Biello Christian
Bonino Leonardo

DISTRIBUZIONI DI VARIABILI ALEATORIE DI INTERESSE FISICO

Corso di Esperimentazioni I
Proff. Bertaina Mario Edoardo, Bonino Raffaella
CdL in Fisica (A.A. 2017-2018)
Università degli Studi di Torino

Indice

Capitolo 1

Teoria delle distribuzioni derivate

Se si considera una variabile aleatoria, il principale obiettivo è calcolare la probabilità che tale variabile assuma un valore in un determinato range. In statistica si definisce una particolare funzione, detta densità di probabilità, la quale risulta estremamente utile quando si va a valutare l'area sottesa al grafico di essa in un determinato intervallo: l'integrale definito di tale funzione rappresenta la probabilità che la variabile aleatoria osservata possa assumere un valore compreso tra i due estremi di integrazione. Per trovare la distribuzione di una variabile aleatoria è quindi necessario determinare la funzione densità di probabilità. Quest'ultima per definizione è normalizzata a 1 e tale proprietà deriva dall'Assioma della Certezza:

$$\int_D f(x)dx = 1, \text{ dove } x \text{ è una variabile aleatoria e } D \text{ è il rispettivo dominio}$$

Di seguito si analizza un metodo per ricavare tale funzione a partire dalla distribuzione nota di una seconda variabile aleatoria.

È nota la distribuzione della variabile x ($f(x)$) e la relazione $y = \varphi(x)$ con φ funzione monotona derivabile. φ è invertibile, quindi si può determinare la funzione inversa:

$$x = \psi(y)$$

Si definisce la funzione integrale $F(x)$, la quale indica la probabilità di avere come valore della variabile uno minore di x:

$$F(x) = \int_{-\infty}^{x} f(x)\, dx$$

Analogamente si definisce la funzione integrale $G(y)$ che rappresenta la probabilità di osservare un valore inferiore a y della variabile con densità di probabilità $g(y)$ non nota:

$$G(y) = \int_{-\infty}^{y} g(y)\, dy$$

Si distinguono ora due casi:

- se φ e ψ sono funzioni **crescenti**, allora la probabilità di ottenere un valore minore di y deve coincidere con la probabilità di ottenere un valore minore di $x = \psi(y)$

  $$G(y) = F(x)$$

  Si determina la funzione $g(y)$

  $$\begin{aligned} g(y) &= \frac{dG(y)}{dy} = \frac{dF(x)}{dy} \\ &= \frac{dF(x)}{dx}\frac{dx}{dy} = f(x)\frac{dx}{dy} \\ &= f(\psi(y))\frac{d\psi(y)}{dy} \end{aligned} \tag{1.1}$$

- se φ e ψ sono funzioni **decrescenti**, allora la probabilità di ottenere un valore minore di y deve essere pari alla probabilità di ottenere un valore maggiore di $x = \psi(y)$, il quale corrisponde all'evento contrario rappresentato dalla funzione $F(x)$, quindi

  $$G(y) = 1 - F(x)$$

  Si determina la funzione $g(y)$

  $$\begin{aligned} g(y) &= \frac{dG(y)}{dy} = \frac{d(1 - F(x))}{dy} = -\frac{dF(x)}{dy} \\ &= -\frac{dF(x)}{dx}\frac{dx}{dy} = -f(x)\frac{dx}{dy} \\ &= -f(\psi(y))\frac{d\psi(y)}{dy} \end{aligned} \tag{1.2}$$

Nel prossimo capitolo si otterrà la distribuzione di Gauss a partire da uno studio sugli errori che possono influenzare la misura di una grandezza fisica. Ottenuta tale funzione si applicheranno le relazioni appena determinate (1.1, 1.2) per ricavare la densità di probabilità delle principali variabili aleatorie di interesse fisico.

Capitolo 2

Distribuzione di Gauss

2.1 Deduzione della funzione tramite Bernoulli

La distribuzione di Bernoulli è utile per valutare la probabilità che si verifichi un evento in un certo numero di tentativi, come nel caso del lancio ripetuto di monete o dadi. L'espressione si basa sul calcolo combinatorio, posta p la probabilità che si verifichi il singolo evento; nel caso in cui si voglia calcolare qual è la probabilità che si verifichi k volte l'evento singolo su un totale di lanci pari a n, l'ordine non conta; è quindi necessario calcolare il numero di combinazioni possibili C_k^n e moltiplicarla per la probabilità che si verifichi una precisa successione:

$$f(k,n) = C_k^n\, p^k(1-p)^{n-k} = \frac{n!}{k!(n-k)!}\, p^k(1-p)^{n-k} \tag{2.1}$$

Di seguito sono riportate proprietà utili della distribuzione di Bernoulli:

$$\begin{aligned} \bar{k} &= \sum_{k=0}^{n} k\, f(k,n) = np \\ \sigma^2 &= \bar{k^2} - \left(\bar{k}\right)^2 = np(1-p) \end{aligned}$$

Consideriamo ora una grandezza fisica con valor vero x^*; nella misurazione è possibile commettere errori in eccesso o in difetto di entità pari a ϵ, i quali rappresentano una combinazioni di errori degli sperimentatori e degli strumenti. La rappresentazione grafica 2.1 mostra in presenza di tre diversi errori i possibili valori osservati; da ciò si può intuire che lo studio dell'influenza degli errori è riconducibile alla distribuzione di Bernoulli in cui l'evento complessivo da studiare è il verificarsi di k errori in positivo. Ad ogni passaggio, si ipotizza di avere uguale probabilità di commettere un errore in eccesso o in difetto, dunque il singolo evento che rappresenta l'applicazione di un errore positivo ha probabilità $p = \frac{1}{2}$.

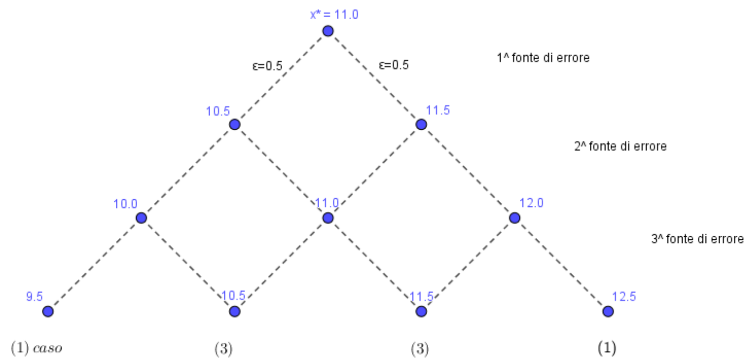

Figura 2.1: Effetti di fonti di errore di entità $\epsilon = 0.5$ sulla misurazione

Si studia il caso in cui vi siano infinite fonti di errore di entità infinitesime ($n \to +\infty$ e $\epsilon \to 0$); se si suppone di aver svolto k errori in positivo, allora $n-k$ saranno gli errori in negativo. La discrepanza tra il valore misurato e il valor vero è:

$$x_k - x^* = k\epsilon - (n-k)\epsilon = (2k-n)\epsilon$$

Si determina il valore atteso della discrepanza:

$$\begin{aligned}
E\left[x_k - x^*\right] &= \sum_k (x_k - x^*) \underbrace{P(k)}_{\text{probabilità}} \\
&= \sum_k (2k-n)\epsilon\, P(k) \\
&= \sum_k 2\epsilon k P(k) - \epsilon n \underbrace{\sum_k P(k)}_{\substack{=1 \text{ per l'Assioma} \\ \text{della Certezza}}} \\
&= 2\epsilon\ \bar{k}\ - \epsilon n \\
&= 2\epsilon n p - \epsilon n \\
&= \epsilon n - \epsilon n = 0
\end{aligned}$$

Si determina la varianza:

$$\begin{aligned}
\sigma^2 = E\left[(x_k - x^*)^2\right] &= \sum_k (x_k - x^*)^2 P(k) \\
&= \sum_k [(2k-n)\epsilon]^2 P(k) = \sum_k (4k^2 + n^2 - 4kn)\epsilon^2 P(k) \\
&= 4\epsilon^2 \sum_k \left[k^2 P(k)\right] + n^2\epsilon^2 \sum_k \left[P(k)\right] - 4n\epsilon^2 \sum_k \left[kP(k)\right] \\
&= 4\epsilon^2\ \overline{k^2}\ + n^2\epsilon^2 - 4n\epsilon^2\ \bar{k}
\end{aligned}$$

$$
\begin{aligned}
&= 4\epsilon^2(\sigma_k^2 + \bar{k}^2) + n^2\epsilon^2 - 4n\epsilon^2\,\bar{k} \\
&= 4\epsilon^2[np(1-p) + (np)^2] + n^2\epsilon^2 - 4n\epsilon^2 np \\
&= 4\epsilon^2\left[n\frac{1}{2}\frac{1}{2} + n^2\frac{1}{4}\right] + n^2\epsilon^2 - 4n^2\epsilon^2\frac{1}{2} \\
&= n\epsilon^2 + n^2\epsilon^2 + n^2\epsilon^2 - 2n^2\epsilon^2 = n\epsilon^2
\end{aligned}
$$

La deviazione standard è quindi:

$$\sigma = \sqrt{n}\,\epsilon$$

Anche se $n \to +\infty$ e $\epsilon \to 0$, il prodotto $\sqrt{n}\,\epsilon$ dovrà esssere finito e non nullo. Per comodità di seguito vengono definite le variabili ausiliarie r e x.

$$
\begin{aligned}
&x_k - x^* = (2k-n)\epsilon = 2r\epsilon \equiv x \\
&2r = 2k - n
\end{aligned}
$$

$$
\begin{cases} k = \frac{n}{2} + r \\ n - k = \frac{n}{2} - r \end{cases}
$$

$$k \in [0, n] \Rightarrow r \in \left[-\frac{n}{2}, \frac{n}{2}\right]$$

Si parte dalla distribuzione di Bernoulli (2.1) per $p = \frac{1}{2}$

$$
\begin{aligned}
f(k) &= \frac{n!}{k!(n-k)!}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{n-k} = \frac{n!}{k!(n-k)!}\left(\frac{1}{2}\right)^n \\
f(r) &= \frac{n!}{(\frac{n}{2}+r)!\,(\frac{n}{2}-r)!}\,\frac{1}{2^n} \equiv f(x)
\end{aligned}
$$

Si valuta la differenza tra due risultati successivi:

$$\Delta x = 2(r+1)\epsilon - 2r\epsilon = 2\epsilon$$

La funzione $f(r+1)$ vale

$$f(r+1) = \frac{n!}{(\frac{n}{2}+r+1)!\,(\frac{n}{2}-r-1)!}\,\frac{1}{2^n} \equiv f(x+2\epsilon)$$

Si determina il rapporto

$$
\begin{aligned}
\frac{f(x+2\epsilon)}{f(x)} &= \frac{f(r+1)}{f(r)} = \frac{(\frac{n}{2}+r)!\,(\frac{n}{2}-r)!}{(\frac{n}{2}+r+1)!\,(\frac{n}{2}-r-1)!} \\
&= \frac{(\frac{n}{2}+r)!\,(\frac{n}{2}-r)\,(\frac{n}{2}-r-1)!}{(\frac{n}{2}+r+1)\,(\frac{n}{2}+r)!\,(\frac{n}{2}-r-1)!} \\
&= \frac{\frac{n}{2}-r}{\frac{n}{2}+r+1} = \frac{n-2r}{n+2r+2} \qquad (2.2)
\end{aligned}
$$

Si applica lo sviluppo di Taylor arrestato al primo ordine:

$$\frac{f(x+2\epsilon)}{f(x)} = \frac{f(x)+2\epsilon f'(x)}{f(x)} = 1+2\epsilon\frac{f'(x)}{f(x)} \tag{2.3}$$

Combiando le relazioni 2.2 e 2.3 si ha:

$$\begin{aligned}
1+2\epsilon\frac{f'(x)}{f(x)} &= \frac{n-2r}{n+2r+2} \\
1+2\epsilon\frac{f'(x)}{f(x)} &= \frac{n-2\frac{x}{2\epsilon}}{n+2\frac{x}{2\epsilon}+2} = \frac{n-\frac{x}{\epsilon}}{n+\frac{x}{\epsilon}+2} \\
1+2\frac{\sigma}{\sqrt{n}}\frac{f'(x)}{f(x)} &= \frac{n-\sqrt{n}\frac{x}{\sigma}}{n+\sqrt{n}\frac{x}{\sigma}+2} \\
1+2\frac{\sigma}{\sqrt{n}}\frac{f'(x)}{f(x)} &= \frac{\sigma n\left(1-\frac{x}{\sqrt{n}\sigma}\right)}{\sigma n\left(1+\frac{x}{\sqrt{n}\sigma}+\frac{2}{n}\right)} \\
2\frac{\sigma}{\sqrt{n}}\frac{f'(x)}{f(x)} &= \frac{1-\frac{x}{\sqrt{n}\sigma}}{1+\frac{x}{\sqrt{n}\sigma}+\frac{2}{n}} - 1 \\
2\frac{\sigma}{\sqrt{n}}\frac{f'(x)}{f(x)} &= 2\frac{-\frac{x}{\sigma\sqrt{n}}-\frac{1}{n}}{1+\frac{x}{\sigma\sqrt{n}}+\frac{2}{n}} \\
\frac{f'(x)}{f(x)} &= \frac{\sqrt{n}}{\sigma}\frac{-\frac{x}{\sigma\sqrt{n}}-\frac{1}{n}}{1+\frac{x}{\sigma\sqrt{n}}+\frac{2}{n}} \\
\frac{f'(x)}{f(x)} &= \frac{-\frac{x}{\sigma^2}-\frac{1}{\sigma\sqrt{n}}}{1+\frac{x}{\sigma\sqrt{n}}+\frac{2}{n}}
\end{aligned}$$

Si applica il limite per $n \to +\infty$.

$$\lim_{n\to+\infty}\frac{f'(x)}{f(x)} = \lim_{n\to+\infty}\frac{-\frac{x}{\sigma^2}-\overbrace{\frac{1}{\sigma\sqrt{n}}}^{\to 0}}{1+\underbrace{\frac{x}{\sigma\sqrt{n}}}_{\to 0}+\underbrace{\frac{2}{n}}_{\to 0}} = -\frac{x}{\sigma^2}$$

Entrambi i membri vengono integrati in modo indefinito per risolvere l'equazione differenziale.

$$\begin{aligned}
&\int \frac{f'(x)}{f(x)}dx = -\int \frac{x}{\sigma^2}dx \\
&\ln f(x) = -\frac{x^2}{2\sigma^2} + \ln A \\
&f(x) = Ae^{-\frac{x^2}{2\sigma^2}}
\end{aligned}$$

Per calcolare la costante A si impone la condizione di normalizzazione.

$$\int_{-\infty}^{+\infty} f(x)dx = 1 = A\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma^2}}dx$$

Per risolvere l'integrale si applica la sostituzione $y = \frac{x}{\sqrt{2}\sigma}$:

$$\begin{aligned}
1 &= A\int_{-\infty}^{+\infty} e^{-y^2}\sqrt{2}\sigma dy \\
1 &= A\sigma\sqrt{2} \underbrace{\int_{-\infty}^{+\infty} e^{-y^2}dy}_{\text{integrale di Gauss } (\sqrt{\pi})} = A\sigma\sqrt{2\pi} \\
&\Rightarrow A = \frac{1}{\sqrt{2\pi}\sigma}
\end{aligned}$$

È quindi definita la funzione

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}$$

La variabile x rappresenta la discrepanza tra la misura (che si indicherà con t) e il valor vero della grandezza (rappresentato da μ). Si può quindi esprimere la funzione distribuzione di probabilità della grandezza fisica dipendente da due parametri (μ e σ):

