Probabilità e incertezza di misura

G. D'Agostini
Dipartimento di Fisica, Università "La Sapienza", Roma

30 ottobre 2001

Indice

Parte IV

Applicazioni di statistica inferenziale

Capitolo 11

Impostazione del problema. Caso di verosimiglianza gaussiana

11.1 Introduzione

Il problema dell'inferenza probabilistica è stato ampiamente discusso nel capitolo **??**. In particolare, partendo da esempi intuitivi, è stata illustrata la potenza del teorema di Bayes per riaggiornare la probabilità di ipotesi alla luce di nuovi fatti. Successivamente, abbiamo visto, rispettivamente nei paragrafi 7.16 e 9.3.3, come il teorema di Bayes si applicchi sia a funzioni di probabilità che a funzioni di densità di probabilità. Siccome in questa parte ci occuperemo essenzialmente di grandezze il cui valore può assumere valori con continuità, faremo uso della formula (9.12), che riscriviamo qui nel seguente modo

$$f(\mu\,|\,\text{dati},\,I) \propto f(\text{dati}\,|\,\mu,\,I)\cdot f_\circ(\mu\,|\,I)\,, \tag{11.1}$$

avendo indicato con μ il *valore vero* della grandezza che vogliamo inferire, con 'dati' i valori osservati e avendo omesso l'inessenziale fattore di normalizzazione (come ampiamente discusso nel capitolo **??**). La condizione di sfondo I tiene conto della globalità del nostro stato di informazione sulla misura (grandezza da misurare, comportamento degli apparati e delle procedure, condizioni ambientali, etc.). Come discusso nell'introduzione alle variabili casuali (paragrafo **??**, in particolare gli ultimi due punti), nell'impostazione seguita in questo testo osservazioni e valori veri giocano un ruolo simmetrico.

- Posso affermare quanto credo all'osservazione di certi dati sperimenatali, subordinatamente ad ogni ipotesi di valore vero e delle condizioni sperimentali, ossia esplicito la mia *verosimiglianza* $f(\text{dati}\,|\,\mu\, I)$. E in questo caso sono i valori dei dati ad essere numeri incerti.

- Posso affermare quanto credo ai possibili valori di μ, subordinatamente all'osservazione di certi valori dei dati. In questo caso è il valore vero μ ad essere un numero incerto, mentre i dati sperimentali sono per defi-

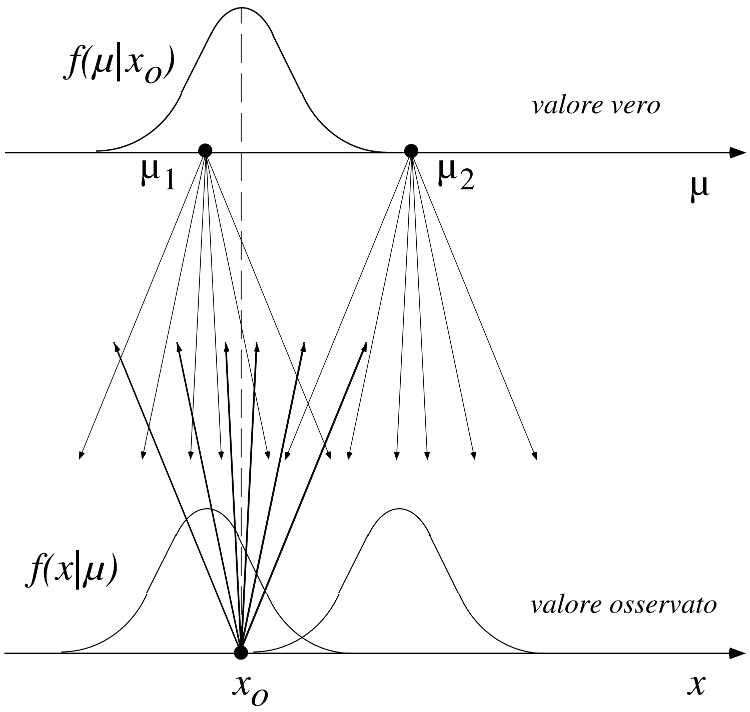

Figura 11.1: Schema generale del modello modello inferenziale basato sulla probabilità condizionata. L'asse delle x rappresenta il mondo reale delle osservazioni. L'asse dei μ rappresenta invece i possibili valori dei parametri della teoria, che, come tali, non si manifestano direttamente ai nostri sensi.

> nizione certi, in quanto 'dati'.[1] Il risultato dell'esperimento sarà quindi della forma $f(\mu\,|\,\text{dati},\, I)$, dove con I sono state indicate, ripetiamo, tutte le informazioni di contorno su grandezza da misurare, strumenti di misura e fattori di influenza.

Lo schema generale di inferenza è illustrato in figura 11.1 L'importanza concettuale della *prior* $f_\circ(\mu\,|\,I)$ è stato discusso nel capitolo **??** e verrà ripreso nel seguito, mostrando come sia praticamente irrilevante nei casi di routine. Si ricorda, infine, che il ruolo della verosimiglianza è quello di modificare il grado di fiducia in ciascun valore di μ, come ovvio dalla formula (11.1).

Il resto del capitolo consisterà nell'applicazione sistematica della formula (11.1) a diverse situazioni che si verificano frequentemente nella pratica di laboratorio. La condizione di contorno I verrà in genere sottintesa, a meno che non la si voglia esplicitare per tenere conto dell'incertezza su grandezze di influenza (i cosiddetti effetti sistematici).

[1] Come discuteremo nel seguito esiste una corrente di pensiero, fortunatamente in declino fra coloro che si occupano dei fondamenti dell'inferenza statistica, ma tuttora in auge per quanto riguarda le applicazioni, la quale nega che si possa parlare di probabilità dei valori veri. Questo è il motivo per cui si sente ancora parlare di "incertezza dei dati", una vera assurdità che fa pensare che si può essere in stato di incertezza di fronte al numero letto sul display di uno strumento. L'incertezza è invece sul valore della grandezza che con quello strumento si vuole misurare, ed è compito dell'inferenza statistica esprimere correttamente tale incertezza.

Vedremo anche come usare delle approssimazioni per semplicare i conti nei semplici casi di routine.

11.2 Verosimiglianza normale con σ nota

L'interesse per la distribuzione di Gauss è giustificato dal 'fatto' che essa descrive abbastanza bene la distribuzione degli errori di misura.[2] Le ragioni per la quale adottiamo tale modello sono le seguenti.

Innanzitutto, dal confronto fra i singoli valori osservati e la loro media (o un valore vero convenzionale, nei casi fortunati in cui si riesca a calibrare le misure su una grandezza di riferimento) si nota che la distribuzione degli scarti ("errori") somiglia ad una gaussiana nel senso visto a proposito dell'approssimazione normale di distribuzioni di variabili casuali discrete (vedi paragrafo 10.10). Come in quel caso la gaussiana si estendeva per valori al di fuori del range della variabile (tipicamente valori negativi, o valori maggiori di n nel caso della binomiale), anche in questo caso bisogna prestare attenzione a non prendere troppo sul serio le code della gaussiana.

La seconda ragione è legata al teorema del limite centrale, in base ad argomenti simili a quelli per descrivere il cammino casuale incontrato nel paragrafo 10.15. Ammettiamo che nel processo di misura ci siano molti effetti di disturbo, ovvero dei piccoli errori e_i di valore atteso nullo e deviazione standard σ_i. L'errore totale è dato da $e = \sum_i e_i$, di varianza $\sigma_e^2 = \sum_i \sigma_i^2$. Se valgono le condizioni del teorema del limite centrale, ci aspettiamo che e sia descritto da una gaussiana di valor medio zero e deviazione standard σ_e. In genere questa approssimazione è abbastanza buona, soprattutto perché è frequente il caso in cui sono le stesse e_i ad essere descritte con buona approssimazione da una gaussiana.

Infine, a volte l'inferenza non è fatta a partire da una singola osservazione, ma dalla media di molte osservazioni, come se si trattasse di una *singola osservazione equivalente*. Questa operazione è legata al concetto statistico di *sufficienza*, su cui torneremo nel seguito (vedi paragrafo ***). Anche in questo caso, interviene il teorema del limite centrale che ci fa credere che questa 'osservazione equivalente' può essere descritta da una gaussiana.

Stanti queste ipotesi (o "credenze" o "assunzioni ragionevoli" che dir si vogliano), tendiamo a credere che la singola osservazione (o osservazione equivalente) X sarà descritta da

$$X \sim \mathcal{N}(\mu, \sigma_e)\,, \tag{11.2}$$

Applicando il teorema di Bayes, includendo il fattore di normalizzazione, abbiamo per $f(\mu)$:

$$f(\mu\,|\,x) = \frac{\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}}\, f_\circ(\mu)}{\int \frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}}\, f_\circ(\mu)\,\mathrm{d}\mu}\,. \tag{11.3}$$

[2]'Fatto' è fra virgolette perché non c'è nessuna necessità logica che ci obblighi ad adottare tale modello; abbiamo soltanto ottimi argomenti, sia di natura teorica che empirica che ci fanno pensare che molto probabilmente sarà così.

Resta ora da discutere cosa usare per $f_\circ(\mu)$. Come si può immaginare, la presenza della prior crea molto imbarazzo fra coloro ritengono che la probabilità sia 'oggettiva', e non semplicemente una descrizione del nostro stato di incertezza, nella quale i possibili valori dei numeri incerti sono flassificati in plausibilità. Simili reazioni provengono da coloro (più o meno gli stessi) che considerano l'inferenza un esercizio matematico, e non, come deve essere, una costante modifica delle nostre credenze alla luce di nuovi fatti sperimentali. Se ci si mette nell'ottica di considerare esclusivamente situazioni pratiche, di modellizzare onestamente quello che si sa e prendere i modelli 'cum grano salis' non ci sono problemi seri, a parte quello di trovare le approssimazioni adatte per semplificare i conti quando i conti diventano complessi.

Facciamo un esempio pratico. Il lettore vuole misurare la temperatura della stanza dove sta in quest'istante. Se vuole effettuare la misura, vuol dire che vuole conoscerla meglio di quanto ne sa in quest'istante, ma non che lui non ne sappia assolutamente niente. Qualsiasi persona è in genere in grado di indovinare la temperatura entro qualche grado, e per questo ci si stupisce quando un termometro indica valori molto diversi da quello che ci si aspetta (tanto da far sorgere il dubbio che il termometro sia rotto!). Quindi, nell'eseguire la misura si sceglierà lo strumento adatto, tale da migliorare la propria conoscenza. Se un amico ci desse uno strumento formidabile (per le sue potenzialità) in grado di misurare la temperatura assoluta da 0 a un milione di gradi entro un errore di 5 gradi, non lo accetteremmo per l'applicazione specifica: qualsiasi cosa leggeremo sullo strumento, crederemo ancora e soltanto a quanto possiamo indovinare dalla nostra sensazione fisiologica.

Riassumendo, nelle condizioni ordinarie delle misure, l'informazione riassunta in $f_\circ(\mu)$ è molto più *vaga* di quella che ci può fornire lo strumento, ovvero $f_\circ(\mu)$ è molto più larga di $f(x \,|\, \mu)$, letta in funzione di μ per $X = x$ fissato (dato). Detto ancora in altri termini: nella regione in cui la verosimiglianza varia rapidamente ("su e giu") intorno a $X = x$, la $f_\circ(\mu)$ è circa costante, mentre per valori lontani da esso, le code gaussiane della verosimiglianza rendono l'esatta forma di $f_\circ(\mu)$ irrilevante. Un esempio è mostrato in figura 11.2.

Quindi, quando valgono queste condizioni, si puù modellizzare la conoscenza a priori come $f_\circ(\mu) = k$, costante in un intervallo "abbastanza ampio". Per sola convenienza matematica, estendiamo tale intervallo a coprire l'intero asse reale. Ne segue che nella (11.3) la $f_\circ(\mu) = k$, si semplifica. Abbiamo allora:

$$f(\mu \,|\, x) = \frac{\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(x-\mu)^2}{2\,\sigma_e^2}}}{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(x-\mu)^2}{2\,\sigma_e^2}}\, \mathrm{d}\mu}\,. \qquad (11.4)$$

L'integrale di normalizzazione è pari a 1, in quanto l'integrale della funzione di Gauss rispetto a μ è uguale all'integrale rispetto a x, che sappiamo valere 1. Il risultato finale è quindi

$$f(\mu \,|\, x) = \frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(\mu-x)^2}{2\,\sigma_e^2}} \quad : \qquad (11.5)$$

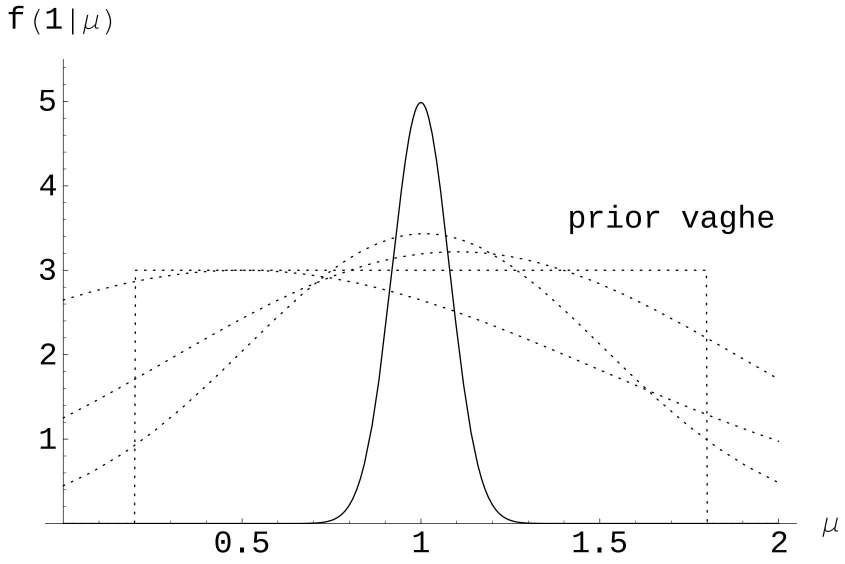

Figura 11.2: Esempio di prior (curve tratteggiate) molto più vaghe della verosimiglianza dello strumento (curva continua). Per questioni di resa grafica, le prior sono a meno di una irrilevante costante di normalizzazione.

Si noti la banale (formalmente) inversione di x e μ all'esponente della funzione di Gauss, a ricordare che ora è μ ad essere il numero incerto. Dalle note proprietà della gaussiana, arriviamo alle seguenti conclusioni.

- La previsione di μ è pari a x: $\mathrm{E}[\mu] = x$.
- L'incertezza (standard) di previsione è pari a $\sigma_\mu = \sigma_e$.
- Il risultato può essere riportato nella forma

$$\mu = x \pm \sigma_e,$$

a meno che non lo si voglia dare sotto forma di intervallo di probabilità. In questo caso va specificato chiaramente, dindicando anche il livello di probabilità prescelto. Avremmo quindi:

Livello di Probabilità (Livello di confidenza)	Intervallo di probabilità (Intervallo di confidenza)
68.3%	$x \pm \sigma_e$
90.0%	$x \pm 1.65\,\sigma_e$
95.0%	$x \pm 1.96\,\sigma_e$
99.0%	$x \pm 2.58\,\sigma_e$
99.73%	$x \pm 3\,\sigma_e$

Abbiamo usato anche le espressioni "livello di confidenza" e "intervallo di confidenza", con il significato che si dà a loro nel linguaggio naturale (e non quello astrusa della prassi statistica detta "frequentista").

