Diverse derivazioni della distribuzione normale

S. Bonaccorsi

Department of Mathematics
University of Trento

Corso di *Mathematical model for the Physical, Natural and Social Sciences*

Introduzione

Nel calcolo delle probabilità, sembra esistere una distribuzione centrale, universale, verso cui tutte le altre distribuzioni tendono – sotto un ampio spettro di possibili operazioni – e che, una volta raggiunta, rimane stabile per un insieme di operazioni ancora più vasto: la distribuzione gaussiana

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}x^2\right)$$

• Il famoso teorema limite centrale rappresenta sicuramente il più famoso esempio di attrazione della legge normale. Già Laplace (1783) aveva osservato che una legge binomiale tende asintoticamente a una legge gaussiana quando il numero di tentativi diventa grande.

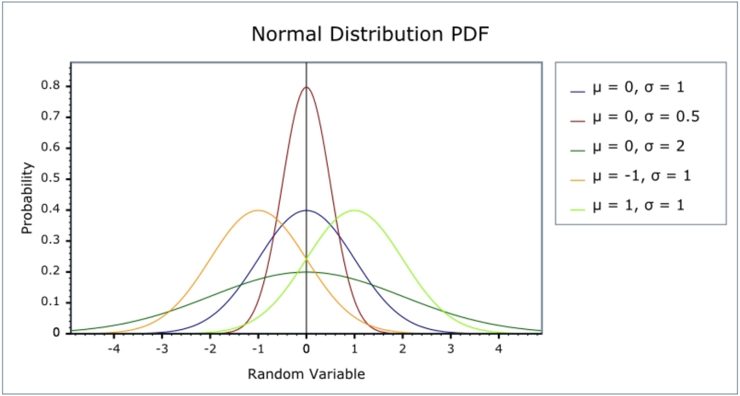
In fisica, queste proprietà di stabilità e di attrazione hanno reso la distribuzione gaussiana la base per la teoria cinetica e per la meccanica statistica.

• Nel campo della biologia, si è dimostrato lo strumento migliore per discutere le dinamiche delle popolazioni e le loro evoluzioni.

• Senza dubbio, risulta essere altrettanto importante in economia (si pensi ai sondaggi oppure alla formula di Black e Scholes); sorprendentemente, il suo uso ovunque sembra essere accompagnato da dubbi e necessità di giustificazioni.

La legge gaussiana

Il grafico sottostante mostra l'andamento della legge gaussiana, anche se la forma "a campana" è sicuramente ben nota.



La funzione di ripartizione sarà indicata dalla funzione

$$\Phi(x) = \int_{-\infty}^{x} \varphi(y)\,\mathrm{d}y.$$

Introduzione

My own impression . . . is that the mathematical results have outrun their interpretation and that some simple explanation of the force and meaning of the celebrated integral . . . will one day be found . . . which will at once render useless all the works hitherto written.

Augustus de Morgan (1838)

In queste parole, de Morgan esprimeva la sua sorpresa al sorprendentemente vasto successo dei metodi di inferenza basati sulla legge degli errori gaussiana, o normale, anche nei casi in cui questa legge non sembra assolutamente essere un modello plausibile della reale frequenza degli errori.

My own impression ... is that the mathematical results have outrun their interpretation and that some simple explanation of the force and meaning of the celebrated integral ... will one day be found ... which will at once render useless all the works hitherto written.

Augustus de Morgan (1838)

Il punto cruciale di queste note sarà mettere in evidenza che la distribuzione di probabilità gaussiana ha un ruolo centrale nel calcolo delle probabilità non tanto per le possibili (ma, per quanto vedremo, irrilevanti) connessioni con le frequenze (relative o assolute) ma per il contenuto di informazione ad essa associato.

Introduzione

$$\frac{1}{2\pi\sigma^2} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Sommario

- Panoramica storica
- Derivazione di Gauss
- Derivazione di Herschel
- Derivazione di Landon
- Derivazione di Shannon-Rao

...Physicists believe that the Gaussian law has been proved in mathematics while mathematicians think that it was experimentally established in physics.

H. Poincaré

La storia in breve

Abraham de Moivre (1667-1754)

- Accidentalmente.
- Scarso interesse.

Pierre-Simon Laplace (1749-1827)

- Note proprietà.
- $e^{-|x|}$.

Carl Friedrich Gauss (1777-1855)

- *Normale*.
- Derivazione (1809).

Sir John Herschel (1792- 1871)

- Astronomo.
- Derivazione (1850).

Vernon D. Landon (prima metà 1900)

- Ingegnere.
- Derivazione (1941).

Shannon-Rao (metà 1900)

- Entropia.
- Derivazione.

La storia in breve

Abraham de Moivre (1667-1754)

- Accidentalmente.
- Scarso interesse.

Pierre-Simon Laplace (1749-1827)

- Note proprietà.
- $e^{-|x|}$.

Carl Friedrich Gauss (1777-1855)

- *Normale*.
- Derivazione (1809).

Sir John Herschel (1792- 1871)

- Astronomo.
- Derivazione (1850).

Vernon D. Landon (prima metà 1900)

- Ingegnere.
- Derivazione (1941).

Shannon-Rao (metà 1900)

- Entropia.
- Derivazione.

La storia in breve

Abraham de Moivre (1667-1754)

- Accidentalmente.
- Scarso interesse.

Pierre-Simon Laplace (1749-1827)

- Note proprietà.
- $e^{-|x|}$.

Carl Friedrich Gauss (1777-1855)

- *Normale.*
- Derivazione (1809).

Sir John Herschel (1792- 1871)

- Astronomo.
- Derivazione (1850).

Vernon D. Landon (prima metà 1900)

- Ingegnere.
- Derivazione (1941).

Shannon-Rao (metà 1900)

- Entropia.
- Derivazione.

La storia in breve

Abraham de Moivre (1667-1754)

- Accidentalmente.
- Scarso interesse.

Pierre-Simon Laplace (1749-1827)

- Note proprietà.
- $e^{-|x|}$.

Carl Friedrich Gauss (1777-1855)

- *Normale*.
- Derivazione (1809).

Sir John Herschel (1792- 1871)

- Astronomo.
- Derivazione (1850).

Vernon D. Landon (prima metà 1900)

- Ingegnere.
- Derivazione (1941).

Shannon-Rao (metà 1900)

- Entropia.
- Derivazione.

La storia in breve

Abraham de Moivre (1667-1754)

- Accidentalmente.
- Scarso interesse.

Pierre-Simon Laplace (1749-1827)

- Note proprietà.
- $e^{-|x|}$.

Carl Friedrich Gauss (1777-1855)

- *Normale.*
- Derivazione (1809).

Sir John Herschel (1792- 1871)

- Astronomo.
- Derivazione (1850).

Vernon D. Landon (prima metà 1900)

- Ingegnere.
- Derivazione (1941).

Shannon-Rao (metà 1900)

- Entropia.
- Derivazione.

La storia in breve

Abraham de Moivre (1667-1754)

- Accidentalmente.
- Scarso interesse.

Pierre-Simon Laplace (1749-1827)

- Note proprietà.
- $e^{-|x|}$.

Carl Friedrich Gauss (1777-1855)

- *Normale.*
- Derivazione (1809).

Sir John Herschel (1792- 1871)

- Astronomo.
- Derivazione (1850).

