Momenti pesati in probabilità

(Probability Weighted Moments - PWM)

Assunto: $F(x(u)) = u$

Si ha: $$\mathrm{E}(X) = \int_{-\infty}^{\infty} x dF(x) = \int_{-\infty}^{\infty} x f(x) dx$$

ovvero: $$\mathrm{E}(X) = \int_{0}^{1} x(u) du \ .$$

Essendo $x(u)$ il quantile della variabile x in corrispondenza della probabilità cumulata u, il PWM di ordine r è definito come:

$$\beta_r = \int_0^1 x(u) u^r du$$

Si ha pertanto: $\beta_0 = E(X)$

Lo stimatore campionario di β_r, detto b_r è:

$$b_r = n^{-1} \sum_{j=r+1}^{n} \frac{(j-1)(j-2)\ldots(j-r)}{(n-1)(n-2)\ldots(n-r)} x_{j:n}$$

dove $x_{j:n}$ è il j-esimo elemento del campione ordinato in senso crescente.

L-momenti

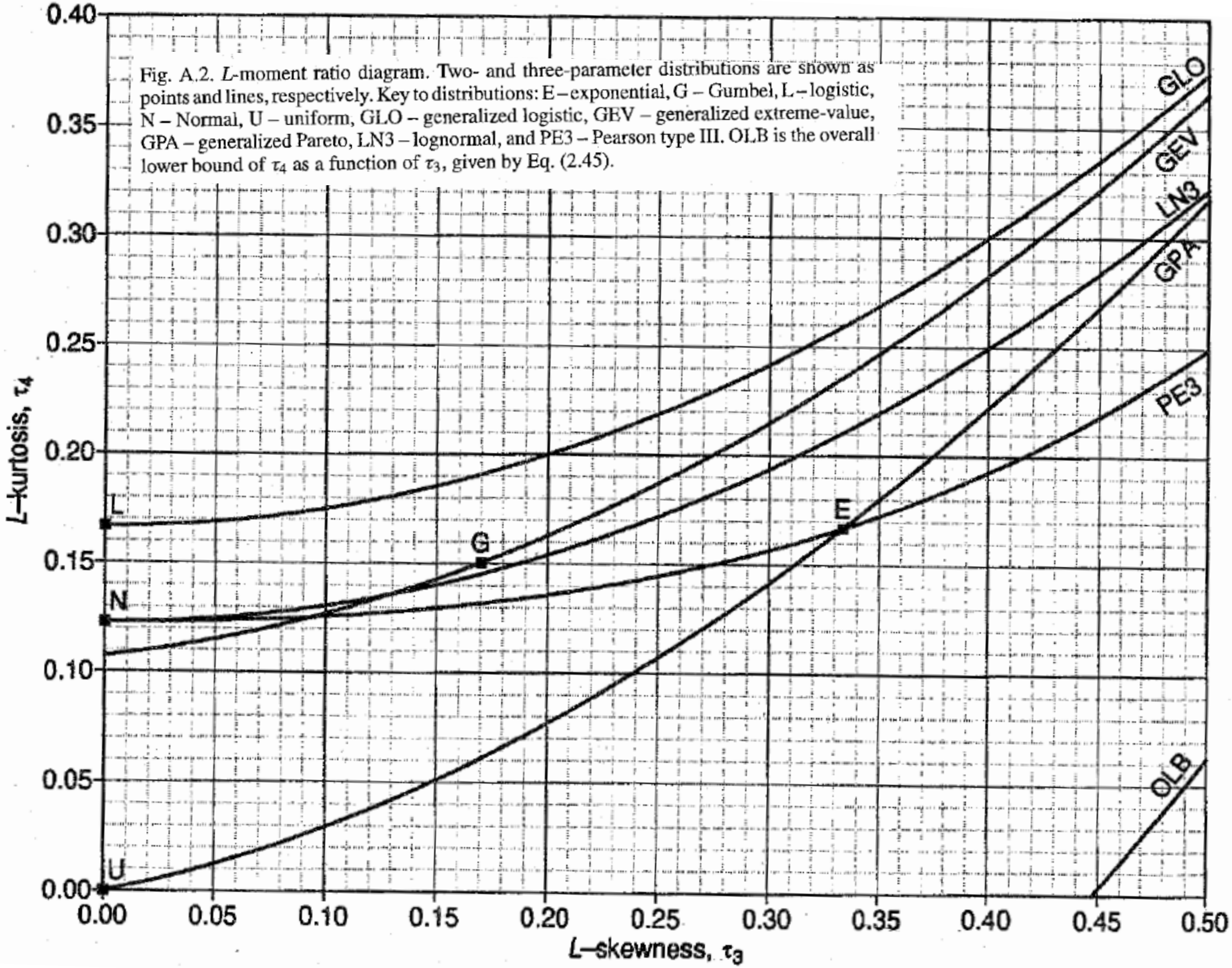
Gli L-momenti sono combinazioni lineari di PWM e sono statistiche basate su combinazioni lineari dei dati.