11.3 Effetto di una prior rilevante: combinazione di risultati

A questo punto, una domanda naturale è cosa succede se la prior non è proprio talmente vaga da essere ininfluente sulla distribuzione di probabilità finale. Per semplificare i conti, modellizziamo la nostra conoscenza a priori con una gaussiana centrata in $\mu_\circ$ e di deviazione standard $\sigma_\circ$. Ad esempio, tale stato di conoscenza potrebbe derivare da una precedente misura effettuata nelle condizioni del paragrafo precedente. Come discusso a lungo nel capitolo 5, nello schema bayesiano il riaggiornamento della probabilità si effettua usando come prior la distribuzione finale dell'inferenza precedente. Nel nostro caso, abbiamo

$$f(\mu\,|\,x,\sigma_e,\mu_\circ,\sigma_\circ) \;=\; \frac{\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\,e^{-\frac{(x-\mu)^2}{2\,\sigma_e^2}}\,\frac{1}{\sqrt{2\,\pi}\,\sigma_\circ}\,e^{-\frac{(\mu-\mu_\circ)^2}{2\,\sigma_\circ^2}}}{\int\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\,e^{-\frac{(x-\mu)^2}{2\,\sigma_e^2}}\,\frac{1}{\sqrt{2\,\pi}\,\sigma_\circ}\,e^{-\frac{(\mu-\mu_\circ)^2}{2\,\sigma_\circ^2}}\,\mathrm{d}\mu}\,. \tag{11.6}$$

L'integrale è un po' più complicato del caso precedente. Con le opportune semplificazoni[3] il risultato dell'inferenza è:

$$f(\mu\,|\,x,\sigma_e,\mu_\circ,\sigma_\circ) = \frac{1}{\sqrt{2\,\pi}\,\sigma_A}\,e^{-\frac{(\mu-\mu_A)^2}{2\,\sigma_A^2}}\,, \tag{11.7}$$

con

$$\mu_A \;=\; \frac{x/\sigma_e^2+\mu_\circ/\sigma_\circ^2}{1/\sigma_e^2+1/\sigma_\circ^2}\,, \tag{11.8}$$

$$\frac{1}{\sigma_A^2} \;=\; \frac{1}{\sigma_e^2}+\frac{1}{\sigma_\circ^2}\,. \tag{11.9}$$

I possibili valori di μ sono ancora descritti da una gaussiana centrata in corrispondenza della *media pesata* fra x e $\mu_\circ$, con pesi pari all'inverso delle varianze. Previsione e incertezza di previsione valgono $\mathrm{E}(X)=\mu_A$ e $\sigma(\mu)=\sigma_A$. Il caso di prior vaga è recuperato per $\sigma_\circ\to\infty$ (con $\mu_\circ$ 'ragionevole').

Poiché il risultato $\mathrm{E}(\mu)=x\pm\sigma_e$ è quello che si ottiene quando la prior è ininfluente, mentre la previsione precedente $\mathrm{E}_\circ(\mu)=\mu_\circ\pm\sigma_\circ$ può essere

[3] A parte fattori moltiplicativi, abbiamo

$$\begin{aligned}
f(\mu) &\propto \exp\left[-\frac{1}{2}\left(\frac{-2\,\mu\,x\sigma_\circ^2+\mu^2\sigma_\circ^2+-2\,\mu\,\mu_\circ\sigma_e^2+\mu^2\sigma_e^2}{\sigma_e^2+\sigma_\circ^2}\right)\right]\\
&= \exp\left[-\frac{1}{2}\left(\frac{\mu^2-2\,\mu\left(\frac{x\,\sigma_\circ^2+\mu_\circ\,\sigma_e^2}{\sigma_e^2+\sigma_\circ^2}\right)}{(\sigma_e^2\cdot\sigma_\circ^2)/(\sigma_e^2+\sigma_\circ^2)}\right)\right]\\
&= \exp\left[-\frac{1}{2}\left(\frac{\mu^2-2\mu\,\mu_A}{\sigma_A^2}\right)\right]\\
&\propto \exp\left[-\frac{(\mu-\mu_A)^2}{2\,\sigma_A^2}\right]
\end{aligned}$$

In particolare, nell'ultimo passaggio abbiamo "complementato" l'esponenziale moltiplicando e dividendo per $\exp[-2\,\mu_A^2/\sigma_A^2]$. Normalizzando, otteniamo la 11.7.

pensato come dovuto ad una precedente inferenza, le (11.8) e (eq:waver2) ci mostrano come combinare due risultati parziali. In particolare, interpretando l'inverso della varianza come *peso statistico*, la (11.9) ci dice che il peso statistico risultante dall'inferenza globale è pari alla somma dei pesi statistici delle inferenze parziali.

A questo punto, la combinazione di molti risultati parziali *indipendenti*, ciascuno ottenuto da una prior vaga è abbastanza ovvio. È istruttivo ragionare in due modi diversi.

- Possiamo immaginare una catena di inferenze, del tipo

$$\begin{array}{rcl}\{x_1, \sigma_1\} & \longrightarrow & f_1(\mu) = f(\mu \,|\, x_1, \sigma_1) \\ \{f_1(\mu), x_2, \sigma_2\} & \longrightarrow & f_2(\mu) = f(\mu \,|\, x_1, \sigma_1, x_2, \sigma_2) \\ \ldots & & \ldots \\ \{f_{n-1}(\mu), x_n, \sigma_n\} & \longrightarrow & f(\mu \,|\, \mathbf{x}, \sigma)\,,\end{array}$$

  ove abbiamo indicato con $\mathbf{x}$ e σ l'insieme dei valori osservati e delle deviazioni standard legate alla verosimiglianza di ciascuna di esse. Applicando iterativamente il teorema di Bayes abbiamo:

$$f(\mu \,|\, \mathbf{x}) \propto f(x_1 \,|\, \mu) \cdots f(x_n \,|, \mu) = \prod_i f(x_i \,|\, \mu)\,. \tag{11.10}$$

- Il secondo modo consiste nel pensare alla verosimiglianza congiunta di osservare $\mathbf{x}$ per ogni ipotesi di μ:

$$f(\text{dati} \,|\, \mu) = f(\mathbf{x} \,|\, \mu) = \prod_i f(x_i \,|\, \mu)\,. \tag{11.11}$$

Otteniamo lo stesso risultato indipendentemente dal percorso seguito, il ché è confortante, visto che entrambi i ragionamenti sono legittimi (vedi anche discussioni in proposito nel capitolo 5).

In conclusione, abbiamo la seguente regola di combinazione:

$$\mathrm{E}(\mu) = \frac{\sum_i \mathrm{E}_i(\mu)/\sigma_i^2(\mu)}{\sum_i 1/\sigma_i^2(\mu)} \tag{11.12}$$

$$\frac{1}{\sigma^2(\mu)} = \sum_i \frac{1}{\sigma_i^2(\mu)}, \tag{11.13}$$

che possiamo riscrivere, facendo riferimento agli n valori osservati x_i e alle deviazioni standard delle relative verosimiglianze come

$$\mathrm{E}(\mu) = \frac{\sum_i x_i/\sigma_i^2}{\sum_i 1/\sigma_i^2} \tag{11.14}$$

$$\frac{1}{\sigma^2(\mu)} = \sum_i \frac{1}{\sigma_i^2}\,. \tag{11.15}$$

Si noti il carattere più generale delle (11.12) e (11.13) rispetto a queste ultime, in quanto quelle possono far riferimento a situazioni più complicate delle semplici n osservazioni individuali. Ad esempio, ciascuna previsone $\mathrm{E}_i(\mu)$ può derivare essa stessa da una precedente combinazione o da un'analisi complicata. Un esempio di inferenza combinata è mostrata in figura 11.3.

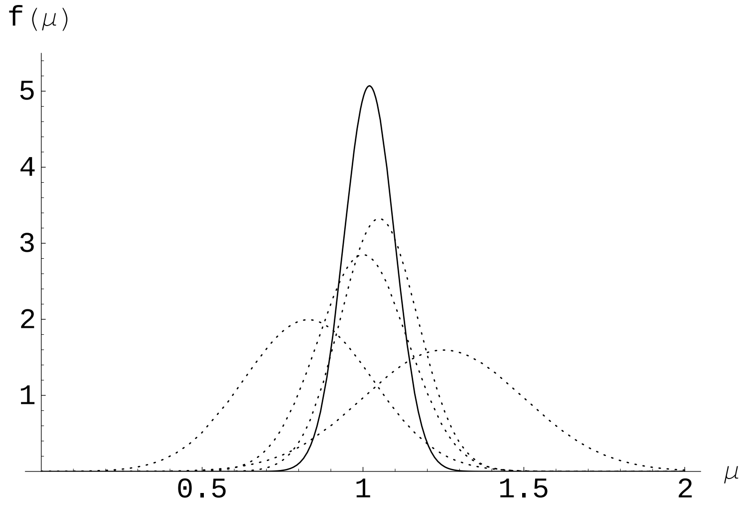

Figura 11.3: Esempio di combinazione di quattro inferenze indipendenti (curve tratteggiate) risultanti in un'unica inferenza globale (curva tratteggiata).

Tornando alle n osservazioni individuali indipendenti, vediamo il caso in cui la deviazione standard sia la stessa per tutte le osservazioni, ovvero $\sigma_i = \sigma$ $\forall\, i$. Le (11.14) e (**??**) diventano

$$\mathrm{E}(\mu) = \frac{1}{n}\sum_i x_i \tag{11.16}$$

$$\sigma(\mu) = \frac{\sigma}{\sqrt{n}}\,. \tag{11.17}$$

La previsione di μ è pari alla media aritmetica delle osservazioni. Inoltre, si vede come l'insieme delle n osservazioni indipendenti hanno un peso statistico di n volte quello di una singola osservazione.

11.4 ❄ Derivazione di Gauss della gaussiana

A questo punto è interessante mostrare con quali argomenti Gauss introdusse la funzione che porta il suo nome[4]. Essa era, infatti giaà nota molto primo di Gauss, come limite della binomiale (vedi paragrafo 8.10). L'interesse nella derivazione di Gauss è che essa fu ottenuta nel contesto dell'inferenza probabilistica, ma come soluzione di un problema inverso: qual'è la forma più generale della verosimiglianza tale che il massimo di probabilità di μ coincida con la media aritmetica? Nel risolvere questo problema, dapprima Gausss derivò la formula della probabilità delle ipotesi, assumendo equiprobabilità iniziale

[4]Vedi ad esempio F. Sibirani, *"Calcolo delle probabilità"*, in Enciclopedia delle Matematiche Elementari e Complementari, a cura di L. Berzolati, Hoepli, 1949 (ristampa anastatica 1987), Volume III, parte 2^a, pp. 234-236. Le citazioni in latino sono dallo scritto originale di Gauss *Theoria motus corporum coelestium in sectionibus conicis solem ambientium"* del 1809. Esse sono riportate per mostrare come l'impostazione di questo testo, seppur moderna, se confrontata con la prassi statistica del XX secolo, si rifà al modo di pensare originario sulla probabilità di Gauss, Laplace, Bernoulli etc.

delle ipotesi. In pratica ottenne il teorema di Bayes nel caso particolare di probabilità iniziale delle ipotesi, concetto a lui molto chiaro (*"ante eventum cognitum"*, contrapposto a *"post eventum cognitum"*). Quindi, passando alle osservazioni sperimentali x_i (usando la nostra simbologia), cerca la forma della funzione incognita φ che descrive la probabilità di ottenere x_i dal valore vero μ (sempre nella nostra simbologia). La (funzione densità di) probabilità del campione $\mathbf{x}$ è allora data da

$$f(\mathbf{x}\,|\,\mu) = \varphi(x_1-\mu)\cdot\varphi(x_2-\mu)\cdot\;\cdots\;\cdot\varphi(x_n-\mu)\,. \tag{11.18}$$

A questo punto intervengono due ipotesi:

1. Tutti i valori di μ sono ritenuti a priori (*"ante illa observationes"*) ugualmente probabili (*"... aeque probabilia fuisse"*).
2. Il massimo di probabilità a posteriori (*"post illas observationes"*) si ottiene per $\mu=\overline{x}$, media aritmetica degli n valori osservati.

Dalla prima ipotesi segue

$$f(\mu\,|\,\mathbf{x}) \propto f(\mathbf{x}\,|\,\mu) = \varphi(x_1-\mu)\cdot\varphi(x_2-\mu)\cdot\;\cdots\;\cdot\varphi(x_n-\mu)\,.$$

Per far uso della seconda, si impone che la derivata prima si annulli in $\mu=\overline{x}$:

$$\left.\frac{\mathrm{d}f(\mu\,|\,\mathbf{x})}{\mathrm{d}\mu}\right|_{\mu=\overline{x}} = 0 \;\Longrightarrow\; \left(\frac{\mathrm{d}}{\mathrm{d}\mu}\prod\varphi(x_i-\mu)\right)_{\mu=\overline{x}} = 0\,,$$

ovvero

$$\sum_i \frac{\varphi'(x_i-\overline{x})}{\varphi(xi-\overline{x})} = 0\,,$$

ove φ' sta per la funzione derivata di ϕ rispetto a μ. Chiamando ψ la funzione φ'/φ e indicando con $z_i = x_i - \overline{x}$ gli scarti dalla media, i quali devono soddisfare la condizione $\sum_i z_i = 0$, abbiamo

$$\left\{\begin{array}{l}\sum_i \psi(z_i) = 0\\ \sum_i z_i = 0\end{array}\right.\,, \tag{11.19}$$

la quale, dovendo essere valida indipendentemente da n e dal valore degli scarti, dà il seguente vincolo[5] alla forma funzionale di $\psi(z)$:

$$\frac{1}{z}\,\psi(z) = k\,, \tag{11.20}$$

[5] Il modo più semplice di usare la condizione di vincolo quello di passare ai differenziali di ψ: $\sum_i \psi'(z_i)\mathrm{d}z_i = 0$ (a), con la condizione $\sum_i z_i$ che si riflette in un'analoga condizione sui differenziali: $\sum_i \mathrm{d}z_i = 0$. Affinché (a) sia sempre valida, $\psi'(z)$ deve essere costante, da cui segue $\psi(z) \propto z$ e quindi la (11.20).

Un modo più vicino alla dimostrazione originale di Gauss è di pensare al caso in cui k degli n scarti acquistino il valore a e gli altri $(n-k)$ acquistino il valore b, con $b = -a\,k/(n-k)$, in virtù degli vincolo della media aritmetica. Abbiamo allora

$$k\,\psi(a) + (n-k)\,\psi(b) \quad = \quad 0$$

con k costante (il limite $z- > 0$ non è un problema, in quanto la derivata di φ in $z = 0$ si annulla e la condizione $\psi(z)/z = k$ implica che numeratore e denominatore ci devono tendere con la stessa "rapidità). Ne segue

$$\frac{\mathrm{d}\varphi}{\varphi} = k\,z\,\mathrm{d}z\,,$$

Ovvero

$$\varphi(z) \propto e^{\frac{k}{2}\,z^2} = e^{-h^2\,z^2}\,, \qquad (11.21)$$

ove $k/2$ è stato chiamato $-h^2$ per imporre che esso deve essere negativo in quanto φ ha il massimo in $z = 0$. Normalizzando la funzione mediante il suo integrale da $-\infty$ a ∞, conto dovuto a Laplace (*"ab ill. Laplace inventum"*), la funzione "di Gauss" è (*"funtio nostra fiet"*):

$$\varphi(z) = \frac{h}{\sqrt{\pi}} e^{-h^2\,z^2}$$

.

