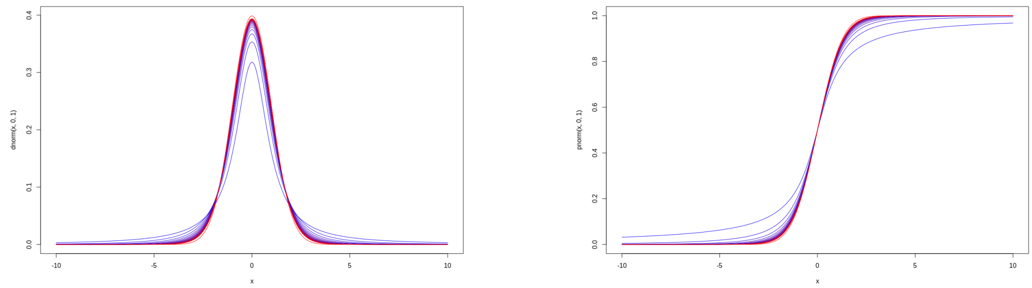

Figura 3.5: Si può dimostrare che se $(X_n)_{n\in\mathbb{N}}$ è una successione di v.a. $\mathbb{P}_{X_n} = t(n)$, allora X_n converge in legge ad una v.a. gaussiana standard. In figura le densità e le leggi delle distribuzioni $t(n)$, $n = 1, \dots, 20$ e della distribuzione gaussiana standard

3.4 Quantili

3.4.1 Distribuzione gaussiana standard

Sia Z una v.a. con distribuzione gaussiana standard. Poiché la densità associata

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right), \qquad \forall x \in \mathbb{R},$$

è strettamente positiva, abbiamo che la legge associata $\Phi(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x$ è una funzione strettamente crescente a valori in $(0, 1)$. Dunque $\Phi\colon \mathbb{R} \to (0, 1)$ è invertibile:

$$\forall \alpha \in (0, 1) \exists! z = z_\alpha \in \mathbb{R}\colon \Phi(z_\alpha) = \alpha.$$

Il numero reale z_α si dice *quantile di livello* α della distribuzione gaussiana standard. Osserviamo che poiché la densità $f(x)$ è una funzione pari, la legge Φ soddisfa la relazione

$$\Phi(-t) = 1 - \Phi(t) \qquad \forall t \in \mathbb{R}.$$

Di conseguenza $z_{1-\alpha} = -z_\alpha$. In particolare $z_{1/2} = 0$, $z_\alpha > 0$ se e solo se $\alpha > \dfrac{1}{2}$.

3.4.2 Distribuzione di Pearson a n gradi di libertà

Sia V una v.a. con distribuzione $\mathbb{P}_V = \chi^2_n$. Poiché la densità associata

$$f(x) = \begin{cases} \dfrac{x^{\frac{n}{2}-1} \exp\left(\frac{-x}{2}\right)}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} & x > 0, \\ 0 & x \le 0 \end{cases}$$

è nulla sulla semiretta $(-\infty, 0]$ e strettamente positiva sulla semiretta $(0, +\infty)$, abbiamo che la legge associata $F_V(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x$ è identicamente nulla sulla semiretta $(-\infty, 0]$ mentre la sua restrizione alla semiretta $(0, +\infty)$ è una funzione strettamente crescente a valori in $(0, 1)$. Dunque $F_V\colon (0, +\infty) \to (0, 1)$ è invertibile:

$$\forall \alpha \in (0, 1) \exists! t = \chi^2_{n,\alpha} \in \mathbb{R}\colon F_V(\chi^2_{n,\alpha}) = \alpha.$$

Il numero reale $\chi^2_{n,\alpha}$ si dice *quantile di livello* α della distribuzione χ^2_n.