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(t-\mu)^2}{2\sigma^2}} \tag{2.4}$$

2.2 Caratteristiche della distribuzione

Nota l'espressione della distribuzione gaussiana, è possibile valutare alcune caratteristiche di particolare interesse della funzione. Secondo un semplice studio di funzione ne si può comprendere l'andamento.
La funzione di tipo esponenziale è sempre positiva e mai nulla. Si valutano i limiti agli estremi:

$$\begin{aligned}
\lim_{t\to-\infty} f(t) &= \lim_{t\to-\infty} \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma}e^{-\infty} = 0 \\
\lim_{t\to+\infty} f(t) &= \lim_{t\to+\infty} \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma}e^{-\infty} = 0
\end{aligned}$$

La funzione avrà un massimo valutabile ponendo:

$$\begin{aligned}
\frac{d}{dt}\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}}\right) &= 0 \\
-\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}}\frac{(t-\mu)}{\sigma^2} = 0 \quad \rightarrow \quad t-\mu &= 0 \quad \rightarrow \quad t = \mu
\end{aligned}$$

Solitamente la misura di una grandezza fisica è un valore positivo, per cui si può intendere l'andamento della funzione come una funzione a campana, simmetrica rispetto al valore medio μ in corrispondenza del quale si ha un massimo. È utile anche calcolare i flessi, ricercando i punti in cui si annulla la derivata seconda:

$$\frac{d^2}{dt^2}\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}}\right) = \frac{d}{dt}\left(-\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}}\frac{(t-\mu)}{\sigma^2}\right) = 0$$

$$-\frac{1}{\sqrt{2\pi}\sigma}\left(-e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}}\frac{(t-\mu)^2}{\sigma^4} + \frac{e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}}}{\sigma^2}\right) = 0$$

$$e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}}\frac{(t-\mu)^2}{\sigma^4} = \frac{e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}}}{\sigma^2}$$

$$\frac{(t-\mu)^2}{\sigma^2} = 1 \quad \rightarrow \quad \left(\frac{t-\mu}{\sigma}\right)^2 = 1 \quad \rightarrow \quad \frac{t-\mu}{\sigma} = \pm 1 \quad \rightarrow \quad t_{1,2} = \mu \pm \sigma$$

La funzione presenta due flessi, che si trovano a distanza $\pm\sigma$ da μ.
Disegnando la funzione mediante *GeoGebra* si confermano le caratteristiche mostrate analiticamente.

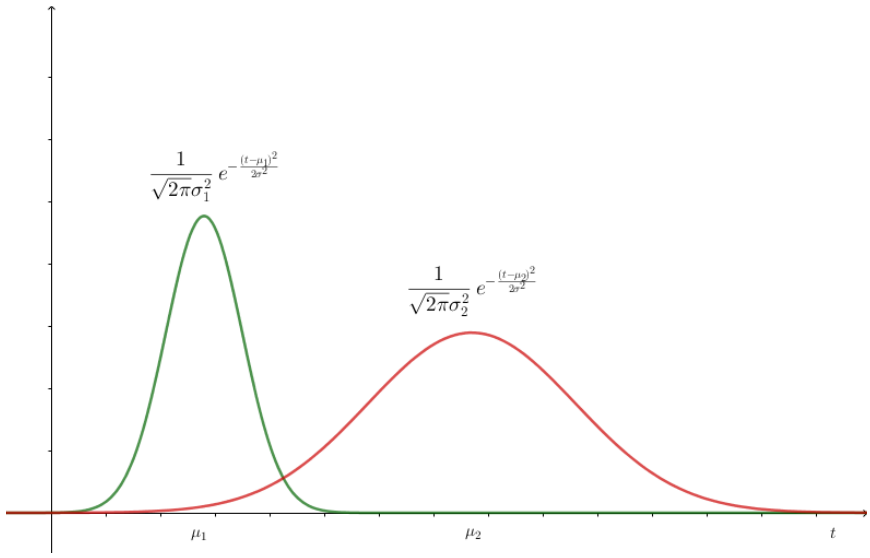

Figura 2.2: Sono rappresentate due distribuzioni gaussiane rispettivamente con valor veri μ_1 e μ_2, tali per cui $\mu_1 < \mu_2$, e deviazione standard σ_1 e σ_2 con $\sigma_1 < \sigma_2$.

Nella fisica sperimentale è importante andare a definire dei parametri secondo i quali è possibile accettare o scartare alcune misurazioni. Nel caso di misure ripetute di una grandezza, dopo aver raccolto una consistente serie di dati, è possibile valutare se alcune misurazioni siano affette da grossolani

errori che le discostano quindi signficativamente dal valor vero.
Non bisogna però ritenere le misure che si decide di scartare dalla popolazione come errate. I valori scartati tipicamente sono valori che si discostano signficativamente dal valor medio (che rappresenta la miglior stima della grandezza) e quindi vi è una *bassa* probabilità che rappresentino effettimente il valore della grandezza misurata.
La valutazione la si fa quindi tenendo conto della probabilità di ottenere quel valore, effettuando una successiva misurazione della stessa grandezza.
La funzione $f(t)$ rappresenta una probabilità puntuale rispetto ad una misurazione di ottenere un certo valore; la probabilità infinitesima di misurare una grandezza sarà invece ottenuta valutando la probabilità puntuale rispetto ad un intervallo infinitesimo all'interno del quale oscilla la misura (una misurazione fisica è sempre fornita con un intevallo di significatività):

$$dP(t) = f(t)dt \qquad \text{da cui si ricava} \qquad P(t) = \int_I f(t)dt \tag{2.5}$$

La probabilità è rappresentata quindi dall'area sottesa alla curva gaussiana.
Valutando alcuni valori di probabilità, si può stabilire un criterio per poter scartare misurazioni con una certa sicurezza che non siano rappresentative, infatti:

$$\begin{aligned}
\int_{-\sigma}^{+\sigma} f(t)dt &= 0,683 \to 68,3\% \\
\int_{-2\sigma}^{+2\sigma} f(t)dt &= 0,954 \to 95,4\% \\
\int_{-3\sigma}^{+3\sigma} f(t)dt &= 0,997 \to 99,7\%
\end{aligned} \tag{2.6}$$

Dall'integrale 2.6 si può ricavare un criterio per escludere alcune misurazioni dalla serie di dati. Se si considera infatti un intevallo compreso tra μ e $\pm\ 3\sigma$, i valori che stanno al di fuori hanno una probabilità molto bassa di essere rimisurati. Nel caso in cui si siano raccolti 100 dati relativi a misure ripetute della stessa grandezza fisica, risulta estremamente improbabile l'aver ottenuto un dato distante più di 3σ dalla miglior stima del valor vero, cioè il valor medio delle misure. Un dato fuori dal 3σ è rappresentato da una probabilità dello $0,3\%$, la quale indicherebbe che solo $0,3$ misure su 100 avrebbero dovuto distanziarsi cosiì tanto dal valor medio, perciò è possibile rigettare i dati al di fuori del range $\mu \pm 3\sigma$ in quanto probabilmente durante la loro raccolta sono stati commessi errori sistematici.
Questo criterio, detto **criterio del** 3σ, è applicabile solo tenendo conto del numero complessivo di dati raccolti: se le misure sono dell'ordine del centinaio allora statisticamente è decisamente improbabile che una di queste possa cadere al di fuori del 3σ, mentre per un numero di dati superiori è probabile; per esempio su 1000 misure la distribuzione di Gauss prevede che 3 di queste non rientrino nel range $\mu \pm 3\sigma$.

Capitolo 3

Distribuzione della variabile Z

3.1 Definizione della variabile Z

La gaussiana presenta alcune difficoltà pratiche di utilizzo, dato che la funzione cambia in base al variare dei parametri (μ e σ). Le grandezze fisiche che si misurano possono assumere dei valori anche molto ridotti, rendendo difficoltoso l'utilizzo di valori medi e deviazioni standard nella distribuzione densità di probabilità. Risulta quindi utile definire una nuova variabile, detta *Z di Gauss* o *Variabile Standardizzata*, deifinita come:

$$Z = \frac{x - \mu}{\sigma}$$

Si nota come una variabile di questo tipo possa semplificare la trattazione di una variabile nell'ottica della valutazione dell'accettazione di un valor medio ottenuto sperimentalmente in relazione al valore teorico che ci si aspetta. Il valore sperimentale x, risulterà compatibile con il valore teorico μ se la loro differenza risulta essere dell'ordine di grandezza dell'incertezza σ. In tal caso sarà quindi $|Z| \leq 1$.
Poiché Z è una grandezza derivata da una variabile x aleatoria distribuita normalmente, anche alla Z è possibile attribuire una variabilità maggiore, tenendo conto delle fluttazioni statistiche della variabile x che la compone. È possibile quindi valutare la distrubuzione di probabilità della variabile Z, a partire dalla distribuzione di probabilità della variabile aleatoria x, mediante il medoto delle Distribuzioni Derivate.

3.2 Deduzione della distribuzione

A partire quindi dalla Distribuzione di Gauss (2.4) si può comporre la $g(Z)$. Noto che

$$\begin{aligned} Z &= \varphi(x) \\ x &= \psi(Z) = \mu + \sigma Z \end{aligned} \tag{3.1}$$

sono due funzioni strettamente monotone e crescenti, per 1.1

$$\begin{aligned} g(Z) &= f(\psi(Z))\frac{d(\psi(Z))}{dZ} \\ &= \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{\mu+\sigma Z-\mu}{\sigma}\right)^2}\sigma \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{Z^2}{2}} \end{aligned} \tag{3.2}$$

che rappresenta la distribuzione di probabilità della variabile Z standardizzata. La distribuzione ottenuta seguirà l'andamento di una curva a campana poichè combinazione lineare di una variabile gaussiana. La distribuzione di probabilità della variabile ottenuta fornisce una curva che descrive la distanza tra valore sperimentale e teorico in relazione all'incertezza.

3.3 Caratteristiche della distribuzione

È possibile valutare le caratteristiche della funzione a partire dalla normalizzazione, condizione necessaria per una distribuzione di probabilità, verificando che la sommatoria delle probabilità infinitesime ai limiti del continuo dia 1:

$$\int_{-\infty}^{+\infty} g(Z)dZ = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{\frac{Z^2}{2}}$$

Effettuando un cambio di variabile e ponendo

$$\frac{Z^2}{2} = y^2 \rightarrow Z = \sqrt{2}y \rightarrow dZ = \sqrt{2}dy$$

dalla sostituzione si ottiene

$$\frac{1}{\sqrt{2\pi}}\sqrt{2}\int_{-\infty}^{+\infty} e^{-y^2}dy = \frac{\sqrt{\pi}}{\sqrt{\pi}} = 1 \tag{3.3}$$

noto essere l'integrale di Gauss pari a

$$\int_{-\infty}^{+\infty} e^{-y^2}dy = \sqrt{\pi}$$

Essendo μ il valore che fornisce la densità di probabilità massima sulla $f(x)$, ci si aspetta che $g(Z)$, poiché

$$Z = \frac{x-\mu}{\sigma}$$

sia centrata sullo zero. Infatti, essendo le due densità di probabilità legate linearmente, dove una è massima lo dovrà essere anche l'altra. Per cui se la gaussiana è massima in corrispondenza di $f(\mu)$ ci si aspetta che la distribuzione standardizzata sia massima per $Z = 0$. Per conferma si valuta a partire dalla funzione densità di probabilità:

$$\int_{-\infty}^{+\infty} Zg(Z)dZ$$

È necessario, prima di procedere al calcolo, valutare l'integrale $\int_{-\infty}^{+\infty} y^m e^{-y^2} dy$ rispetto a due casi (m intero)

- **m dispari**, la funzione f è dispari e vale che: $f(-y) = -f(y)$

$$\int_{-\infty}^{+\infty} y^m e^{-y^2} dy = 0$$

- **m pari,** $m = 2n$, si avrà invece

$$\int_{-\infty}^{+\infty} y^m e^{-y^2} dy = \int_{-\infty}^{+\infty} y^{2n} e^{-y^2} dy = \int_{-\infty}^{+\infty} y^{2n-1} y e^{-y^2} dy$$

  noto che

$$ye^{-y^2} = \frac{d\left(-\frac{1}{2}e^{-y^2}\right)}{dy} \rightarrow \int ye^{-y^2} dy = -\frac{1}{2}e^{-y^2}$$

  e integrando per parti

$$\begin{aligned} \int_{-\infty}^{\infty} y^{2n-1} y e^{-y^2} dy &= \left[y^{2n-1}\left(-\frac{1}{2}\right)e^{-y^2}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{2}(2n-1)y^{2n-2}e^{-y^2} dy \\ &= \frac{2n-1}{2}\int_{-\infty}^{\infty} y^{2n-2}e^{-y^2} dy \end{aligned}$$

Si può quindi procedere al calcolo del *valor medio* e della *varianza* mediante gli integrali appena calcolati

- **Valor medio**

$$E[Z] = \int_{-\infty}^{\infty} \frac{Z}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} dZ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} Z e^{-\frac{Z^2}{2}} dZ = 0$$

  poiché si tratta di un integrale simmetrico di funzione dispari.

- **Varianza**

$$\begin{aligned} \sigma^2 &= E[(Z-\mu)^2] = E[Z^2] - E[Z]^2 = E[Z^2] \\ &= \int_{-\infty}^{\infty} \frac{Z^2}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} dZ \end{aligned} \tag{3.4}$$

ed effettuado il cambio di variabile

$$Z^2 = 2y^2 \rightarrow z = \sqrt{2}y \rightarrow dZ = \sqrt{2}dy$$

si ottiene, noto il valore dell'integrale pari (m=2, n=1)

$$\begin{aligned}\sigma^2 &= \frac{1}{\sqrt{2\pi}}\sqrt{2}\int_{-\infty}^{\infty} 2y^2 e^{-y^2} dy = \frac{1}{\sqrt{2\pi}}\sqrt{2}\frac{2n-1}{2}2\int_{-\infty}^{\infty} y^0 e^{-y^2} \\ &= \frac{1}{\sqrt{\pi}}\sqrt{\pi} = 1 \qquad (3.5)\end{aligned}$$

Come ci si poteva aspettare, la distribuzione della variabile standardizzata è centrata sullo zero, dove presenta un massimo.

3.4 Test di Gauss o Test Zeta

In generale si pensa di dover considerare rappresentativa una misura effettuata se la discrepanza tra valore atteso e valore sperimentale è dell'ordine della deviazione standard. Per come è definita la variabile Z (3.1), ci si aspetta quindi che essa debba essere compresa tra -1 e 1 se si tratta di una misura coerente con il valore atteso. Essendo però la variabile x una variabile aleatoria, soggetta a fluttuazioni statistiche casuali, è necessario considerare un range più ampio di accettabilità di un risultato nel test Z.
È possibile infatti commettere una serie di errori durante la misurazione (sia da strumenti che da operatori) tutti in "positivo" o in "negativo". La misura ottenuta si discosterà così più di una deviazione standard dal valore teorico, essendo comunque una stima dello stesso.
Di fatto, quando si effettua un Test di Gauss, si sviluppa un'**ipotesi nulla** (H_0) . Questa consiste nell'assumere che il valor medio (ricavato sperimentalmente) concordi con il valore atteso e che la loro discrepanza sia unicamente dovuta ad una $fluttuazione\ statistica$. È necessario quindi determinare un criterio, secondo il quale è possibile accettare o rigettare H_0.
Per far ciò si considera un intervallo detto $intervallo\ di\ fiducia$, che definisce un range all'interno del quale un valore di probabilità garantisce di poter considerare la misurazione sufficientemente descrittiva del valor vero. I valori di riferimento usati in laboratorio sono principalmente due e si ottengono mediante un procedimento di integrazione della distribuzione $g(Z)$, tenendo sempre conto che l'area sottesa al grafico ha un significato di probabilità. Si definisce z_x il valore tale per cui:

$$\int_0^{z_x} g(Z)dZ = x$$

Si ricorda che per la proprietà di simmetria della distribuzione standardizzata si ha che:

$$\int_0^{z_x} g(Z)dZ = \int_{-z_x}^{0} g(Z)dZ$$

Rispetto a questa definizione si stabiliscono due livelli di significatività: al 5% e all'1%.