11.5 ❄ Caso di forte vincolo dato dalla prior

Anche se, per evitare ogni pedanteria, è stato suggerito che, per i casi di routine, una prior uniforme è più che ragionevole, un ricercatore esperto ha sempre delle prior in mente, e con esse valuta il risultato. Lo può accettare tranquillamente, diffidando quasi sempre delle code di probabilità perché è ben cosciente dei modelli matematici che usa; oppure si può insospettire e controllare meglio lo strumento; oppure può decidere di ripetere le misure, e così via. Ad esempio, se si guarda al volo verso ora di pranzo un orologio che non si sa essere rotto e si legge 12:27 si prende per buona tale osservazione (immagino che questa o analoghe esperienze sia capitata a molti), mentre se avesse indicato 17:43 la si sarebbe rifiutata. Lo stesso vale per voltmetri rotti, termometro starati e così via. Lo sperimentatore esperto ha sempre delle prior che, pur nella loro vaghezza, sono molto solide e filtrano le piccole disavventure che capitano in laboratorio. Uno strumento rotto, un procedimento di misura sbagliato o un errore di calcolo possono produrre effetti drammatici nel risultato. Lo sperimentatore che ha buone prior, sviluppate con anni di ricerca può sbagliare del 30%, del 50% o addirittura di un fattore due, ma raramente di fattori 10, 100 o 1000.

Purtroppo questo discorso esula da questa trattazione e niente può sostituire l'esperienza diretta. Vogliamo mostrare qui soltanto un caso, molto

da cui

$$\begin{aligned}
\psi(a) &= -\frac{n-k}{k}\psi\left(-\frac{k\,a}{n-k}\right) \\
\frac{1}{a}\psi(a) &= -\frac{n-k}{k\,a}\psi\left(-\frac{k\,a}{n-k}\right).
\end{aligned}$$

Poichè questa relazione non deve dipendere dal valore di a scelto, e nemmeno da k (nella dimostrazione originale di Gauss viene preso in considerazione soltanto il caso $k = n - 1$), la condizione $\frac{\psi(z)}{z} = k$ è assunta universale, e questa porta alla gaussiana.

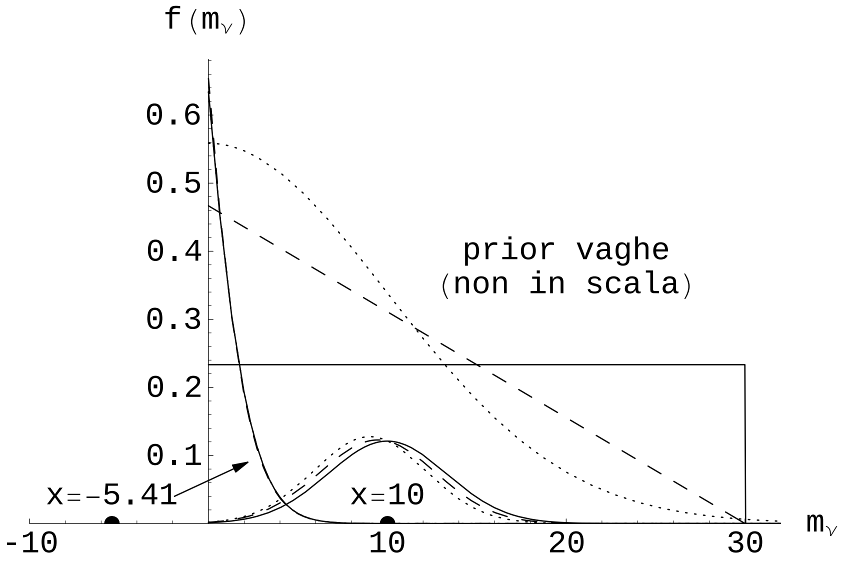

Figura 11.4: Esempio di inferenza in prossimità del limite fisico di una grandezza (massa del neutrino in unità di eV/c^2) con tre diverse prior 'motivate' e due diverse osservazioni. Si noti come, nel caso dell'osservazione negativa ($x = -5.41$) le tre distribuzioni finali sono praticamente coincidenti.

schematizzato, di come comportarsi quando il risultato ottenuto da una rapida applicazione delle formule precedenti cozza violentemente con le proprie convinzioni.

Immaginiamo un esperimento progettato per misurare la massa del neutrino (di tipo elettronico). Assumiamo che, la conoscenza dettagliata dell'esperimento induca gli sperimentatori a credere che le possibili osservazioni 'equivalenti' (nel senso specificato sopra, ovvero il numero al quale si arriva avendo elaborato opportunamente la massa di dati) siano descritte da una gaussiana centrata intorno alla massa vera, con deviazione standard 3.3 eV/c^2 (per avere un'idea dell'ordine di grandezza, in questa unità di misura la massa dell'elettrone è 511 mila, quella del protone 938 milioni). Alla fine dell'esperimento si ottiene $x = -5.41\,$eV/c^2. Cosa dobbiamo concluderne? Che il neutrino abbia, con probabilità del 95% una massa negativa? Oppure che l'esperimento sia "sbagliato"? Se prima di questo esperimento i fisici erano convinti che la massa del neutrino può essere al più qualche decine di eV/c^2, come viene modificata la loro conoscenza? (Si noti che sotto l'ipotesi "esperimento sbagliato" non si impara niente).

In questo caso, la prior forte è che la massa debba essere non negativa. Inoltre si è praticamente convinti che essa non può essere troppo grande (altrimenti avrebbe dato effetti in altre osservazioni precedenti). Con queste considerazioni, tenendo conto che l'esperimento è stato proposto, finanziato e costruito con lo scopo di vedere qualcosa, si può modellizzare l'incertezza in diversi modi (vedi figura 11.4):

- Una distribuzione uniforme fra 0 e 30 eV/c^2:

$$f_{\circ K}(m) = k = 1/30 \qquad (0 \leq m \leq 30)\,; \tag{11.22}$$

- una mezza gaussiana centrata su zero con deviazione standard $\sigma_\circ = 10\,\mathrm{eV/c^2}$:

$$f_{\circ N}(m) = \frac{2}{\sqrt{2\,\pi}\,\sigma_\circ}\,\exp\left[-\frac{m^2}{2\,\sigma_\circ^2}\right] \qquad (m \geq 0)\,; \tag{11.23}$$

- una distribuzione triangolare

$$f_{\circ T}(m) = \frac{1}{450}\,(30 - m) \qquad (0 \leq m \leq 30)\,. \tag{11.24}$$

La "mezza-gaussiana" ha il vantaggio di ammettere anche valori di massa molto grandi, seppur molto poco plausibili. Consideriamo per semplicità la distribuzione uniforme. Inserendola nella formula di Bayes otteniamo

$$f(m\,|\,x, f_{\circ K}) = \frac{\frac{1}{\sqrt{2\,\pi}\,\sigma}\,\exp\left[-\frac{(x-m)^2}{2\,\sigma^2}\right]\,k}{\int_0^{30}\frac{1}{\sqrt{2\,\pi}\,\sigma}\,\exp\left[-\frac{(x-m)^2}{2\,\sigma^2}\right]\,k\,\mathrm{dm}} \tag{11.25}$$

$$= \frac{19.8}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(m-x)^2}{2\,\sigma^2}\right] \qquad (0 \leq m \leq 30) \tag{11.26}$$

Come conclusione dell'esperimento, il valore al quale crediamo di più è zero (moda), ma $f(m)$ è diversa da zero fino a 30 eV/c^2. In questi casi è conveniente dare il risultato come limite superiore ad un certo livello di probabilità. Facendo i conti otteniamo:

$$m < 3.9\,\mathrm{eV}/c^2 \qquad \text{al } 0.95\,\% \text{ di probabilità}\,. \tag{11.27}$$

Se avessimo usato le altre prior avremmo avuto

$$m < 3.7\,\mathrm{eV}/c^2 \qquad \text{al } 0.95\,\% \text{ di probabilità}\,, \tag{11.28}$$

praticamente lo stesso valore (soprattutto se confrontato con la risoluzione sperimentale di $3.3\,\mathrm{eV}/c^2$).

Ben diverso è il caso in cui si osserva un valore positivo ben distanziato dallo zero, esemplificato in figura 11.4 dal punto $x = 10$. Sebbene l'inferenza dipende leggermente dalla prior (ma in modo irrrilevante dal punto di vista pratico) si vede che anche la banale inversione intuitiva di probabilità dà un risultato soddisfacente.

11.6 Caso di σ ignota

Nei paragrafi precedenti abbiamo assunto di avere già delle idee sulla deviazione standard del modello gaussiano di verosimiglianza. Ovvero, ci siamo posti nelle condizioni di avere già acquisito una certa esperienza, o comunque una certa confidenza, sul comportamento della singola osservazione (o osservazione equivalente). È preferibile parlare, in termini generali, di confidenza piuttosto che di esperienza, in quanto il nostro modello (e i sui parametri) possono derivare, e è spesso il caso, da simulazioni, estrapolazioni e analogie, piuttosto che da una vera sperimentazione su *quella* grandezza fisica, in *quell'*intervallo di valori possibili e in *quelle* condizioni sperimentali. Vediamo ora come valutare σ del modello gaussiano avendo registrato n osservazioni in condizioni apparentemente identiche (nei limiti di un controllo reale dell'esperimento e non dal punto di vista di un diavoletto di Maxwell in grado di percepire qualsiasi variazione microspopica) e avendo osservato una dispersione di valori, intorno a un valore medio $\overline{x}$. Quantifichiamo la dispersione dei valori con la deviazione standard del campione statistico, che per evitare confusione con il parametro σ della gaussiana, chiameremo s. Abbiamo quindi:

$$\overline{x} = \frac{1}{n}\sum_i x_i \tag{11.29}$$

$$s = \sqrt{\frac{1}{n}\sum_i (x_i - \overline{x})^2} = \sqrt{\overline{x^2} - \overline{x}^2}\,. \tag{11.30}$$

Avendo introdotto questi riassunti statistici del campione in esame, procediamo per passi. Infatti questo è uno degli argomenti nei quali ci si può far prendere la mano dalla matematica dimenticando il proprio stato di conoscenza sulla natura delle cose e facendo perdere di mira gli obiettivi pratici.

11.6.1 Ragionamento intuitivo

Il modo più semplice di ragionare, senz'altro valido quando n è abbastanza grande, è il seguente.

- Abbiamo visto che, assumendo il parametro σ identico in tutte le osservazioni, la previsione di $\mathrm{E}(\mu)$ è uguale alla media aritmetica, indipendentemente dal valore di σ, con incertezza di previsione $\sigma/\sqrt{n}$.
- Possiamo allora dire che il valore vero di μ sia approssimativamente $\overline{x}$, e quindi gli scarti $x_i - \overline{x}$ siano circa uguali agli scarti delle osservazioni dal valore vero.
- La media dei quadrati degli scarti rispetto alla media è, di consequenza, approssimativamente ugale alla media degli scarti rispetto al valore vero, ovvero $\sigma^2 \approx s^2$.

In conclusione, abbiamo

$$\mathrm{E}(\mu) = \overline{x} \tag{11.31}$$

$$\sigma_e \approx s \tag{11.32}$$

$$\sigma(\mu) \approx \frac{s}{\sqrt{n}} \tag{11.33}$$

$$\mu \sim \approx \mathcal{N}(\overline{x}, \frac{s}{\sqrt{n}}) \tag{11.34}$$

Si noti il simbolo "=" per la prima uguaglianza e "$\approx$" per le altre. In fatti mentre l'espressione della previsione di μ è "esatta" (nel senso di previsione probabilistica e con la sua incertezza) le alte dipendono dall'incertezza di previsione di μ e quindi sono "esatte" soltanto nel caso di n "molto grande".

Nel seguito vedremo, più formalmente, l'origine e il limite di queste approssimazioni. Per ora possiamo assicurare che questo ragionamento va abbastanza bene, ai fini dei risultati quantitativi e interessandoci soltanto alla ragione di μ dove è condensata la massa di probabilità, per $n \geq \mathcal{O}(10)$.

11.6.2 Possibili dubbi sul modello normale

A questo punto è naturale la domanda di come si faccia ad essere "sicuri" del modello gaussiano. Come discusso più volte nel testo e, più specificatamente nel paragrafo 11.2 introducendo l'inferenza con modello gaussiano, non siamo sicuri di niente, altrimenti non avremmo affrontato questo lungo discorso sulla logica dell'incerto. In particolare, nel paragrafo 11.2 abbiamo passato in rassegna le credenze che ci portano ad adottare, almeno in prima istanza, il modello gaussiano. Se, da un'occhiata ai dati sperimentali, sorge il sospetto che un modello alternativo possa descrivere meglio la dispersione delle osservazioni lo si può prendere in considerazione, soppesando opportunamente l'abilità del nuovo modello nel descrivere i dati sperimentali con la sua plausibilità, secondo lo schema descritto nel capitolo 5 e sul quale torneremo nel capitolo ***. Ad esempio, è chiaro che un modello ad hoc che segue esattamente la forma dell'istogramma di frequenza osservato è quello che massimizza la descrizione dei dati (per definizione!), ma è anche quello al quale non si crede per niente! Come esempio di (*** ritrovare o rifare la figura con le 20 gaussiane e le venti distribuzioni uniformi non n picolo ***).

11.6.3 ❄ Inferenza simultanea su μ e σ

Affrontiamo ora il problema dal punto di vista generale. Per inferire μ e σ da un insieme di n osservazioni che riteniamo indipendenti (condizionatamente ad ogni ipotesi di mu e σ) e descritte da verosimiglianza normale, dobbiamo semplicemente applicare il teorema di Bayes a due numeri incerti anziché ad uno solo. Successivamente si tratta di marginalizzare la distribuzione congiunta sulla variabile che non ci interessa:

$$f(\mu, \sigma \,|\, \mathbf{x}, I) \propto f(\mathbf{x} \,|, \mu, \sigma, I) \times f_\circ(\mu, \sigma) \tag{11.35}$$

$$f(\mu \,|\, \mathbf{x}, I) = \int_0^\infty f(\mu, \sigma \,|\, \mathbf{x}, I)\, \mathrm{d}\sigma \tag{11.36}$$

$$f(\sigma \,|\, \mathbf{x}, I) = \int_{-\infty}^\infty f(\mu, \sigma \,|\, \mathbf{x}, I)\, \mathrm{d}\mu\,, \tag{11.37}$$

ove abbiamo ricordato, ancora una volta, che tutta l'inferenza dipende da tutte le condizioni di contorno I. Come al solito cè il "problema" di che prior utilizzare. Senza ripetere il lungo discorso fatto a proposito di μ, è chiaro che $f_\circ(\mu, \sigma)$ dovrebbe modellizzare, pur nella sua vaghezza, quello che ci si aspetta su μ e σ. In alcune misure si è abbastanza sicuri l'ordine di grandezza della deviazione standard ("si è stupiti se venisse oltre un valore di un ordine di grandezza in più o in meno di quanto ci si aspetta). In altre misure si aspettano valori che possono differire "tranquillamente" di uno o due ordini di grandezza rispetto a quello atteso, ma quasi certamente non di un fattore 1000 o più. In altri tipi di misure, forse è il caso più generale, le aspettazioni sull'ordine di grandezza non sono su σ ma sul coefficiente di variazione $v = \sigma/\mu$, ovvero sulla precisione, in modo largamente indipendente dal valore μ. E, infatti, le misure vengono classificate in "come ordine di grandezza", "al percento", "al per mille", e così via. Quindi, ricordando che incertezza sull'ordine di grandezza significa incertezza sul logaritmo della variabile, otteniamo le seguenti possibilità:

$$\sigma \sim \mathcal{N}(\ln \sigma_\circ, \sigma_{\ln \sigma_\circ}) \tag{11.38}$$
$$v = \frac{\sigma}{\mu} \sim \mathcal{N}(\ln v_\circ, \sigma_{\ln v_\circ})\,. \tag{11.39}$$

Si capisce come, una volta combinate queste prior su σ o su v con quella su μ e inserite nella formula di Bayes, i conti diventano complicati e, come vedremo, non vale la pena di farli, a meno che non si tratti di un problema cruciale. Vediamo un paio di modi di modellizzare la vaghezza su σ in modo da semplificare i conti e, in base ai risultati ottenuti, di capire se vale la pena di fare di meglio.