3.4.3 Distribuzione t di Student a n gradi di libertà

Sia T una v.a. con distribuzione $t(n)$. Poiché la densità associata

$$\tau_n(x) = \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{\frac{-(n+1)}{2}} \qquad x \in \mathbb{R}.$$

è strettamente positiva, abbiamo che la legge associata $F_T(t) = \int_{-\infty}^{t} \tau_n(x)\mathrm{d}x$ è una funzione strettamente crescente a valori in $(0,1)$. Dunque $F_T \colon \mathbb{R} \to (0,1)$ è invertibile:

$$\forall \alpha \in (0,1) \exists! t = t_{n,\alpha} \in \mathbb{R} \colon F_T(t_{n,\alpha}) = \alpha.$$

Il numero reale $t_{n,\alpha}$ si dice *quantile di livello* α della distribuzione t di Student a n gradi di libertà. Osserviamo che poiché la densità $\tau_n(x)$ è una funzione pari, la legge F_T soddisfa la relazione

$$F_T(-t) = 1 - F_T(t) \qquad \forall t \in \mathbb{R}.$$

Di conseguenza $t_{n,1-\alpha} = -t_{n,\alpha}$. In particolare $t_{n,1/2} = 0$, $t_{n,\alpha} > 0$ se e solo se $\alpha > \dfrac{1}{2}$.

4. Stimatori di massima versosimiglianza

Sia $X_1, \ \ldots, \ X_n$ un campione statistico e sia $Y = \varphi\left(X_1, \ \ldots, \ X_n\right)$ una sua statistica. Se Y ha lo scopo di stimare un parametro θ della distribuzione del campione, diciamo che Y è uno *stimatore del parametro θ.*

Supponiamo di conoscere la distribuzione del campione a meno di un parametro θ e supponiamo che tale distribuzione sia discreta o assolutamente continua e dunque dotata di densità (discreta o meno). Tale densità dipenderà dal parametro θ e la indico col simbolo $g(x|\theta)$. La densità congiunta si indica col simbolo $f(x_1, \ \ldots, \ x_n|\theta)$ e sappiamo che, grazie all'indipendenza delle v.a. che costituiscono il campione, si ha

$$f(x_1, \ \ldots, \ x_n|\theta) = g(x_1|\theta) \cdot \ldots \cdot g(x_n|\theta) = \prod_{i=1}^{n} g(x_i|\theta).$$

Interpreto $f(x_1, \ \ldots, \ x_n|\theta)$ come la *plausibilità* che la n-upla $x_1, \ \ldots, \ x_n$ si realizzi nel campione empirico quando il parametro incognito prende il valore θ. Consideriamo infatti i due casi, discreto e assolutamente continuo

- Se il campione ha distribuzione discreta, allora

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) = \prod_{i=1}^{n} g(x_i|\theta).$$

- Se il campione ha distribuzione assolutamente continua, e se la densità g è continua, allora, per $\delta > 0$ e sufficientemente piccolo si ha

$$\begin{aligned}\mathbb{P}\left(|X_1 - x_1| < \frac{\delta}{2}, \ldots, |X_n - x_n| < \frac{\delta}{2}\right) &= \prod_{i=1}^{n} \mathbb{P}\left(|X_i - x_i| < \frac{\delta}{2}\right) \\ &= \prod_{i=1}^{n} \mathbb{P}\left(X_i \in \left(x_i - \frac{\delta}{2}, x_i + \frac{\delta}{2}\right)\right) \simeq \prod_{i=1}^{n} \left(g\left(x_i|\theta\right)\delta\right)^n \\ &= \delta^n \prod_{i=1}^{n} g\left(x_i|\theta\right) = \delta^n f(x_1, \ \ldots, \ x_n|\theta)\end{aligned}$$

La funzione f, vista come funzione di θ, si dice *funzione di verosimiglianza.*

Dunque: dato il campione empirico $x_1, \ \ldots, \ x_n$, cerco $\widehat{\theta} = \widehat{\theta}(x_1, \ \ldots, \ x_n)$ che massimizza la funzione $f(x_1, \ \ldots, \ x_n|\theta)$. La statistica $\widehat{\theta} = \widehat{\theta}(X_1, \ \ldots, \ X_n)$ si dirà *stimatore di massima verosimiglianza del parametro θ.*

