Disuguaglianza di Chebyshev

Scelto un valore d,

$$\frac{\#\{x_i \mid |x_i - \overline{x}| > d\}}{n} \leq \frac{\mathrm{var}_e(x)}{d^2}.$$

Il termine sinistro indica la percentuale di dati che differiscono dalla media più di d. Questo valore può essere tanto più alto quanta maggiore è la varianza, che appunto misura la dispersione dei dati.

Dimostrazione

La disuguaglianza è equivalente a:

$$\#\{x_i \mid |x_i - \overline{x}| > d\} \leq \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{d^2},$$

che vale perché:

$$\begin{aligned}
\sum_{i=0}^{n}(x_i - \overline{x})^2 &\geq \sum_{\substack{i \text{ t.c.} \\ |x_i - \overline{x}| > d}} (x_i - \overline{x})^2 \\
&\geq \sum_{\substack{i \text{ t.c.} \\ |x_i - \overline{x}| > d}} d^2 \\
&= d^2 \cdot \#\{x_i \mid |x_i - \overline{x}| > d\}
\end{aligned}$$