- Per il livello di significatività del 5%, a cui corrisponde un livello di confidenza del 95% si può valutare che:

$$\begin{aligned} \int_{-z_{0.475}}^{z_{0.475}} g(Z)dZ &= \int_{-z_{0.475}}^{0} g(Z)dZ + \int_{0}^{z_{0.475}} g(Z)dZ \\ &= 0.475 + 0.475 = 0.95 = 95\% \end{aligned} \tag{3.6}$$

  da cui si determina il valore di $|z_{0.475}| = 1,96$.
  Se vale che:

$$-z_{0.475} < Z_{spe} < z_{0.475} \quad \rightarrow \quad -1,96 < Z_{spe} < 1,96 \quad \rightarrow \quad |Z_{spe}| < 1,96$$

  allora vi sarà una probabilità del 95% che il valore sperimentale x, che fornisce uno z_{spe} compreso in questo intorno, sia compatibile con il valor vero, tenendo conto dell'incertezza. Nel caso in cui Z_{spe} non appartenesse al range stabilito, allora l'evento osservato risulterebbe poco probabile se davvero la misura fosse compatibile con il valor vero μ, infatti vi era una probabilità pari a 5% di ottenere un valore di Z in valore assoluto così elevato.

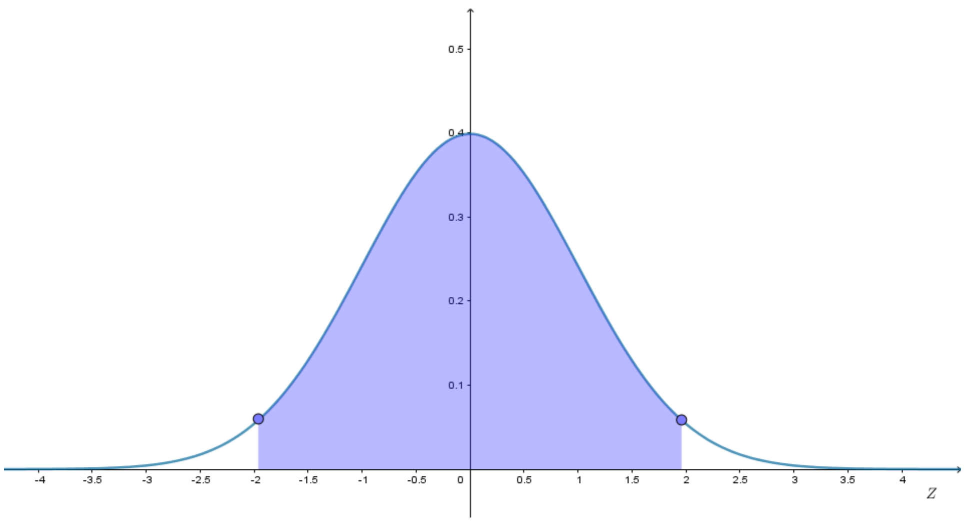

Figura 3.1: Distribuzione della variabile Z: in azzurro è rappresentata la regione sottesa alla curva con area pari al 95%

- In egual modo, per il livello di significatività dell'1%, a cui corrisponde un livello di confidenza del 99% si può valutare che:

$$\int_{-z_{0.495}}^{z_{0.495}} g(Z)dZ = \int_{-z_{0.495}}^{0} g(Z)dZ + \int_{0}^{z_{0.495}} g(Z)dZ$$

$$= \quad 0.495 + 0.495 = 0.99 = 99\%$$

da cui si determina il valore di $|z_{0.495}| = 2,58$. Se vale che:

$$-z_{0.995} < Z_{spe} < z_{0.995} \quad \rightarrow \quad -2,58 < Z_{spe} < 2,58 \quad \rightarrow \quad |Z_{spe}| < 2,58$$

allora l'ipotesi nulla di compatibilità verrà accetta in quanto vi era una probabilità del 99% che il valore sperimentale x, che ha fornito uno z_{spe} compreso in questo intorno, sia compatibile con il valor vero, tenendo conto dell'incertezza.

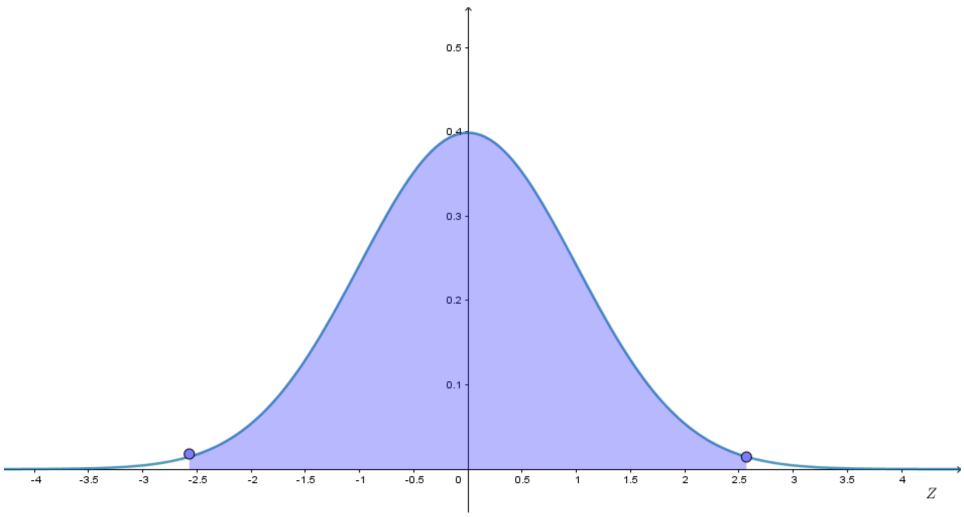

Figura 3.2: Distribuzione della variabile Z: in azzurro è rappresentata la regione sottesa alla curva con area pari al 99%

I valori che, in base alla scelta del test, forniscono Z_{spe} nelle code devono essere scartati poiché la probabilità che siano rappresentativi è molto bassa (più probabilmente si trovano nelle code a causa di errori durante la presa dati).
Se si effettua un test Z rispetto, per esempio, ad un valor medio sperimentale, in caso di fallimento, diventa necessario rifiutare l'ipotesi H_0: il valore sperimentale è in disaccordo con il valore teorico, tenendo anche conto delle fluttuazioni statistiche del caso. A seguito di questo risultato si possono trarre le dovute conclusioni riguardo alla procedura sperimentale.
Al contrario, se il test è verificato, si dice che il valor sperimentale è compatibile con il valor vero con una significatività del 5% o dell'1%.
Di norma si fissa il livello di significatività al 5% in quanto l'1% risulta essere troppo concessivo. All'1% vengono tipicamente scartati pochissimi dati e di conseguenza si ritengono accettabili valori anche significativamente distanti dal valore teorico; accettando molti più valori si rischia di non scartare le misure meno accurate.

Capitolo 4

Distribuzione della variabile χ^2

4.1 Definizione della variabile χ^2

La variabile aleatoria χ^2 rappresenta un parametro in grado di comparare i risultati ottenuti sperimentalmente con la rispettiva previsione teorica. Se si ha a disposizione un campione di dati $\{y_{1s}, y_{2s}, ..., y_{ns}\}$ e i corrisponenti valori attesi sono $\{y_{1t}, y_{2t}, ..., y_{nt}\}$, allora si può definire la variabile χ^2:

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_{is} - y_{it})^2}{\sigma_i^2} \tag{4.1}$$

Il parametro n è detto numero di gradi di libertà. Nel caso in cui per la determinazione dei valori attesi si faccia riferimento ai dati sperimentali, i gradi di libertà diminuiranno in base al numero di parametri che sono stati stimati.
Il χ^2 può essere definito anche per verificare se una data grandezza fisica segua una precisa distribuzione (per esempio quella di Poisson o di Gauss). In tal caso le misure ripetute vengono suddivise in gruppi, ciascuno rappresentante un piccolo range tale da permettere di associare come varianza il valore atteso, infatti il singolo gruppo segue con buona approssimazione la distribuzione di Poisson.

$$\sigma_i^2 = e_i$$

Se nell'i-esimo gruppo, il valore sperimentale è o_i e il valore previsto dalla distribuzione ipotizzata è e_i, allora il χ^2 può essere calcolato come segue:

$$\chi^2 = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i} \tag{4.2}$$

A partire dall'espressione 4.1, si ricava una definzione alternativa in cui si riconosce la variabile normale standardizzata Z:

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{y_{is} - y_{it}}{\sigma_i} \right)^2 = \sum_{i=1}^{n} Z_i^2 \tag{4.3}$$

4.2 Funzione Gamma

Per dedurre la distribuzione del χ^2 è necessario definire una particolare funzione di rilevante importanza in matematica, chiamata $funzioneGamma$. Essa è definita come segue:

$$\Gamma(m+1) = \int_0^{+\infty} x^m e^{-x} dx \tag{4.4}$$

Durante questo studio m assumerà sempre un valore intero o semi-intero. La funzione Gamma gode di una importante proprietà:

$$\Gamma(m+1) = m\Gamma(m) \tag{4.5}$$

Tale proprietà può essere dimostrata applicando l'integrazione per parti:

$$\begin{aligned}
\Gamma(m+1) &= \int_0^{+\infty} x^m e^{-x} dx \\
&= -\int_0^{+\infty} x^m \left(-e^{-x}\right) dx \\
&= -\left\{ \underbrace{\left[x^m e^{-x}\right]_0^{+\infty}}_{\to 0} - \int_0^{+\infty} m\, x^{m-1} e^{-x} dx \right\} \\
&= \int_0^{+\infty} m\, x^{m-1} e^{-x} = m \int_0^{+\infty} x^{m-1} e^{-x} = m\,\Gamma(m)
\end{aligned}$$

Si valuta ora $\Gamma(1)$:

$$\Gamma(1) = \int_0^{+\infty} x^0 e^{-x} dx = \int_0^{+\infty} e^{-x} dx = \left[-e^{-x}\right]_0^{+\infty} = 1 \tag{4.6}$$

Dalle relazioni 4.5 e 4.6 si conclude che la funzione Gamma con argomento intero naturale rappresenta un fattoriale, e in particolare:

$$\Gamma(m+1) = m!\,, \ \text{ se } m \in \mathbf{N} \tag{4.7}$$

Per la valutazione della funzioni Gamma con argommenti semi-interi si ricorre all'utilizzo della proprietà 4.5 fino a esplicitare $\Gamma\left(\frac{1}{2}\right)$ il cui valore si può ricavare noto l'integrale di Gauss:

$$\begin{aligned}
\Gamma\left(\frac{1}{2}\right) &= \int_0^{+\infty} x^{-1/2} e^{-x} dx \qquad\qquad \text{Posto } x = t^2, dx = 2tdt \\
&= \int_0^{+\infty} t^{-1} e^{-t^2} 2tdt \\
&= 2\int_0^{+\infty} e^{-t^2} dt = 2\frac{\sqrt{\pi}}{2} = \sqrt{\pi} \qquad (4.8)
\end{aligned}$$

A titolo di esempio, si calcolano i valori della funzione Gamma per due argomenti semi-interi:

$$
\begin{aligned}
\Gamma\left(\frac{3}{2}\right) &= \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2} \\
\Gamma\left(\frac{5}{2}\right) &= \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3}{2}\left[\frac{1}{2}\Gamma\left(\frac{1}{2}\right)\right] = \frac{3\sqrt{\pi}}{4}
\end{aligned}
$$

4.3 Deduzione per 1 g.d.l. e generalizzazione

Si considera inizialmente il caso con un unico grado di libertà. La variabile χ^2 sarà pari al quadrato della variabile normale standardizzata: $\chi^2 = Z^2$. Si espicita la variabile con distribuzione nota (Z) in funzione del chi quadro:

$$Z = \pm\sqrt{\chi^2}$$

La funzione trovata non è iniettiva ma può essere suddivisa in due tratti:

- Z=$\psi(\chi^2)$=$\sqrt{\chi^2}$,
  il tratto è monotono crescente quindi, secondo la relazione 1.1, la distribuzione derivata sarà pari a:

  $$g(\chi^2) = f\left(\psi(\chi^2)\right)\frac{dZ}{d\chi^2} = f\left(\psi(\chi^2)\right)\frac{1}{2\sqrt{\chi^2}}$$

- Z=$\psi(\chi^2)$=$-\sqrt{\chi^2}$,
  il tratto è monotono decrescente quindi, secondo la relazione 1.2, la distribuzione derivata sarà pari a:

  $$g(\chi^2) = -f\left(\psi(\chi^2)\right)\frac{dZ}{d\chi^2} = -f\left(\psi(\chi^2)\right)\left(-\frac{1}{2\sqrt{\chi^2}}\right) = f\left(\psi(\chi^2)\right)\frac{1}{2\sqrt{\chi^2}}$$

Considerando i contributi di entrambi i rami, si conclude che la distribuzione del χ^2 per $n = 1$ è:

$$
\begin{aligned}
g(\chi^2) &= \mathbf{2}\, f\left(\psi(\chi^2)\right)\frac{1}{2\sqrt{\chi^2}} \\
&= 2\,\frac{1}{\sqrt{2\pi}}\,e^{-\frac{\chi^2}{2}}\,\frac{1}{2\sqrt{\chi^2}} \\
&= \frac{1}{\sqrt{2\pi}}\left(\chi^2\right)^{-\frac{1}{2}}\,e^{-\frac{\chi^2}{2}} \qquad (4.9)
\end{aligned}
$$

Si riscrive la densità di probabilità per $n = 1$ mettendo in evidenza la funzione Gamma:

$$g(\chi^2_1) = \frac{1}{2^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)}\left(\chi^2\right)^{\frac{1}{2}-1}\,e^{-\frac{\chi^2}{2}}$$

Si deduce quindi la distribuzione del χ^2 con n gradi di libertà:

$$g(\chi_n^2) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \left(\chi^2\right)^{\frac{n}{2}-1} e^{-\frac{\chi^2}{2}} \tag{4.10}$$

4.4 Caratteristiche della distribuzione

Normalizzazione

Si considera la distribuzione per un unico grado di libertà (4.9) e si verifica che tale funzione sia normalizzata a 1.

$$\begin{aligned}
\int_0^{+\infty} g(\chi^2) d\chi^2 &= \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} \left(\chi^2\right)^{-\frac{1}{2}} e^{-\frac{\chi^2}{2}} d\chi^2 \\
&= \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} (2x)^{-\frac{1}{2}} e^{-x}\, 2dx \\
&= \frac{1}{\sqrt{\pi}} \int_0^{+\infty} (x)^{-\frac{1}{2}} e^{-x}\, dx \\
&= \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right) = \frac{1}{\sqrt{\pi}}\sqrt{\pi} = 1
\end{aligned} \tag{4.11}$$

L'integrale è stato svolto tra gli estremi 0 e $+\infty$ in quanto il χ^2 può assumere valori solo positivi. Nel passaggio 4.11 è stata effettuata la sostituzione $x = \frac{\chi^2}{2}$.