Vernon D. Landon (prima metà 1900)

- Ingegnere.
- Derivazione (1941).

Shannon-Rao (metà 1900)

- Entropia.
- Derivazione.

Premessa storica

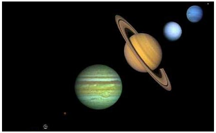
In astronomia, prendere la media di tante osservazioni ha il grande vantaggio che gli errori casuali tendono a cancellarsi.
Ipparco, nel secondo secolo a.C., aveva stimato la precessione degli equinozi facendo la media tra le misure di varie stelle.

Nel corso del XVI secolo,
prendere la media di numerose osservazioni era prassi comune per Tycho Brahe. Molto prima che vi fosse una giustificazione teorica, l'intuizione diceva agli astronomi professionisti che questo metodo di lavoro era la cosa giusta da fare per trattare dati soggetti a errori casuali.

La grande diseguaglianza di Giove e Saturno

Giove e Saturno

Newton, nei *Principia* (1687) aveva potuto calcolare la massa di Giove e Saturno utilizzando il periodo e le distanze dei loro satelliti e concludeva che "l'azione di Giove su Saturno non si può trascurare": l'attrazione di Giove su Saturno alla congiunzione è circa 1:215 volte quella del sole. Tuttavia, Newton non era in grado di determinare come si muoveva il sistema completo.



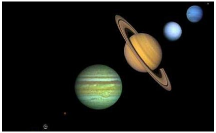
Eulero fu il primo a studiare le perturbazioni dell'orbita della Terra, in una memoria che vinse il premio dell'Accademia di Parigi nel 1756. Nel 1771, Lalande pubblica delle nuove tavole delle perturbazioni periodiche delle orbite di Giove e Saturno ed egli, evidenziando lo strano andamento che aveva notato nel moto di Saturno, afferma che non poteva sperare che le sue tavole fossero corrette anche nel futuro.

Se questo andamento fosse stato confermato, infatti, il sistema solare sarebbe stato instabile e si sarebbe distrutto.

La grande diseguaglianza di Giove e Saturno

Giove e Saturno

Newton, nei *Principia* (1687) aveva potuto calcolare la massa di Giove e Saturno utilizzando il periodo e le distanze dei loro satelliti e concludeva che "l'azione di Giove su Saturno non si può trascurare": l'attrazione di Giove su Saturno alla congiunzione è circa 1:215 volte quella del sole. Tuttavia, Newton non era in grado di determinare come si muoveva il sistema completo.

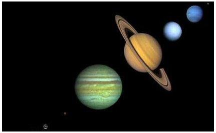


Eulero fu il primo a studiare le perturbazioni dell'orbita della Terra, in una memoria che vinse il premio dell'Accademia di Parigi nel 1756. Nel 1771, Lalande pubblica delle nuove tavole delle perturbazioni periodiche delle orbite di Giove e Saturno ed egli, evidenziando lo strano andamento che aveva notato nel moto di Saturno, afferma che non poteva sperare che le sue tavole fossero corrette anche nel futuro.
Se questo andamento fosse stato confermato, infatti, il sistema solare sarebbe stato instabile e si sarebbe distrutto.

La grande diseguaglianza di Giove e Saturno

Giove e Saturno

Nella seconda metà del XVIII secolo, quindi, gli astronomi si trovavano nella condizione di non sapere se la teoria Newtoniana fosse sufficiente per calcolare la posizione dei pianeti. Eulero e Lagrange, tra gli altri, avevano affrontato il problema, ma pur essendo in contraddizione tra loro, non erano riusciti trovare una soluzione.

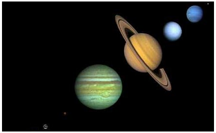


Eulero (1749), si era trovato di fronte a questo problema: come stimare gli otto parametri orbitali da 75 osservazioni che coprivano un periodo di 164 anni. Eulero non fu sconfitto dalla massa di calcoli necessaria: non fu proprio in grado di capire come si poteva risolvere il problema. Invece di vedere come, combinando tante osservazioni, gli errori tendono a cancellarsi, pensava che le cose tendono a peggiorare: in qualche modo, si concentrava sul caso peggiore nella certezza che fosse quello che poi accadeva in realtà . . . which makes him perhaps the first really devout believer in Murphy's Law (Jaynes).

La grande diseguaglianza di Giove e Saturno

Giove e Saturno

Nella seconda metà del XVIII secolo, quindi, gli astronomi si trovavano nella condizione di non sapere se la teoria Newtoniana fosse sufficiente per calcolare la posizione dei pianeti. Eulero e Lagrange, tra gli altri, avevano affrontato il problema, ma pur essendo in contraddizione tra loro, non erano riusciti trovare una soluzione.



Eulero (1749), si era trovato di fronte a questo problema: come stimare gli otto parametri orbitali da 75 osservazioni che coprivano un periodo di 164 anni. Eulero non fu sconfitto dalla massa di calcoli necessaria: non fu proprio in grado di capire come si poteva risolvere il problema. Invece di vedere come, combinando tante osservazioni, gli errori tendono a cancellarsi, pensava che le cose tendono a peggiorare: in qualche modo, si concentrava sul caso peggiore nella certezza che fosse quello che poi accadeva in realtà . . . which makes him perhaps the first really devout believer in Murphy's Law (Jaynes).

La grande diseguaglianza di Giove e Saturno

Théorie de Jupiter et de Saturne

La svolta arriva nel 1785: Laplace presenta all'Accademia di Parigi una memoria dove mostra che le anomalie nel moto di Giove e Saturno possono essere ricondotte all'interno della teoria della gravitazione universale. Laplace "salva il mondo" mostrando che la teoria della probabilità può essere usata per stimare i parametri in maniera più accurata.

Laplace, invece, capì che doveva esistere una legge empirica degli errori come limite delle distribuzioni binomiali e ne aveva derivato le prime proprietà, pur avendo mancato la forma corretta (che sarebbe stata trovata 24 anni dopo da Gauss) e avendo concentrato i suoi sforzi sulla distribuzione $f(x) \propto \exp(-\alpha|x|)$. In questo modo, usando il principio qualitativo che la combinazione delle osservazioni migliora l'accuratezza delle stime, fu in grado di risolvere il problema di Giove e Saturno.

La grande diseguaglianza di Giove e Saturno

Théorie de Jupiter et de Saturne

La svolta arriva nel 1785: Laplace presenta all'Accademia di Parigi una memoria dove mostra che le anomalie nel moto di Giove e Saturno possono essere ricondotte all'interno della teoria della gravitazione universale. Laplace "salva il mondo" mostrando che la teoria della probabilità può essere usata per stimare i parametri in maniera più accurata.

Laplace, invece, capì che doveva esistere una legge empirica degli errori come limite delle distribuzioni binomiali e ne aveva derivato le prime proprietà, pur avendo mancato la forma corretta (che sarebbe stata trovata 24 anni dopo da Gauss) e avendo concentrato i suoi sforzi sulla distribuzione $f(x) \propto \exp(-\alpha|x|)$. In questo modo, usando il principio qualitativo che la combinazione delle osservazioni migliora l'accuratezza delle stime, fu in grado di risolvere il problema di Giove e Saturno.

La derivazione di Gauss

Vogliamo stimare la media θ di una popolazione a partire da $(n+1)$ osservazioni $(x_0, \ldots, x_n)$ con una stima di massima verosimiglianza.

