LabC: memorandum sulla *normale* ('gaussiana')

- Funzione **densità di probabilità** ('pdf' – *probability density function*):

$$\begin{aligned} f(x \mid \mu, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{2\,\sigma^2}} && (1) \\ P(x_1 \le X \le x_2 \,|\, \mu, \sigma) &= \int_{x_1}^{x_2} f(x \mid \mu, \sigma)\, \mathrm{d}x && (2) \end{aligned}$$

  Nota: X è il nome della variabile di interesse; x è un suo possibile valore.
  Nel linguaggio R la pdf normale si chiama `dnorm()`.

- Funzione **cumulativa** (CDF – *Cumulative distribution function*):

$$\begin{aligned} F(x \mid \mu, \sigma) &= P(X \le x \,|\, \mu, \sigma) = \int_{-\infty}^{x} f(x' \mid \mu, \sigma)\, \mathrm{d}x' \\ P(x_1 \le X \le x_2 \,|\, \mu, \sigma) &= \int_{x_1}^{x_2} f(x \mid \mu, \sigma)\, \mathrm{d}x \\ &= \int_{-\infty}^{x_2} f(x \mid \mu, \sigma)\, \mathrm{d}x - \int_{-\infty}^{x_1} f(x \mid \mu, \sigma)\, \mathrm{d}x \\ &= F(x_2 \mid \mu, \sigma) - F(x_1 \mid \mu, \sigma) \end{aligned}$$

  Nel linguaggio R la CDF normale si chiama `pnorm()`. Ad esempio

$$P(1 \le X \le 5 \,|\, \mu = 3, \sigma = 2) = \texttt{pnorm}(5,3,2) - \texttt{pnorm}(1,3,2)$$

  il cui risultato dovrebbe essere ben noto.

- **Normale standardizzata**: indicata tipicamente con Z, è caratterizzata da $\mu = 0$ e $\sigma = 1$:

$$\begin{aligned} \varphi(z) &= f(z \mid \mu = 1, \sigma = 1) \\ &= \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}} \end{aligned}$$

  Si passa dalla generica X normale alla normale standardizzata Z, e viceversa, mediante le trasformazioni

$$\begin{aligned} Z &= \frac{X-\mu}{\sigma} && (3) \\ X &= \sigma\, Z + \mu\,. && (4) \end{aligned}$$

- Uso della **funzione speciale** `erf()`.
  Un caso speciale di importanza pratica è quello con $\mu = 0$ e $\sigma^2 = 1/2$. La pdf si riduce a

$$f(x \mid \mu = 0, \sigma = 1/\sqrt{2}) = \frac{1}{\sqrt{\pi}}\, e^{-\frac{(x-\mu)^2}{\sigma^2}}$$

  e la probabilità che la variabile sia compresa fra $-x$ e x è data da

$$P(-x \le X \le x \,|\, \mu, \sigma) = \int_{-x}^{x} \frac{1}{\sqrt{\pi}}\, e^{-\frac{(x'-\mu)^2}{\sigma^2}} \mathrm{d}x' = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-\frac{(x'-\mu)^2}{\sigma^2}} \mathrm{d}x'$$

Per la simmetria della funzione integranda, quest'ultimo integrale è pari a due volte l'integrale fra 0 e x, ovvero

$$\frac{1}{\sqrt{\pi}}\int_{-x}^{x} e^{-\frac{(x'-\mu)^2}{\sigma^2}}\mathrm{d}x' = \frac{2}{\sqrt{\pi}}\int_{0}^{x} e^{-\frac{(x'-\mu)^2}{\sigma^2}}\mathrm{d}x'$$

Questi integrali definiscono la funzione speciale `erf` (a parte il nome della variabile di integrazione che tipicamente è t). Quindi

$$\begin{aligned}\mathrm{erf}(x) &= \frac{1}{\sqrt{\pi}}\int_{-x}^{x} e^{-t^2}\,\mathrm{d}t \\ &= \frac{2}{\sqrt{\pi}}\int_{0}^{x} e^{-t^2}\,\mathrm{d}t\end{aligned}$$

- Valutazione della **cumulativa mediante** `erf()`. L'interesse della funzione erf (*error function* per 'ovvi' motivi storici) è che essa permette di valutare la CDF della normale, e quindi le probabilità di interesse, mediante l'uguaglianza (ottenibile con un paio di trasformazioni)

$$F(x\mid\mu,\sigma) = \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \tag{5}$$

  - Nel linguaggio R per usare la funzione `erf()` bisogna prima caricare il pacchetto `VGAM`, ad esempio
    > `library VGAM`
    > `erf(1)`
  - Nel linguaggio C per usare la funzione `erf()` bisogna includere `math.h` (argomento e valore 'ritornato' sono double)
    Nota: erf non compare in `C-reference_card.pdf`, essendo concisa: → vedi file aggiuntivo `C-special_functions.pdf`.

- **Generatore** approssimato di numeri **'random gaussiani'**.
  Un vecchio trucco per costruirsi, a partire da numeri (pseudo-)random nell'intervallo $[0,1]$, un generatore di numeri (pseudo-)casuali che seguono con buona approssimazione una distribuzione normale entro alcune sigma dalla media si basa sul cosidetto *teorema del limite centrale*.
  Se indichiamo con U_i dei numeri casuali indipendenti distribuiti uniformemente nell'intervallo fra 0 e 1, è ben noto come il loro valore atteso sia 0.5 con varianza 1/12. Ne segue che se ne sommiamo 12, il valore atteso della somma vale 6 e la varianza 1. Definendo quindi

$$R = \sum_{i=1}^{12} U_i - 6\,, \tag{6}$$

  questo numero casuale ha media nulla e varianza unitaria. Inoltre, per il teorema del limite centrale, R è distribuito circa normalmente (si ricorda che il teorema parla di "infiniti termini" e quindi va usato con cautela! In particolare, è chiaro che R sarà compreso fra -6 e 6.)
  Quindi la (6) fornisce un semplice algoritmo per produrre (entro alcune sigma dal valore centrale) numeri normali standardizzati, ovvero aventi $\mu = 0$ e $\sigma = 1$. Per ottenere numeri gaussiani con μ e σ arbitrari è sufficiente moltiplicare R per la σ e aggiungere μ, come da formula (4), essendo R 'una Z', per dirlo alla buona.