Valore atteso

Si calcola ora il valore atteso della variabile per un grado di libertà.

$$\begin{aligned}
E\left[\chi_1^2\right] &= \int_0^{+\infty} \chi^2 g(\chi^2) d\chi^2 \\
&= \int_0^{+\infty} \chi^2 \frac{1}{\sqrt{2\pi}} \left(\chi^2\right)^{-\frac{1}{2}} e^{-\frac{\chi^2}{2}} d\chi^2 \\
&= \int_0^{+\infty} 2x\, \frac{1}{\sqrt{2\pi}} (2x)^{-\frac{1}{2}} e^{-x}\, 2dx \\
&= \frac{2}{\sqrt{\pi}} \int_0^{+\infty} (x)^{\frac{1}{2}} e^{-x} dx \\
&= \frac{2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) = \frac{2}{\sqrt{\pi}} \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \frac{2}{\sqrt{\pi}} \frac{1}{2} \sqrt{\pi} = 1
\end{aligned} \tag{4.12}$$

Nel passaggio 4.12 è stata svolta la sostituzione $x = \frac{\chi^2}{2}$.
Il risultato ottenuto si può generalizzare per un generico numero n di gradi di libertà:

$$E\left[\chi_n^2\right] = n$$

Varianza

Per determinare la varianza del χ_1^2 si ricorre ad una importante relazione dove la variabile aleatoria per comodità viene sostituita ($\chi^2 = y$).

$$\sigma_{\chi^2}^2 = E\left[\left(\chi^2 - \mu\right)^2\right] = E\left[(y-\mu)^2\right] = E[y^2] - (E[y])^2 \tag{4.13}$$

Il problema si riconduce quindi alla determinazione del valore atteso di y^2.

$$\begin{aligned}
E[y^2] &= \int_0^{+\infty} y^2\, g(y)\, dy \\
&= \int_0^{+\infty} y^2 \frac{1}{\sqrt{2\pi}}\, y^{-\frac{1}{2}}\, e^{-\frac{y}{2}}\, dy \\
&= \int_0^{+\infty} 4x^2 \frac{1}{\sqrt{2\pi}}\, (2x)^{-\frac{1}{2}}\, e^{-x}\, 2dx \qquad (4.14)\\
&= \frac{4}{\sqrt{\pi}} \int_0^{+\infty} x^{\frac{3}{2}}\, e^{-x}\, dx \\
&= \frac{4}{\sqrt{\pi}}\, \Gamma\left(\frac{5}{2}\right) = \frac{4}{\sqrt{\pi}}\, \frac{3}{2}\, \Gamma\left(\frac{3}{2}\right) \\
&= \frac{4}{\sqrt{\pi}}\, \frac{3}{2}\, \frac{1}{2}\, \Gamma\left(\frac{1}{2}\right) = \frac{4}{\sqrt{\pi}}\, \frac{3}{2}\, \frac{1}{2}\, \sqrt{\pi} = 3
\end{aligned}$$

Nel passaggio 4.14 è stata svolta la sostituzione $x = \frac{y}{2}$.
Applicando la relazione 4.13, si ottiene la varianza:

$$\sigma_{\chi_1^2}^2 = E[y^2] - (E[y])^2 = 3 - 1^2 = 2$$

Si può dimostrare che per un generico numero n di gradi di libertà,

$$\sigma_{\chi_n^2}^2 = E\left[\left(\chi^2 - \mu\right)^2\right] = 2n$$

Limiti della funzione al variare dei gradi di libertà

Si vuole studiare analiticamente il comportamento della distribuzione agli estremi. Si ricorda innanzitutto l'espressione della funzione:

$$g(\chi_n^2) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \left(\chi^2\right)^{\frac{n}{2}-1} e^{-\frac{\chi^2}{2}}$$

Si valuta il limite di tale funzione per diversi gradi di libertà:

- Per $n = 1$

$$\begin{aligned}
\lim_{\chi^2 \to 0} g(\chi_1^2) &= \lim_{\chi^2 \to 0} \frac{1}{\sqrt{2\pi}} \overbrace{\left(\chi^2\right)^{-\frac{1}{2}}}^{\to +\infty} \overbrace{e^{-\frac{\chi^2}{2}}}^{\to 1} = +\infty \\
\lim_{\chi^2 \to +\infty} g(\chi_1^2) &= \lim_{\chi^2 \to +\infty} \frac{1}{\sqrt{2\pi}} \overbrace{\left(\chi^2\right)^{-\frac{1}{2}}}^{\to 0} \overbrace{e^{-\frac{\chi^2}{2}}}^{\to 0} = 0
\end{aligned}$$

- Per $n = 2$

$$
\begin{aligned}
\lim_{\chi^2 \to 0} g(\chi_2^2) &= \lim_{\chi^2 \to 0} \frac{1}{2\,\Gamma(1)} \left(\chi^2\right)^0 e^{-\frac{\chi^2}{2}} \\
&= \lim_{\chi^2 \to 0} \frac{1}{2 \cdot 0!} \underbrace{e^{-\frac{\chi^2}{2}}}_{\to 1} = \frac{1}{2} \\
\lim_{\chi^2 \to +\infty} g(\chi_2^2) &= \lim_{\chi^2 \to +\infty} \frac{1}{2} \underbrace{e^{-\frac{\chi^2}{2}}}_{\to 0} = 0
\end{aligned}
$$

- Per $n = 3$

$$
\begin{aligned}
\lim_{\chi^2 \to 0} g(\chi_3^2) &= \lim_{\chi^2 \to 0} \frac{1}{2^{\frac{3}{2}}\Gamma(\frac{3}{2})} \overbrace{\left(\chi^2\right)^{\frac{1}{2}}}^{\to 0} \overbrace{e^{-\frac{\chi^2}{2}}}^{\to 0} = 0 \\
\lim_{\chi^2 \to +\infty} g(\chi_3^2) &= \lim_{\chi^2 \to +\infty} \frac{1}{2^{\frac{3}{2}}\Gamma(\frac{3}{2})} \left(\chi^2\right)^{\frac{1}{2}} e^{-\frac{\chi^2}{2}} \\
&= \lim_{\chi^2 \to +\infty} \frac{1}{2^{\frac{3}{2}}\Gamma(\frac{3}{2})} \underbrace{\frac{(\chi^2)^{\frac{1}{2}}}{e^{\frac{\chi^2}{2}}}}_{\substack{\to 0 \text{ per la} \\ \text{gerarchia degli } \infty}} = 0
\end{aligned}
$$

Per $n \geq 3$ il comportamento agli estremi è analogo poiché l'esponente del secondo fattore risulta sempre maggiore di zero $(\frac{n}{2} - 1 > 0)$: per $\chi^2 \to 0$ e per $\chi^2 \to +\infty$ la funzione tende a zero.
Di seguito sono riportati i grafici della distribuzione per diversi gradi di libertà in cui si riconosco i limiti determinati analiticamente.

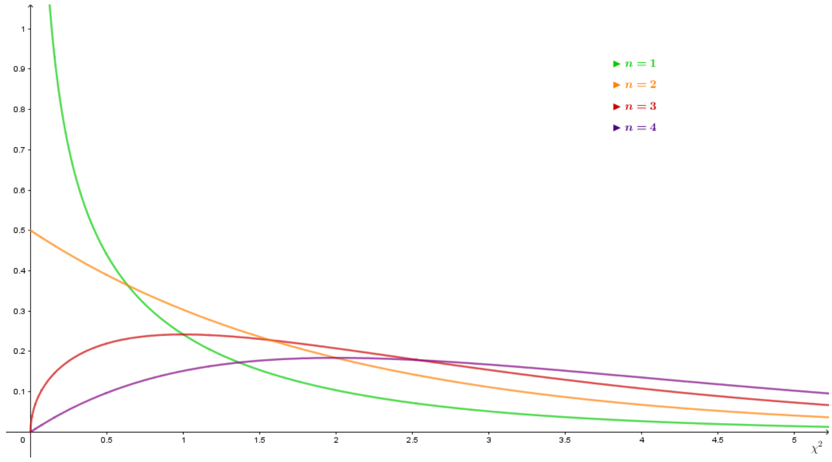

Figura 4.1: Grafico della distribuzione del χ^2 per $n = 1, 2, 3, 4$

Punto di massimo

Per valutare il punto di massimo si calcola il valore del χ^2 in grado di annullare la derivata prima della funzione.

$$\begin{aligned}\frac{d}{d\chi^2}\left\{g(\chi^2)\right\} &= \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\left[\frac{n-2}{2}\left(\chi^2\right)^{\frac{n-2}{2}-1}e^{-\frac{\chi^2}{2}}-\frac{1}{2}\left(\chi^2\right)^{\frac{n-2}{2}}e^{-\frac{\chi^2}{2}}\right]\\ &= \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\left(\chi^2\right)^{\frac{n-2}{2}}e^{-\frac{\chi^2}{2}}\left[\frac{n-2}{2}\left(\chi^2\right)^{-1}-\frac{1}{2}\right]\end{aligned}$$

$$\begin{aligned}\frac{d}{d\chi^2}\left\{g(\chi^2)\right\}=0 \quad &\Leftrightarrow \quad \frac{n-2}{2}\left(\chi^2\right)^{-1}-\frac{1}{2}=0\\ &\Leftrightarrow \quad (n-2)\left(\chi^2\right)^{-1}=1\\ &\Leftrightarrow \quad \left(\chi^2\right)^{-1}=(n-2)^{-1}\\ &\Leftrightarrow \quad \chi^2_{max}=n-2\end{aligned}$$

La soluzione è valida per $n > 2$ e il massimo risulta sempre traslato rispetto al valore atteso n. Per $n = 1$ la funzione non assume un valore massimo infatti tende a $+\infty$ per $\chi^2 \to 0$, mentre per $n = 2$ la funzione è sempre decrescente e in assenza di un punto stazionario il valore massimo viene assunto all'estremo per $\chi^2 = 0$ come si può notare dalla rappresentazione grafica.

4.5 Test del χ^2

Per svolgere un test sull'attendibilità di una previsione teorica o di un'ipotetica distribuzione è necessario innazitutto valutare la variabile χ^2 secondo la definizione 4.1 oppure 4.2. Successivamente si deve determinare il numero di gradi di libertà:

- nel caso in cui si voglia confrontare dati sperimentali con una previsione teorica, si deve considerare il numero di paragoni svolti equivalente al numero dei dati y_i (N) a cui va sottratto il numero di parametri della funzione teorica calcolati a partire dai dati sperimentali. Per esempio, nel caso di un fit lineare vengono calcolati due parametri (il coefficiente angolare e l'ordinata all'origine), dunque il numero di gradi di libertà risulta pari a $N - 2$.

- nel caso in cui si voglia studiare la distribuzione di una grandezza, si deve considerare il numero di dati sperimentali della grandezza (N) a cui vanno sottratti i parametri della distribuzione ipotizzata calcolati a partire dalle misure. Per il confronto con la gaussiana, è necessario stimare il valor vero μ e la deviazione standard σ, perciò i gradi di libertà del χ^2 saranno $N-2$; se invece si ipotizza una distribuzione di Poisson è necessario stimare unicamente il parametro λ quindi in tal caso il numero dei gradi sarà $N-1$.

Si deve quindi considerare la distribuzione della variabile χ^2 per i gradi di libertà calcolati. Minore è il valore della variabile e maggiore è l'accordo tra i dati sperimentali e la previsione teorica, perciò partendo dall'estremo $\chi^2 = 0$ si considera l'area sottesa al grafico della distribuzione rappresentante una probabilità. Solitamente si considera un'area pari al 95% oppure al 99% e si determina il corrispondente estremo destro del range ($\chi^2_{0.95}$ oppure $\chi^2_{0.99}$).

$$\int_0^{\chi^2_{0.95}} g(\chi^2)\, d\chi^2 = 0.95 = 95\%$$
$$\int_0^{\chi^2_{0.99}} g(\chi^2)\, d\chi^2 = 0.99 = 99\%$$

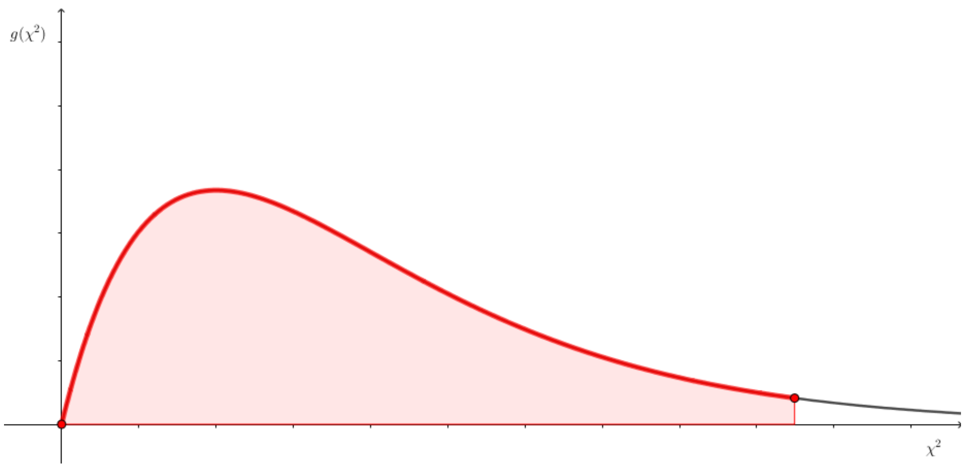

Figura 4.2: Regione sinistra della distribuzione con un'area pari al 95%

La funzione densità di probabilità è stata tabulata e i valori degli estremi $\chi^2_{0.95}$ e $\chi^2_{0.99}$ sono riportati in apposite tabelle. Si svolge dunque il test confrontando il valore di χ^2 ottenuto con il valore dell'estremo: il test darà riscontro positivo nel caso in cui il primo risulti minore del secondo. In tal caso si dice che l'ipotesi nulla H_0, secondo cui la discrepanza tra i valori sperimentali e la previsione teorica è dovuta unicamente al caso, è accettata con un livello di confidenza del 95% (o del 99%) oppure con un livello di significatività del 5% (o dell'1%).