Osservazione 4.0.1. Poiché la funzione $\ln\colon (0, +\infty) \to \mathbb{R}$ è strettamente monotona crescente, massimizzare $f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^n g(x_i|\theta)$ equivale a massimizzare la funzione

$$h(x_1, \ldots, x_n|\theta) := \ln f(x_1, \ldots, x_n|\theta) = \sum_{i=1}^n \ln g(x_i|\theta)$$

Si ha

$$\frac{\partial}{\partial\theta} h(x_1, \ldots, x_n|\theta) = \frac{\partial}{\partial\theta} \sum_{i=1}^n \ln g(x_i|\theta) = \sum_{i=1}^n \frac{\partial}{\partial\theta} \ln g(x_i|\theta) = \sum_{i=1}^n \frac{1}{g(x_i|\theta)} \frac{\partial g(x_i|\theta)}{\partial\theta}$$

4.1 Distribuzione di Bernoulli

In questo caso la distribuzione dipende dal solo parametro $p = \mathbb{P}(X_i = 1) = \mathbb{E}[X_i]$.
Sia dunque $X_1, \ldots, X_n$ un campione statistico di Bernoulli di parametro incognito $p \in [0, 1]$. Realizzo n prove di Bernoulli e ottengo il campione empirico $x_1, \ldots, x_n$, $x_i \in \{0, 1\}$. Sia $k = k(x_1, \ldots, x_n) := \sum_{i=1}^n x_i$. Abbiamo

$$f(x_1, \ldots, x_n|p) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = p^k(1-p)^{n-k},$$
$$h(x_1, \ldots, x_n|p) = \ln(p^k(1-p)^{n-k}) = k \ln p + (n-k)\ln(1-p).$$
$$\frac{\partial h}{\partial p} = \frac{k}{p} - \frac{n-k}{1-p} = \frac{k - np}{p(1-p)} \gtreqless 0 \iff k - np \gtreqless 0 \iff p \lesseqgtr \frac{k}{n}.$$

Poiché $k = \sum_{i=1}^n x_i$, lo stimatore di massima verosimiglianza per il parametro p è $\frac{\sum_{i=1}^n X_i}{n}$ cioè la media campionaria $\overline{X}$. Ricordiamo che la media campionaria è uno stimatore corretto di $\mathbb{E}[X_i] = p$.

4.2 Distribuzione di Poisson

La distribuzione di Poisson è concentrata sugli interi nonnegativi e dipende da un solo parametro:

$$g(x|\lambda) = \mathrm{e}^{-\lambda}\frac{\lambda^x}{x!}, \qquad x = 0, 1, \ldots.$$

Dunque

$$f(x_1, \ldots, x_n|\lambda) = \prod_{i=1}^n \left(\mathrm{e}^{-\lambda}\frac{\lambda^{x_i}}{x_i!}\right),$$

$$\begin{aligned} h(x_1, \ldots, x_n|\lambda) &= \ln f(x_1, \ldots, x_n|\lambda) = \sum_{i=1}^n \ln\left(\mathrm{e}^{-\lambda}\frac{\lambda^{x_i}}{x_i!}\right) \\ &= \sum_{i=1}^n (-\lambda + x_i \ln(\lambda) - \ln(x_i!)) = -n\lambda + n\overline{x}\ln(\lambda) - \sum_{i=1}^n \ln(x_i!) \end{aligned}$$

Da cui

$$\frac{\partial}{\partial\lambda} h(x_1, \ldots, x_n|\lambda) = n\left(-1 + \frac{\overline{x}}{\lambda}\right) \gtreqless 0 \iff \lambda \lesseqgtr \overline{x}.$$

Quindi anche in questo caso lo stimatore di massima verosimiglianza per il parametro λ è la media campionaria $\overline{X}$ (che è uno stimatore corretto).

