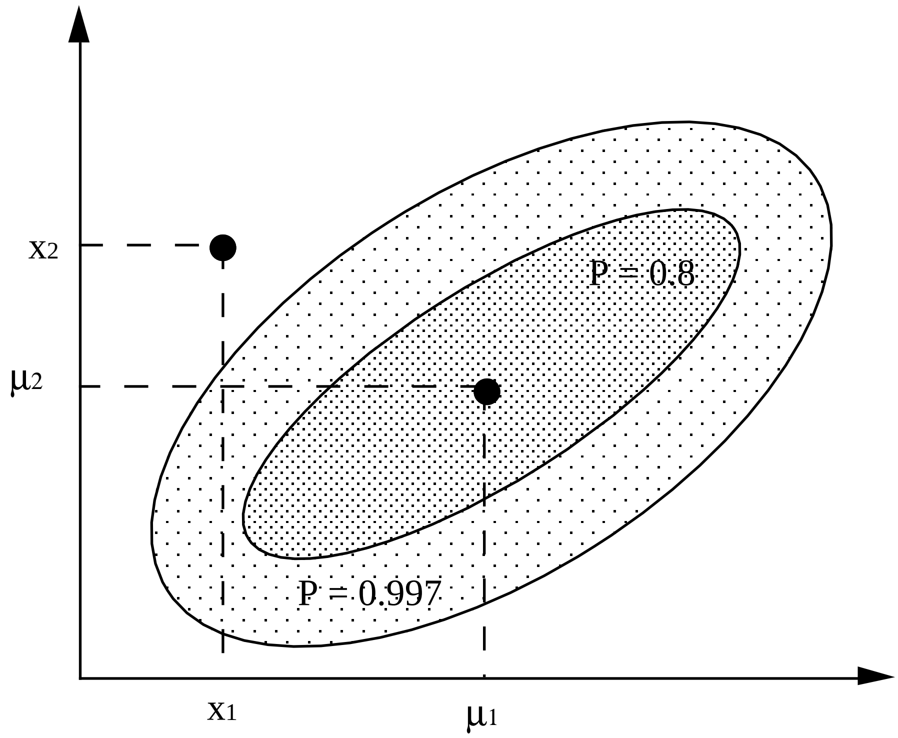
Distanza di Mahalanobis

Fatto: Le linee di livello della d.d.p. gaussiana multivariata (ellissi) individuano le regioni che, a parità di area, sono più probabili (o, viceversa, le regioni che, a parità di probabilità, hanno area minima)

$$\Downarrow$$

giusto allarmarsi quando osservo un punto su un ellissoide "esterno" (= a probabilità bassa)



Osservazione: Un punto è tanto più probabile quanto più l'ellisse sui cui giace è vicina al centro (le cui coordinate sono i valori medi $[\mu_1,\ldots,\mu_n]' = E[x]$).

Idea: Introduco una misura di distanza ("improbabilità") che cresce a mano a mano che x si sposta su ellissi più esterne

Distanza di Mahalanobis:

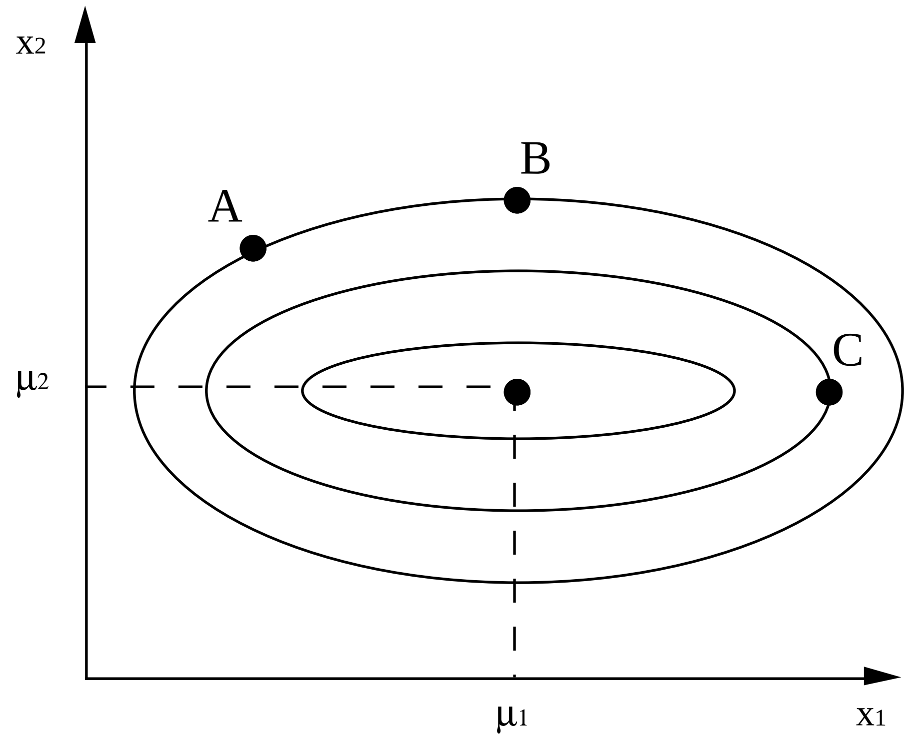
$$Q := (x-\mu)'\Sigma^{-1}(x-\mu)$$

$$\Sigma = Var[x] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \ldots & \sigma_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_n^2 \end{bmatrix}$$

Osservazioni:

- Caso particolare: $n = 2$, $r = 0$

$$Q := \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2}$$



Peso di meno le variabili più disperse

- Q è l'esponente della gaussiana multivariata:

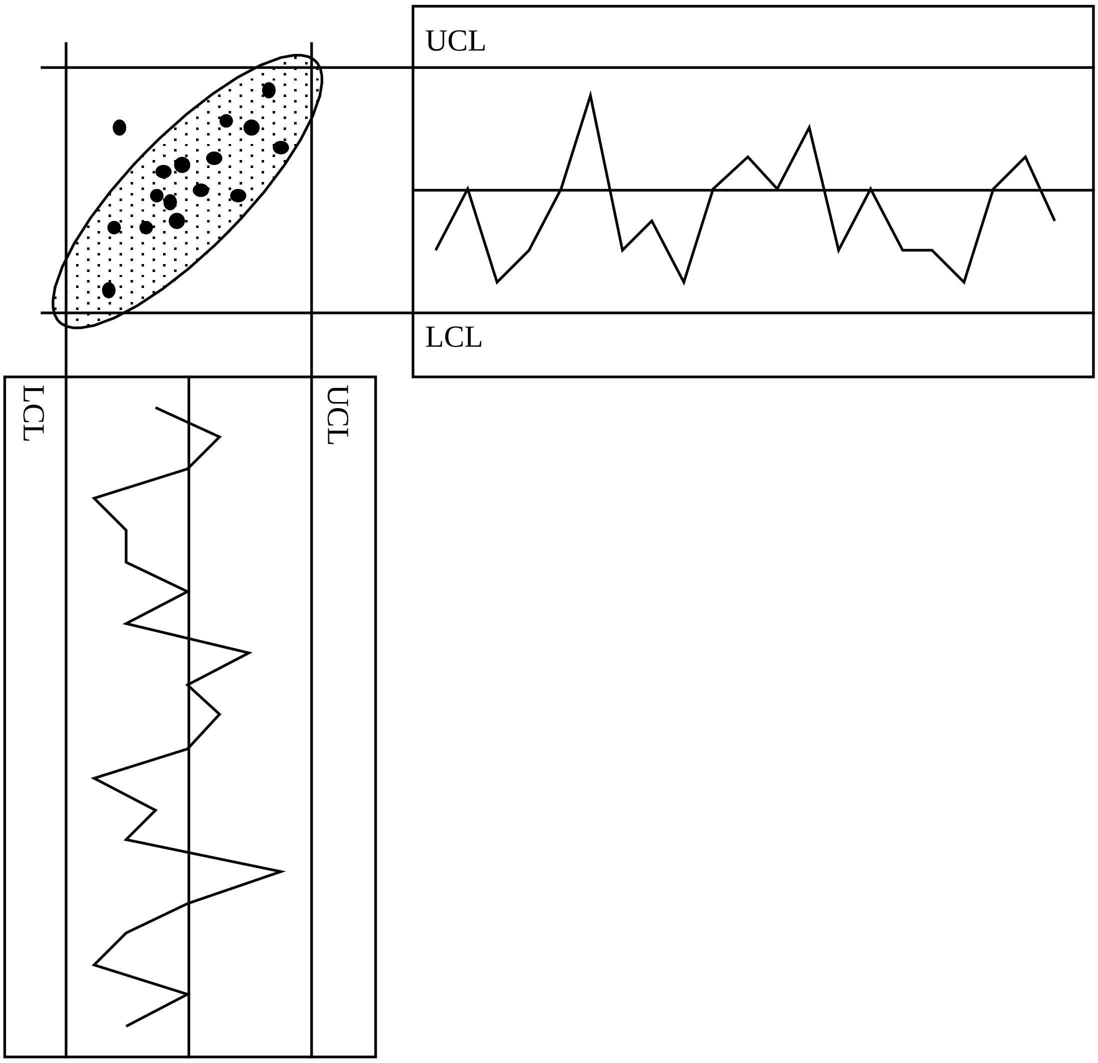
$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n det(\Sigma)}} e^{-\frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{2}}$$

