Distribuzione Normale
Premesse

1 Calcoliamo $I = \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}\, dt$.

Calcoliamo $\iint_{R^2} e^{-\frac{x^2+y^2}{2}} dx\, dy$ con un cambiamento di coordinate

(1)
$$\iint_{R^2} e^{-\frac{x^2+y^2}{2}} dx\, dy = \int_0^{+\infty} \int_0^{2\pi} \rho e^{-\frac{\rho^2}{2}} d\theta d\rho = 2\pi[-e^{-\frac{\rho^2}{2}}]_0^{+\infty} = 2\pi(0+1) = 2\pi\,.$$

Inoltre (R^2 é dominio normale)

$$\iint_{R^2} e^{-\frac{x^2+y^2}{2}} dx\, dy = \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \cdot \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy = I \cdot I = I^2\,. \tag{2}$$

Da (1) e (2) $I^2 = 2\pi$, e

$$\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}\,. \tag{3}$$

2 Siano $\mu \in R$, $\sigma > 0$. Studiamo la funzione $f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$. f è definita su R, é simmetrica rispetto alla retta di equazione $t = \mu$, $lim_{t\to\pm\infty} f(t) = 0$.

$$f'(t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \left(-\frac{2(t-\mu)}{2\sigma^2}\right) e^{-\frac{(t-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma^3\sqrt{2\pi}} \cdot (\mu - t)\, e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,.$$

$f'(t) > 0 \Leftrightarrow \mu - t > 0 \Leftrightarrow t < \mu$. $t = \mu$ é punto di max relativo e assoluto e $f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$. Al crescere di σ il massimo diminuisce.

$$f''(t) = \frac{1}{\sigma^3\sqrt{2\pi}} \cdot \left[-e^{-\frac{(t-\mu)^2}{2\sigma^2}} - (t-\mu) \cdot \left(-\frac{2(t-\mu)}{2\sigma^2}\right) e^{-\frac{(t-\mu)^2}{2\sigma^2}}\right] =$$
$$\frac{1}{\sigma^3\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \cdot \left(-1 + \frac{(t-\mu)^2}{\sigma^2}\right)\,.$$

$$f''(t) > 0 \Leftrightarrow \frac{(t-\mu)^2}{\sigma^2} > 1 \Leftrightarrow \frac{t-\mu}{\sigma} > 1 \vee \frac{t-\mu}{\sigma} < -1\,.$$

I punti $t = \mu + \sigma$ e $t = \mu - \sigma$ sono punti di flesso. $f(\mu \pm \sigma) = \frac{1}{\sigma\sqrt{2\pi e}}$. Al crescere di σ la "campana" si allarga.
Vedere figura 1 e figura 2.
Valutiamo ora

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt\,.$$

Facciamo il cambiamento di variabile $\tau = \frac{t-\mu}{\sigma}$. Se $t \to -\infty$ anche $\tau \to -\infty$. Se $t \to +\infty$ anche $\tau \to +\infty$. $d\tau = \frac{dt}{\sigma}$ e l'integrale si scrive:

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\tau^2}{2}} d\tau = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = 1\,. \tag{4}$$

Possiamo affermare che f é una densità di probabilità, associata ad una variabile aleatoria X. Chiamiamo X distribuzione normale di media μ e varianza σ^2:

$$X \sim \mathcal{N}(\mu, \sigma^2)\,.$$

Verifichiamo che in effetti μ e σ^2 sono media e variamza:

$$\begin{aligned} \mu_X = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} t e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (t-\mu+\mu) e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \\ \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{(t-\mu)}{\sigma^2} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt + \frac{\mu}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \\ = \frac{\sigma}{\sqrt{2\pi}} \left[-e^{-\frac{(t-\mu)^2}{2\sigma^2}} \right]_{-\infty}^{+\infty} + \mu = \mu\,. \end{aligned} \tag{5}$$

Analogamente si prova che

$$\sigma_X^2 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (t-\mu)^2 e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \sigma^2\,. \tag{6}$$

$$P(X \le x) = \int_{-\infty}^{x} f(t)dt \ \forall x \in R,\ P(-x \le X \le x) = \int_{-x}^{x} f(t)dt\,.$$

$$P(x_1 \le X \le x_2) = P(X \le x_2) - P(X \le x_1),\ P(X \ge x) = \ldots.$$

Gli integrali delle formule precedenti non sono elementarmente calcolabili.
La distribuzione $\mathcal{N}(0,1)$ si chiama normale standardizzata.