4.3 Distribuzione gaussiana

In questo caso la densità dipende da due parametri, $\mu \in \mathbb{R}$ e $\sigma > 0$:

$$\begin{aligned} f(x_1,\ \ldots,\ x_n|\mu,\sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x_i-\mu)^2}{2\sigma^2}\right) \\ &= (2\pi)^{\frac{-n}{2}}(\sigma)^{-n} \exp\left(\frac{-\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right) \end{aligned}$$

cosicché

$$\begin{aligned} h(x_1,\ \ldots,\ x_n|\mu,\sigma) = \ln f(x_1,\ \ldots,\ x_n|\mu,\sigma) &= -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\ &= -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2. \end{aligned}$$

Si ha quindi

$$\begin{aligned} \frac{\partial}{\partial\mu} h(x_1,\ \ldots,\ x_n|\mu,\sigma) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = n\left(\overline{x}-\mu\right), \\ \frac{\partial}{\partial\sigma} h(x_1,\ \ldots,\ x_n|\mu,\sigma) &= \frac{-n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2 = \frac{1}{\sigma^3}\left(-n\sigma^2 + \sum_{i=1}^{n}(x_i-\mu)^2\right). \end{aligned}$$

Dunque le due derivate parziali si annullano contemporaneamente se e solo se

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

Dunque la media campionaria $\overline{X}$ è uno stimatore di massima verosimiglianza (ed è uno stimatore corretto) per il valore atteso μ mentre $\dfrac{n-1}{n}S^2$ è uno stimatore di massima verosimiglianza per la varianza σ^2.

4.4 Distribuzione uniforme su un intervallo

Se (a,b) è l'intervallo, allora la densità del campione è

$$g(x|a,b) = \begin{cases} \frac{1}{b-a} & x \in [a,b], \\ 0 & \text{altrimenti} \end{cases}$$

da cui

$$f(x_1,\ \ldots,\ x_n|a,b) = \begin{cases} \frac{1}{(b-a)^n} & x_i \in [a,b] \quad \forall i = 1,\ldots,n, \\ 0 & \text{altrimenti.} \end{cases}$$

Devo massimizzare $\frac{1}{(b-a)^n}$ con il vincolo $a \le x_i \le b$ per ogni $i = 1,\ldots,n$. Devo dunque minimizzare la lunghezza dell'intervallo $b-a$ con il vincolo $a \le x_i \le b$ per ogni $i = 1,\ldots,n$. È dunque

$$a = \min\{x_1,\ \ldots,\ x_n\}, \qquad b = \max\{x_1,\ \ldots,\ x_n\}.$$

Dunque

$$\min\{X_1,\ \dots,\ X_n\}\,, \qquad \max\{X_1,\ \dots,\ X_n\}$$

sono stimatori di massima verosimiglianza rispettivamente per l'estremo inferiore e per l'estremo superiore dell'intervallo.

5. Intervalli di confidenza

La media campionaria e la varianza campionaria ci offrono una stima dei parametri valore atteso e varianza del campione statistico in esame. Abbiamo però bisogno di sapere *quanto ci si possa fidare di questa stima* ovvero quale sia la probabilità che il *vero* valore del parametro incognito non sia *troppo distante* dalla stima trovata.

Diamo perciò la seguente definizione:

Definizione 5.0.1 (Intervallo di confidenza)**.** Sia $X_1, \ldots, X_n$ un campione statistico e sia θ un parametro (ignoto) che caratterizza la distribuzione del campione.
Siano $L_i = l_i(X_1, \ldots, X_n)$ e $L_s = l_s(X_1, \ldots, X_n)$ due statistiche del campione e sia $\alpha \in (0,1)$.