Se le osservazioni sono supposte indipendenti, la distribuzione congiunta delle osservazioni si può fattorizzare:
$p(x_0, \ldots, x_n \mid \theta) = f(x_0 \mid \theta) \cdots f(x_n \mid \theta)$

la stima di massima verosimiglianza diventa

$$\sum_{i=0}^{n} \frac{\partial}{\partial \theta} \log(f(x_i \mid \theta)) = 0 \tag{1}$$

ossia, scrivendo $\log(f(x \mid \theta)) = g(\theta - x) = g(u)$,
lo stimatore di massima verosimiglianza $\hat{\theta}$ deve soddisfare

$$\sum_{i=0}^{n} g'(\hat{\theta} - x_i) = 0. \tag{2}$$

La derivazione di Gauss

Ora, l'intuizione suggerisce che la stima deve essere anche la media aritmetica delle osservazioni:

$$\hat{\theta} = \bar{x} = \frac{1}{n+1} \sum_{i=0}^{n} x_i, \tag{3}$$

ma queste due equazioni sono, in generale, incompatibili: il valore ottenuto in (3) non è una soluzione dell'equazione (2) per una funzione arbitraria g.

La derivazione di Gauss

Necessità

Per vedere se esistono distribuzioni di probabilità per cui la stima di massima verosimiglianza coincide con la media aritmetica, consideriamo un possibile insieme di osservazioni, in cui una sola osservazione x_0 è diversa da zero: poniamo

$$x_0 = (n+1)u,\ x_1 = \cdots = x_n = 0,$$

per un qualche parametro u; allora risulta $\bar{x} = \hat{\theta} = u$ e $\hat{\theta} - x_0 = -nu$ mentre $\hat{\theta} - x_i = u$ per ogni $i = 1, \ldots, n$.

Sostituendo in (2) si ottiene $g'(-n\,u) + n\,g'(u) = 0$, per ogni $n = 1, 2, 3, \ldots$.
Il caso $n = 1$ ci mostra che $g'(u)$ deve essere una funzione antisimmetrica: $g'(-u) = -g'(u)$, da cui otteniamo l'equazione

$$g'(n\,u) = n\,g'(u),$$
$$u \in \mathbb{R}, \quad n = 1, 2, 3, \ldots.$$

La derivazione di Gauss

Necessità

Per vedere se esistono distribuzioni di probabilità per cui la stima di massima verosimiglianza coincide con la media aritmetica, consideriamo un possibile insieme di osservazioni, in cui una sola osservazione x_0 è diversa da zero: poniamo

$$x_0 = (n+1)u, \; x_1 = \cdots = x_n = 0,$$

per un qualche parametro u; allora risulta $\bar{x} = \hat{\theta} = u$ e $\hat{\theta} - x_0 = -nu$ mentre $\hat{\theta} - x_i = u$ per ogni $i = 1, \ldots, n$.

Sostituendo in (2) si ottiene $g'(-n\,u) + n\,g'(u) = 0$, per ogni $n = 1, 2, 3, \ldots$.
Il caso $n = 1$ ci mostra che $g'(u)$ deve essere una funzione antisimmetrica: $g'(-u) = -g'(u)$, da cui otteniamo l'equazione

$$g'(n\,u) = n\,g'(u),$$
$$u \in \mathbb{R}, \quad n = 1, 2, 3, \ldots .$$

La derivazione di Gauss

Osservazione.

La soluzione continua dell'equazione $h(nu) = nh(u)$ (per $u \in \mathbb{R}$ e $n \in \mathbb{N}$) è la funzione lineare $h(x) = \alpha x$.
Infatti, possiamo vedere che $h(n\frac{1}{n}) = nh(\frac{1}{n})$ ossia $h(\frac{1}{n}) = \frac{1}{n}h(1)$ per ogni n; allora per ogni $m \in \mathbb{N}$ $h(\frac{m}{n}) = mh(\frac{1}{n}) = \frac{m}{n}h(1)$ e quindi per ogni $x \in \mathbb{Q}$, $h(x) = xh(1)$. Per continuità, ponendo $h(1) = \alpha$, si ha la tesi.

Se $g'(u) = \alpha u$ allora $g(u) = \frac{1}{2}\alpha u^2 + b$.

Usiamo di nuovo l'identità $\log(f(x \mid \theta)) = g(\theta - x) = g(u)$. Perché f sia una distribuzione di probabilità, deve essere α negativo e la normalizzazione determina la costante b.

La derivazione di Gauss

Necessità

Abbiamo visto – con la scelta di un particolare insieme di dati – che condizione necessaria per l'eguaglianza massima verosimiglianza = media aritmetica è che la distribuzione abbia la forma

$$f(x \mid \theta) = \sqrt{\frac{\alpha}{2\,\pi}} \exp\left(-\tfrac{1}{2}\,\alpha\,(x-\theta)^2\right) . \tag{4}$$

Sufficienza

Se la distribuzione f ha la legge gaussiana (4), allora l'equazione di verosimiglianza

$$\sum_{i=0}^{n} \frac{\partial}{\partial\theta} \log(f(x_i \mid \theta)) = 0$$

ha sempre la soluzione $\hat{\theta} = \bar{x}$.

Derivazione

Abbiamo mostrato che condizione necessaria e sufficiente per l'eguaglianza di media aritmetica e stima di massima verosimiglianza è che la densità sia gaussiana. La sola libertà è nella scelta del parametro di scala α.

La derivazione di Gauss

Necessità

Abbiamo visto – con la scelta di un particolare insieme di dati – che condizione necessaria per l'eguaglianza massima verosimiglianza = media aritmetica è che la distribuzione abbia la forma

$$f(x \mid \theta) = \sqrt{\frac{\alpha}{2\,\pi}}\, \exp\left(-\tfrac{1}{2}\,\alpha\,(x-\theta)^2\right). \tag{4}$$

Sufficienza

Se la distribuzione f ha la legge gaussiana (4), allora l'equazione di verosimiglianza

$$\sum_{i=0}^{n} \frac{\partial}{\partial\theta} \log(f(x_i \mid \theta)) = 0$$

ha sempre la soluzione $\hat{\theta} = \bar{x}$.

Derivazione

Abbiamo mostrato che condizione necessaria e sufficiente per l'eguaglianza di media aritmetica e stima di massima verosimiglianza è che la densità sia gaussiana. La sola libertà è nella scelta del parametro di scala α.

La derivazione di Gauss

Necessità

Abbiamo visto – con la scelta di un particolare insieme di dati – che condizione necessaria per l'eguaglianza massima verosimiglianza = media aritmetica è che la distribuzione abbia la forma

$$f(x \mid \theta) = \sqrt{\frac{\alpha}{2\,\pi}}\, \exp\left(-\tfrac{1}{2}\,\alpha\,(x-\theta)^2\right). \tag{4}$$

Sufficienza

Se la distribuzione f ha la legge gaussiana (4), allora l'equazione di verosimiglianza

$$\sum_{i=0}^{n} \frac{\partial}{\partial\theta} \log(f(x_i \mid \theta)) = 0$$

ha sempre la soluzione $\hat{\theta} = \bar{x}$.