Prior uniforme in σ

Come per μ, il primo modello matematicamente semplice di vaghezza che salta in mente è una distribuzione uniforme per valori positivi di σ. Ovviamente, questo modello va preso con cautela, come anche nel caso di μ, in quanto certamente non crediamo allo stesso modo a tutti i valori di σ, specialmente quelli molto prossimi a zero o che tendono a infinito (rispetto alla scala che abbiamo in mente di valori plausibili). Dal modello otteniamo (assumendo implicito il condizionante di contorno I):

$$f(\mu, \sigma \,|\, \mathbf{x}) \propto f(\mathbf{x} \,|, \mu, \sigma) \tag{11.40}$$
$$\propto \prod_i \frac{1}{\sigma} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] \tag{11.41}$$
$$\propto \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2\right] \tag{11.42}$$
$$\propto \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2}\left(n\,(\overline{x}-\mu)^2 + n\,s^2\right)\right], \tag{11.43}$$

ove nell'ultimo passaggio abbiamo utilizzato l'uguaglianza

$$\sum_{i=1}^n (x_i-\mu)^2 = n(\overline{x}-\mu)^2 + \sum_{i=1}^n (x_i-\overline{x})^2 = n(\overline{x}-\mu)^2 + n\,s^2\,,$$

con s pari alla deviazione standard calcolata sui dati sperimentali:

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^n (x_i-\overline{x})^2} \tag{11.44}$$

Marginalizzando su σ abbiamo: [6]

$$
\begin{aligned}
f(\mu \,|\, \mathbf{x}) &= \int_0^\infty f(\mu, \sigma)\, \mathrm{d}\sigma \\
&\propto \left(n\, (\overline{x} - \mu)^2 + n\, s^2\right)^{-(n-1)/2} \\
&\propto \left(1 + \frac{(\mu - \overline{x})^2}{s^2}\right)^{-(n-1)/2} \qquad (11.45)\\
&\propto \left(1 + \frac{(\mu - \overline{x})^2}{(n-2)\, s^2/(n-2)}\right)^{-((n-2)+1)/2}
\end{aligned}
$$

Avendo moltiplicato e diviso s^2 per $(n-2)$, si riconosce una forma del tipo

$$\propto \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2} \qquad (11.46)$$

con

$$\nu = n - 2 \qquad (11.47)$$

$$t = \frac{\mu - \overline{x}}{s/\sqrt{n-2}}\,, \qquad (11.48)$$

ovvero

$$\mu = \overline{x} + \frac{s}{\sqrt{n-2}}\, t\,, \qquad (11.49)$$

ove t è la variabile t di Student con $\nu = n - 2$. Applicando le note proprietà della distribuzione di Student (vedi paragrafo 8.14.4), otteniamo:

$$\frac{\mu - \overline{x}}{s/\sqrt{n-2}} \sim \text{Student}(\nu = n - 2) \qquad (11.50)$$

$$\mathrm{E}(\mu) \overset{(n \geq 3)}{=} \overline{x} \qquad (11.51)$$

$$\sigma(\mu) \overset{(n \geq 4)}{=} \frac{s}{\sqrt{n-4}}\,. \qquad (11.52)$$

Come si vede, l'incertezza su σ produce una ulteriore incertezza su μ, tale da rendere più plausibili valori molto lontani dalla media (caratteristica delle t di Student rispetto alla normale standardizzata). Questo effetto è, come è ragionevole che sia, più importante per n piccolo e si attenua immediatamente quando n supera il valore di qualche decina. Per n al di sotto di $n = 5$ le code della distribuzione sono talmente pronunciate che la varianza è infinita e per $n < 4$ è addirittura il calcolo della media a non convergere. Questo va bene dal punto di vista matematico, ma non vuol dire che "non si può dire niente su μ". Queste divergenze non sono altro che il risultato della ipersemplicità del modello. Nessuna persona ragionevole crederà mai che avendo letto su un

[6] L'integrale di interesse è

$$\int_0^\infty z^{-n} \exp\left[-\frac{C}{2\, z^2}\right] \mathrm{d}z = 2^{(n-3)/2}\, \Gamma\left[\frac{1}{2}(n-1)\right] C^{-(n-1)/2}.$$

voltmetro 6.25 V, 6.32 V e 6.29 V crederà mai che il valore vero della tensione sia compatibile con valori "infiniti", sia positivi che negativi (è quello che ci dice la t di Student, e anche la più tranquilla gaussiana, seppur con gli infiniti "un po' meno probabili", per dirla alla buona).

Quando n diventa molto grande otteniamo gli stessi risultati delle argomentazioni intuitive discusse precedentemente (basate giustappunto su tale limite):

$$\mathrm{E}(\mu) \xrightarrow{n\to\infty} \overline{x} \tag{11.53}$$

$$\sigma(\mu) \xrightarrow{n\to\infty} \frac{s}{\sqrt{n}} \tag{11.54}$$

$$\mu \xrightarrow{n\to\infty} \sim \mathcal{N}(\overline{x}, s/\sqrt{n}). \tag{11.55}$$

Seguitiamo ora con questa parte formale, ritornando successivamente a raccomandazioni su come comportarsi in pratica.

Prior uniforme in $\log\sigma$

Vediamo ora cosa succede se si sceglie una posizione di assoluta indifferenza sugli ordini di grandezza di σ, posizione assurda quanto quella precedente, ma se non altro un po' più ragionevole della precedente e con il vantaggio pratico di smorzare un po' gli eccessivamente grandi valori di σ responsabili delle divergenze.

Assumere che $f_\circ(\ln\sigma) = k$ è equivalente a $f_\circ(\sigma) = 1/\sigma$. Inserendo questa prior nei conti precedenti, l'effetto è di diminuire di 1 la potenza di σ nell'integrando. L'effetto sulla $f(\mu)$ è che la potenza dell'espressione finale diventa $-n/2$ anziché $-(n-1)/2$. Di consequenza, abbiamo ancora una t di Student, ma con $\nu = n-1$ e nella variabile $(\mu-\overline{x})/(s/\sqrt{n-1})$, da cui:

$$\frac{\mu-\overline{x}}{s/\sqrt{n-1}} \sim \text{Student}(\nu = n-1) \tag{11.56}$$

$$\mathrm{E}(\mu) \stackrel{(n>2)}{=} \overline{x} \tag{11.57}$$

$$\sigma(\mu) \stackrel{(n>3)}{=} \frac{s}{\sqrt{n-3}}\,. \tag{11.58}$$

Per n piccoli questo modello produce una incertezza su μ minore di quella del modello precedente, ma anche questa è da considerarsi molto conservativa (e quindi non coerente!) perché usa una prior su σ irragionevole per qualsiasi applicazione pratica. Quando n aumenta abbiamo una più rapida convergenza al modello normale in quanto l'osservazione "solida" di s esclude valori troppo fantasiosi per σ.

Per un confronto quantitativo fra i diversi modelli, riportiamo in tabella 11.1 i valori del semiampiezza Δ, in unità di s, tale che $P(\overline{x}-\Delta \le \mu \le \overline{x}+\Delta)$ sia uguale al 50%, al 90%, a; 95% e al 99%. Ad esempio, con $n = 10$ osservazioni dalle quali abbiamo ricavato $\overline{x}$ e s, abbiamo: $P(\overline{x} - 0.66\,s \le \mu \le \overline{x} + 0.66\,s\,|\,f_\circ(\sigma) = k) = 90\%$, $P(\overline{x} - 0.61\,s \le \mu \le \overline{x} + 0.61\,s\,|\,f_\circ(\ln\sigma) = k) = 90\%$ e così via. Per i ragionamenti fatti sull'eccessiva prudenza di entrambi i modelli matematicamente abbordabili, si possono considerare valori ragionevoli per Δ quelli circa intermedi fra il modello $f_\circ(\ln\sigma) = k$

		$P_{50\%}$	$P_{90\%}$	$P_{95\%}$	$P_{99\%}$
$n=4$	$f(\sigma)=k$	0.58	2.06	3.04	7.03
	$f(\ln\sigma)=k$	0.44	1.36	1.83	3.38
	Normale $\sigma=s$	0.34	0.82	0.98	1.29
$n=6$	$f(\sigma)=k$	0.37	1.07	1.40	2.30
	$f(\ln\sigma)=k$	0.33	0.90	1.15	1.81
	Normale $\sigma=s$	0.28	0.67	0.80	1.05
$n=8$	$f(\sigma)=k$	0.29	0.79	1.00	1.52
	$f(\ln\sigma)=k$	0.27	0.72	0.90	1.32
	Normale $\sigma=s$	0.24	0.58	0.69	0.92
$n=10$	$f(\sigma)=k$	0.25	0.66	0.82	1.20
	$f(\ln\sigma)=k$	0.23	0.61	0.75	1.09
	Normale $\sigma=s$	0.21	0.52	0.62	0.82
$n=20$	$f(\sigma)=k$	0.16	0.41	0.49	0.68
	$f(\ln\sigma)=k$	0.16	0.40	0.48	0.66
	Normale $\sigma=s$	0.15	0.37	0.44	0.58
$n=50$	$f(\sigma)=k$	0.10	0.24	0.29	0.39
	$f(\ln\sigma)=k$	0.10	0.24	0.29	0.38
	Normale $\sigma=s$	0.10	0.23	0.28	0.36
$n=100$	$f(\sigma)=k$	0.07	0.17	0.20	0.26
	$f(\ln\sigma)=k$	0.07	0.17	0.20	0.26
	Normale $\sigma=s$	0.07	0.16	0.20	0.26

Tabella 11.1: Semiampiezza in unità di s Δ dell'intervallo intorno al valore medio tale che racchiuda con probabilità $P_{xx}\%$ il valore vero di μ. Nel caso di σ ignota la probabilità dipende dalla prior $f(\sigma)$. Per confronto è riportato il caso limite gaussiano nell'ipotesi che σ sia esattamente uguale a quella osservata.

e quello normale in cui si assume $\sigma = s$. in questo esempio avremmo: $P(\overline{x}- \approx 0.57\,s \le \mu \le \overline{x}+ \approx 0.57\,s\,|\,f_\circ(\ln\sigma) = k) = 90\%$. Come si vede, tenendo conto degli arrotondamenti con i quali si forniscono le incertezze, già con $n = 10$ possiamo considerarci in approssimazione normale, a meno di non essere interessati a valori molto lontani da dove si concentra la massa di probabilità.

Incertezza su σ

È interessante, a questo punto marginalizzare la funzione congiunta $f(\mu,\sigma)$ rispetto a μ per inferire i valori di σ. Facciamo i conti nei due modelli per $f_\circ(\sigma)$.

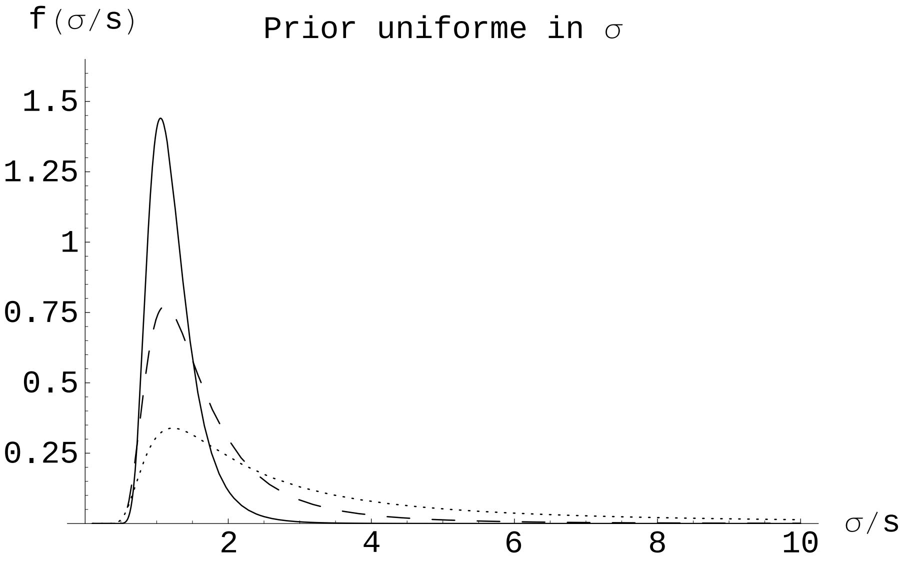

Figura 11.5: Funzione densità di probabilità del parametro σ della gaussiana, in unità della deviazione standard sui valori osservati, assumendo $f_\circ(\sigma) = k$ per un numero n di osservazioni pari 3 (curva puntinata), 5 (tratteggiata) e 10 (continua).

- Prior uniforme in σ:

$$\begin{aligned} f(\sigma\,|\,\mathbf{x}) &= \int_{-\infty}^{+\infty} f(\mu,\sigma\,|\,\mathbf{x})\,\mathrm{d}\mu \\ &\propto \sigma^{-n}\exp\left[-\frac{n\,s^2}{2\,\sigma^2}\right]\int_{-\infty}^{+\infty}\exp\left[-\frac{n\,(\overline{x}-\mu)^2}{2\sigma^2}\right]\mathrm{d}\mu \\ &\propto \sigma^{-(n-1)}\exp\left[-\frac{n\,s^2}{2\,\sigma^2}\right], \end{aligned} \tag{11.59}$$

in quanto l'integrale è pari a $\sqrt{2\,\pi}\sigma/\sqrt{n}$ (la ben nota condizione di normalizzazine della gaussiana).