- Dico che l'intervallo (L_i, L_s) è un *intervallo di confidenza* (o di fiducia) di livello $1-\alpha$ se $\mathbb{P}(\theta \in (L_i, L_s)) \geq 1-\alpha$, ovvero che (L_i, L_s) è un intervallo di confidenza (o di fiducia) di errore α se $\mathbb{P}(\theta \notin (L_i, L_s)) \leq \alpha$.
- Dico che la semiretta $(L_i, +\infty)$ è un *intervallo di confidenza unilaterale superiore* di livello $1-\alpha$ se $\mathbb{P}(\theta > L_i) \geq 1-\alpha$
- Dico che la semiretta $(-\infty, L_s)$ è un *intervallo di confidenza unilaterale inferiore* di livello $1-\alpha$ se $\mathbb{P}(\theta < L_s) \geq 1-\alpha$

Osservazione 5.0.1.

1. La scelta dei nomi delle due statistiche non è casuale: L_i sta per limitazione inferiore mentre L_s sta per limitazione superiore.
2. Di solito si è interessati a *piccoli* valori di α, più precisamente a $\alpha \in (10^{-2}, 10^{-1})$.
3. La disuguaglianza di Chebychev ci ha fornito un intervallo di confidenza per il valore atteso μ del campione nel caso in cui la varianza σ^2 sia nota

$$\mathbb{P}\left(\left|\overline{X} - \mu\right| \geq t\right) \leq \frac{\sigma^2}{n\,t^2} \qquad \forall t > 0$$

ovvero

$$\mathbb{P}\left(\left|\overline{X} - \mu\right| < t\right) \geq 1 - \frac{\sigma^2}{n\,t^2} \qquad \forall t > 0$$

cioè

$$\mathbb{P}\left(\overline{X} - t < \mu < \overline{X} + t\right) \geq 1 - \frac{\sigma^2}{n\,t^2} \qquad \forall t > 0.$$

Fissato $\alpha \in (0,1)$ scelgo $t = \dfrac{\sigma}{\sqrt{n\,\alpha}}$. La disuguaglianza di Chebychev si legge allora

$$\mathbb{P}\left(\overline{X} - \frac{\sigma}{\sqrt{n\,\alpha}} < \mu < \overline{X} + \frac{\sigma}{\sqrt{n\,\alpha}}\right) \geq 1 - \alpha \qquad \forall \alpha \in (0,1).$$

Dunque l'intervallo $\left(\overline{X} - \frac{\sigma}{\sqrt{\alpha}}, \overline{X} + \frac{\sigma}{\sqrt{\alpha}}\right)$ è un intervallo di confidenza di livello $1-\alpha$ per il valore atteso μ del campione.

5.1 Stima per intervalli del valore atteso di campioni esponenziali

Sia $X_1, \ldots, X_n$ un campione statistico con distribuzione $\mathbb{P}_{X_i} = \exp(\lambda)$. Cioè $\mathbb{P}_{X_i} = f(x)\mathrm{d}x$ con

$$f(x) = \begin{cases} 0 & x \le 0, \\ \lambda \mathrm{e}^{-\lambda x} & x > 0. \end{cases}$$

Abbiamo già osservato che $\exp(\lambda) = \Gamma(1,\lambda)$, vedi Osservazione 3.3.3. Di conseguenza, vedi Lemma 3.3.1, $S_n := \sum_{i=1}^n X_i$ ha distrubuzione $\Gamma(n,\lambda)$ (detta anche *distribuzione di Erlangen di parametri n e λ*) cioè $\mathbb{P}_{S_n} = g(x)\mathrm{d}x$ con

$$g(x) = \begin{cases} 0 & x \le 0, \\ \dfrac{\lambda^n x^{n-1}}{\Gamma(n)} \mathrm{e}^{-\lambda x} & x > 0, \end{cases} = \begin{cases} 0 & x \le 0, \\ \dfrac{\lambda^n x^{n-1}}{(n-1)!} \mathrm{e}^{-\lambda x} & x > 0. \end{cases}$$

Consideriamo ora la v.a. $Y := 2\lambda S_n$. Sappiamo che $\mathbb{P}_Y = h(x)\mathrm{d}x$ con

$$h(x) = \frac{1}{2\lambda} g\left(\frac{x}{2\lambda}\right) = \begin{cases} 0 & x \le 0, \\ \dfrac{x^{n-1}}{2^n \Gamma(n)} \mathrm{e}^{-\frac{x}{2}} & x > 0. \end{cases}$$

Ovvero, abbiamo provato che $\mathbb{P}_Y = \chi^2_{2n}$.