Derivazione

Abbiamo mostrato che condizione necessaria e sufficiente per l'eguaglianza di media aritmetica e stima di massima verosimiglianza è che la densità sia gaussiana. La sola libertà è nella scelta del parametro di scala α.

La derivazione di Gauss

Questa derivazione della legge normale fu trovata da Gauss in un passaggio di un lavoro dedicato all'astronomia.
Avrebbe potuto passare inosservata, se Laplace non l'avesse notata e non avesse richiamato l'attenzione su di essa pubblicando una serie di lavori (1810, 1812) in cui dimostrava molte utili proprietà della legge normale (4) come distribuzione di probabilità, tra cui il teorema centrale del limite e la soluzione generale del problema della riduzione delle osservazioni, che è il modo in cui ancora oggi affrontiamo il problema.

Da Laplace prese il nome di "distribuzione gaussiana". In effetti, la stessa distribuzione era stata trovata, in maniera più o meno accidentale, da de Moivre (1733), che non ne aveva compreso a l'importanza e non l'aveva poi utilizzata.

Not until the time of Einstein did such a simple mathematical argument again have such a great effect on scientific practice (Jaynes).

La derivazione di Gauss

Questa derivazione della legge normale fu trovata da Gauss in un passaggio di un lavoro dedicato all'astronomia.
Avrebbe potuto passare inosservata, se Laplace non l'avesse notata e non avesse richiamato l'attenzione su di essa pubblicando una serie di lavori (1810, 1812) in cui dimostrava molte utili proprietà della legge normale (4) come distribuzione di probabilità, tra cui il teorema centrale del limite e la soluzione generale del problema della riduzione delle osservazioni, che è il modo in cui ancora oggi affrontiamo il problema.

Da Laplace prese il nome di "distribuzione gaussiana". In effetti, la stessa distribuzione era stata trovata, in maniera più o meno accidentale, da de Moivre (1733), che non ne aveva compreso a l'importanza e non l'aveva poi utilizzata.

Not until the time of Einstein did such a simple mathematical argument again have such a great effect on scientific practice (Jaynes).

La derivazione di Gauss

Questa derivazione della legge normale fu trovata da Gauss in un passaggio di un lavoro dedicato all'astronomia.
Avrebbe potuto passare inosservata, se Laplace non l'avesse notata e non avesse richiamato l'attenzione su di essa pubblicando una serie di lavori (1810, 1812) in cui dimostrava molte utili proprietà della legge normale (4) come distribuzione di probabilità, tra cui il teorema centrale del limite e la soluzione generale del problema della riduzione delle osservazioni, che è il modo in cui ancora oggi affrontiamo il problema.

Da Laplace prese il nome di "distribuzione gaussiana". In effetti, la stessa distribuzione era stata trovata, in maniera più o meno accidentale, da de Moivre (1733), che non ne aveva compreso a l'importanza e non l'aveva poi utilizzata.

Not until the time of Einstein did such a simple mathematical argument again have such a great effect on scientific practice (Jaynes).

La derivazione di Herschel

In questa sezione mostriamo, seguendo il lavoro dell'astronomo John Herschel (1850), come la distribuzione gaussiana sia lo strumento naturale nello studio della distribuzione degli errori di misura nella posizione di una stella.

Sia allora x l'errore nella direzione longitudinale (est – ovest) e y l'errore nella declinazione (nord – sud) e ci chiediamo qual è la distribuzione congiunta $\rho(x, y)$.

Herschel fece due ipotesi (che chiameremo postulati *(P1), (P2)* che sembrano essere intuitivamente richiesti da considerazioni geometriche:

Postulato (P1)

La conoscenza di x non ci dice nulla su y.

In questo modo, richiediamo che gli errori di misura nelle due direzioni ortogonali sono indipendenti. Sappiamo, dalla teoria, che questo implica che la distribuzione congiunta deve avere la forma

$$\rho(x, y)\, \mathrm{d}x\, \mathrm{d}y = f(x)\, \mathrm{d}x \times f(y)\, \mathrm{d}y.$$

La derivazione di Herschel

In questa sezione mostriamo, seguendo il lavoro dell'astronomo John Herschel (1850), come la distribuzione gaussiana sia lo strumento naturale nello studio della distribuzione degli errori di misura nella posizione di una stella.
Sia allora x l'errore nella direzione longitudinale (est – ovest) e y l'errore nella declinazione (nord – sud) e ci chiediamo qual è la distribuzione congiunta $\rho(x, y)$.
Herschel fece due ipotesi (che chiameremo postulati *(P1), (P2)* che sembrano essere intuitivamente richiesti da considerazioni geometriche:

Postulato (P1)

La conoscenza di x non ci dice nulla su y.

In questo modo, richiediamo che gli errori di misura nelle due direzioni ortogonali sono indipendenti. Sappiamo, dalla teoria, che questo implica che la distribuzione congiunta deve avere la forma

$$\rho(x, y)\,\mathrm{d}x\,\mathrm{d}y = f(x)\,\mathrm{d}x \times f(y)\,\mathrm{d}y.$$

La derivazione di Herschel

Possiamo anche scrivere la distribuzione in coordinate polari ρ, θ definite dalle relazion $x = r\cos(\theta)$, $y = r\sin(\theta)$:

$$\rho(x, y)\,\mathrm{d}x\,\mathrm{d}y = g(r, \theta)\,r\,\mathrm{d}r\,\mathrm{d}\theta.$$

Postulato (P2)

Questa probabilità deve essere indipendente dall'angolo: $g(r, \theta) = g(r)$.

Allora, le diseguaglianze precedenti portano all'equazione funzionale

$$f(x)f(y) = g(\sqrt{x^2 + y^2}). \tag{5}$$

Ponendo $y = 0$ si ottiene $\boxed{g(x) = f(x)f(0)}$.

Dividiamo ora ambo i lati di (5) per $f(0)^2$ e utilizziamo la relazione precedente: otteniamo allora

$$\log\left(\frac{f(x)}{f(0)}\right) + \log\left(\frac{f(y)}{f(0)}\right) = \log\left(\frac{f(\sqrt{x^2 + y^2})}{f(0)}\right).$$

La derivazione di Herschel

Possiamo anche scrivere la distribuzione in coordinate polari ρ, θ definite dalle relazion $x = r\cos(\theta)$, $y = r\sin(\theta)$:

$$\rho(x, y)\,\mathrm{d}x\,\mathrm{d}y = g(r, \theta)\,r\,\mathrm{d}r\,\mathrm{d}\theta.$$

Postulato (P2)

Questa probabilità deve essere indipendente dall'angolo: $g(r, \theta) = g(r)$.