Osservando la distribuzione del χ^2 si può notare che per gradi di libertà superiori a 3, la densità di probabilità tende a zero al tendere a zero della variabile, dunque è poco probabile ottenere un valore troppo piccolo. Per verificare ciò si può svolgere un ulteriore test considerando una regione destra con area sempre pari al 95% oppure al 99%.

$$\int_{\chi^2_{0.05}}^{+\infty} g(\chi^2)\, d\chi^2 = \int_0^{+\infty} g(\chi^2)\, d\chi^2 - \int_0^{\chi^2_{0.05}} = 1 - 0.05 = 0.95 = 95\%$$
$$\int_{\chi^2_{0.01}}^{+\infty} g(\chi^2)\, d\chi^2 = \int_0^{+\infty} g(\chi^2)\, d\chi^2 - \int_0^{\chi^2_{0.01}} = 1 - 0.01 = 0.99 = 99\%$$

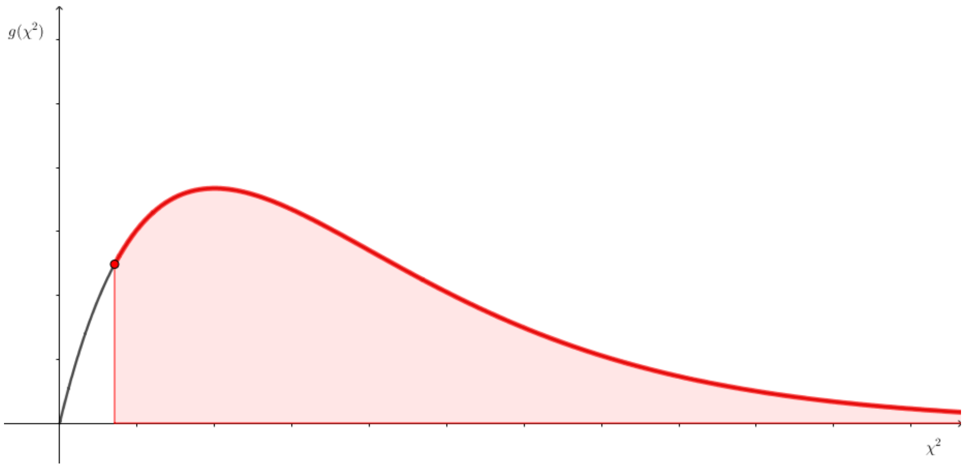

Figura 4.3: Regione destra della distribuzione con area pari al 95%

Si stabilisce il valore dell'estremo sinistro del range ($\chi^2_{0.05}$ oppure $\chi^2_{0.01}$) determinabile attraverso la relativa tabella e si verifica che il χ^2 calcolato con i dati sperimentali sia superiore a tale numero. Se questo non fosse verificato allora il χ^2 ottenuto sarebbe fortemente sospetto: la causa di ciò potrebbe essere riconducibile ad una errata stima dell'incertezza dei dati sperimentali (σ_i) infatti una sua sovrastima, nel calcolo del χ^2, comporterebbe valori minori degli addendi:

$$\frac{(y_{is} - y_{it})^2}{\sigma_i^2}$$

Capitolo 5

Distribuzione della variabile t

5.1 Definizione della variabile t

La variabile Z è utile per confrontare il risultato sperimentale di una grandezza fisica con il rispettivo valore teorico, nota la deviazione standard. Nella realtà si ha sempre a disposizione un campione con un numero limitato di dati attraverso cui si può stimare l'errore da associare alla singola misura e alla media. Per svolgere test di compatibilità si considera la discrepanza tra il valor medio dei dati sperimentali e il valor vero in termini di errore standard; in questo modo è definita la variabile aleatoria t di Student:

$$t = \frac{\bar{x} - \mu}{s_{media}}$$

A partire dai dati del campione si ricava la deviazione standard empirica s da cui segue la valutazione dell'errore standard s_{media}:

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

$$s_{media} = \frac{s}{\sqrt{N}}$$

Essendo una stima, s è affetta da un errore relativo dipendente dal numero di dati e questo errore ricade anche nella determinazione dell'incertezza da associare alla media:

$$\frac{\sigma_s}{s} = \frac{1}{\sqrt{2(N-1)}}$$

Di conseguenza, se si ha a disposizione un ristretto numero di dati, bisogna considerare il possibile errore sulla stima della deviazione standard. La variabile t di Student tiene conto di ciò e assume una distribuzione differente dalla Z di Gauss.

La variabile t può essere definita in un modo differente, utile per la determinazione della densità di probabilità:

$$\begin{aligned} t &= \frac{\bar{x}-\mu}{s_m} = \frac{\bar{x}-\mu}{\frac{s_m}{\sigma_m}\sigma_m} = \frac{Z}{\frac{s_m}{\sigma_m}} = \frac{Z}{\frac{s}{\sigma}} = \frac{Z}{\sqrt{\frac{1}{N-1}\frac{1}{\sigma^2}\sum(x_i-\bar{x})^2}} \\ &= \frac{Z}{\sqrt{\frac{1}{N-1}\sum(\frac{x_i-\bar{x}}{\sigma})^2}} = \frac{Z}{\sqrt{\frac{1}{N-1}\chi^2}} = \frac{Z}{\sqrt{\tilde{\chi}^2}} \end{aligned} \tag{5.1}$$

dove si è riconosciuta una variabile di tipo chi quadro con gradi di libertà pari a $N-1$ in quanto, a partire dai dati, si è calcolata la media campionaria; il rapporto tra la variabile e i rispettivi gradi viene indicato come *chi quadro ridotto* ($\tilde{\chi}^2$).

5.2 Deduzione della distribuzione

Si parte dalla relazione 5.1 per ricavare la distribuzione della variabile di Student. Si definisce innanzitutto r come segue:

$$r = \sqrt{\tilde{\chi}^2} = \sqrt{\frac{\chi^2}{n}}, \text{ con } n \text{ gradi di libertà}$$

Si ricava ora il chi quadro in funzione di r e si valuta la derivata.

$$r^2 = \frac{\chi^2}{n} \Rightarrow \chi^2 = nr^2$$

$$\frac{d\chi^2}{dr} = 2nr$$

Nota ora la distribuzione del chi quadro (4.10), si ricava la densità di probabilità di r attraverso la relazione 1.1.

$$\begin{aligned} g(r) &= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}(nr^2)^{\frac{n}{2}-1}e^{-\frac{nr^2}{2}}\frac{d\chi^2}{dr} \\ &= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}(nr^2)^{\frac{n}{2}-1}e^{-\frac{nr^2}{2}}2nr \\ &= \underbrace{\frac{1}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})}n^{\frac{n}{2}}}_{C(n)} r^{n-1}e^{-\frac{nr^2}{2}} = C(n)\ r^{n-1}e^{-\frac{nr^2}{2}} \end{aligned} \tag{5.2}$$

La variabile t è definita come rapporto tra due variabili aleatorie con distribuzioni note. Consideriamo inizialmente r costante e ricaviamo la funzione densità di probabilità a partire da quella relativa a Z (3.2):

$$\begin{aligned} h(t,|r|) &= \frac{1}{\sqrt{2\pi}}e^{-\frac{r^2t^2}{2}}\frac{\partial Z}{\partial t} \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{r^2t^2}{2}}r \end{aligned} \tag{5.3}$$

Per ricavare la distribuzione di t è necessario considerare la densità a un r fissato ($h(t,|r|)$) e moltiplicarla per la probabilità infinitesima di ottenere un valore compreso nel range $[r, r+dr]$ rappresentata dal prodotto tra la distribuzione 5.2 e il differenziale dr. La funzione ricavata andrà integrata su tutto il campo di variabilità di r ($\mathbf{R}^+$) per ottenere la distribuzione di Student.

$$\begin{aligned} h(t) &= \int_0^{+\infty} h(t,|r|)\ g(r)dr \\ &= \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{r^2t^2}{2}} r\ C(n)\ r^{n-1} e^{-\frac{nr^2}{2}} dr \\ &= \frac{C(n)}{\sqrt{2\pi}} \int_0^{+\infty} r^n e^{-\frac{nr^2}{2}(1+\frac{t^2}{n})} dr \end{aligned}$$

Per risolvere l'integrale si applica il metodo di integrazione per parti.

$$\begin{aligned} h(t) &= \frac{C(n)}{\sqrt{2\pi}} \int_0^{+\infty} r^{n-1}\ r \frac{n(1+\frac{t^2}{n})}{n(1+\frac{t^2}{n})} e^{-\frac{nr^2}{2}(1+\frac{t^2}{n})} dr \\ &= \frac{C(n)}{\sqrt{2\pi}} \left\{ \left[-r^{n-1} \frac{1}{n(1+\frac{t^2}{n})} e^{-\frac{nr^2}{2}(1+\frac{t^2}{n})} \right]_0^{+\infty} + \int_0^{+\infty} \frac{1}{n(1+\frac{t^2}{n})} e^{-\frac{nr^2}{2}(1+\frac{t^2}{n})} (n-1) r^{n-2}\, dr \right\} \end{aligned}$$

Il primo addendo risulta nullo: se $r \to +\infty$ per la gerarchia degli infiniti l'esponenziale convergente a zero rende nulla l'espressione mentre per $r \to 0$ il primo fattore r^{n-1} tende a zero.

$$h(t) = \frac{C(n)}{\sqrt{2\pi}} \int_0^{+\infty} \frac{n-1}{n(1+\frac{t^2}{n})} r^{n-2} e^{-\frac{nr^2}{2}(1+\frac{t^2}{n})}\, dr$$

Per capire come risolvere l'integrale riapplichiamo l'integrazione per parti.

$$\begin{aligned} h(t) &= \frac{C(n)}{\sqrt{2\pi}} \int_0^{+\infty} \frac{n-1}{n(1+\frac{t^2}{n})} r^{n-3}\ r \frac{n(1+\frac{t^2}{n})}{n(1+\frac{t^2}{n})} e^{-\frac{nr^2}{2}(1+\frac{t^2}{n})} dr \\ &= \frac{C(n)}{\sqrt{2\pi}} \left\{ \left[\frac{-r^{n-3}(n-1)}{n^2(1+\frac{t^2}{n})^2} e^{-\frac{nr^2}{2}(1+\frac{t^2}{n})} \right]_0^{+\infty} + \int_0^{+\infty} \frac{n-1}{n^2(1+\frac{t^2}{n})^2} e^{-\frac{nr^2}{2}(1+\frac{t^2}{n})} (n-3) r^{n-4}\, dr \right\} \end{aligned}$$

Si ottiene quindi a seguito della seconda integrazione:

$$h(t) = \frac{C(n)}{\sqrt{2\pi}} \frac{(n-1)(n-3)}{n^2(1+\frac{t^2}{n})^2} \int_0^{+\infty} r^{n-4} e^{-\frac{nr^2}{2}(1+\frac{t^2}{n})}\, dr \tag{5.4}$$

Si distinguono ora due casi in base al numero di gradi di libertà.
Se n è **pari** allora iterando il processo per un numero di volte pari a $\frac{n}{2}$ si ottiene:

$$h_p(t) = \frac{C(n)}{\sqrt{2\pi}} \frac{(n-1)(n-3)(n-5)...}{n^{\frac{n}{2}}(1+\frac{t^2}{n})^{\frac{n}{2}}} \int_0^{+\infty} e^{-\frac{nr^2}{2}(1+\frac{t^2}{n})} dr$$

Quest'ultimo integrale si risolve per sostituzione ponendo $x^2 = \frac{nr^2}{2}(1+\frac{t^2}{n})$:

$$\begin{aligned}
h_p(t) &= \frac{C(n)}{\sqrt{2\pi}} \frac{(n-1)(n-3)(n-5)...}{n^{\frac{n}{2}}(1+\frac{t^2}{n})^{\frac{n}{2}}} \int_0^{+\infty} e^{-x^2} \frac{dx}{\sqrt{\frac{n}{2}(1+\frac{t^2}{n})}} \\
&= \frac{C(n)}{\sqrt{2\pi}} \frac{(n-1)(n-3)(n-5)...}{n^{\frac{n+1}{2}}(1+\frac{t^2}{n})^{\frac{n+1}{2}}} \sqrt{2} \int_0^{+\infty} e^{-x^2} dx \\
&= \frac{C(n)}{\sqrt{2\pi}} \frac{(n-1)(n-3)(n-5)...}{n^{\frac{n+1}{2}}(1+\frac{t^2}{n})^{\frac{n+1}{2}}} \sqrt{2} \frac{\sqrt{\pi}}{2} \qquad (5.5) \\
&= \underbrace{\frac{C(n)}{2} \frac{(n-1)(n-3)(n-5)...}{n^{\frac{n+1}{2}}}}_{K_p(n)} \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}
\end{aligned}$$

L'espressione 5.5 è stata ottenuta riconoscendo l'integrale di Gauss svolto tra gli estremi 0 e $+\infty$. Si studia ora la costante $K_p(n)$ per la distribuzione di Student con gradi pari ricordando il valore di $C(n)$ dalla (5.2).

$$\begin{aligned}
K_p(n) &= \frac{1}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} n^{\frac{n}{2}} \frac{1}{2} \frac{(n-1)(n-2)(n-5)...}{n^{\frac{n+1}{2}}} \\
&= \frac{(n-1)(n-3)(n-5)...}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})\sqrt{n}} \\
&= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})\sqrt{n}} \sum_{k=1}^{\frac{n}{2}} (n-2k+1)
\end{aligned}$$

Si osserva che per $k = \frac{n}{2}$ si ottiene come ultimo fattore 1.

$$\begin{aligned}
K_p(n) &= \frac{1}{\Gamma(\frac{n}{2})\sqrt{n}} \frac{(n-1)}{2} \frac{(n-3)}{2} \frac{(n-5)}{2} ... \frac{1}{2} \\
&= \frac{1}{\Gamma(\frac{n}{2})\sqrt{\pi n}} \frac{(n-1)}{2} \frac{(n-3)}{2} \frac{(n-5)}{2} ... \frac{1}{2}\sqrt{\pi} \\
&= \frac{1}{\Gamma(\frac{n}{2})\sqrt{\pi n}} \left(\frac{n}{2}-\frac{1}{2}\right)\left(\frac{n}{2}-\frac{3}{2}\right)\left(\frac{n}{2}-\frac{5}{2}\right) ... \frac{1}{2}\Gamma\left(\frac{1}{2}\right) \\
&= \frac{1}{\Gamma(\frac{n}{2})\sqrt{\pi n}} \left(\frac{n}{2}-\frac{1}{2}\right)\left(\frac{n}{2}-\frac{3}{2}\right)\Gamma\left(\frac{n}{2}-\frac{3}{2}\right) \\
&= \frac{1}{\Gamma(\frac{n}{2})\sqrt{\pi n}} \left(\frac{n}{2}-\frac{1}{2}\right)\Gamma\left(\frac{n}{2}-\frac{1}{2}\right) \\
&= \frac{1}{\Gamma(\frac{n}{2})\sqrt{\pi n}} \Gamma\left(\frac{n+1}{2}\right)
\end{aligned}$$

Negli ultimi passaggi si sono applicate le proprietà della funzione Gamma (4.5). L'espressione della funzione $h_p(t)$ risulta quindi:

$$h_p(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})} \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \qquad (5.6)$$

Si considera ora il caso con gradi di liberta n **dispari**.
Dopo un numero di iterazioni pari a $\frac{n-1}{2}$ si ottiene:

$$
\begin{aligned}
h_d(t) &= \frac{C(n)}{\sqrt{2\pi}} \frac{\overbrace{(n-1)(n-3)(n-5)...}^{\frac{n-1}{2}\text{ fattori}}}{n^{\frac{n-1}{2}}(1+\frac{t^2}{n})^{\frac{n-1}{2}}} \int_0^{+\infty} r\, e^{-\frac{nr^2}{2}(1+\frac{t^2}{n})} dr \\
&= \frac{C(n)}{\sqrt{2\pi}} \frac{(n-1)(n-3)(n-5)...}{n^{\frac{n-1}{2}}(1+\frac{t^2}{n})^{\frac{n-1}{2}}} \int_0^{+\infty} \frac{1}{n^1\left(1+\frac{t^2}{n}\right)^1} nr \left(1+\frac{t^2}{n}\right) e^{-\frac{nr^2}{2}(1+\frac{t^2}{n})} dr \\
&= \frac{C(n)}{\sqrt{2\pi}} \frac{(n-1)(n-3)(n-5)...}{n^{\frac{n+1}{2}}(1+\frac{t^2}{n})^{\frac{n+1}{2}}} \left[-e^{-\frac{nr^2}{2}(1+\frac{t^2}{n})}\right]_0^{+\infty} \\
&= \frac{C(n)}{\sqrt{2\pi}} \frac{(n-1)(n-3)(n-5)...}{n^{\frac{n+1}{2}}(1+\frac{t^2}{n})^{\frac{n+1}{2}}} [0-(-1)] \\
&= \frac{C(n)}{\sqrt{2\pi}} \frac{(n-1)(n-3)(n-5)...}{n^{\frac{n+1}{2}}\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} \\
&= \underbrace{\frac{C(n)}{\sqrt{2\pi}} \frac{(n-1)(n-3)(n-5)...}{n^{\frac{n+1}{2}}}}_{K_d(n)} \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}
\end{aligned}
$$

Si semplifica la costante $K_d(n)$:

$$
\begin{aligned}
K_d(n) &= \frac{1}{\sqrt{2\pi}} \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} \frac{(n-1)(n-3)(n-5)...}{n^{\frac{n+1}{2}}} \\
&= \frac{1}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \frac{\overbrace{(n-1)(n-3)(n-5)\dots 4\cdot 2}^{\frac{n-1}{2}\text{ fattori}}}{2^{\frac{n-1}{2}}} \\
&= \frac{1}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(\frac{n-1}{2}\right)\left(\frac{n-3}{2}\right)\left(\frac{n-5}{2}\right) ... \frac{4}{2}\cdot\frac{2}{2} \\
&= \frac{1}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(\frac{n-1}{2}\right)\left(\frac{n-1}{2}-1\right)\left(\frac{n-1}{2}-2\right) ... 2\cdot 1 \\
&= \frac{1}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(\frac{n-1}{2}\right)! \\
&= \frac{1}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \Gamma\left(\frac{n+1}{2}\right)
\end{aligned}
$$

L'ultimo passaggio è stato svolto considerando la definizione della funzione Gamma per argomento intero (4.7). L'espressione della funzione $h_d(t)$ risulta quindi:

$$h_d(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})} \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \tag{5.7}$$

Confrontando le funzioni 5.6 e 5.7, si conclude che, indipendentemente dalla natura dei gradi di libertà, la densità di probabilità della variabile t di Student è:

$$h(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})}\left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \tag{5.8}$$

5.3 Caratteristiche della distribuzione

5.3.1 Grafico della funzione

La distribuzione di Student assume una forma diversa in base al numero di gradi di libertà.