Per ogni $\alpha \in (0,1)$ abbiamo dunque

$$\begin{aligned} 1-\alpha = \mathbb{P}\left(\chi^2_{\frac{\alpha}{2},2n} < Y < \chi^2_{1-\frac{\alpha}{2},2n}\right) &= \mathbb{P}\left(\chi^2_{\frac{\alpha}{2},2n} < 2\lambda n \overline{X} < \chi^2_{1-\frac{\alpha}{2},2n}\right) \\ &= \mathbb{P}\left(\frac{1}{\chi^2_{1-\frac{\alpha}{2},2n}} < \frac{1}{2\lambda n\overline{X}} < \frac{1}{\chi^2_{\frac{\alpha}{2},2n}}\right) = \mathbb{P}\left(\frac{2n\overline{X}}{\chi^2_{1-\frac{\alpha}{2},2n}} < \frac{1}{\lambda} < \frac{2n\overline{X}}{\chi^2_{\frac{\alpha}{2},2n}}\right) \end{aligned}$$

e dunque abbiamo l'intervallo di confidenza di livello $1-\alpha$ per il valore atteso della distribuzione

$$(L_i, L_s) = \left(\frac{2n\overline{X}}{\chi^2_{1-\frac{\alpha}{2},2n}}, \frac{2n\overline{X}}{\chi^2_{\frac{\alpha}{2},2n}}\right)$$

5.2 Stima per intervalli del valore atteso di campioni gaussiani

5.2.1 Campione gaussiano di cui è nota la varianza

Intervallo bilaterale

Sia $X_1, \ldots, X_n$ un campione gaussiano di valore atteso μ incognita e varianza σ^2 nota.

Sia Z una v.a. gaussiana standard, sia $\alpha \in (0,1)$ e sia $z_{1-\frac{\alpha}{2}}$ il quantile della gaussiana standard di livello $1-\frac{\alpha}{2}$: $\Phi\left(z_{1-\frac{\alpha}{2}}\right) = 1-\frac{\alpha}{2}$. Calcolo $\mathbb{P}\left(|Z| \leq z_{1-\frac{\alpha}{2}}\right)$:

$$\begin{aligned}
&\mathbb{P}\left(|Z| \leq z_{1-\frac{\alpha}{2}}\right) = \mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}\right) \\
&= \mathbb{P}\left(Z \leq z_{1-\frac{\alpha}{2}}\right) - \mathbb{P}\left(Z \leq -z_{1-\frac{\alpha}{2}}\right) = \mathbb{P}\left(Z \leq z_{1-\frac{\alpha}{2}}\right) - \mathbb{P}\left(Z \leq z_{\frac{\alpha}{2}}\right) \\
&= \Phi\left(z_{1-\frac{\alpha}{2}}\right) - \Phi\left(z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1-\alpha.
\end{aligned} \tag{5.1}$$

Sappiamo che $\mathbb{P}_{\overline{X}} = N\left(\mu, \frac{\sigma^2}{n}\right)$ e che dunque $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ ha distribuzione $N(0,1)$. Applichiamo quindi la disuguaglianza (5.1) a $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$. Si ha:

$$\begin{aligned}
1-\alpha &= \mathbb{P}\left(\frac{\left|\overline{X}-\mu\right|}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = \mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\mu-\overline{X}}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\frac{\alpha}{2}}\right) \\
&= \mathbb{P}\left(\frac{-\sigma\, z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leq \mu - \overline{X} \leq \frac{\sigma\, z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(\overline{X} - \frac{\sigma\, z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leq \mu \leq \overline{X} + \frac{\sigma\, z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right)
\end{aligned}$$

L'intervallo

$$(L_i, L_s) = \left(\overline{X} - \frac{\sigma\, z_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \overline{X} + \frac{\sigma\, z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right)$$

è dunque un intervallo di confidenza di livello $1-\alpha$ per il valore atteso μ del campione.