Allora, le diseguaglianze precedenti portano all'equazione funzionale

$$f(x)f(y) = g(\sqrt{x^2 + y^2}). \tag{5}$$

Ponendo $y = 0$ si ottiene $\boxed{g(x) = f(x)f(0)}$.
Dividiamo ora ambo i lati di (5) per $f(0)^2$ e utilizziamo la relazione precedente: otteniamo allora

$$\log\left(\frac{f(x)}{f(0)}\right) + \log\left(\frac{f(y)}{f(0)}\right) = \log\left(\frac{f(\sqrt{x^2 + y^2})}{f(0)}\right).$$

La derivazione di Herschel

La soluzione generale dell'equazione funzionale

$$\log\left(\frac{f(x)}{f(0)}\right) + \log\left(\frac{f(y)}{f(0)}\right) = \log\left(\frac{f(\sqrt{x^2+y^2})}{f(0)}\right).$$

è ovvia: dato che la funzione calcolata in x sommata alla funzione calcolata in y deve essere funzione di $x^2 + y^2$, l'unica possibilità è che sia
$\log\left[\frac{f(x)}{f(0)}\right] = a\,x^2.$
Perché questa sia una distribuzione di probabilità, deve essere $a < 0$; allora, normalizzare porta a determinare il valore di $f(0)$ e la soluzione deve avere la forma

$$f(x) = \sqrt{\frac{\alpha}{\pi}}\exp(-\alpha x^2),$$

dove rimane un parametro indeterminato.
La sola densità di probabilità nel piano che soddisfa i postulati di invarianza di Herschel è una legge gaussiana simmetrica:

$$\rho(x,y) = \frac{\alpha}{\pi}\exp\left(-\alpha(x^2+y^2)\right).$$

La derivazione di Herschel

Dieci anni dopo, James Clerk Maxwell (1860) diede una versione tri-dimensionale dello stesso argomento per trovae la distribuzione di probabilità $\rho(v_x, v_y, v_z) \propto \exp\left(-\alpha\,|v|^2\right)$ per la velocità delle molecole in un gas, che divenne in seguito nota ai fisici come la legge di Maxwell di distribuzione delle velocità fondamentale nella teoria cinetica e nella meccanica statistica.

La dimostrazione di Herschel – Maxwell è interessante perchè due condizioni qualitative, in generale incompatibili, diventano compatibili per una precisa distribuzione di probabilità, che quindi determinano univocamente. Einstein (1905) ha usato la stessa costruzione per dedurre la legge di trasformazione di Lorentz dai suoi postulati della teoria della relatività.

La dimostrazione di Herschel – Maxwell è economica nel senso che non fa uso di nessuno strumento della teoria delle probabilità ma solo di proprietà geometriche di invarianza che possono essere applicate altrettanto bene in altri contesti.
Le distribuzioni gaussiane sono oggetti speciali per pure ragioni matematiche.

La derivazione di Herschel

Dieci anni dopo, James Clerk Maxwell (1860) diede una versione tri-dimensionale dello stesso argomento per trovae la distribuzione di probabilità $\rho(v_x, v_y, v_z) \propto \exp\left(-\alpha\,|v|^2\right)$ per la velocità delle molecole in un gas, che divenne in seguito nota ai fisici come la legge di Maxwell di distribuzione delle velocità fondamentale nella teoria cinetica e nella meccanica statistica.

La dimostrazione di Herschel – Maxwell è interessante perchè due condizioni qualitative, in generale incompatibili, diventano compatibili per una precisa distribuzione di probabilità, che quindi determinano univocamente. Einstein (1905) ha usato la stessa costruzione per dedurre la legge di trasformazione di Lorentz dai suoi postulati della teoria della relatività.

La dimostrazione di Herschel – Maxwell è economica nel senso che non fa uso di nessuno strumento della teoria delle probabilità ma solo di proprietà geometriche di invarianza che possono essere applicate altrettanto bene in altri contesti.
Le distribuzioni gaussiane sono oggetti speciali per pure ragioni matematiche.

La derivazione di Herschel

Dieci anni dopo, James Clerk Maxwell (1860) diede una versione tri-dimensionale dello stesso argomento per trovae la distribuzione di probabilità $\rho(v_x, v_y, v_z) \propto \exp\left(-\alpha\,|v|^2\right)$ per la velocità delle molecole in un gas, che divenne in seguito nota ai fisici come la legge di Maxwell di distribuzione delle velocità fondamentale nella teoria cinetica e nella meccanica statistica.

La dimostrazione di Herschel – Maxwell è interessante perchè due condizioni qualitative, in generale incompatibili, diventano compatibili per una precisa distribuzione di probabilità, che quindi determinano univocamente. Einstein (1905) ha usato la stessa costruzione per dedurre la legge di trasformazione di Lorentz dai suoi postulati della teoria della relatività.

La dimostrazione di Herschel – Maxwell è economica nel senso che non fa uso di nessuno strumento della teoria delle probabilità ma solo di proprietà geometriche di invarianza che possono essere applicate altrettanto bene in altri contesti.
Le distribuzioni gaussiane sono oggetti speciali per pure ragioni matematiche.

Derivazione di Landon

Ipotesi

- Osservazione empirica: la variabilità del rumore elettrico $v(t)$ osservato in un circuito sembra avere una forma generale, indipendente dal tipo di sorgente, possibilmente diversa per livello.
- Distribuzioni gerarchiche a un parametro $p(v \mid \sigma^2)$ dove σ^2 rappresenta la varianza.
- Livello di rumore aumenta $\longrightarrow$ la distribuzione avrà stessa forma ma spostata ad un nuovo livello σ^2.

Nelle parole di Landon: "samples of electrical noise produced by widely different sources ... cannot be distinguished one from the other by any known test."

Derivazione di Landon

Ipotesi

- Osservazione empirica: la variabilità del rumore elettrico $v(t)$ osservato in un circuito sembra avere una forma generale, indipendente dal tipo di sorgente, possibilmente diversa per livello.
- Distribuzioni gerarchiche a un parametro $p(v \mid \sigma^2)$ dove σ^2 rappresenta la varianza.
- Livello di rumore aumenta $\longrightarrow$ la distribuzione avrà stessa forma ma spostata ad un nuovo livello σ^2.

Nelle parole di Landon: "samples of electrical noise produced by widely different sources ... cannot be distinguished one from the other by any known test."

Derivazione di Landon

Input

- Supponiamo che il rumore v' sia descritto dalla distribuzione di probabilità $p(v' \mid \sigma)$.
- Aumentiamo v' di una quantità ε, quindi otteniamo un rumore $v = v' + \varepsilon$, dove ε è piccolo rispetto a σ^2 e ha una distribuzione di probabilità $q(\varepsilon)\mathrm{d}\varepsilon$, indipendente da $p(v \mid \sigma^2)$.
- La distribuzione di v è data dal prodotto di convoluzione

$$f(v) = \int p(v - \varepsilon \mid \sigma^2) q(\varepsilon) \, \mathrm{d}\varepsilon$$

Derivazione di Landon

Input

- Supponiamo che il rumore v' sia descritto dalla distribuzione di probabilità $p(v' \mid \sigma)$.
- Aumentiamo v' di una quantità ε, quindi otteniamo un rumore $v = v' + \varepsilon$, dove ε è piccolo rispetto a σ^2 e ha una distribuzione di probabilità $q(\varepsilon)\mathrm{d}\varepsilon$, indipendente da $p(v \mid \sigma^2)$.
- La distribuzione di v è data dal prodotto di convoluzione

$$f(v) = \int p(v - \varepsilon \mid \sigma^2) q(\varepsilon) \, \mathrm{d}\varepsilon$$

Derivazione di Landon

Input

- Supponiamo che il rumore v' sia descritto dalla distribuzione di probabilità $p(v' \mid \sigma)$.
- Aumentiamo v' di una quantità ε, quindi otteniamo un rumore $v = v' + \varepsilon$, dove ε è piccolo rispetto a σ^2 e ha una distribuzione di probabilità $q(\varepsilon)\mathrm{d}\varepsilon$, indipendente da $p(v \mid \sigma^2)$.
- La distribuzione di v è data dal prodotto di convoluzione

$$f(v) = \int p(v - \varepsilon \mid \sigma^2) q(\varepsilon) \, \mathrm{d}\varepsilon$$