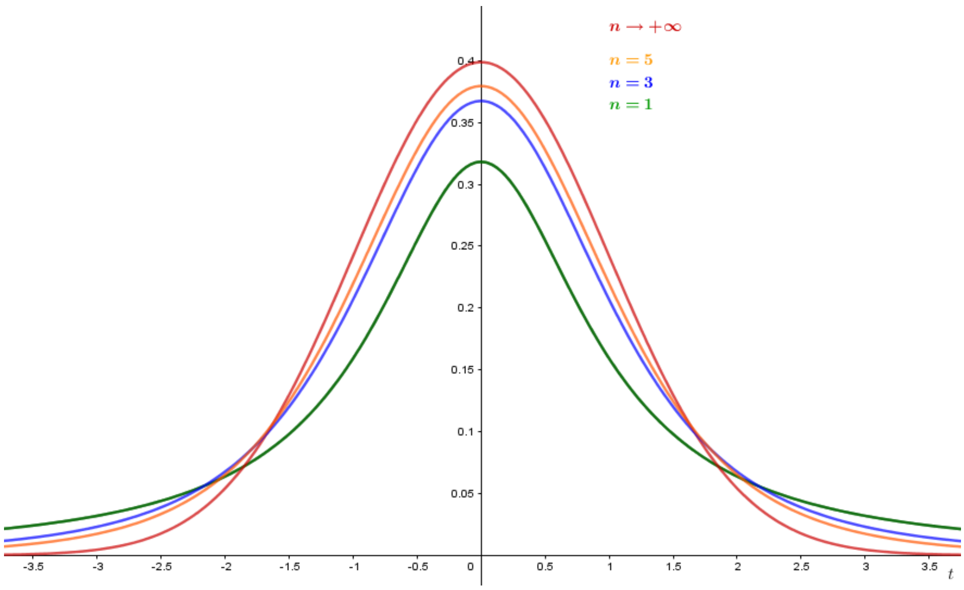

Figura 5.1: Grafico della distribuzione per $n = 1, 3, 5$ e per $n \to +\infty$

Come si può osservare dalla rappresentazione grafica, la distribuzione di Student assume un massimo in corrispondenza di $t_{max} = 0$; analiticamente questo risultato si ottiene ponendo $\left.\frac{dh(t)}{dt}\right|_{t_{max}} = 0$.

All'aumentare dei gradi di libertà si osserva un aumento del picco della distribuzione; al tempo stesso le code risultano appiattirsi maggiormente. Infatti la distribuzione tiene conto del fatto che la variabile t è definita attraverso una deviazione standard empirica, quest'ultima potrebbe essere affetta da un errore maggiore in corrispondenza di un minore numero di dati e di gradi di libertà; è quindi da considerare più probabile l'acquisizione di un valore maggiore di t. Il test di Student risulterà meno restrittivo del test di Gauss in quanto tiene conto anche del fatto che l'errore è una stima più o meno

precisa della deviazione standard in base ai gradi di libertà.
Graficamente si osserva una tendenza della distribuzione alla normalità, per $n \to +\infty$ la curva si sovrappone alla curva gaussiana corrispondente alla densità di probabilità di Z. Infatti all'aumentare dei gradi di libertà il valore di s tende all'effettiva deviazione standard σ e quindi le definizioni di t e Z vengono a coincidere.

5.3.2 Tendenza alla normalità

Di seguito si studia analiticamente il limite della distribuzione di Student per un numero di gradi di libertà tendente ad infinito.

$$\begin{aligned}
\lim_{n\to+\infty} h(t) &= \lim_{n\to+\infty} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \\
&= \underbrace{\lim_{n\to+\infty} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})}}_{K_\infty} \lim_{n\to+\infty}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \\
&= K_\infty \lim_{n\to+\infty}\left(1+\frac{1}{\frac{n}{t^2}}\right)^{-\frac{n+1}{2}} \\
&= K_\infty \lim_{n\to+\infty}\left(1+\frac{1}{\frac{n}{t^2}}\right)^{\frac{n}{t^2}\frac{t^2}{n}(-\frac{n+1}{2})} \\
&= K_\infty \lim_{n\to+\infty}\left[\underbrace{\left(1+\frac{1}{\frac{n}{t^2}}\right)^{\frac{n}{t^2}}}_{\to e}\right]^{-\frac{t^2}{2}\overbrace{\left(\frac{n+1}{n}\right)}^{\to 1}} \\
&= K_\infty\, e^{-\frac{t^2}{2}}
\end{aligned} \tag{5.9}$$

Per calcolare la costante K_∞, si ricorre all'approssimazione di Stirling:

$$\Gamma(\alpha) = (\alpha-1)! \sim \sqrt{2\pi(\alpha-1)}\left(\frac{\alpha-1}{e}\right)^{\alpha-1} \quad \text{per } \alpha \to +\infty$$

Si determina quindi la costante applicando il limite.

$$\begin{aligned}
K_\infty &= \lim_{n\to+\infty} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})} \\
&= \lim_{n\to+\infty} \frac{\sqrt{2\pi(\frac{n+1}{2}-1)}\left(\frac{\frac{n+1}{2}-1}{e}\right)^{\frac{n+1}{2}-1}}{\sqrt{\pi n}\sqrt{2\pi(\frac{n}{2}-1)}\left(\frac{\frac{n}{2}-1}{e}\right)^{\frac{n}{2}-1}}
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to+\infty} \frac{\sqrt{\pi n-\pi}\left(\frac{n-1}{2e}\right)^{\frac{n}{2}-\frac{1}{2}}}{\sqrt{\pi n}\sqrt{\pi n-2\pi}\left(\frac{n-2}{2e}\right)^{\frac{n}{2}-1}} \\
&= \lim_{n\to+\infty} \frac{\sqrt{\pi n-\pi}}{\sqrt{\pi n-2\pi}} \frac{(2e)^{\frac{n}{2}-1}}{\sqrt{\pi n}(n-2)^{\frac{n}{2}-1}} \frac{(n-1)^{\frac{n}{2}-\frac{1}{2}}}{(2e)^{\frac{n}{2}-\frac{1}{2}}} \\
&= \lim_{n\to+\infty} \frac{\sqrt{n-1}}{\sqrt{n-2}} \frac{(2e)^{\frac{n}{2}-\frac{1}{2}}(2e)^{-\frac{1}{2}}}{\sqrt{\pi n}(n-2)^{\frac{n}{2}}(n-2)^{-1}} \frac{(n-1)^{\frac{n}{2}}(n-1)^{-\frac{1}{2}}}{(2e)^{\frac{n}{2}-\frac{1}{2}}} \\
&= \lim_{n\to+\infty} \frac{(n-1)^{\frac{1}{2}}}{(n-2)^{\frac{1}{2}}} \frac{(2e)^{-\frac{1}{2}}}{\sqrt{\pi n}(n-2)^{\frac{n}{2}}(n-2)^{-1}}(n-1)^{\frac{n}{2}}(n-1)^{-\frac{1}{2}} \\
&= \lim_{n\to+\infty} \frac{1}{(n-2)^{-\frac{1}{2}}} \frac{(n-1)^{\frac{n}{2}}}{\sqrt{\pi}n^{\frac{1}{2}}(n-2)^{\frac{n}{2}}(2e)^{\frac{1}{2}}} \\
&= \lim_{n\to+\infty} \frac{(n-2)^{\frac{1}{2}}}{n^{\frac{1}{2}}} \frac{1}{\sqrt{\pi}} \frac{1}{(2e)^{\frac{1}{2}}} \left(\frac{n-1}{n-2}\right)^{\frac{n}{2}} \\
&= \lim_{n\to+\infty} \left(\frac{n-2}{n}\right)^{\frac{1}{2}} \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{2}\sqrt{e}} \left(\frac{n-2+1}{n-2}\right)^{\frac{n}{2}} \\
&= \lim_{n\to+\infty} \left(\frac{n-2}{n}\right)^{\frac{1}{2}} \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{2}\sqrt{e}} \left(1+\frac{1}{n-2}\right)^{\frac{n-2}{2}} \left(1+\frac{1}{n-2}\right)^{1} \\
&= \lim_{n\to+\infty} \left(\underbrace{\frac{n-2}{n}}_{\to 1}\right)^{\frac{1}{2}} \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{2}\sqrt{e}} \left[\underbrace{\left(1+\frac{1}{n-2}\right)^{n-2}}_{\to e}\right]^{\frac{1}{2}} \left(\underbrace{1+\frac{1}{n-2}}_{\to 1}\right) \\
&= \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{2}\sqrt{e}} \sqrt{e} \\
&= \frac{1}{\sqrt{2\pi}} \qquad (5.10)
\end{aligned}
$$

Combinando le relazioni 5.9 e 5.10 si conclude che la distribuzione della variabile t di Student per un numero di gradi di libertà tendente ad infinito viene a coincidere con la distribuzione normale standardizzata di Z.

$$\lim_{n\to+\infty} h(t) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}} \qquad (5.11)$$

5.4 Test di Student

Il test sulla variabile t è utile per verificare la compatibilità tra un risultato sperimentale e il valore teorico previsto nel caso in cui l'errore della misura sperimentale sia stato determinato in modo statistico, per esempio tramite misure ripetute. La procedura del test di Student è analoga a quella del test normale, ma in questo caso è necessario stimare il numero di gradi di libertà.

Se l'errore statistico è stato ottenuto a partire da misure ripetute, dalla relazione 5.1, il numero dei gradi risulterà pari a $N-1$, dove N rappresenta il numero di dati raccolti per stabilire la media campionaria. Si può invece dimostrare che, se si associa un errore a posteriori derivante da un fit lineare, il numero di gradi di libertà si riduce a $N-2$.
Come nel caso del test Z, il valore atteso è nullo in corrispondenza del quale si avrebbe una coincidenza tra il risultato sperimentale e il valore teorico. Si deve quindi stabilire un intervallo della variabile t simmetrico rispetto al valor zero all'interno del quale i risultati possono essere considerati compatibili in quanto la discrepanza potrebbe essere dovuta unicamente al caso.
Si considera una probabilità pari al 95% oppure al 99% e si determinano gli estremi opposti della regione sottesa alla distribuzione tale da rappresentare la probabilità scelta.
Si definisce t_x il valore tale per cui $\int_{-\infty}^{t_x} h(t)dt = x$

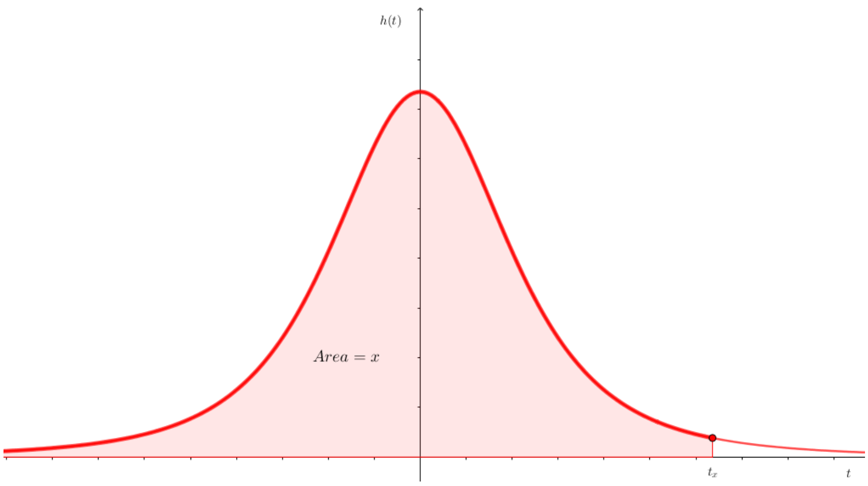

Figura 5.2: Rappresentazione grafica della definizione di t_x

Si stabiliscono quindi gli estremi da considerare nel test di Student

$$\begin{aligned}
\int_{-t_{0.975}}^{t_{0.975}} h(t)dt &= \int_{-\infty}^{t_{0.975}} h(t)dt - \int_{-\infty}^{-t_{0.975}} h(t)dt \\
&= \int_{-\infty}^{t_{0.975}} h(t)dt - \int_{t_{0.975}}^{+\infty} h(t)dt \\
&= 0.975 - (1-0.975) = 0.95 = 95\% \\
\int_{-t_{0.995}}^{t_{0.995}} h(t)dt &= \int_{-\infty}^{t_{0.995}} h(t)dt - \int_{-\infty}^{-t_{0.995}} h(t)dt \\
&= \int_{-\infty}^{t_{0.995}} h(t)dt - \int_{t_{0.995}}^{+\infty} h(t)dt \\
&= 0.995 - (1-0.995) = 0.99 = 99\%
\end{aligned}$$

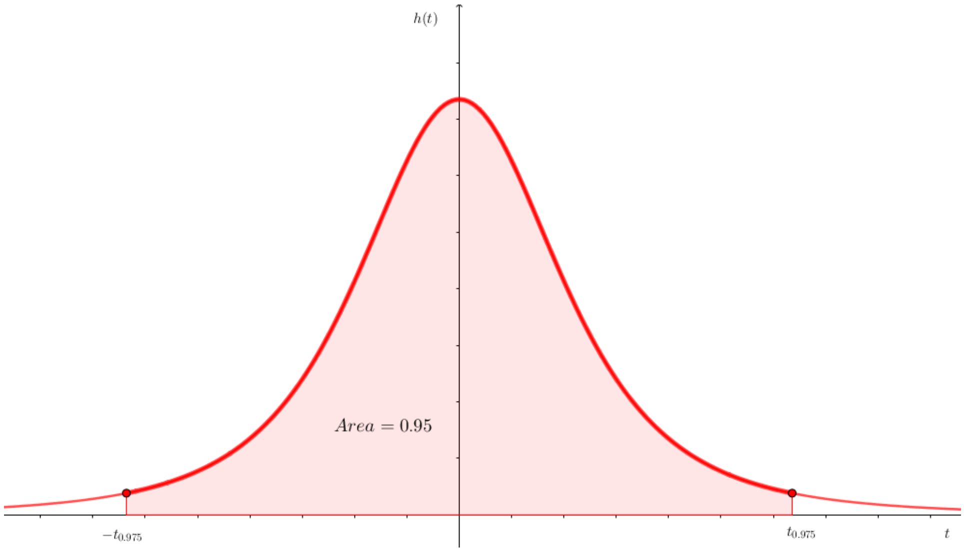

Figura 5.3: Rappresentazione della regione simmetrica di area pari al 95%

Determinati gli estremi da considerare si può svolgere il test di Student:

- con un livello di significatività del 5% (oppure con un livello di confidenza del 95%), l'ipotesi nulla di compatibilità è accettata se il valore di t determinato sperimentalmente rispetta la disuguaglianza:

$$-t_{0.975} \leq t_{spe} \leq t_{0.975} \Leftrightarrow |t_{spe}| \leq t_{0.975}$$

- con un livello di significatività del 1% (oppure con un livello di confidenza del 99%), l'ipotesi nulla di compatibilità è accettata se il valore di t determinato sperimentalmente rispetta la disuguaglianza:

$$-t_{0.995} \leq t_{spe} \leq t_{0.995} \Leftrightarrow |t_{spe}| \leq t_{0.995}$$

Come già osservato graficamente (5.3.1), la distribuzione di Student presenta al variare dei gradi di libertà una forma diversa: al crescere dei gradi, il picco aumenta e le code si appiattiscono fino a tendere al grafico della distribuzione normale. Fissato un livello di confidenza, i test di Student e di Gauss vengono applicati considerando la stessa area sottesa alla distribuzione ma gli estremi di accettabilità sono differenti e in base alla forma funzionale è sempre vero che l'intervallo della variabile t risulta essere superiore all'intervallo della variabile Z. Daltronde la variabile di Student tiene conto del fatto che l'errore associato è una stima statistica e potrebbe non rappresentare la vera deviazione standard della media campionaria.

