.

Esempio: *Completezza dell'uniforme.* Consideriamo il campione casuale

$$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{U}_{[0,\vartheta]},$$

e la statistica $T = X_{(n)}$. La sua densità è data da:

$$f(\vec{x},\vartheta) = n\frac{t^{n-1}}{\vartheta^n}\mathbb{I}_{[0,\vartheta]}(t) \tag{1.3}$$

Consideriamo infine g tale che il suo valore atteso sia nullo:

$$\mathbb{E}_\vartheta[g(T)] = 0 \ \forall \vartheta \tag{1.4}$$

Possiamo riscrivere 1.4 utilizzando la densità della statistica 1.3:

$$0 = \int_0^\vartheta g(t) n \frac{t^{n-1}}{\vartheta^n} \mathrm{d}t$$

Derivando in ϑ si ottiene:

$$\left(\frac{\mathrm{d}}{\mathrm{d}\vartheta}\frac{1}{\vartheta^n}\right) \cdot \left(\int_0^\vartheta n t^{n-1} g(t) dt\right) = \frac{1}{\vartheta^n} n \vartheta^{n-1} g(\vartheta) = 0 \iff g(\vartheta) = 0$$

$X_{(n)}$ è quindi statistica sufficiente minimale e completa.

Teorema 1.10: di Bahadur
Una statistica sufficiente e completa è anche minimale.

Teorema 1.11
Sia $X_1, \ldots, X_n$ un campione casuale nella famiglia esponenziale (come da definizione 1.7):

$$f(x, \vartheta) = h(x) g(\vartheta) \exp\left[\sum_{j=1}^{K} t_j(x) w_j(\vartheta)\right]$$

Se l'insieme

$$\{w_n(\vartheta), \ldots, w_k(\vartheta), \quad \vec{\vartheta} \in \Theta\}$$

contiene un aperto di $\mathbb{R}^k$, allora la statistica

$$\vec{T} = \left(\sum_{i=1}^{n} t_1(x_i), \ldots, \sum_{i=1}^{n} t_k(x_i)\right)$$

è completa. Poiché $\vec{T}$ è anche sufficiente, segue dal teorema di Bahadur che sia minimale.

Esercizio. Dimostrare il teorema 1.11. Suggerimento: basta verificare che lo spazio immagine contenga almeno un aperto.