Derivazione di Landon

Espansione asintotica

Facciamo una espansione asintotica in termini di ε

$$f(v) = p(v \mid \sigma^2) - \frac{\partial}{\partial v} p(v \mid \sigma^2) \int \varepsilon q(\varepsilon)\, \mathrm{d}\varepsilon + \frac{1}{2} \frac{\partial^2}{\partial v^2} p(v \mid \sigma^2) \int \varepsilon^2 q(\varepsilon)\, \mathrm{d}\varepsilon + \dots$$

e indicando con $\langle \varepsilon \rangle$ la media di ε:

$$f(v) = p(v \mid \sigma^2) - \frac{\partial}{\partial v} p(v \mid \sigma^2) \langle \varepsilon \rangle + \frac{1}{2} \frac{\partial^2}{\partial v^2} p(v \mid \sigma^2) \langle \varepsilon^2 \rangle + \dots$$

Media

se gli incrementi possono assumere qualunque segno, assumeremo che $\langle \varepsilon \rangle = 0$.

Varianza

La media di v^2 è spostata al livello $\sigma^2 + \langle \varepsilon^2 \rangle$: allora la proprietà di invarianza implica che $f(v)$ deve essere uguale a

$$f(v) = p(v \mid \sigma^2 + \langle \varepsilon^2 \rangle) = p(v \mid \sigma^2) + \langle \varepsilon^2 \rangle \frac{\partial}{\partial (\sigma^2)} p(v \mid \sigma^2) + \dots$$

Derivazione di Landon

Espansione asintotica

Facciamo una espansione asintotica in termini di ε

$$f(v) = p(v \mid \sigma^2) - \frac{\partial}{\partial v} p(v \mid \sigma^2) \int \varepsilon q(\varepsilon)\, \mathrm{d}\varepsilon + \frac{1}{2} \frac{\partial^2}{\partial v^2} p(v \mid \sigma^2) \int \varepsilon^2 q(\varepsilon)\, \mathrm{d}\varepsilon + \dots$$

e indicando con $\langle\varepsilon\rangle$ la media di ε:

$$f(v) = p(v \mid \sigma^2) - \frac{\partial}{\partial v} p(v \mid \sigma^2) \langle\varepsilon\rangle + \frac{1}{2} \frac{\partial^2}{\partial v^2} p(v \mid \sigma^2) \langle\varepsilon^2\rangle + \dots$$

Media

se gli incrementi possono assumere qualunque segno, assumeremo che $\langle\varepsilon\rangle = 0$.

Varianza

La media di v^2 è spostata al livello $\sigma^2 + \langle\varepsilon^2\rangle$: allora la proprietà di invarianza implica che $f(v)$ deve essere uguale a

$$f(v) = p(v \mid \sigma^2 + \langle\varepsilon^2\rangle) = p(v \mid \sigma^2) + \langle\varepsilon^2\rangle \frac{\partial}{\partial(\sigma^2)} p(v \mid \sigma^2) + \dots$$

Derivazione di Landon

Espansione asintotica

Facciamo una espansione asintotica in termini di ε

$$f(v) = p(v \mid \sigma^2) - \frac{\partial}{\partial v} p(v \mid \sigma^2) \int \varepsilon q(\varepsilon)\,\mathrm{d}\varepsilon + \frac{1}{2}\frac{\partial^2}{\partial v^2} p(v \mid \sigma^2) \int \varepsilon^2 q(\varepsilon)\,\mathrm{d}\varepsilon + \ldots$$

e indicando con $\langle\varepsilon\rangle$ la media di ε:

$$f(v) = p(v \mid \sigma^2) - \frac{\partial}{\partial v} p(v \mid \sigma^2)\langle\varepsilon\rangle + \frac{1}{2}\frac{\partial^2}{\partial v^2} p(v \mid \sigma^2)\langle\varepsilon^2\rangle + \ldots$$

Media

se gli incrementi possono assumere qualunque segno, assumeremo che $\langle\varepsilon\rangle = 0$.

Varianza

La media di v^2 è spostata al livello $\sigma^2 + \langle\varepsilon^2\rangle$: allora la proprietà di invarianza implica che $f(v)$ deve essere uguale a

$$f(v) = p(v \mid \sigma^2 + \langle\varepsilon^2\rangle) = p(v \mid \sigma^2) + \langle\varepsilon^2\rangle \frac{\partial}{\partial(\sigma^2)} p(v \mid \sigma^2) + \ldots$$

Derivazione di Landon

- Confrontando le due equazioni, si ottiene che per avere invarianza deve essere

$$\frac{\partial}{\partial(\sigma^2)} p(v \mid \sigma^2) = \frac{1}{2} \frac{\partial^2}{\partial v^2} p(v \mid \sigma^2)$$

- con le condizioni iniziali $p(v \mid \sigma^2 = 0) = \delta(v)$
- l'unica soluzione è data dalla distribuzione gaussiana

$$p(v \mid \sigma^2) = \frac{1}{2\pi\sigma^2} e^{-\frac{v^2}{2\sigma^2}}$$

Conclusioni

La derivazione di Landon rappresenta una estensione del teorema centrale del limite; invece di aggiungere tutti i piccoli contributi tutti insieme, li consideriamo uno alla volta, richiedendo che ad ogni passo la distribuzione di probabilità abbia la stessa forma funzionale (fino al secondo ordine in ε).
Ma questo è proprio il processo in cui il rumore è prodotto in natura – attraverso la somma di molti piccoli contributi, uno alla volta (ad esempio, la collisione di un elettrone con un atomo, che produce un piccolo impulso di onde elettromagnetiche, la cui somma è il rumore osservato).

Derivazione di Landon

- Confrontando le due equazioni, si ottiene che per avere invarianza deve essere

$$\frac{\partial}{\partial(\sigma^2)} p(v \mid \sigma^2) = \frac{1}{2} \frac{\partial^2}{\partial v^2} p(v \mid \sigma^2)$$

- con le condizioni iniziali $p(v \mid \sigma^2 = 0) = \delta(v)$
- l'unica soluzione è data dalla distribuzione gaussiana

$$p(v \mid \sigma^2) = \frac{1}{2\pi\sigma^2} e^{-\frac{v^2}{2\sigma^2}}$$

Conclusioni

La derivazione di Landon rappresenta una estensione del teorema centrale del limite; invece di aggiungere tutti i piccoli contributi tutti insieme, li consideriamo uno alla volta, richiedendo che ad ogni passo la distribuzione di probabilità abbia la stessa forma funzionale (fino al secondo ordine in ε).
Ma questo è proprio il processo in cui il rumore è prodotto in natura – attraverso la somma di molti piccoli contributi, uno alla volta (ad esempio, la collisione di un elettrone con un atomo, che produce un piccolo impulso di onde elettromagnetiche, la cui somma è il rumore osservato).