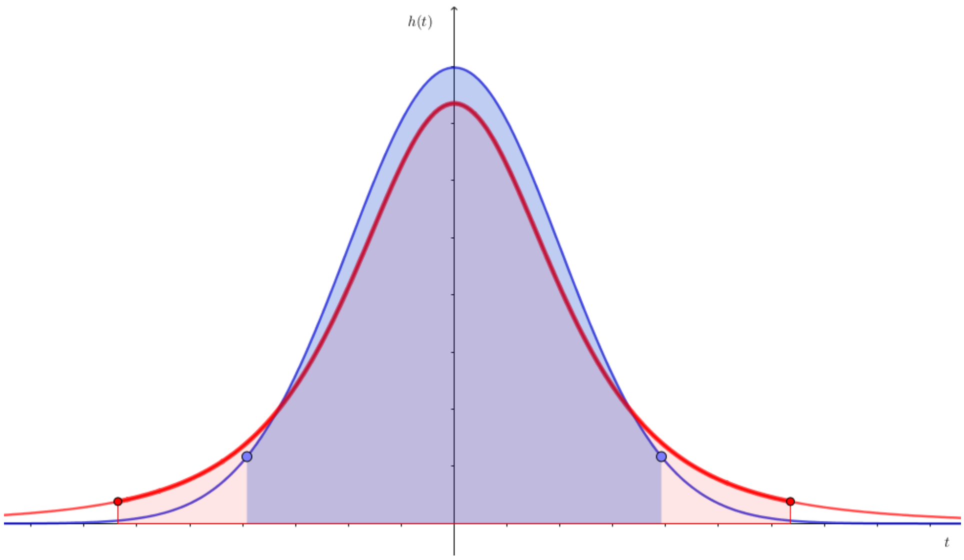

Figura 5.4: Grafico rappresentante una distribuzione di Student (in rosso) e la distribuzione normale standardizzata (in blu). Sono evidenziate due regioni sottese alle funzioni entrambe di area pari a $0,95$.

Il test di Gauss risulta quindi sempre più restrittivo del test di Student. Questa osservazione può essere utile nei casi in cui in presenza di errori statistici risulta difficile la determinazione dei numero di gradi di libertà: se lo studio di compatibilità sulla variabile Z (ponendo a denominatore la *stima* della deviazione standard) dà riscontro positivo allora sicuramente sarà anche soddisfatto il test sulla variabile t; se invece non è accettata l'ipotesi nulla tramite il test normale, è necessario verificare la compatibilità o meno con lo specifico studio sulla variabile t.

Capitolo 6

Distribuzione della variabile F

6.1 Definizione della variabile F

È possibile introdurre una nuova variabile, detta *F di Fisher*, definita come rapporto tra due chi quadri ridotti:

$$F = \frac{\frac{\chi_1^2}{\nu_1}}{\frac{\chi_2^2}{\nu_2}}$$

Mediante questo rapporto (nota la distribuzione di probabilità della F) è possibile confrontare due relazioni funzionali diverse associate agli stessi dati. Si può quindi scegliere tra due fit diversi valutando i valori di χ^2 ottenuti dagli stessi dati.
Si può ricavare una nuova definizione della variabile a partire dalla precedente, a partire dall'espressione dell'espressione del χ^2 e moltiplicando ambo i membri per $\nu = (N-1)$ gradi di libertà:

$$\chi^2 = \underbrace{\frac{\sum_{i=1}^{n}(x_i - x_m)^2}{(N-1)}}_{S^2} \frac{(N-1)}{\sigma^2} = \frac{S^2}{\sigma^2}(N-1)$$

$$\frac{\chi^2}{\nu} = \frac{S^2}{\sigma^2}$$

A partire dalla definizione di F e ipotizzando che appartengano alla stessa popolazione ($\sigma_1 = \sigma_2 = \sigma$):

$$F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} = \frac{S_1^2}{S_2^2} \tag{6.1}$$

Questa nuova definizione è utile per verificare le ipotesi di appartenenza di campioni differenti alla medesima popolazione.

Si può quindi valutare la densità di probabilità della F che indica la probabilità puntuale, inifitesima, di ottenere un valore di F pari a quello calcolato, sotto le ipotesi funzionali effettuate in partenza.

6.2 Deduzione della distribuzione

Essendo la F valutata dai rapporti dei χ^2, nota la distribuzione di probabilità del χ^2 (4.10) è possibile determinare la distribuzione della F ricorrendo alla teoria delle Distribuzioni Derivate (1.1). É necessario però tener conto che la F è una funzione di due variabili di tipo chi quadro, che per comodità vengono chiamate r ed s:

$$F = \frac{\frac{\chi_1^2}{\nu_1}}{\frac{\chi_2^2}{\nu_2}} = \frac{\nu_2}{\nu_1}\frac{\chi_1^2}{\chi_2^2} = \frac{\nu_2}{\nu_1}\frac{r}{s} \tag{6.2}$$

Il procedimento che si utilizza per il calcolo è in breve il seguente: si considera inzialmente $|s|$, ovvero la s costante, si valuta la $g(F, |s|)$, per poi moltiplicare per la probabilità infinitesima della variabile s; si ottiene infine la distribuzione di F integrando rispetto ad s.
Fissando s si ottiene:

$$\begin{aligned}
|s| &\rightarrow F(r) \\
F &= \frac{\nu_2}{\nu_1 s} r \\
r &= \frac{\nu_1 s}{\nu_2}
\end{aligned}$$

Tenendo conto per la costruzione di $g(F, |s|)$ che, essendo r variabile, si deve considerare ν_1 gradi di libertà, si procede con la determinazione di $g(F, |s|)$ come distribuzione derivata.

$$\begin{aligned}
g(F, |s|) &= f\left(\frac{\nu_1 s}{\nu_2}F\right)\frac{\partial\left[\frac{\nu_1 s}{\nu_2}F\right]}{\partial F} \\
&= \frac{1}{2^{\frac{\nu_1}{2}}\Gamma\left(\frac{\nu_1}{2}\right)}\left[\frac{\nu_1 s}{\nu_2}F\right]^{\frac{\nu_1-2}{2}} e^{-\frac{\frac{\nu_1 s}{\nu_2}F}{2}}\frac{\nu_1 s}{\nu_2} \\
&= \frac{1}{2^{\frac{\nu_1}{2}}\Gamma\left(\frac{\nu_1}{2}\right)}\left[\frac{\nu_1 s}{\nu_2}F\right]^{\frac{\nu_1-2}{2}} e^{-\frac{\nu_1 s F}{\nu_2 2}}\frac{\nu_1 s}{\nu_2} && (6.3) \\
g(F) &= \int_s g(F, |s|)dP(s) = \int_s g(F, |s|)f(s)ds && (6.4)
\end{aligned}$$

dove

$$f(s) = \frac{1}{2^{\frac{\nu_2}{2}}\Gamma\left(\frac{\nu_2}{2}\right)} s^{\frac{\nu_2-2}{2}} e^{-\frac{s}{2}} \tag{6.5}$$

Si valuta a questo punto $g(F)$ mediante 6.4 ricorrendo all'uso delle espressioni 6.3 e 6.5.

$$
\begin{aligned}
g(F) &= \int_s \frac{1}{2^{\frac{\nu_1}{2}}\Gamma\left(\frac{\nu_1}{2}\right)}\left[\frac{\nu_1 s}{\nu_2}F\right]^{\frac{\nu_1-2}{2}}\frac{\nu_1 s}{\nu_2}e^{-\frac{\nu_1 sF}{\nu_2 2}}\frac{\nu_1 s}{\nu_2}\frac{1}{2^{\frac{\nu_2}{2}}\Gamma\left(\frac{\nu_2}{2}\right)}s^{\frac{\nu_2-2}{2}}e^{-\frac{s}{2}}ds \\
&= \int_s \underbrace{\frac{1}{2^{\left(\frac{\nu_1}{2}+\frac{\nu_2}{2}\right)}\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\frac{\nu_1}{\nu_2}\left[\frac{\nu_1}{\nu_2}F\right]^{\frac{\nu_1-2}{2}}}_{C(\nu_1,\nu_2,F)} s^{\frac{\nu_1-2}{2}+\frac{\nu_2-2}{2}+1}e^{-\frac{s}{2}\left[\frac{\nu_1}{\nu_2}F+1\right]}ds \\
&= C(\nu_1,\nu_2,F)\int_0^{+\infty} s^{\frac{\nu_1+\nu_2-2}{2}}e^{-\frac{s}{2}\left[\frac{\nu_1}{\nu_2}F+1\right]}ds \qquad (6.6)
\end{aligned}
$$

dove $C(\nu_1,\nu_2,F)$ è una costante rispetto all'integrazione lungo s, e perciò portata fuori dall'integrale. Tenendo presente che da (4.4) si ha:

$$\Gamma(m+1) = \int_0^{+\infty} y^m e^{-y}dy$$

Si effettua un cambio di variabile ponendo:

$$
\begin{aligned}
\phi = \frac{s}{2}\left[\frac{\nu_1}{\nu_2}F+1\right] &\rightarrow s = \frac{2\phi}{\frac{\nu_1}{\nu_2}F+1} \\
d\phi = \left[\frac{\nu_1}{\nu_2}F+1\right]\frac{ds}{2} &\rightarrow ds = \frac{2d\phi}{\frac{\nu_1}{\nu_2}F+1} \\
\phi \rightarrow (0,+\infty) &\qquad s \rightarrow (0,+\infty)
\end{aligned}
$$

Effettuando la sostituzione si sviluppa $g(F)$:

$$
\begin{aligned}
g(F) &= C(\nu_1,\nu_2,F)\int_0^{+\infty}\left[\frac{2\phi}{\frac{\nu_1}{\nu_2}F+1}\right]^{\frac{\nu_1+\nu_2-2}{2}} e^{-\phi}\frac{2d\phi}{\frac{\nu_1}{\nu_2}F+1} \\
&= C(\nu_1,\nu_2,F)\left[\frac{2}{\frac{\nu_1}{\nu_2}F+1}\right]^{\frac{\nu_1+\nu_2-2}{2}}\frac{2}{\frac{\nu_1}{\nu_2}F+1}\int_0^{+\infty}\phi^{\frac{\nu_1+\nu_2-2}{2}}e^{-\phi}d\phi \\
&= \underbrace{C(\nu_1,\nu_2,F)\left[\frac{2}{\frac{\nu_1}{\nu_2}F+1}\right]^{\frac{\nu_1+\nu_2}{2}}}_{K(\nu_1,\nu_2,F)}\int_0^{+\infty}\phi^{\frac{\nu_1+\nu_2-2}{2}}e^{-\phi}d\phi \\
&= K(\nu_1,\nu_2,F)\int_0^{+\infty}\phi^{\frac{\nu_1+\nu_2-2}{2}}e^{-\phi}d\phi \qquad (6.7)
\end{aligned}
$$

Si considera quindi

$$m = \frac{\nu_1+\nu_2-2}{2} \rightarrow m+1 = \frac{\nu_1+\nu_2}{2}$$

per cui, essendo

$$\int_0^{+\infty} \phi^{\frac{\nu_1+\nu_2-2}{2}} e^{-\phi} d\phi = \Gamma\left(\frac{\nu_1+\nu_2}{2}\right)$$

è possibile riscrivere $g(F)$ utilizzando la Funzione Gamma e sviluppando la costante:

$$\begin{aligned}
g(F) &= K(\nu_1, \nu_2, F)\Gamma\left(\frac{\nu_1}{2}+\frac{\nu_2}{2}\right) \\
&= \frac{2^{\frac{\nu_1+\nu_2}{2}}}{\left[\frac{\nu_1}{\nu_2}F+1\right]^{\frac{\nu_1+\nu_2}{2}}} \frac{\Gamma\left(\frac{\nu_1}{2}+\frac{\nu_2}{2}\right)}{2^{\frac{\nu_1+\nu_2}{2}}\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \frac{\nu_1}{\nu_2}\left(\frac{\nu_1}{\nu_2}F\right)^{\frac{\nu_1-2}{2}} \\
&= \frac{\Gamma\left(\frac{\nu_1}{2}+\frac{\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \frac{\nu_2^{\frac{\nu_1+\nu_2}{2}}}{(\nu_1 F+\nu_2)^{\frac{\nu_1+\nu_2}{2}}} \frac{\nu_1}{\nu_2}\left(\frac{\nu_1}{\nu_2}F\right)^{\frac{\nu_1}{2}}\left(\frac{\nu_1}{\nu_2}\right)^{-1} F^{-1} \\
&= \frac{\Gamma\left(\frac{\nu_1}{2}+\frac{\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \frac{(\nu_1 F)^{\frac{\nu_1}{2}}\nu_2^{\frac{\nu_2}{2}}}{(\nu_1 F+\nu_2)^{\frac{\nu_1+\nu_2}{2}}} \frac{1}{F} \\
g(F) &= \frac{\Gamma\left(\frac{\nu_1}{2}+\frac{\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{(\nu_1 F)^{\nu_1}\nu_2^{\nu_2}}{(\nu_1 F+\nu_2)^{\nu_1+\nu_2}}\right)^{\frac{1}{2}} \frac{1}{F} \qquad (6.8)
\end{aligned}$$

6.3 Applicazioni della variabile F e test di Fisher

In molti esperimenti è necessario assumere una particolare ipotesi riguardo ai dati raccolti: per svolgere una media campionaria tra dati raccolti da sperimentatori differenti o attraverso procedure sperimentali diverse è necessario verificare che i dati raccolti provengano verosimilmente dalla medesima popolazione. Si deve quindi valutare se trattamenti differenti applicati ai campioni non alterino i risultati a livelli significativi (dell'1% o del 5%). Per far ciò si vanno a valutare le varianze dei campioni in modalità differenti, per poi confrontarle con un test, come verrà spiegato in seguito.
Si considerino n campioni ciascuno con numero N uguale di dati; per ogni campione si calcola varianza e valor medio ottenendo:

C_{m1}	C_{m2}	C_{m3}	...	C_{mn}	valori medi
S_1^2	S_2^2	S_3^2	...	S_n^2	varianze

Si considerano quindi le varianze complessive, calcolate in due modalità differenti:

- **Varianza ENTRO i gruppi**

$$S^2_{entro} = \frac{1}{n}\sum_{i=1}^{n} S_i^2 \qquad \text{media delle varianze}$$

- **Varianza TRA i gruppi** dove si calcola inizialmente la *varianza delle medie campionarie* $S^2_{C_m}$ utilizzando la $Media\ delle\ Medie$ C^m_m:

$$
\begin{aligned}
C^m_m &= \frac{1}{n}\sum_{i=1}^{n} C_{mi} \\
S^2_{C_m} &= \frac{\sum_{i=1}^{n}(C_{mi} - C^m_m)^2}{n_g - 1} \\
&= \frac{(C_{m1} - C^m_m)^2 + (C_{m2} - C^m_m)^2 + (C_{m3} - C^m_m)^2 + ... + (C_{mn} - C^m_m)^2}{n_g - 1}
\end{aligned}
$$

  da cui la varianza della popolazione è ottenuta mediante la dispersione delle medie secondo

$$S^2_{tra} = NS^2_{C_m} \qquad \text{infatti} \qquad S^2_{C_m} = \frac{S^2_{tra}}{N}$$

  Per il confronto si calcola la variabile di $Fisher$ che, in base alla definizione 6.1 e ai valori appena valutati, assumerà la forma:

$$F = \frac{S^2_{tra}}{S^2_{entro}} \qquad \text{con} \qquad \nu_{tra} = n_g - 1 \qquad \nu_{entro} = n_g(N-1)$$

È chiaro che, se i gruppi non appartengono alla stessa popolazione, F sarà sensibilmente diverso da 1.