Osservazione 5.2.1 (Dimensionamento del campione)**.** Fissato il livello di confidenza $1-\alpha$, supponiamo di voler controllare l'ampiezza dell'intervallo di confidenza $L_s - L_i$. Nel caso in esame l'ampiezza dell'intervallo di confidenza è $\frac{2\,\sigma\, z_{1-\frac{\alpha}{2}}}{\sqrt{n}}$. Se fissiamo una limitazione superiore $2\,\delta$ per l'ampiezza di tale intervallo, deve dunque essere

$$\frac{2\,\sigma\, z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leq 2\,\delta$$

ovvero

$$n \geq \left(\frac{\sigma\, z_{1-\frac{\alpha}{2}}}{\delta}\right)^2.$$

Intervallo unilaterale superiore

Sia Z una v.a. tale che $\mathbb{P}_Z = N(0,1)$. Sappiamo che

$$\mathbb{P}(Z \leq t) = 1-\alpha \quad \text{se e solo se} \quad t = z_{1-\alpha}.$$

Abbiamo dunque

$$1-\alpha = \mathbb{P}\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\alpha}\right) = \mathbb{P}\left(\overline{X}-\mu \leq \frac{\sigma\, z_{1-\alpha}}{\sqrt{n}}\right) = \mathbb{P}\left(\mu \geq \overline{X} - \frac{\sigma\, z_{1-\alpha}}{\sqrt{n}}\right).$$

Quindi la semiretta

$$(L_i, +\infty) = \left(\overline{X} - \frac{\sigma\, z_{1-\alpha}}{\sqrt{n}}, +\infty\right)$$

è un intervallo di confidenza unilaterale superiore di livello $1 - \alpha$.

Intervallo unilaterale inferiore

Sia Z una v.a. tale che $\mathbb{P}_Z = N(0,1)$. Sappiamo che

$$\mathbb{P}(Z \geq t) = 1 - \alpha \quad \text{se e solo se} \quad \mathbb{P}(Z \leq t) = \alpha \quad \text{se e solo se} \quad t = z_\alpha.$$

Abbiamo dunque

$$1 - \alpha = \mathbb{P}\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq z_\alpha\right) = \mathbb{P}\left(\overline{X} - \mu \geq \frac{\sigma\, z_\alpha}{\sqrt{n}}\right) = \mathbb{P}\left(\mu \leq \overline{X} - \frac{\sigma\, z_\alpha}{\sqrt{n}}\right).$$

Quindi la semiretta

$$(-\infty, L_s) = \left(-\infty, \overline{X} - \frac{\sigma\, z_\alpha}{\sqrt{n}}\right) = \left(-\infty, \overline{X} + \frac{\sigma\, z_{1-\alpha}}{\sqrt{n}}\right)$$

è un intervallo di confidenza unilaterale inferiore di livello $1 - \alpha$.

5.2.2 Campione gaussiano di cui non è nota la varianza

Intervallo bilaterale

Sia $X_1, \ldots, X_n$ un campione gaussiano di valore atteso μ varianza σ^2, entrambi incogniti.

Sappiamo che la v.a. $T := \dfrac{(\overline{X} - \mu)\sqrt{n}}{S}$ segue la distribuzione t di Student con $n - 1$ gradi di libertà:

$$\mathbb{P}_T = t(n-1).$$

Sia $t_{n-1,1-\frac{\alpha}{2}}$ il relativo quantile di livello $1 - \frac{\alpha}{2}$:

$$\mathbb{P}\left(T \leq t_{n-1,1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}.$$