$$Z \sim \mathcal{N}(0,1)\,.$$

Vediamo come si passa da una Normale a una Normale standardizzata. Sia $X \sim \mathcal{N}(\mu, \sigma^2)$. Allora

$$P(X \le x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \tag{7}$$

Facciamo il cambiamento di variabile $\tau = \frac{t-\mu}{\sigma}$. Se $t \to -\infty$ anche $\tau \to -\infty$. Se $t = x$, $\tau = \frac{x-\mu}{\sigma}$. $d\tau = \frac{dt}{\sigma}$ e (7) si scrive:

$$\begin{aligned} P(X \le x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \\ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{\tau^2}{2}} d\tau = P\left(Z \le \frac{x-\mu}{\sigma}\right) \end{aligned} \tag{8}$$

Poniamo

$$\mathbf{\Phi(z) = P(Z \le z)}.$$

Allora

$$P(X \le x) = \Phi\left(\frac{x-\mu}{\sigma}\right)\,.$$

$$P(X \ge x) = 1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\,.$$

$$P(x_1 \le X \le x_2) = \ldots\ldots = \Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right)\,.$$

$\Phi(0) = \frac{1}{2}$, $\Phi(z) = 1 - \Phi(-z)$. Inoltre (vedi figure) se $X \sim \mathcal{N}(0, \sigma^2)$

$$\begin{aligned} & P(-\sigma \le X \le \sigma) = P(-1 \le Z \le 1) \simeq 0,6826\,, \\ & P(-2\sigma \le X \le 2\sigma) = P(-2 \le Z \le 2) \simeq 0,9544\,, \\ & P(-3\sigma \le X \le 3\sigma) = P(-3 \le Z \le 3) \simeq 0,9973\,. \end{aligned} \tag{9}$$

Utilizzo delle tavole (vd file pdf allegati) e di Excel.
Esercizi
1) Il diametro di certe biglie segue una $\mathcal{N}(6,2mm, (0,05mm)^2)$. Qual è la percentuale di biglie con diametro compreso tra $6,3mm$ e $6,35mm$?
$P(6,3 \le X \le 6,35) = P(X \le 6,35) - P(X \le 6,3) = 0,0214.$
2) Un test per il Q.I. produce punteggi con distribuzione normale di media 100 e deviazione standard 14,2. Che intervallo di punteggi raggiunge l'1% della popolazione formato dalle persone col Q.I. più alto?
Abbiamo una distribuzione $X \sim \mathcal{N}(100, 201,64)$ ($\sigma = 14,2$, $\sigma^2 = 201,64$). Uso delle tavole: riconduciamoci a una $Z \sim \mathcal{N}(0, 1)$:

$$P(X \ge x) = 0,01 \Rightarrow P\left(\frac{X-100}{14,2} \ge \frac{x-100}{14,2}\right) = 0,01 \Rightarrow$$
$$\Rightarrow P\left(Z \ge \frac{x-100}{14,2}\right) = 0,01.$$

Dalle tavole $\frac{x-100}{14,2} \approx 2,33$ da cui $x = 100 + 2.33 \cdot 14,2 = 133,08$.
3) La statura delle donne adulte in una certa nazione, ha una distribuzione normale con media 163 cm e deviazione standard 6 cm. Qual è la probabilità che una donna scelta a caso sia alta meno di 160 cm? meno di 178? tra i 160 e i 178? Alice è alta 183 cm. Che percentuale della popolazione femminile adulta è più bassa di lei?
Abbiamo una distribuzione $X \sim \mathcal{N}(163, 36)$. Per rispondere alle domande ci riconduciamo a una $Z \sim \mathcal{N}(0, 1)$:

$$P(X \le 160) = P\left(\frac{X-163}{6} \le \frac{-1}{2}\right) = P\left(Z \le \frac{-1}{2}\right) = P\left(Z \ge \frac{1}{2}\right) \approx 0,31.$$
$$P(X \le 178) = P\left(Z \le 2,5\right) \approx 0,99.$$
$$P(160 \le X \le 178) = P(X \le 178) - P(X \le 160) \approx 0,68.$$
$$P(X \le 183) = P\left(Z \le 3,\overline{3}\right) \approx 0,999.$$