Test di Fisher
Ricorrendo alla distribuzione di Fisher è possibile determinare se campioni differenti appartengono alla stessa popolazione attraverso il confronto tra due varianze, oppure se tra due ipotesi funzionali diverse relative agli stessi dati sperimentali è preferibile un'assunzione rispetto all'altra, ricorrendo al paragone dei due chi quadri. Per far ciò si determina il valore sperimentale della variabile F attraverso una delle due definizioni, ponendo a numeratore il valore maggiore; dunque F sarà necessariamente maggiore di 1. Successivamente si valutano i gradi di libertà delle due variabili poste a numeratore e denominatore (ν_{tra} e ν_{entro} oppure i gradi dei χ^2).
Si determina l'estremo destro della variabile F tale per cui la probabilità di ottenere un valore inferiore a F_{max} corrisponda al livello di confidenza stabilito (di norma 95% oppure 99%). La probabilità si ricava integrando la distribuzione di Fisher unicamente determinata se noti i gradi di libertà.

$$\int_0^{F_{max}} g(F)dF = 0.95 \text{ oppure } \int_0^{F_{max}} g(F)dF = 0.99$$

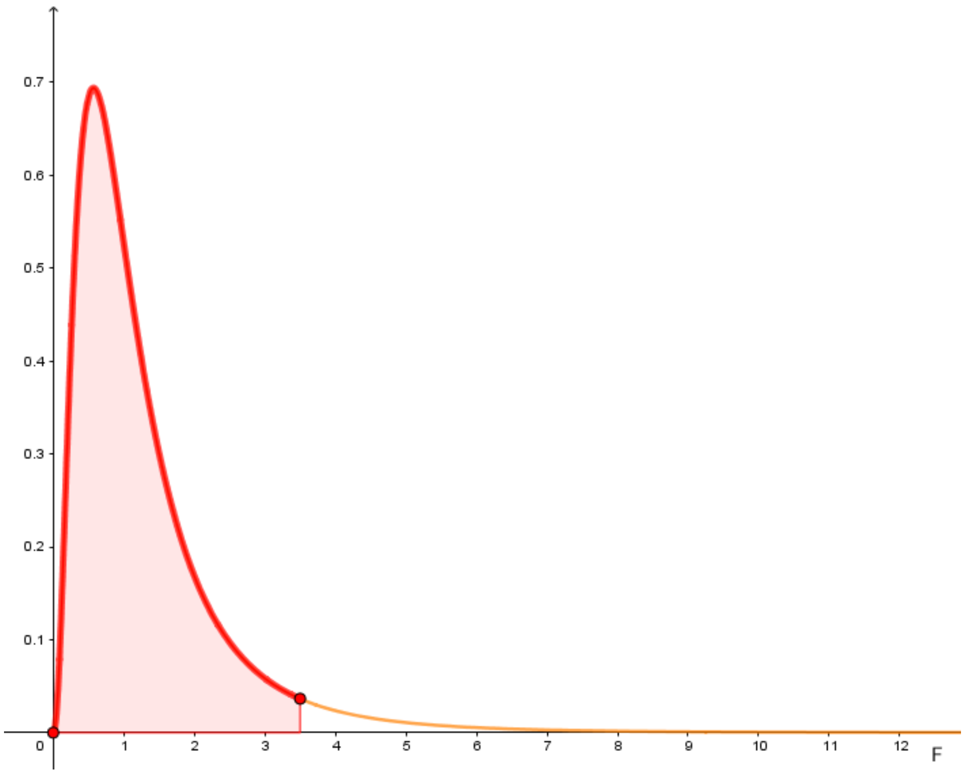

Figura 6.1: Distribuzione della variabile F per gradi di libertà pari a $\nu_1 = 7$ e $\nu_2 = 8$: in rosso è rappresentata la regione sottesa alla curva con area pari al 95%

La distribuzione di Fisher è stata tabulata e l'estremo F_{max} si ricava dall'apposita tabella in base ai gradi di libertà. Confrontando il valore sperimentale di F con quello massimo si possono stabilire le conlusioni del test:

- Se si sta svolgendo il test sulle varianze, si può sostenere che al livello di confidenza del 95% o del 99% (oppure, in modo equivalente, con un livello di significatività del 5% o dell'1%) i campioni studiati appartengono alla stessa popolazione se:

  $$F_{spe} < F_{max}$$

  In tal caso è possibile svolgere una media complessiva dei dati raccolti nei diversi campioni.

- Se si sta svolgendo il test sui chi quadri, nel caso in cui

  $$F_{spe} < F_{max}$$

si può sostenere che al livello di confidenza del 95% o del 99% la differenza tra i valori dei chi quadri non è così significativa da poter preferire l'ipotesi funzionale da cui deriva il chi quadro minore. Nel caso in cui il test dia riscontro negativo, allora è lecito sostenere che una delle due funzioni descriva meglio l'andamento dei dati rispetto all'altra.

6.4 Relazione tra le variabili t^2 e F

6.4.1 Verifica della coincidenza delle due variabili

É possibile valutare che, per determinate condizioni al contorno, si ha una coincidenza tra le variabili t e F. Prima di poter sviluppare i calcoli è necessario valutare alcune caratteristiche della variabile di Student. Noto da 5.1:

$$t = \frac{x_m - \mu}{S_{x_m}} = \frac{x_m - \mu}{\frac{S}{\sqrt{N}}} \tag{6.9}$$

Si dispone di **due** piccoli campioni di misure della stessa grandezza, con equivalente numero di dati N, con x_{m1} x_{m2} medie e σ_1 σ_2 deviazioni standard. Noto il valore atteso parti a:

$$E[x_{m1} - x_{m2}] = \mu = 0 \tag{6.10}$$

Per valutare la t di Student si studia quindi l'errore da porre a denominatore, per il confronto tra due medie campionarie:

$$\sigma^2_{(x_{m1}-x_{m2})} = \sigma^2_{x_{m1}} + \sigma^2_{x_{m2}} \tag{6.11}$$

Se i dati arrivano dalla stessa popolazione, si può stimare la varianza per entrambi i campioni:

$$\sigma^2 \simeq S^2 = \frac{\sum(x_i - x_{m1})^2 + \sum(x_i - x_{m2})^2}{(N-1)+(N-1)} = \frac{(N-1)(S_1^2 + S_2^2)}{2(N-1)} = \frac{S_1^2 + S_2^2}{2}$$

Applicando la definizione di t si ottiene:

$$\begin{aligned} t &= \frac{(x_{m1} - x_{m2}) - 0}{\sigma_{(x_{m1}-x_{m2})}} = \frac{x_{m1} - x_{m2}}{\sqrt{\frac{S^2}{N} + \frac{S^2}{N}}} = \frac{x_{m1} - x_{m2}}{\sqrt{\frac{2S^2}{N}}} = \frac{x_{m1} - x_{m2}}{\sqrt{\frac{2}{N}\frac{S_1^2+S_2^2}{2}}} \\ t &= \frac{x_{m1} - x_{m2}}{\sqrt{\frac{S_1^2+S_2^2}{N}}} \end{aligned} \tag{6.12}$$

Si considera a questo punto la variabile F nel caso di **due** soli gruppi. Si valuta quindi per questo caso specifico S^2_{tra} e S^2_{entro}:

x_{m1}	x_{m2}	siano le medie campionarie
S_1^2	S_2^2	siano le varianze

$$
\begin{aligned}
S^2_{entro} &= \frac{S_1^2 + S_2^2}{2} \\
S^2_{tra} &= \frac{N}{n_g - 1}[(x_{m1} - x^m_m)^2 + (x_{m2} - x^m_m)^2] \qquad \text{dove} \qquad x^m_m = \frac{x_{m1} + x_{m2}}{2} \\
&= \frac{N}{2-1}\left[\left(x_{m1} - \frac{x_{m1}}{2} - \frac{x_{m2}}{2}\right)^2 + \left(x_{m2} - \frac{x_{m1}}{2} - \frac{x_{m2}}{2}\right)^2\right] \\
&= N\left[\frac{1}{4}(x_{m1} - x_{m2})^2 + \frac{1}{4}(x_{m2} - x_{m1})^2\right] \\
&= \frac{N}{2}(x_{m1} - x_{m2})^2
\end{aligned}
$$

da cui si può valutare la F mediante 6.3:

$$
F = \frac{\frac{N}{2}(x_{m1} - x_{m2})^2}{\frac{S_1^2 + S_2^2}{2}} = \frac{(x_{m1} - x_{m2})^2}{\frac{S_1^2}{N} + \frac{S_2^2}{N}} = \left[\frac{x_{m1} - x_{m2}}{\sqrt{\frac{S_1^2}{N} + \frac{S_2^2}{N}}}\right]^2 = t^2 \tag{6.13}
$$

Confrontando le definizioni ottenute per la variabile di Student (6.12) e la variabile di Fisher, si nota come l'espressioni coincidano sotto le ipotesi imposte.

6.4.2 Verifica della coincidenza delle distribuzioni

La stessa proprietà può essere verificata mediante le distribuzioni, evidenziando la coincidenza tra le funzioni densità di probabilità per i parametri sopra citati e tenendo conto di come è espressa la variabile F.
Per la distribuzione 6.8 è necessario valutare a cosa corrispondono ν_1 e ν_2. Se si pensa a come è stata considerata la F in 6.3 allora è facile dedurre che (tenendo conto che resta sempre $n_g = 2$):

$$
\nu_1 = \nu_{tra} = n_g - 1 = 1
$$

$$
\nu_2 = \nu_{entro} = 2(N - 1) = n
$$

Il processo utilizzato è di questo tipo: inizialmente si valuta il valore della distriubuzione della variabile t^2 ed in seguito quello della F, per verificare poi che le distribuzioni ottenute coincidono.

1. Noto il valore della $h(t)$ da (5.8), ponendo $t^2 = y$ si nota che la funzione non è monotona crescente; è necessario dividerla in due funzioni monotone, una crescente e una decrescente per poter utilizzare 1.1 e 1.2 e sommare le due funzioni ottenute per poter avere la distribuzione derivata di t^2.

$$
t^2 = y \qquad \to t = \pm\sqrt{y}
$$

   - ψ monotona crescente:

$$
\zeta(y) = h(\sqrt{y})\frac{d\sqrt{y}}{dy} = \frac{1}{2\sqrt{y}}h(\sqrt{y}) \tag{6.14}
$$

- ψ monotona decrescente:

$$\zeta(y) = -h(-\sqrt{y})\frac{d(-\sqrt{y})}{dy} = \frac{1}{2\sqrt{y}}h(-\sqrt{y}) = \frac{1}{2\sqrt{y}}h(\sqrt{y}) \quad (6.15)$$

da cui si ricava che:

$$\zeta(y) = 2h(\sqrt{y})\frac{1}{2\sqrt{y}} = \frac{h(\sqrt{y})}{\sqrt{y}} \quad (6.16)$$

E quindi, noto il valore della $h(\sqrt{y})$ si ricava la distribuzione della variabile t^2:

$$\zeta(y) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}\sqrt{y}}\left(1+\frac{y}{n}\right)^{-\left(\frac{n+1}{2}\right)} \equiv \zeta(t^2) \quad (6.17)$$

2. È possibile a questo punto valutare la distribuzione della F, considerando $\nu_1 = \nu_{tra}$ e $\nu_2 = \nu_{entro}$ ricorrendo alle proprietà (4.4 4.8) delle funzioni Gamma:

$$\begin{aligned}
g(F) &= \frac{\Gamma\left(\frac{1}{2}+\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}\frac{1}{F}\left[\frac{Fn^n}{(F+n)^{1+n}}\right]^{\frac{1}{2}} \\
&= \frac{\Gamma\left(\frac{1}{2}+\frac{n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)}\frac{1}{F^{\frac{1}{2}}}\left[\frac{n^{n+1}n^{-1}}{(F+n)^{1+n}}\right]^{\frac{1}{2}} \\
&= \frac{\Gamma\left(\frac{1}{2}+\frac{n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)}\frac{1}{\sqrt{F}}n^{-\frac{1}{2}}\left[\frac{n^{n+1}}{(F+n)^{n+1}}\right]^{\frac{1}{2}} \\
&= \frac{\Gamma\left(\frac{1}{2}+\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}\sqrt{F}}\left[\frac{n}{F+n}\right]^{\frac{n+1}{2}} \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}\sqrt{F}}\left[\frac{F+n}{n}\right]^{-\frac{n+1}{2}} \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}\sqrt{F}}\left[1+\frac{F}{n}\right]^{-\frac{n+1}{2}} \qquad (6.18) \\
&= \zeta(t^2)|_{t^2=F}
\end{aligned}$$

Si nota che 6.17 e 6.18 sono la medesima espressione, da cui si deduce che per due gruppi, sotto le particolari condizioni precedentemente enunciate, la distribuzione di Fisher coincide con quella della variabile t^2.

La medesima proprietà è stata mostrata sia confrontando le definizioni delle variabili aleatorie sia studiando le distribuzioni funzionali.

Bibliografia

[1] Gaetano Cannelli, *Metodologie sperimentali in Fisica – Introduzione al metodo scientifico*, EdiSES, 2010.

[2] Appunti delle *lezioni di Statistica* tenute dal prof. Mario Edoardo Bertaina durante il corso di Esperimentazioni I (A.A. 2017-2018).