Calcolo $\mathbb{P}\left(|T| \leq t_{n-1,1-\frac{\alpha}{2}}\right)$:

$$\begin{aligned}
\mathbb{P}\left(|T| \leq t_{n-1,1-\frac{\alpha}{2}}\right) &= \mathbb{P}\left(-t_{n-1,1-\frac{\alpha}{2}} \leq T \leq t_{n-1,1-\frac{\alpha}{2}}\right) \\
&= \mathbb{P}\left(T \leq t_{n-1,1-\frac{\alpha}{2}}\right) - \mathbb{P}\left(T \leq -t_{n-1,1-\frac{\alpha}{2}}\right) \\
&= \mathbb{P}\left(T \leq t_{n-1,1-\frac{\alpha}{2}}\right) - \mathbb{P}\left(T \leq t_{n-1,\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha.
\end{aligned}$$

Abbiamo dunque

$$
\begin{aligned}
1-\alpha = \mathbb{P}\left(|T| \le t_{n-1,1-\frac{\alpha}{2}}\right) &= \mathbb{P}\left(\frac{\left|\overline{X}-\mu\right|\sqrt{n}}{S} \le t_{n-1,1-\frac{\alpha}{2}}\right)\\
&= \mathbb{P}\left(\left|\overline{X}-\mu\right| \le \frac{S}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}\right)\\
&= \mathbb{P}\left(\frac{-S}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}} \le \mu - \overline{X} \le \frac{S}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}\right)\\
&= \mathbb{P}\left(\overline{X} - \frac{S}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}} \le \mu \le \overline{X} + \frac{S}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}\right)
\end{aligned}
$$

L'intervallo

$$
(L_i, L_s) = \left(\overline{X} - \frac{S\, t_{n-1,1-\frac{\alpha}{2}}}{\sqrt{n}}, \overline{X} + \frac{S\, t_{n-1,1-\frac{\alpha}{2}}}{\sqrt{n}}\right)
$$

è dunque un intervallo di confidenza di livello $1-\alpha$ per il valore atteso μ del campione.

Intervallo unilaterale superiore

Sappiamo che

$$
\mathbb{P}\left(T \le t\right) = 1-\alpha \quad \text{se e solo se} \quad t = t_{n-1,1-\alpha}.
$$

Abbiamo dunque

$$
\begin{aligned}
1-\alpha = \mathbb{P}\left(\frac{(\overline{X}-\mu)\sqrt{n}}{S} \le t_{n-1,1-\alpha}\right) &= \mathbb{P}\left(\overline{X}-\mu \le \frac{S\, t_{n-1,1-\alpha}}{\sqrt{n}}\right)\\
= \mathbb{P}\left(\mu \ge \overline{X} - \frac{S\, t_{n-1,1-\alpha}}{\sqrt{n}}\right).
\end{aligned}
$$

Quindi la semiretta

$$
(L_i, +\infty) = \left(\overline{X} - \frac{S\, t_{n-1,1-\alpha}}{\sqrt{n}}, +\infty\right)
$$

è un intervallo di confidenza unilaterale superiore di livello $1-\alpha$.

Intervallo unilaterale inferiore

Sappiamo che

$$
\mathbb{P}\left(T \ge t\right) = 1-\alpha \quad \text{se e solo se} \quad \mathbb{P}\left(T \le t\right) = \alpha \quad \text{se e solo se} \quad t = t_{n-1,\alpha}.
$$

Abbiamo dunque

$$
1-\alpha = \mathbb{P}\left(\frac{(\overline{X}-\mu)\sqrt{n}}{S} \ge t_{n-1,\alpha}\right) = \mathbb{P}\left(\overline{X}-\mu \ge \frac{S\, t_{n-1,\alpha}}{\sqrt{n}}\right) = \mathbb{P}\left(\mu \le \overline{X} - \frac{S\, t_{n-1,\alpha}}{\sqrt{n}}\right).
$$

Quindi la semiretta

$$
(-\infty, L_s) = \left(-\infty, \overline{X} - \frac{S\, t_{n-1,\alpha}}{\sqrt{n}}\right) = \left(-\infty, \overline{X} + \frac{S\, t_{n-1,1-\alpha}}{\sqrt{n}}\right)
$$

è un intervallo di confidenza unilaterale inferiore di livello $1-\alpha$.