La forma della funzione è molto asimmetrica per n piccoli, mentre tende ad una gaussiana per $n \to \infty$. La figura 11.5 mostra degli esempi. Queste sono le espressioni di moda, valore atteso e deviazione standard in unità di s

$$\mathrm{Moda}(\sigma)/s = \sqrt{\frac{n}{n-1}} \tag{11.60}$$

$$\mathrm{E}(\sigma)/s \overset{(n>3)}{=} \underbrace{\sqrt{\frac{n}{2}}\frac{\Gamma[(n-3)/2]}{\Gamma[n/2-1]}}_{\alpha_0} \tag{11.61}$$

$$\mathrm{DevSt}(\sigma)/s \overset{(n>4)}{=} \frac{1}{\sqrt{2\,n}}\,n\underbrace{\sqrt{\frac{2}{n-4}-\frac{\Gamma^2\left[(n-3)/2\right]}{\Gamma^2[n/2-1]}}}_{\beta_0} \tag{11.62}$$

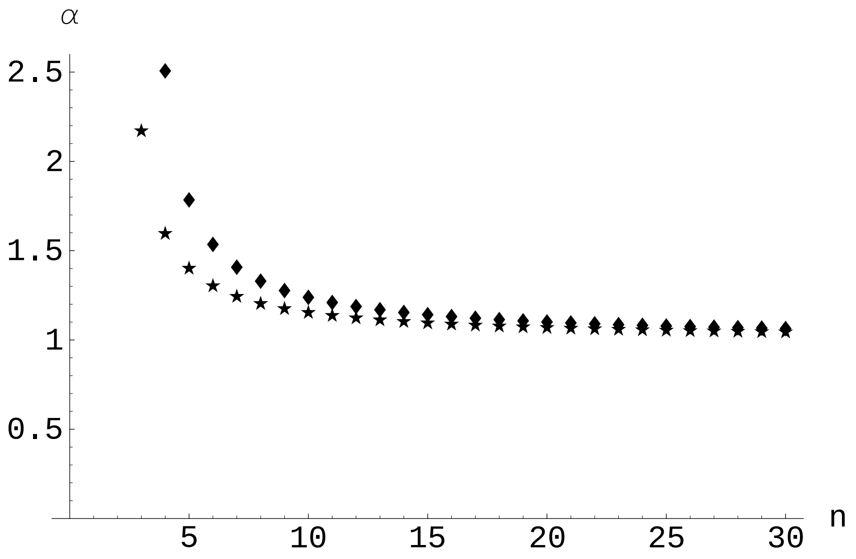

Figura 11.6: Fattore α fra valore atteso di σ e deviazione standard delle osservazioni sperimentali in funzione del numero di osservazioni. I rombi si riferiscono al caso di prior uniforme in σ, le stelle al caso di uniforme in $\ln\sigma$.

ove abbiamo designato con "DevSt" la deviazione standard di σ, per ovvi motivi. I due fattori complicati $\alpha_\circ$ e $\beta_\circ$ sono rilevanti soltanto per piccoli valori n. Per $n \to \infty$ essi tendono a 1 e già per $n = 10$ differiscono dall'unità per meno del 50% (vedi figure 11.6 e 11.7). Quindi, per n grandi abbiamo i seguenti valori asintotici:

$$\mathrm{E}(\sigma) \xrightarrow[n\to\infty]{} s \tag{11.63}$$

$$\mathrm{DevSt}(\sigma) \xrightarrow[n\to\infty]{} \frac{s}{\sqrt{2\,n}} \tag{11.64}$$

$$\sigma \xrightarrow[n\to\infty]{} \sim \mathcal{N}(s, s/\sqrt{2\,n})\,. \tag{11.65}$$

(Si noti come il limite a normale è, per ora, una congettura basata sull'osservazione delle curve. Nel paragrafo 11.6.4 vedremo un altro argomento più formale, basato sulla distribuzione di probabilità di $1/\sigma^2$.)

- Prior uniforme in $\ln\sigma$: la prior $f_\circ(\sigma) \propto 1/\sigma$ abbassa di un grado la potenza di σ nella funzione finale:

$$f(\sigma\,|\,\mathbf{x}) \;=\; \sigma^{-n}\exp\left[-\frac{n\,s^2}{2\,\sigma^2}\right], \tag{11.66}$$

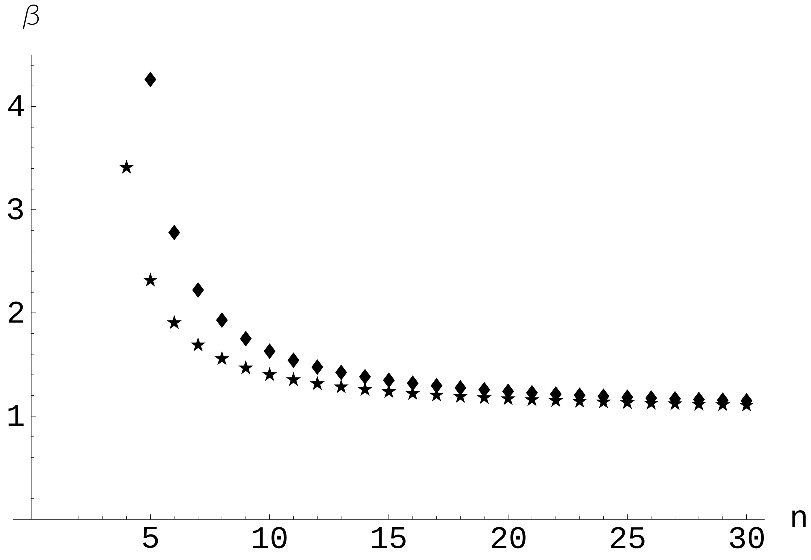

Figura 11.7: Fattore β fra incertezza standard su σ e $s/\sqrt{2\,n}$ in funzione del numero di osservazioni. I rombi si riferiscono al caso di prior uniforme in σ, le stelle al caso di uniforme in $\ln\sigma$.

da cui seguono

$$\mathrm{Moda}(\sigma)/s \;=\; 1 \tag{11.67}$$

$$\mathrm{E}(\sigma)/s \;\stackrel{(n\geq 2)}{=}\; \underbrace{\sqrt{\frac{n}{2}}\frac{\Gamma[(n/2-1]}{\Gamma[(n-1)/2]}}_{\alpha_1} \tag{11.68}$$

$$\mathrm{DevSt}(\sigma)/s \;\stackrel{(n\geq 3)}{=}\; \frac{1}{\sqrt{2\,n}}\underbrace{n\sqrt{\frac{2}{n-3}-\frac{\Gamma^2\,[n/2-1]}{\Gamma^2[(n-1)/2]}}}_{\beta_1} \tag{11.69}$$

Come succedeva per la previsione di μ, questo secondo modello dà delle curve un po' più strette per n piccoli, mentre le differenze diventano immediatamente irrilevanti quando n arriva all'ordine della decina. In particolare, gli andamenti asintotici

$$\mathrm{E}(\sigma) \;\xrightarrow[n\rightarrow\infty]{}\; s \tag{11.70}$$

$$\mathrm{DevSt}(\sigma) \;\xrightarrow[n\rightarrow\infty]{}\; \frac{s}{\sqrt{2\,n}} \tag{11.71}$$

$$\sigma \;\xrightarrow[n\rightarrow\infty]{}\; \sim\mathcal{N}(s, s/\sqrt{2\,n})\,. \tag{11.72}$$

sono gli stessi. Degli esempi sono mostrati in figura 11.8. Le funzioni α_1 e β_1 sono mostrate nelle figure 11.6 e 11.7.

11.6.4 Distribuzione di $1/\sigma^2$

Abbiamo visto come, seppur con delle scelte di convenienza formale delle prior, le funzioni distribuzioni finali di σ sono alquanto inconsuete. È

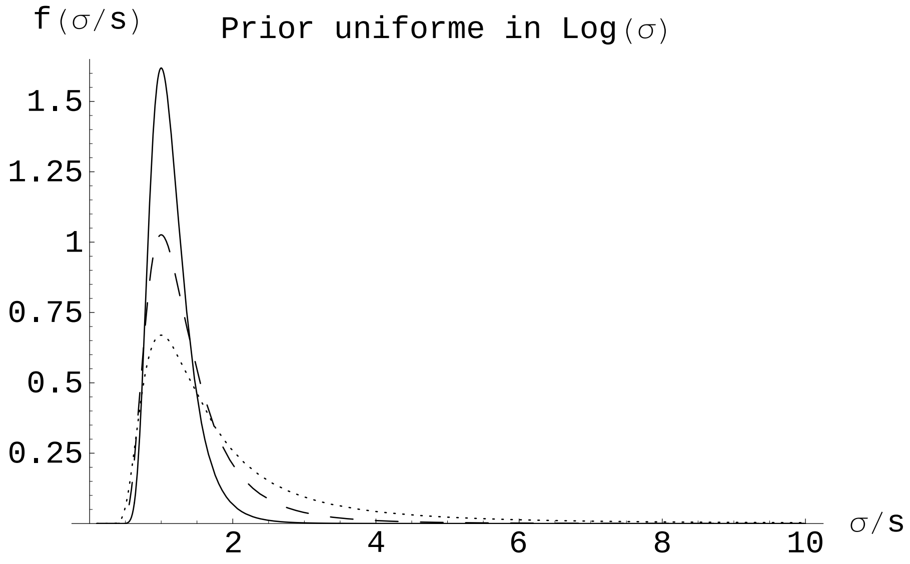

Figura 11.8: Come figura 11.5, con prior uniforme in $\ln\sigma$.

interessante studiare la forma di $1/\sigma^2$. Chiamando

$$z = \frac{n\,s^2}{\sigma^2}\,,$$

abbiamo dalle (11.59) e (11.66) rispettivamente

$$f(z\,|\,\text{prior uniforme in }\sigma) \;\propto\; z^{(n-2)/2-1}\,e^{-z} \qquad (z\geq 0) \tag{11.73}$$
$$f(z\,|\,\text{prior uniforme in }\ln\sigma) \;\propto\; z^{(n-1)/2-1}\,e^{-z} \qquad (z\geq 0) \tag{11.74}$$

Confrontando con la (8.45) si riconoscono nelle due funzioni delle distribuzioni di χ^2 con un numero di gradi di libertà rispettivamente $\nu = n-2$ e $\nu = n-1$. Possiamo utilizzare questa osservazione per ricavare in modo approssimato previsione e incertezza di previsione di σ, sotto ipotesi che la linearizzazione sia soddisfacente, ovvero l'incertezza relativa su z sia piccola, ovvero ancora che n sia grande. Vediamo il solo caso di prior uniforme in $\ln\sigma$, essendo l'altro caso simile e, comunque, praticamente indistinguibile nei casi in cui vale la linearizzazione.

$$s = \frac{\sqrt{n}\,s}{\sqrt{z}} \tag{11.75}$$
$$\mathrm{E}(\sigma) \approx \frac{\sqrt{n}\,s}{\sqrt{z}} = \sqrt{\frac{n}{n-2}}\,s \approx s \tag{11.76}$$
$$\mathrm{DevSt}(\sigma) \approx \left|\frac{\partial\sigma}{\partial z}\right| \mathrm{DevSt}(z) = \frac{\sqrt{n}\,s}{\sqrt{2}\,(n-2)} \approx \frac{s}{\sqrt{2\,n}}\,, \tag{11.77}$$

riottendo i risultati asintotici calcolati direttamente. In particolare, per $n\rightarrow\infty$ abbiamo che $z \sim \mathcal{N}(n, \sqrt{2\,n})$, con coefficiente di variazione $v = \sqrt{2}/\sqrt{n}$. Quindi la linearizzazione diventa una buona approssimazione e anche $f(\sigma)$ diventa gaussiana.

Terminiamo questo paragrafo facendo notare un altro modo di scrivere il risultato ottenuto.

$$\text{prior uniforme in } \sigma \quad : \quad \frac{\sum_i (x_i - \overline{x})^2}{\sigma^2} \sim \chi^2_{n-2} \qquad (11.78)$$

$$\text{prior uniforme in } \log\sigma \quad : \quad \frac{\sum_i (x_i - \overline{x})^2}{\sigma^2} \sim \chi^2_{n-1} \,. \qquad (11.79)$$

11.6.5 Conclusioni e raccomandazioni

A questo punto è importante riordinare brevemente le idee, per non lasciarsi prendere dal formalismo e perdere di mira gli obiettivi dei metodi che stiamo studiando. All'inizio del paragrafo 11.6.3 abbiamo illustrato la procedura corretta che permetterebbe di implementare lo stato di informazione dello sperimentatore. Purtroppo gli integrali diventano complicati. Successivamente abbiamo visto due modelli che semplificano i conti, pur risultando in formule non banali. Le formule che ne derivano vanno intese come ordine di grandezza di incertezze molto prudenti. Comunque, esse indicano pure che i ragionamenti intuitivi descritti nel paragrafo 11.6.1 vanno nella direzione giusta quando si ha a disposizione oltre una decina di osservazioni. Quindi, in sostanza, le raccomandazioni seguono più o meno la prassi degli sperimentali esperti.

- Se si hanno oltre una decina di osservazioni, i risultati non differiscono molto da quelli ottenuti da un modello gaussiano con σ_e nota pari a s.

- Se i dati sperimentali sono veramente pochi, dell'ordine dell'unità, è più importante la conoscenza a priori su σ_e data dall'esperienza acquisita precedentemente che quanto si ottiene dalla dispersione dei valori. In questo caso il giudizio dell'esperto è più affidabile di qualsiasi argomento teorico. Come esempio, consideriamo il caso di 4 osservazioni, assumendo una prior su σ uniforme in $\log\sigma$. L'incertezza relativa su σ vale $1/\sqrt{2\times 4}\times(\beta_1(4)/\alpha_1(4)) \approx 50\%$. Quest'incertezza è confrontabile con quanto può stimare uno sperimentatore esperto prima ancora di eseguire le misure.

Nel seguito assumeremo di essere approssimativamente in regime di normalità.

11.7 Distribuzione predittiva

Nei paragrafi precedenti abbiamo visto la funzione $f(x\,|\,\mu)$, interpretata come funzione di verosimiglianza di ottenere un certo valore osservato data l'ipotesi μ. Interessiamoci ora al problema di determinare veramente la probabilità di un valore ignoto. Ci rendiamo conto che $f(x\,|\,\mu)$ non è sufficente. Infatti non siamo interessati alla probabilità di x per ogni ipotesi μ, bensì alla probabilità che tenga conto di tutti i valori possibili di μ, pesati con la loro plausibilità.

$$f(x\,|\,I) = \int f(x\,|\,\mu, I)\cdot f(\mu\,|\,I)\,\mathrm{d}\mu\,. \qquad (11.80)$$

Nel caso in cui la conoscenza di μ ci deriva dall'aver osservato il singolo (o valore equivalente) valore (passato) x_p con un esperimento avente deviazione

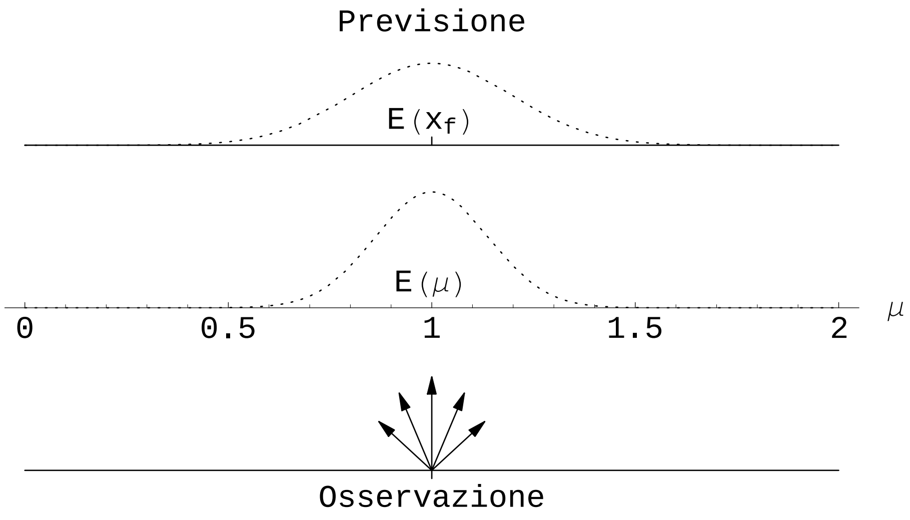

Figura 11.9: Schema di inferenza predittiva attraverso lo stadio intermedio "metafisico" μ.

standard $\sigma_e = \sigma_p$, e il valore (futuro) x_f che andremo ad osservare deriverà da un esperimento con $sigma_e = \sigma_f$, abbiamo:

$$f(x_f \,|\, x_p) = \int \frac{1}{\sqrt{2\,\pi}\sigma_f}\, e^{-\frac{(x_f-\mu)^2}{2\,\sigma_f^2}} \frac{1}{\sqrt{2\,\pi}\sigma_p}\, e^{-\frac{(\mu-x_p)^2}{2\,\sigma_p^2}}\, \mathrm{d}\mu \quad (11.81)$$

$$= \frac{1}{\sqrt{2\,\pi}\sqrt{\sigma_p^2+\sigma_f^2}}\, e^{-\frac{(x_f-x_p)^2}{2\,(\sigma_p^2+\sigma_f^2)}}\,, \quad (11.82)$$

con

$$\mathrm{E}(X_f) = x_p \quad (11.83)$$

$$\sigma(X_f) = \sqrt{\sigma_p^2+\sigma_f^2}\,. \quad (11.84)$$

Si noti come la distribuzione *predittiva* (11.82) descrive l'incertezza su un valore ignoto condizionata da una precedente osservazione, essendo sparito il "metafisico" valore vero (nel senso di non accessibile ai nostri sensi) μ. La figura 11.9 mostra lo schema di inferenza dall'osservazione certa x_p all'osservazione (futura) incerta x_f attraverso lo stadio intermedio non osservabile μ.