Derivazione di Landon

- Confrontando le due equazioni, si ottiene che per avere invarianza deve essere

$$\frac{\partial}{\partial(\sigma^2)} p(v \mid \sigma^2) = \frac{1}{2} \frac{\partial^2}{\partial v^2} p(v \mid \sigma^2)$$

- con le condizioni iniziali $p(v \mid \sigma^2 = 0) = \delta(v)$
- l'unica soluzione è data dalla distribuzione gaussiana

$$p(v \mid \sigma^2) = \frac{1}{2\pi\sigma^2} e^{-\frac{v^2}{2\sigma^2}}$$

Conclusioni

La derivazione di Landon rappresenta una estensione del teorema centrale del limite; invece di aggiungere tutti i piccoli contributi tutti insieme, li consideriamo uno alla volta, richiedendo che ad ogni passo la distribuzione di probabilità abbia la stessa forma funzionale (fino al secondo ordine in ε).
Ma questo è proprio il processo in cui il rumore è prodotto in natura – attraverso la somma di molti piccoli contributi, uno alla volta (ad esempio, la collisione di un elettrone con un atomo, che produce un piccolo impulso di onde elettromagnetiche, la cui somma è il rumore osservato).

Derivazione di Landon

- Confrontando le due equazioni, si ottiene che per avere invarianza deve essere

$$\frac{\partial}{\partial(\sigma^2)} p(v \mid \sigma^2) = \frac{1}{2} \frac{\partial^2}{\partial v^2} p(v \mid \sigma^2)$$

- con le condizioni iniziali $p(v \mid \sigma^2 = 0) = \delta(v)$
- l'unica soluzione è data dalla distribuzione gaussiana

$$p(v \mid \sigma^2) = \frac{1}{2\pi\sigma^2} e^{-\frac{v^2}{2\sigma^2}}$$

Conclusioni

La derivazione di Landon rappresenta una estensione del teorema centrale del limite; invece di aggiungere tutti i piccoli contributi tutti insieme, li consideriamo uno alla volta, richiedendo che ad ogni passo la distribuzione di probabilità abbia la stessa forma funzionale (fino al secondo ordine in ε).
Ma questo è proprio il processo in cui il rumore è prodotto in natura – attraverso la somma di molti piccoli contributi, uno alla volta (ad esempio, la collisione di un elettrone con un atomo, che produce un piccolo impulso di onde elettromagnetiche, la cui somma è il rumore osservato).

Derivazione di Shannon - Rao

Entropia nel caso discreto

- informazione:
  $I(E) = -\log_2(\mathbb{P}(E))$
- se X è variabile aleatoria discreta:

$$H(X) = \mathbb{E}[I(X)] = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

- vale la stima $0 \leq H(X) \leq \log_2(n)$
- l'eguaglianza vale solo per la distribuzione uniforme.

Entropia nel caso continuo

- X variabile aleatoria con densità $f(x)$
- l'entropia differenziale di X

$$h(X) = -\int f(x) \log_2 f(x)\, dx$$

- buone proprietà (invarianza per traslazioni, chain rule: $h(X_1, \ldots, X_n) \leq \sum h(X_i)$), ma anche
- non è invariante per cambio di variabili e può essere negativa.

Derivazione di Shannon - Rao

Entropia nel caso discreto

- informazione: $I(E) = -\log_2(\mathbb{P}(E))$
- se X è variabile aleatoria discreta:

$$\begin{aligned} H(X) &= \mathbb{E}[I(X)] \\ &= -\sum_{i=1}^{n} p_i \log_2(p_i) \end{aligned}$$

- vale la stima $0 \leq H(X) \leq \log_2(n)$
- l'eguaglianza vale solo per la distribuzione uniforme.

Entropia nel caso continuo

- X variabile aleatoria con densità $f(x)$
- l'entropia differenziale di X

$$h(X) = -\int f(x) \log_2 f(x) \, \mathrm{d}x$$

- buone proprietà (invarianza per traslazioni, chain rule: $h(X_1, \ldots, X_n) \leq \sum h(X_i)$), ma anche
- non è invariante per cambio di variabili e può essere negativa.

Derivazione di Shannon - Rao

Per dimostrare che l'entropia massima di una distribuzione discreta si ottiene nel caso di distribuzione uniforme è necessario utilizzare il seguente risultato.

Lemma

Date due distribuzioni sullo stesso insieme di valori, risulta

$$D_{KL}(p||q) = \sum p_i \log_2 \left(\frac{p_i}{q_i} \right) \geq 0.$$

La quantità $D_{KL}(p||q)$ si chiama divergenza di Kullback–Leibler (anche detta divergenza di informazione, entropia relativa, o KLIC).
È una misura non simmetrica della differenza tra due distribuzioni di probabilità p e q. Anche se è spesso pensata come una distanza, la divergenza KL non è una vera e propria metrica: per esempio, non è simmetrica.

Derivazione di Shannon - Rao

Teorema

Tra tutte le distribuzioni continue di data varianza, l'entropia differenziale è massima per la distribuzione gaussiana.

Idea della dimostrazione

Derivazione di Shannon - Rao

Teorema

Tra tutte le distribuzioni continue di data varianza, l'entropia differenziale è massima per la distribuzione gaussiana.

Idea della dimostrazione

Sia $g(x)$ la densità gaussiana con media μ e varianza σ^2 e sia $f(x)$ una legge di probabilità con la stessa varianza. Dato che l'entropia differenziale è invariante per traslazioni, possiamo supporre che anche f abbia la stessa media μ e per semplicità prendiamo $\mu = 0$.

Derivazione di Shannon - Rao

Teorema

Tra tutte le distribuzioni continue di data varianza, l'entropia differenziale è massima per la distribuzione gaussiana.

Idea della dimostrazione

Consideriamo la divergenza di Kullback–Leibler (una versione continua della diseguaglianza di Gibbs) tra le due distribuzioni

$$0 \leq D_{KL}(f||g) = \int_{-\infty}^{\infty} f(x) \log\left(\frac{f(x)}{g(x)}\right) \mathrm{d}x = -h(f) - \int_{-\infty}^{\infty} f(x) \log(g(x))\, \mathrm{d}x.$$

Si osservi che

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x) \log(g(x)) dx &= \int_{-\infty}^{\infty} f(x) \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}\right) dx \\
&= \int_{-\infty}^{\infty} f(x) \log \frac{1}{\sqrt{2\pi\sigma^2}} dx + \log(e) \int_{-\infty}^{\infty} f(x) \left(-\frac{x^2}{2\sigma^2}\right) dx \\
&= -\tfrac{1}{2} \log(2\pi\sigma^2) - \log(e) \frac{\sigma^2}{2\sigma^2} = -\tfrac{1}{2} \left(\log(2\pi\sigma^2) + \log(e)\right) \\
&= -\tfrac{1}{2} \log(2\pi e\sigma^2) = -h(g)
\end{aligned}$$

Derivazione di Shannon - Rao

Teorema

Tra tutte le distribuzioni continue di data varianza, l'entropia differenziale è massima per la distribuzione gaussiana.

Idea della dimostrazione

Consideriamo la divergenza di Kullback–Leibler (una versione continua della diseguaglianza di Gibbs) tra le due distribuzioni

$$0 \leq D_{KL}(f||g) = \int_{-\infty}^{\infty} f(x) \log\left(\frac{f(x)}{g(x)}\right) \mathrm{d}x = -h(f) - \int_{-\infty}^{\infty} f(x) \log(g(x))\, \mathrm{d}x.$$

Da

$$\int_{-\infty}^{\infty} f(x) \log(g(x)) dx = -h(g)$$

quindi otteniamo

$$h(g) - h(f) \geq 0$$

e l'eguaglianza vale se e solo se $g(x) = f(x)$.