L'ellisse di controllo

Proprietà: $Q = (x-\mu)'\Sigma^{-1}(x-\mu) \sim \chi_n^2$

Ipotesi: $n = 2$

Ellisse di controllo: Fissato α, trovo $\chi^2_{\alpha,n}$ e traccio l'ellisse di equazione $(x-\mu)'\Sigma^{-1}(x-\mu) = \chi^2_{\alpha,n}$. Segnalo un fuori controllo quando un punto cade fuori.



Svantaggi:

- Si perde la sequenza temporale dei punti.
- Cosa fare per $n > 2$?

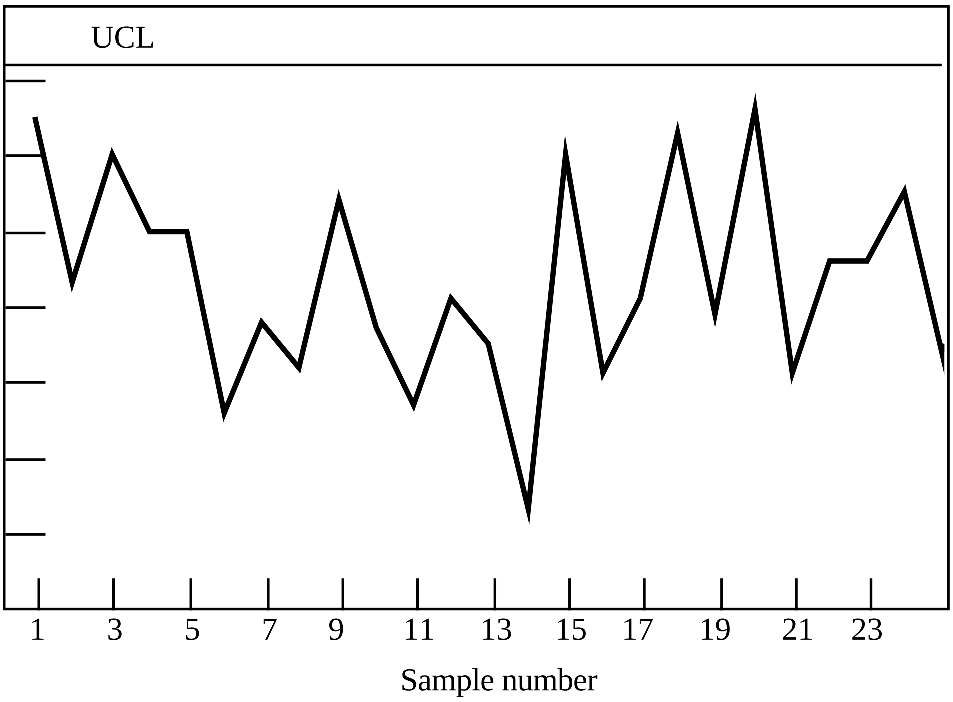
La carta χ^2

Idea: Per ogni campione $x(t)$, calcolo

$$Q(t) = (x(t)-\mu)'\Sigma^{-1}(x(t)-\mu)$$

e lo riporto su una carta apposita.

Limite di controllo: UCL = $\chi^2_{\alpha,n}$



Carta T^2 (Hotelling 1947)

Nella pratica, μ e Σ non sono noti, ma disponiamo di stime $\bar{\bar{X}}$ e $\hat{\Sigma}$.

Punti da disegnare:

$$T^2(t) = (x(t) - \bar{\bar{X}})'\hat{\Sigma}^{-1}(x(t) - \bar{\bar{X}})$$