Si noti in particolare il caso di $\sigma_f = \sigma_p = \sigma_\circ/\sqrt{n}$, ovvero di osservazioni equilavalenti ottenute nelle stesse condizioni sperimentali e con la stessa quantità di osservazioni individuali. Abbiamo $\sigma(X_f) = \sqrt{2}\,\sigma_\circ/\sqrt{n}$. In altri termini, riteniamo probabile al 52% che la nuova media cada entro $\pm\sigma_\circ/\sqrt{n}$ da quella precedente.[7]

[7]Non è raro sentire o leggere che tale probabilità è del 68%. È chiaro l'errore concettuale che si sta commettendo.

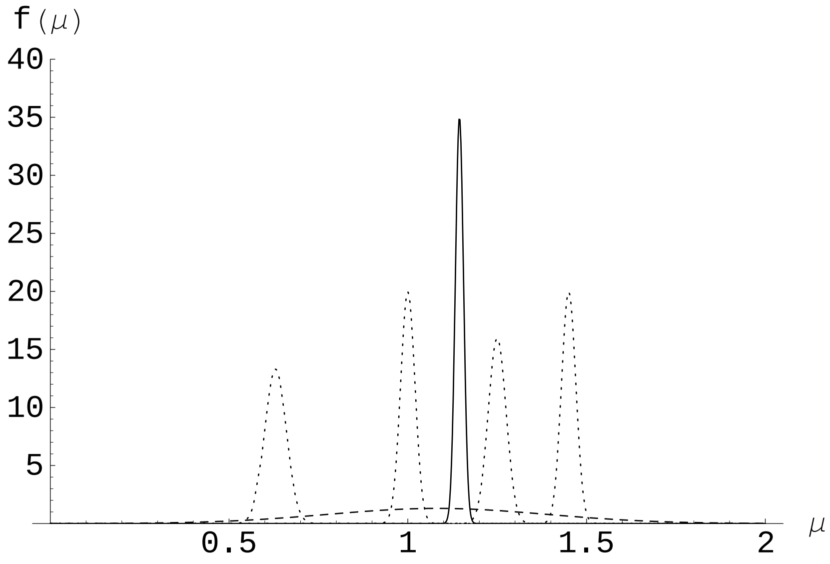

Figura 11.10: Esempio di combinazione di risultati parziali in "palese" (vedi testo) disaccordo fra di loro. La curva continua rappresenta l'inferenza globale ottenuta dalla "formula standard" di combinazione. Quella tratteggiata tiene conto del forte dubbio che ci sia qualche effetto spurio fra i diversi risultati.

11.8 Combinazione scettica

Nel paragrafo 11.3 abbiamo visto come combinare risultati parziali indipendenti in un risultato globale. Consideriamo ora il caso illustrato in figura 11.10 in cui le masse di probabilità indicate da ciascuna inferenza (curve puntinate) non si sovrappongono[8]. Applicando le formule di combinazione (11.12) e ((11.13) si ottiene il risultato indicato con la curva continua di figura 11.10. A questo punto ci sono ottime ragioni per rimanere perplessi: il risultato finale concentra la probabilità in una zona molto più ristretta del singolo risultato parziale, ma non indicata da nessuno di quei risultati, non tenendo conto dell'enorme variabilità da un risultato all'altro. Cerchiamo di capire cosa sta succedendo, elencando le ipotesi alla base della "combinazione standard" data dalle (11.12) e ((11.13).

1. Tutte le misure si riferiscono alla stessa grandezza ("stiamo misurando tutti la stessa cosa").

2. La verosimiglianza di ciascuna misura è descritta da una gaussiana.

3. Il parametro σ della gaussiana è stato stimato correttamente.

4. I risultati sono indipendenti.

5. Infine, la conoscenza iniziale di μ è supposta essere sufficientemente vaga rispetto a quanto forniscono le singole verosimiglianze.

[8]Questa affermazione vale per le regioni in cui si ammassa il più alto grado di fiducia: avendo assunto un modello gaussiano, in cui le variabili possono assumere valori su tutto l'asse reale, le regioni di certezza (da $-\infty$ a $+\infty$) si sovrappongono sempre!

Se crediamo in modo assoluto a tutte queste ipotesi, non possiamo far altro che accettare il risultato ottenuto dalla combinazione standard, in quanto, trattandosi di fenomeni aleatori, l'osservazione di questa configurazione di risultati parziali *non è incompatibile* con con il modello (come non è incompatibile con le leggi della probabilità che una moneta dia testa 100 volte di seguito, o che un processo poissoniano di $\lambda = 1$ produca l'osservazione $x = 1000$. Nonostante questi richiami alle "leggi del caso", qualsiasi persona esperta è disposta a scommettere che c'e' qualcosa che non va nei quattro risultati di figura 11.10. E ha ragione. Il motivo è che un ricercatore esperto ha sviluppato delle fortissime prior[9] sul comportamento di strumentazioni, misure ... e colleghi. Sulla base di queste prior, *sa* che gli strumenti possono essere non calibrati, che le tecniche sperimentali possono essere inadeguate allo misura particolare, che i colleghi si possono sbagliare, e così via. Alla luce di queste considerazioni, è praticamente sicuro che che qualcosa sia *andata male* nelle misure, piuttosto che attribuire la dispersione di risultati al puro caso.

Dopo queste riflessioni, si tratta di rimodellizzare il problema, cercando di modificare una o più delle ipotesi alla base della combinazione normale. Come si capisce, il problema non ha una soluzione unica. Si tratta soltanto di scegliere i modelli che sembrano più ragionevoli e confrontarne i risultati. La sola ipotesi tranquilla è quella sulla prior uniforme su μ. In questo testo non intendiamo affrontare il caso più generale[10], ma ci limitiamo a fornire delle indicazioni su caso illustrato nella figura 11.10 di risultati "molto incompatibili" fra di loro. Ciascuna delle ipotesi 1-3 possono essere soggette a critica. Un modo di schematizzare il problema è quello di concentrarsi sulla 1 e 3. Il ricercatore esperto dirà infatti che "non stiamo misurando la stessa cosa", ovvero "ci siamo dimenticati di effetti sistematici, diversi in entità in ciascun risultato parziale" (il che è più o meno la stessa cosa), ovvero punta il dito sull'ipotesi 1:

$$\mu_i = \mu + \eta_i \,, \tag{11.85}$$

avendo indicato con μ il valore ideale che tutti "volevano misurare", η_i l'errore sistematico di entità ignota, e μ_i il valore vero che essi "hanno in fatti determinato".

Se sapessimo, mediante opportune calibrazioni, quanto vale η_i per ciascun laboratorio, il nostro problema è risolto. Ma se siamo in questa situazione è perché non non riusciamo a intercalibrare i laboratori e, inoltre, non abbiamo alcun motivo di preferire uno di essi. A questo punto non possiamo fare altro che considerare η_i come un errore aleatorio e stimarne la distribuzione dalla dispersione dei dati sperimentali, come abbiamo visto nei paragrafi precedenti, ritenendo il modello gaussiano per semplicità. Il caso numeri co in esame si presta particolarmente bene per quasta modellizzazione, in quanto possiamo ignorare completamente le singole deviazioni standard e concentrarci soltanto sui valori centrali. Il risultato è di **** ed è mostrato nella curva tratteggiata. Come si vede, tale curva descrive abbastanza bene la nostra incertezza, in quanto il valore vero si può trovare con alta probabilità fra il risultato minimo e quello massimo.

[9] Soltanto il "perfetto idiota" (a trovarlo) è esente da prior. Ma non è la persona giusta a cui rivolgersi per aumentare la nostra conoscenza.

[10] Si veda ad esempio G. D'Agostini, "sceptiical .."

Discutere in footnote il caso intermedio: vedi mail a Press.

11.9 Problemi

1. Uno sperimentatore ritiene che la risposta del suo rivelatore sia descritta da una gaussiana centrata nel valore vero, con deviazione standard $\sigma_e = 0.5$ in unità arbitrarie. Esegue tre osservazioni, ottenendo $x_1 = 5.3$, $x_2 = 6.4$ e $x_3 = 5.8$. Quanto vale l'intervallo entro cui lo sperimentatore riterrà contenuto il valore vero al 95 Come presenterà il risultato usando previsione e incertezza standard? determinto
2. Sullo stesso problema: quante misure deve fare per ottenere un'incertezza relativa di circa il 2%?
3. Il risultato $\mu = 5.83 \pm 0.29$ è combinato con altri due risultati ottenuti indipendentemente da altri sperimentatori: 5.72 ± 0.17 e 6.11 ± 0.22. Quanto varrà il risultato combinato?
4.
5. Uno sperimentatore è interessato a misurare una grandezza il cui valore numerico è ritenuto essere di circa 0.15 in opportune unità. L'errore casuale sulla singola osservazione è di circa 0.2. Inoltre, la misura è affetta da una incertezza, di natura additiva di 0.0010. Quante osservazioni deve effettuare affinché le due incertezze siano confrontabili? Quanto varrà l'incertezza relativa complessiva?
6. Sul problema precedente. Supponiamo che per raccogliere 40000 osservazioni siano necessari 3 mesi di raccolta dati. Vale la pena di raccogliere 100000 oservazioni?
7. Uno strumento è caratterizzato da una incertezza legata ad una costante di calibrazione additiva, pari a $\sigma_Z = 0.12$. L'errore casuale dello strumento è pari a $sigma_e = 1.5$. Vengono eseguite 100 misure di due grandezze omogene, ottenendo i seguenti valori medi: $\overline{x}_1 = 12.21$ e $\overline{x}_2 = 14.81$. Dare i risultati delle misure, comprensivi del coefficiente di correlazione.
8. Uno sperimentatore ottiene i seguenti dati sperimentali da un rivelatore fatto operare in condizioni apparentemente identiche: {25.4, 23.9, 20.9, 23.3, 26.8, 18.6, 19.5, 24.6 26.5 21.5}. Lo strumento è ritenuto perfettamente tarato e lo sperimentatore ritiene ragionevoli fluttuazioni dell'ordine del 10% per il tipo di strumento e di misura. Come presenterà il risultato?
9. Uno sperimentatore ritiene, per esperienza acquisita, che in un certo tipo di misura si ottengono fluttuazioni percentuali intorno all'1%. Esegue soltanto due misure ottenendo: 3.017 e 3.033. Quale riterrà essere l'incertezza sul risultato?
10. Uno strumento risponde con $\sigma_e = 0.23$. Uno sperimentatore esegue 25 misure, ottenendo un valore medio di 47.25. Successivamente, un suo collega esegue 10 misure sulla stessa grandezza, con lo stesso strumento e nelle stesse condizioni. Quanto vale la probabilità che la seconda media sia compresa entro 47.20 e 47.30?

Capitolo 12

Verosimiglianza binomiale e poissoniana. Approssimazioni

12.1 Misure di conteggi, di proporzioni e di efficienze

Misurare non è semplicemente confrontare una grandezza fisica con una data unità di misura (si pensi ad una distranza misurata con un metro) o, più in generale, applicare un certo stimolo (ad esempio una temperatura) ad uno strumento calibrato e registrare la sua indicazione. A volte siamo interessati a *contare* il numero di disintegrazioni nucleari o di globuli bianchi per inferire, rispettivamente, all'intensità del processo di decadimento o densità di globuli nel sangue. Altre volte siamo interessanti alla proporzione con la quale un certo *carattere* è contenuto in una data popolazione ad un certo istante. 'Carattere' va inteso in senso lato e può riferirsi ad una certa caratteristica fisica o patologica (si pensi alla percentuale di infetti da epatite B), oppure ad una opinione (quanti cittadini voterebbero per un certo candidato). Oppure ancora possiamo essere interessati alla probabilità dell'occorrenza di eventi futuri di una certa classe, avendo eseguito osservazioni nel passato (tutte a parità apparenti di condizioni e pensando che le stesse condizioni varranno anche per gli eventi futuri) nelle quali quel tipo di eventi si è verificato con una certa frequenza.

In altre parole, questo tipo di misure può essere ricondotto all'inferenza del parametro p di un processo di Bernoulli, oppure al parametro r di un processo di Poisson, ovvero al parametro λ di una poissoniana.

Come fatto nel capitolo precedente, cominciamo con una approccio intuitivo, trattato successivamente il caso generale, analizzando anche alcune le possibili complicazioni del problema e definendo bene quali sono le condizioni per le quali l'approccio intuitivo è soddisfacente.

12.2 Inferenza su p e λ (o r) in condizioni di normalità.

Riprendiamo l'approccio sviluppato nei paragrafi 11.2 e 11.6.1 che ci ha condotto alle inversioni intuitive di probabilità che abbiamo visto valide per la stragrande maggioranza dei casi di routine. Rieffettuiamo il ragionamento sia per i due casi di interesse.