La scoperta di Galton

Consideriamo la ben nota proprietà di stabilità delle leggi Gaussiane rispetto alla somma di variabili aleatorie indipendenti, che per i nostri scopi possiamo scrivere nella forma

$$\int_{\mathbb{R}} \varphi(x - \mu_1 \mid \sigma_1^2)\varphi(y - \mu_2 - x \mid \sigma_2^2)\, \mathrm{d}x = \varphi(y - \mu \mid \sigma^2)$$

con $\mu = \mu_1 + \mu_2$ e $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

Nelle mani di Francis Galton (1886), questa formula divenne la chiave per studiare i meccanismi della evoluzione biologica.

La scoperta di Galton

Come varia l'altezza della popolazione

Consideriamo, ad esempio, la distribuzione di frequenza delle altezze h dei maschi adulti nella popolazione inglese. Galton trovò che questa poteva essere rappresentata bene da una legge Gaussiana

$$\varphi(x - \mu \mid \sigma^2), \qquad \mu = 68.1 in., \ \sigma = 2.6 in.$$

Invece di considerare i due genitori, decise di confrontare l'altezza del figlio con quella di un "genitore medio", la cui altezza era pari a

$$h_{\text{mid}} = \frac{1}{2}\left(h_{\text{padre}} + 1.08 h_{\text{madre}}\right)$$

dove il fattore 1.08 era stato determinato da Galton osservando che l'altezza media degli uomini era 1.08 volte quella delle donne.

La scoperta di Galton

Dai dati raccolti, Galton si accorse che in effetti i figli di genitori alti tendevano ad essere alti, ma che esisteva una variabilità delle altezze, sebbene la deviazione standard calcolata in questo gruppo fosse minore di quella dell'intera popolazione.

Se i figli di ogni gruppo di genitori (alti, bassi...) continuano a disperdersi in altezza, perché la variabilità delle altezze dell'intera popolazione rimane costante da una generazione all'altra?

Perché nell'intera popolazione entra in gioco un fenomeno di reversione: i figli dei genitori alti tendono a essere più alti della media, ma meno dei loro genitori, e lo stesso i figli di genitori bassi.
Se la popolazione nel complesso è stabile, questa tendenza (sistematica) a ritornare verso la media dell'intera popolazione deve bilanciare esattamente la tendenza (casuale) a disperdersi dei dati.

La scoperta di Galton

Dai dati raccolti, Galton si accorse che in effetti i figli di genitori alti tendevano ad essere alti, ma che esisteva una variabilità delle altezze, sebbene la deviazione standard calcolata in questo gruppo fosse minore di quella dell'intera popolazione.

Se i figli di ogni gruppo di genitori (alti, bassi...) continuano a disperdersi in altezza, perché la variabilità delle altezze dell'intera popolazione rimane costante da una generazione all'altra?

Perché nell'intera popolazione entra in gioco un fenomeno di reversione: i figli dei genitori alti tendono a essere più alti della media, ma meno dei loro genitori, e lo stesso i figli di genitori bassi.
Se la popolazione nel complesso è stabile, questa tendenza (sistematica) a ritornare verso la media dell'intera popolazione deve bilanciare esattamente la tendenza (casuale) a disperdersi dei dati.

La scoperta di Galton

Possiamo quantificare la forza del fenomeno di reversione.

- indichiamo con $x \equiv h_{\text{mid}} - \mu$ lo scarto dalla media dell'altezza dei genitori;
- la distribuzione delle altezze nell'intera popolazione è $\varphi(x \mid \sigma^2)$;
- la distribuzione delle altezze dei figli di un "genitore medio" di altezza $x + \mu$ è $y + \mu$, dove lo scarto dalla media y è descritto da $\varphi(y - ax \mid \sigma_2^2)$;
- se $a > 0$, allora l'altezza media dei figli di un "genitore medio" più alto (basso) della media è maggiore (minore) di quella del "genitore medio"; viceversa se $a < 0$.

La scoperta di Galton

Possiamo quantificare la forza del fenomeno di reversione.

- indichiamo con $x \equiv h_{\text{mid}} - \mu$ lo scarto dalla media dell'altezza dei genitori;
- la distribuzione delle altezze nell'intera popolazione è $\varphi(x \mid \sigma^2)$;
- la distribuzione delle altezze dei figli di un "genitore medio" di altezza $x + \mu$ è $y + \mu$, dove lo scarto dalla media y è descritto da $\varphi(y - ax \mid \sigma_2^2)$;
- se $a > 0$, allora l'altezza media dei figli di un "genitore medio" più alto (basso) della media è maggiore (minore) di quella del "genitore medio"; viceversa se $a < 0$.

La scoperta di Galton

Possiamo quantificare la forza del fenomeno di reversione.

- indichiamo con $x \equiv h_{\text{mid}} - \mu$ lo scarto dalla media dell'altezza dei genitori;
- la distribuzione delle altezze nell'intera popolazione è $\varphi(x \mid \sigma^2)$;
- la distribuzione delle altezze dei figli di un "genitore medio" di altezza $x + \mu$ è $y + \mu$, dove lo scarto dalla media y è descritto da $\varphi(y - ax \mid \sigma_2^2)$;
- se $a > 0$, allora l'altezza media dei figli di un "genitore medio" più alto (basso) della media è maggiore (minore) di quella del "genitore medio"; viceversa se $a < 0$.

La scoperta di Galton

Possiamo quantificare la forza del fenomeno di reversione.

- indichiamo con $x \equiv h_{\mathrm{mid}} - \mu$ lo scarto dalla media dell'altezza dei genitori;
- la distribuzione delle altezze nell'intera popolazione è $\varphi(x \mid \sigma^2)$;
- la distribuzione delle altezze dei figli di un "genitore medio" di altezza $x + \mu$ è $y + \mu$, dove lo scarto dalla media y è descritto da $\varphi(y - ax \mid \sigma_2^2)$;
- se $a > 0$, allora l'altezza media dei figli di un "genitore medio" più alto (basso) della media è maggiore (minore) di quella del "genitore medio"; viceversa se $a < 0$.

La scoperta di Galton

La popolazione nel complesso ha distribuzione data da

$$\begin{aligned}\varphi(y \mid \sigma^2) &= \int_{\mathbb{R}} \varphi(x \mid \sigma^2)\varphi(y - ax \mid \sigma_2^2)\,\mathrm{d}x \\ &= \frac{1}{a}\int_{\mathbb{R}} \varphi(x \mid \sigma^2)\varphi(y/a - x \mid (\sigma_2/a)^2)\,\mathrm{d}x \\ &= \varphi(y \mid a^2\sigma^2 + \sigma_2^2)\end{aligned}$$

Perché la varianza sia costante tra le generazioni, deve essere

$$\sigma^2 = a^2\sigma^2 + \sigma_2^2 \qquad \Longrightarrow \qquad a = \pm\sqrt{1 - \frac{\sigma_2^2}{\sigma^2}}$$

quindi condizione necessaria è che sia $|a| \leq 1$ (altrimenti, per avere stabilità dovrei richiedere $\sigma_2^2 < 0$, il che è assurdo).