12.2.1 Caso poissoniano

Se il numero di conteggi osservato è grande, la verosimiglianza $X \sim \mathcal{P}_\lambda$ è pressoché gaussiana, ovvero

$$X \sim\approx \mathcal{N}(\lambda, \sqrt{\lambda})\,.$$

Se si assume una condizione sufficiente vaga su λ si ha

$$\lambda \sim \mathcal{N}(x, \sigma(X))\,,$$

con

$$\begin{aligned} \mathrm{E}(\lambda) &\approx x && (12.1)\\ \sigma(\lambda) &\approx \sigma(x) \approx \sqrt{\mathrm{E}(\lambda)} \approx \sqrt{x}\,, && (12.2) \end{aligned}$$

ovvero

$$\lambda \;\sim\; \approx \mathcal{N}(x, \sqrt{x})\,. \qquad (12.3)$$

A seconda che i conteggi siano nel dominio del tempo o dello spazio si possono inferire le opportune intensità dei processi di Poisson associati dividendo per il tempo o la porzione di spazio nel quale si sono verificati i conteggi. Ad esempio, se i conteggi si sono verificati in un tempo T perfettamente noto, abbiamo

$$r \;\sim\; \mathcal{N}\left(\frac{x}{T}, \frac{\sqrt{x}}{T}\right)\,. \qquad (12.4)$$

12.2.2 Caso binomiale

Nel caso che il numero di successi sia molto maggiore di zero e molto minore di n, la verosimiglianza $X \sim \mathcal{B}_{n,p}$ è approssimativamente gaussiana, ovvero

$$X \sim\approx \mathcal{N}(n\,p, \sqrt{n\,p\,(1-p)})\,.$$

Pensando alla frequenza relativa di successi X/n, abbiamo

$$\frac{X}{n} \sim\approx \mathcal{N}(p, \frac{\sqrt{p\,(1-p)}}{\sqrt{n}})\,.$$

Di nuovo, assumendo la nostra conoscenza su p abbastanza vaga, abbiamo

$$p \;\sim\; \approx \mathcal{N}(\frac{x}{n}, \sigma(\frac{X}{n})\,. \qquad (12.5)$$

Ne segue

$$\begin{aligned} \mathrm{E}(p) &\approx \frac{x}{n} && (12.6)\\ \sigma(p) &\approx \sigma\left(\frac{X}{n}\right) && (12.7)\\ &\approx \frac{\sqrt{\mathrm{E}(p)\,[1-\mathrm{E}(p)]}}{\sqrt{n}} && (12.8)\\ &\approx \sqrt{\frac{x/n(1-x/n)}{n}}\,. && (12.9) \end{aligned}$$

12.3 ↻ Caso generale di inferenza con verosimiglianza binomiale

Affrontiamo ora il caso generale dell'inferenza di p dalla conoscenza di n e di $X = x$ e dall'aver assunto un processo di Bernoulli indipendente per ogni esito, sotto condizione che p valga un certo valore (si potrebbe scrivere "$P = p$", ma, come già visto per μ e σ, per i parametri delle distribuzioni preferiamo utilizzare soltanto la lettera minuscola). Questo è stato il famoso "*problem in the theory of chance*" che aveva indotto il reverendo Bayes a sviluppare formalmente il metodo di inversione di probabilità che porta il suo nome. Assumendo una distribuzione uniforme per p abbiamo:

$$
\begin{aligned}
f(p \,|\, x, n, \mathcal{B}) &= \frac{f(x \,|\, \mathcal{B}_{n,p})\, f_\circ(p)}{\int_0^1 f(x \,|\, \mathcal{B}_{n,p})\, f_\circ(p)\, \mathrm{d}p} \\
&= \frac{\frac{n!}{(n-x)!\,x!}\, p^x\,(1-p)^{n-x}\, f_\circ(p)}{\int_0^1 \frac{n!}{(n-x)!\,x!}\, p^x\,(1-p)^{n-x}\, f_\circ(p)\, \mathrm{d}p} \\
&= \frac{p^x\,(1-p)^{n-x}}{\int_0^1 p^x\,(1-p)^{n-x}\, \mathrm{d}p}\,. \qquad (12.10)
\end{aligned}
$$

Questa volta l'integrale a denominatore è meno banale del caso gaussiano. Il risultato finale è:

$$
f(p \,|\, x, n, \mathcal{B}) = \frac{(n+1)!}{x!\,(n-x)!}\, p^x\,(1-p)^{n-x}\,, \qquad (12.11)
$$

di cui sono mostrati alcuni esempi in Fig. 12.1. Si vede, come a parità di x/n, al crescere di n si è sempre più sicuri su p. Inoltre (semplice riflesso del limite a normale della binomiale) per n grande e x/n lontano da 0 e da 1 la funzione finale $f(p)$ ha la forma gaussiana. Valore atteso, varianza e moda di $f(p)$ sono:[1]

$$
\begin{aligned}
\mathrm{E}(p) &= \frac{x+1}{n+2} \qquad (12.12) \\
\mathrm{Var}(p) &= \frac{(x+1)(n-x+1)}{(n+3)(n+2)^2} \qquad (12.13) \\
&= \frac{x+1}{n+2}\left(\frac{n+2}{n+2} - \frac{x+1}{n+2}\right)\frac{1}{n+3} \\
&= \mathrm{E}(p)\,\left[1-\mathrm{E}(p)\right]\,\frac{1}{n+3} \qquad (12.14) \\
\mathrm{Moda}(p)\ [\equiv p_m] &= \frac{x}{n}. \qquad (12.15)
\end{aligned}
$$

Su questo problema aveva lavorato, oltre che Bayes, anche Laplace e, in particolare, la (12.12) è nota come *formula recursiva di Laplace*. Si noti come essa sia dia valori di $\mathrm{E}(p)$ diversi da quelli di sicurezza (ovvero 0 e 1) anche

[1] Si riconosce nella (12.11) una distribuzione Beta di parametri $r = x+1$ e $s = n - x + 1$ (vedi paragrafo 8.14.1). Questa identificazione sarà usata nel seguito per introdurre la distribuzione Beta come coniugata della binomiale in problemi inferenziali.

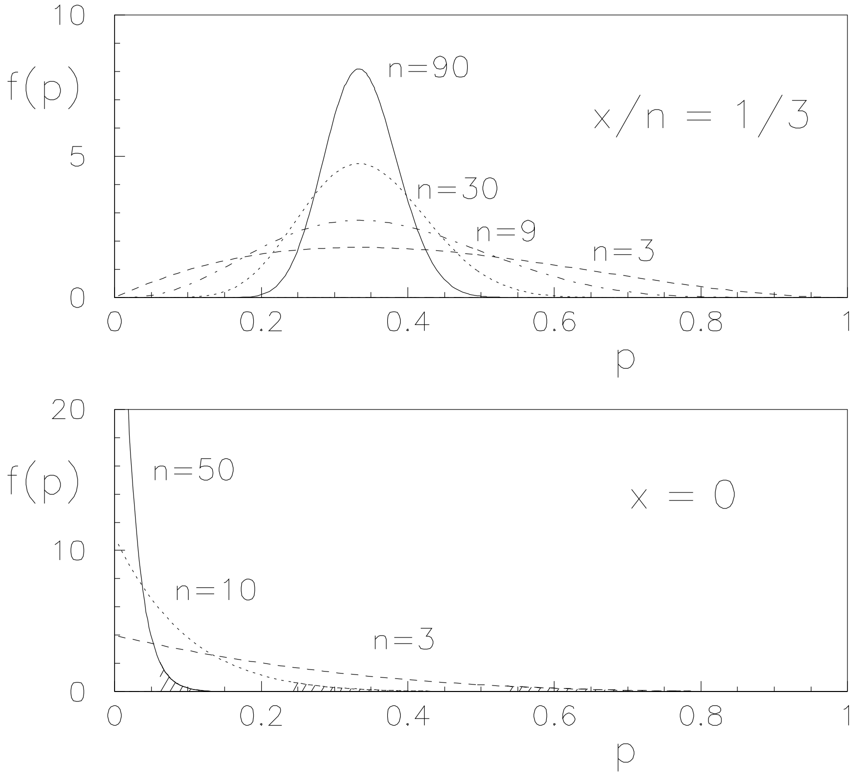

Figura 12.1: Funzione densità di probabilità del parametro p della binomiale, avendo osservato x successi in n prove.

quando X vale 0 o 1, consistente con il fatto che una inferenza asata su processi aleatori non può mai condurre alla sicurezza. Addirittura la formula dà una risposta ragionevole ($\text{E}(p) = 1/2$) per $x = n = 0$, ovvero prima di aver eseguito le misure. Questo non ha niente di magico, ma è il solo riflesso della prior uniforme su p.

12.3.1 Caso di routine

Quando x e n diventano grandi, con $0 \ll x \ll n$, $f(p)$ gode delle seguenti proprietà asintotiche:

$$\begin{aligned} \text{E}(p]) &\approx p_m = \frac{x}{n}\,, & (12.16)\\ \text{Var}(p) &\approx \frac{x}{n}\left(1-\frac{x}{n}\right)\frac{1}{n} = \frac{p_m\,(1-p_m)}{n}\,, & (12.17)\\ \sigma_p &\approx \sqrt{\frac{p_m\,(1-p_m)}{n}}\,, & (12.18)\\ p &\sim \mathcal{N}(p_m,\sigma_p)\,. & (12.19) \end{aligned}$$

Sotto queste condizioni abbiamo ricuperato il ragionamento intuitivo mostrato nel paragrafo precedente, però con una maggiore consapevolezza delle condizioni di validità. Si noti inoltre che, quando $n \to \infty$, siamo "praticamente sicuri" che p valga x/n. In un certo senso, abbiamo recuperato la "definizione" frequentista della probabilità come *regola di valutazione* sotto ben precise condizioni.

12.3.2 Casi critici

Consideriamo ora due situazioni estreme nelle quali non vale l'approssimazione gaussiana: $x = 0$ e $x = n$. Esse sono particolarmente istruttive anche perché i conti si semplificano.

- x = 0: Il caso in cui non si osserva nessun evento favorevole è illustrato in figura 12.1. Sempre assumendo una prior uniforme, abbiamo:

$$f(0 \,|\, \mathcal{B}_{n,p}) = (1-p)^n, \tag{12.20}$$
$$f(p \,|\, x=0,n) = \frac{(1-p)^n}{\int_0^1 (1-p)^n \, dp} = (n+1)\,(1-p)^n, \tag{12.21}$$
$$F(p \,|\, x=0,n) = 1-(1-p)^{n+1}\,. \tag{12.22}$$

  Come si vede, benché il valore al quale si crede di più è sempre zero, quando n è piccolo c'è ancora un ampio intervallo di valori di p compatibili con le (non) osservazioni sperimentali. All'aumentare di n cresce la confidenza che p sia sempre più piccolo. In questi casi è utile riportare il risultato come limite superiore ad un certo livello di probabilità. Tipicamente si sceglie il 95%. Otteniamo quindi:

$$F(p_\circ \,|\, x=0,n) = 0.95,$$

$$p_\circ = 1 - \sqrt[n+1]{0.05}\,, \tag{12.23}$$

  che tende a 0 per $n \to \infty$.

- x = n: il caso di tutti eenti favorevoli è complementare al precedente (basta ragionare su Q anziché su p):

$$f(n \,|\, \mathcal{B}_{n,p}) = p^n, \tag{12.24}$$
$$f(p \,|\, x=n) = \frac{p^n}{\int_0^1 p^n \, \mathrm{dp}} = (n+1)\,p^n, \tag{12.25}$$
$$F(p \,|\, x=n) = p^{n+1}\,. \tag{12.26}$$

  In questi casi si può essere interessati ad un limite inferiore ad un certo livello di probabilità. Usando di nuovo il 95% abbiamo:

$$F(p_\circ \,|\, x=n) = 0.05\,,$$

$$p_\circ = Sqrt[n+1]0.05\,. \tag{12.27}$$

  In questo caso all'aumentare di n il limite inferiore $p_\circ$ viene 'spinto' verso 1.

La seguente tabella fornisce i limiti di probabilità al 95% in funzione di n. Per confronto diamo anche il limite superiore ottenuto in approssimazione poissoniana, che vredremo fra breve.

	Probability level = 95 %		
n	$x = n$	$x = 0$	
	binomial	binomial	Poisson approx. ($p_\circ = 3/n$)
3	$p \geq 0.47$	$p \leq 0.53$	$p \leq 1$
5	$p \geq 0.61$	$p \leq 0.39$	$p \leq 0.6$
10	$p \geq 0.76$	$p \leq 0.24$	$p \leq 0.3$
50	$p \geq 0.94$	$p \leq 0.057$	$p \leq 0.06$
100	$p \geq 0.97$	$p \leq 0.029$	$p \leq 0.03$
1000	$p \geq 0.997$	$p \leq 0.003$	$p \leq 0.003$

12.3.3 Combinazione di misure indipendenti

Usiamo il caso di $x = n$ per mostrare in modo semplice come $f(p)$ viene riaggiornata alla luce di nuove prove. Immaginiamo di aver effettuato due esperimenti, rispettivamente di n_1 e n_2 prove. Ovviamente, la globalità delle informazioni consiste in $n_1 + n_2$ prove con $n_1 + n_2$ successi, ovvero

$$f(p\,|\,x = n, \mathcal{B}) = (n+1)\,p^n = (n_1 + n_2 + 1)\,p^{n_1+n_2}\,. \tag{12.28}$$

Si può procedere in un altro modo, calcolando la distribuzione successiva all'informazione $x_1 = n_1$

$$f(p\,|\,x_1 = n_1, \mathcal{B}) = (n_1 + 1)\,p^{n_1}\,, \tag{12.29}$$

e utilizzandola come prior per l'inferenza successiva:

$$\begin{aligned} f(p\,|\,x_1 = n_1, x_2 = n_2, \mathcal{B}) &= \frac{p^{n_2} f(p\,|\,x_1 = n_1, \mathcal{B})}{\int_0^1 p^{n_2} f(p\,|\,x_1 = n_1, \mathcal{B})\,\mathrm{dp}} & (12.30) \\ &= \frac{p^{n_2}(n_1+1)\,p^{n_1}}{\int_0^1 p^{n_2}(n_1+1)\,p^{n_1}\,\mathrm{dp}} & (12.31) \\ &= (n_1 + n_2 + 1)\,p^{n_1+n_2}\,, & (12.32) \end{aligned}$$

ottenendo lo stesso risultato.

12.3.4 ❄ Uso della prior coniugata Beta

Finora abbiamo basato l'inferenza su una prior uniforme che, come detto e come ripeteremo ancora, va più che bene per la maggior parte delle applicazioni 'tranquille'. Come abbiamo discusso nel caso dei problemi con verosimiglianza gaussiana, la difficoltà nell'introdurre delle prior che corrispondano all'effettivo stato di informazione risiede nelle complicazioni matematiche che ne possono scaturire. Un trucco per semplificare i conti è quello di modellizzare le prior con una funzione formalmente simile alla verosimiglianza (vista come funzione matematica del parametro che si vuole inferire). Si parla allora di *distribuzioni coniugate*. Abbiamo già utilizzato questo trucco per modellizzare la prior in modo gaussiano quando la verosimiglianza è anch'essa gaussiana.

Nel caso caso di verosimiglianza binomiale, distribuzione coniugata Beta (vedi paragrafo 8.14.1) è particolarmente conveniente in quanto, per opportune scelte dei suoi due soli parametri, tale distribuzione può assumere una interessante varità di forme (vedi figura 8.14). Infatti, a parte il fattore di normalizzazione abbiamo:

$$\begin{aligned} f(p\,|\,n,n,f_\circ = \text{Beta}(r,s)) &\propto p^x\,(1-p)^{n-x}\cdot p^{r-1}\,(1-p)^{s-1} \\ &\propto p^{x+r-1}\,(1-p)^{n-x+s-1}\,, \end{aligned}$$

ovvero si ottiene una nuova funzione Beta in cui r viene incrementato del numero dei successi e s del numero degli insuccessi:

$$\begin{aligned} p &\sim \text{Beta}(r+x, s+n-x) && (12.33)\\ \text{E}(p) &= \frac{r+x}{r+s+n} && (12.34)\\ \text{Var}(p) &= \frac{(r+x)\cdot(s+n-x)}{(r+s+1+n)\cdot(r+s+n)^2} && (12.35)\\ &= \text{E}(p)\cdot\frac{s+n-x}{(r+s+1+n)\cdot(r+s+n)}\,. && (12.36) \end{aligned}$$

Per $s = r = 1$ la beta si riduce ad una distribuzione uniforme, riottenendo le (12.12) e (12.14).

Ad esempio, se le prior sono tali da farci credere che p debba essere intorno a 1/2 con un'incertezza del 20%, ovvero $1/5$, dalle (8.38) e (8.39), abbiamo:

$$\left\{ \begin{array}{lcl} \frac{r}{r+s} &=& \frac{1}{2} \\ \frac{rs}{(r+s+1)\cdot(r+s)^2} &=& \left(\frac{1}{5}\frac{1}{2}\right)^2 \end{array} \right. \Longrightarrow r = s = 12\,. \qquad (12.37)$$

In altre parole, lo stato di iniziale informazione che ci fa ritenere $p \approx 0.5 \pm 0.1$ è lo stesso che si acquisirebbe da una dozzina di successi e una dozzina di insuccessi partendo da un precedente stato di informazione che ci faceva ritenere tutti i valori di p ugualmente possibili. Se ora facciamo due prove e vengono due successi, lo stato di conoscenza su p cambia, ma non tanto da farci ritenere che che p sia molto prossimo a 1, come si otterrebbe ongenuamente dalla frequenza relativa di successi. Si ottiene infatti

$$\begin{aligned} \text{E}(p) &= \frac{2+12}{2+24} = \frac{14}{26} = 0.538 && (12.38)\\ \sigma(p) &= \sqrt{\frac{14}{1521}} = 0.096 && (12.39) \end{aligned}$$

Come si vede, questa informazione empirica non ci fa cambiare sostanzialmente la nostra opinione sulla probabilità del prossimo successo, come è ragionevole che debba essere.

12.4 ⭮ Caso generale di inferenza con verosimiglianza poissoniana

Sia ora λ il parametro di interesse, avendo osservato $X = x$ conteggi in un intervallo di tempo T (o avendo osservato contato $X = x$ oggetti nel volume V, e così via). Il problema può essere anche impostato diversamente, inferendo direttamente l'intensità del processo di Poisson r, ovvero il tasso di conteggi, tale che $r \cdot T = \lambda$. Per ragioni didattiche preferiamo provedere un passo alla volta.

La verosimiglianza di osservare $X = x$ per ogni ipotesi λ è data dalla distribuzione di Poisson. Applicando il teorema di Bayes, abbiamo:

$$f(\lambda \,|\, x, \mathcal{P}) \;=\; \frac{\frac{\lambda^x\, e^{-\lambda}}{x!}\, f_\circ(\lambda)}{\int_0^\infty \frac{\lambda^x\, e^{-\lambda}}{x!}\, f_\circ(\lambda)\,\mathrm{d}\lambda}\,. \tag{12.40}$$

Per $f_\circ(\lambda)$ possiamo usare una distribuzione uniforme fino ad un valore massimo molto più grande di x, che per convenienza matematica possiamo estendere fino a infinito. Quindi $f_\circ(\lambda) = k$ a numeratore e denominatore si semplificano e, risolvendo per parti l'integrale di normalizzazione (come per il caso gaussiano, siamo "fortunati" e l'integrale vale 1), otteniamo

$$f(\lambda \,|\, x, \mathcal{P}) \;=\; \frac{\lambda^x\, e^{-\lambda}}{x!} \tag{12.41}$$

$$F(\lambda \,|\, x, \mathcal{P}) \;=\; 1 - e^{-\lambda}\left(\sum_{n=0}^{x} \frac{\lambda^n}{n!}\right). \tag{12.42}$$

Come si vede, questa volta anche la funzione cumulativa è relativamente semplice, ottenuta anch'essa mediante integrazione per parti della funzione densità di probabilità. Alcuni esempi sono mostrati in figura 12.2. Figure 12.2 shows some numerical examples. Si riconosce nella (12.41) una distribuzione Gamma di parametri $c = x + 1$ e $r = 1$ (vedi paragrafo). Ne segue che valore atteso, varianza e moda valgono

$$\mathrm{E}[\lambda] \;=\; x + 1\,, \tag{12.43}$$

$$\mathrm{Var}(\lambda) \;=\; x + 1\,, \tag{12.44}$$

$$\mathrm{Moda}(\lambda)\,[\equiv \lambda_m] \;=\; x\,. \tag{12.45}$$

12.4.1 Caso di routine

Quando x diventa abbastanza grande, otteniamo

$$\mathrm{E}[\lambda] \;\approx\; \lambda_m = x\,, \tag{12.46}$$

$$\mathrm{Var}(\lambda) \;\approx\; \lambda_m = x\,, \tag{12.47}$$

$$\sigma_\lambda \;\approx\; \sqrt{x}\,, \tag{12.48}$$

$$\lambda \;\sim\; \mathcal{N}(x, \sqrt{x})\,. \tag{12.49}$$

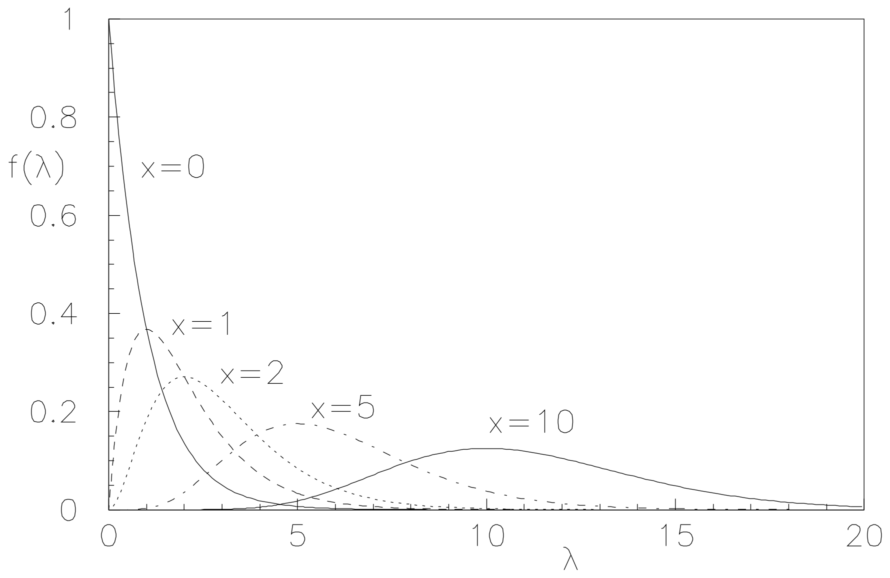

Figura 12.2: Examples of $f(\lambda\,|\,x_i)$.

12.4.2 Caso di $x = 0$ con prior uniforme

Anche in questo caso, vediamo il caso critico in cui non è stato misurato alcun conteggio, pur non essendoci alcuna ragione per cui si debba osservare esattamente zero e "si sarebbero potuti attendere anche numeri diversi da zero con buona probabilità" (questa espressione sarà meglio formalizzata nel seguito, vedi paragrafo ***).

$$f(\lambda\,|\,x=0,\mathcal{P}) = e^{-\lambda} \quad (12.50)$$
$$F(\lambda\,|\,x=0,\mathcal{P}) = 1-e^{-\lambda}\,. \quad (12.51)$$
$$(12.52)$$

La figura 12.3 mostra $f(\lambda)$. Il valore al quale si crede di più è senz'altro 0, ma la massa di probabilità si estende fino a diverse unità. In questo caso è spesso di interesse dare un *limite superiore* tale che la probabilità che λ ecceda il limite sia ritenuta piccola, per esempio il 10%, il 5% o l'1%. Scegliendo il 5% e risolvendo l'equazione $F(\lambda_\circ\,|\,x=0) = 0.95$, otteniamo

$$\lambda < \lambda_\circ = 3 \text{ al } 95\,\% \text{ di probabilità}\,. \quad (12.53)$$

Si noti come in questo caso vale anche $P(x=0\,|\,\lambda=3,\mathcal{P}) = 5\%$, ma questa è una pura coincidenza dovuta alle proprietà dell'esponenziale. Affermare che *"la probabilità che λ sia maggiore di 3, avendo osservato zero, vale il 5% in quanto la probabilità che una poissoniana di $\lambda = 3$ dia zero è del 5%"* è assolutamente scorretto. Questa affermazione è "innocua" quando viene applicata al problema in esame (il risultato è giusto per puro caso), ma è gravemente fuorviante quando viene applicata ad altri problemi.

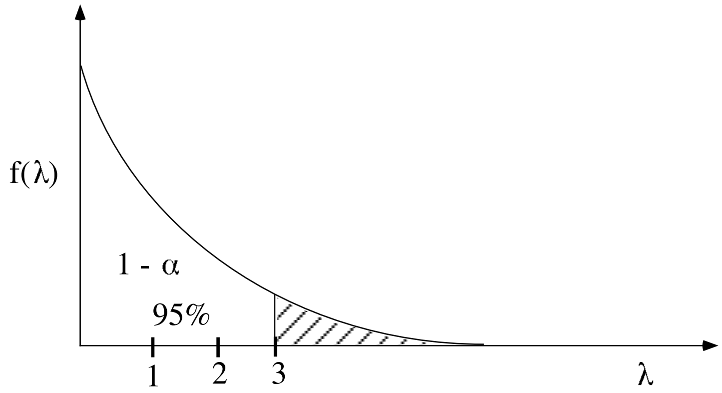

Figura 12.3: Upper limit to λ having observed 0 events.

12.4.3 Combinazione di risultati

Come visto più volte nei paragrafi precedenti, la combinazione di risultati si ottiene iterando l'apprendimento bayesiano. Se abbiamo n osservazioni x_i in intervalli di tempo di *uguale durata* (o in regioni spaziali uguali, se il problema è nel dominio dello spazio) e partiamo da una distribuzione di λ uniforme, abbiamo (a parte irrilevanti fattori moltiplicativi):

$$f(\mu\,|\,\mathbf{x}) \;\propto\; (\lambda^{x_1}\, e^{-\lambda})\cdot(\lambda^{x_n}\, e^{-\lambda})\,\cdots\,(\lambda^{x_n}\, e^{-\lambda}) = \lambda^{-\sum_i x_i}\, e^{-n\lambda} \tag{12.54}$$

$$\mu \;\sim\; \text{Gamma}\left(\sum_i x_i - 1, n\right), \tag{12.55}$$

avendo riconosciuto in $f(\mu)$ una distribuzione Gamma. Otteniamo quindi dalle (8.42)–(8.44)

$$\mathrm{E}(\lambda) = \frac{\sum_i x_i - 1}{n} \tag{12.56}$$

$$\sigma(\lambda) = \frac{\sqrt{\mathrm{E}(\lambda)}}{\sqrt{n}} \tag{12.57}$$

$$\text{Moda}(\lambda) = \frac{\sum_i x_i}{n}\,. \tag{12.58}$$

Come al solito, per $\sum_i x_i \gg 0$ abbiamo una distribuzione gaussiana

$$\mu \;\sim\; \mathcal{N}\left(\frac{\sum_i x_i}{n}, \frac{1}{\sqrt{n}}\sqrt{\frac{\sum_i x_i}{n}}\right). \tag{12.59}$$

12.4.4 ❄ Uso della prior coniugata Gamma

Come si è già visto dai paragrafi precedenti, la verosimiglianza vista come funzione matematica di λ ha la struttura di una distribuzione Gamma. Quindi questa è la prior coniugata della poissoniana. Quindi, a parte i soliti fattori moltiplicativi e considerando il caso generale di n osservazioni, abbiamo:

$$f(\mu \,|\, \mathbf{x}, f_\circ = \text{Gamma}) \propto \lambda^{\sum_i x_i} e^{-n\,\lambda} \cdot \lambda^{c_\circ - 1}\, e^{-r_\circ\,\lambda}$$

$$\propto \lambda^{c_\circ + \sum_i x_i - 1}\, , e^{-(r_\circ + n)\,\lambda} \qquad (12.60)$$

$$\lambda \sim \text{Gamma}\left(c_\circ + \sum_i x_i, r_\circ + n\right). \qquad (12.61)$$

Il caso di prior uniforme viene recuperato quando $c_\circ = 1$ e $r_\circ \to 0$.

Come esempio, vediamo come modellizzare in modo più realistico il caso in cui siano stati osservati 0 conteggi in un esperimento in cui "se attendevano pochi, non escludendo zero". Questo si verifica quando si cercano fenomeni rari. Un modo per esprimere questa vaghezza in termini della Gamma può essere di pensare ad un valore atteso di una decina con una incertezza relativa del 100%, per fissare le idee, diciamo $\mathrm{E}_\circ(\lambda) = \sigma_\circ(\lambda) = 10$, ovvero $c_\circ = 1$ e $r_\circ = 1/10$. Il risultato finale su λ sarà

$$\lambda \sim \text{Gamma}(1, 1.1) \longrightarrow\propto e^{-1.1\,\lambda} \qquad (12.62)$$

$$\mathrm{E}(\lambda) = 1/1.1 = 0.91 \qquad (12.63)$$

$$\sigma(\lambda) = 1/1.1 = 0.91 \qquad (12.64)$$

$$\lambda < 2.7 \ \text{ al } 95\% \text{ di probabilità.} \qquad (12.65)$$

Si noti che pur cambiando drasticamente la funzione a priori (una prior uniforme ha valore atteso "infinito", nel senso che è la metà del limite superiore, arbitrario) l'effetto sul risultato finale è modesto (valore atteso, incertezza di previsione e limite superiore al 95% cambiano di un 10%, che per questo tipo di ricerca di frontiera è irrilevante, quelli che contano sono gli ordini di grandezza).

12.4.5 Inferenza sull'intensità del processo di Poisson da osservazioni effettuate con diversi tempi di osservazione

Nei paragrafi precedenti ci siamo concentrati sull'inferenza di λ. Il passaggio all'intensità del processo di Poisson r è immediata in quanto r e λ sono legati fra di loro linearmente $\lambda = r\,T$, con T tempo di osservazione (non si confonda questo r con l'omonimo parametro della Gamma). A volte lo stesso fenomeno modellizzato come un processo di Poisson è osservato con diversi tempi di osservazione, ottenemdo x_i conteggi negli intervalli di tempo T_i. Avendo visto ancora una volta l'irrilevanza della esatta forma della prior nei problemi di interesse pratico e per evitare confusione con il parametro r della Gamma, usiamo una distribuzione iniziale uniforme. Ottemiamo, a meno dei soliti fattori moltiplicativi:

$$f(r \,|\, \mathbf{x}, \mathbf{T}) \propto \prod_i (r\,T_i)^{x_i} \cdot e^{-r\,T_i} \propto r^{\sum_i x_i} \cdot e^{-r\,\sum_i T_i} \qquad (12.66)$$

$$r \sim \text{Gamma}\left(\sum_i x_i + 1, \sum_i T_i\right), \qquad (12.67)$$

e quindi

$$\mathrm{E}(r) = \frac{x_{tot}+1}{T_{tot}} \xrightarrow[x_{tot}>>1]{} \frac{x_{tot}}{T_{tot}} \tag{12.68}$$

$$\sigma(r) = \frac{\sqrt{x_{tot}+1}}{T_{tot}} \xrightarrow[x_{tot}>>1]{} \frac{1}{\sqrt{T_{tot}}}\sqrt{\mathrm{E}(r)}\,, \tag{12.69}$$

ove con e T_{tot} e x_{tot} sono stati indicati, rispettivamente, il tempo totale di misura e il numero totale di eventi eventi osservati. Ai fini dell'inferenza di r, le tante misure sono equivalenti alla sola misura in T_{tot}.

Capitolo 10

1. E(X) $= 0$; Var(X) $= 1$; $X \sim \mathcal{N}(0, 1)$:
   $P(-1 \leq X \leq 1) = 68.3\,\%$.
2. $0.14\,\%$.
3. $5.8\,\%$.
4. $2.8\,\%$.
5. 0.377, 0.029 , $\ll 10^{-7}$; $P(X \geq 6000 \,|\, X \sim \mathcal{B}_{10000,0.5})$ è talmente piccola che non è calcolabile con l'uso delle usuali tabelle.
6. $52\,\%$.
7. $10.3\,\%$.
8. 25%.

Capitolo 11

1. e la soluzione è: ....

Capitolo 12

1. e la soluzione è: ....

Capitolo 13

1. e la soluzione è: ....