$$l_1 = b_0 \ , \quad = \text{media campionaria}$$
$$l_2 = 2b_1 - b_0 \ ,$$
$$l_3 = 6b_2 - 6b_1 + b_0 \ ,$$
$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0$$

Di particolare interesse sono i rapporti adimensionali di L-momenti:

- L-CV o $\tau = l_2/l_1$
- L-CA o L-skewness o $\tau_3 = l_3/l_2$
- L-kurtosis o $\tau_4 = l_4/l_2$

Diagramma diagnostico di Hosking e Wallis



Stima dei parametri di alcune distribuzioni mediante gli L-momenti

La procedura per la stima dei parametri è analoga a quella utilizzata nel metodo dei momenti: gli L-momenti teorici vengono equiparati a quelli campionari. Invertendo la relazione tra parametri e L-momenti si ottengono i parametri.

Distribuzione Normale

$$f(x) = \frac{1}{\theta_2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\theta_1}{\theta_2}\right)^2}$$

$$x(F) = \theta_1 + \theta_2 \Phi^{-1}(F)$$

L-momenti	*Parametri*
$\lambda_1 = \theta_1$	
$\lambda_2 = 0,5642\theta_2$	$\theta_1 = \lambda_1$
$\tau_3 = 0$	$\theta_2 = \pi^{1/2}\lambda_2$
$\tau_4 = 0.1226$	

Distribuzione di Gumbel

$$F(x) = e^{-e^{-\frac{x-\theta_1}{\theta_2}}}$$

$$f(x) = \frac{1}{\theta_2} e^{-\frac{x-\theta_1}{\theta_2}} e^{-e^{-\frac{x-\theta_1}{\theta_2}}}$$

$$x(F) = \theta_1 - \theta_2 \ln[-\ln(F)]$$

L-momenti	*Parametri*
$\lambda_1 = \theta_1 + 0,5772\theta_2$	
$\lambda_2 = \theta_2 \log 2$	$\theta_1 = \lambda_1 - 0,5772(\lambda_2/\log 2)$
$\tau_3 = 0,1699$	$\theta_2 = \lambda_2/\log 2$
$\tau_4 = 0,1504$	

Stima dei parametri di alcune distribuzioni mediante gli L-momenti

Distribuzione Log Normale

$$f_X(x) = \frac{1}{x\theta_2\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{\ln x - \theta_1}{\theta_2}\right)^2\right]$$

$$x(F) = \exp\left[\theta_1 + \theta_2\Phi^{-1}(F)\right] \qquad \theta_2 > 0$$

L-momenti

$$L_1 = \exp(\theta_1 + \theta_2^2/2)$$

$$L_2 = e^{\theta_1+\theta_2^2/2}\left[2\Phi(\theta_2/\sqrt{2}) - 1\right]$$

Parametri

$$\widehat{\theta_1} = \ln l_1 - \theta_2^2/2$$

$$\widehat{\theta_2} = \sqrt{2}\Phi^{-1}\left(\frac{1+l_2/l_1}{2}\right)$$

Distribuzione GEV

$$F(x) = \exp\left[-\left(1 - \frac{\theta_3}{\theta_2}(x - \theta_1)\right)^{1/\theta_3}\right]$$

$$x(F) = \begin{cases} \theta_1 + \theta_2[1 - (-\ln F)^{\theta_3}]/\theta_3 \ , & \theta_3 \neq 0 \\ \theta_1 - \theta_2 \ln(-\ln F) \ , & \theta_3 = 0 \end{cases}$$

L-momenti

$$L_1 = \theta_1 + \theta_2[1 - \Gamma(1+\theta_3)]/\theta_3$$

$$L_2 = \theta_2(1 - 2^{-\theta_3})\Gamma(1+\theta_3)]/\theta_3$$

$$\tau_3 = 2(1 - 3^{-\theta_3})/(1 - 2^{-\theta_3}) - 3$$

$$\tau_4 = \frac{5(1-4^{-\theta_3})-10(1-3^{-\theta_3})+6(1-2^{-\theta_3})}{(1-2^{-\theta_3})}$$

Parametri

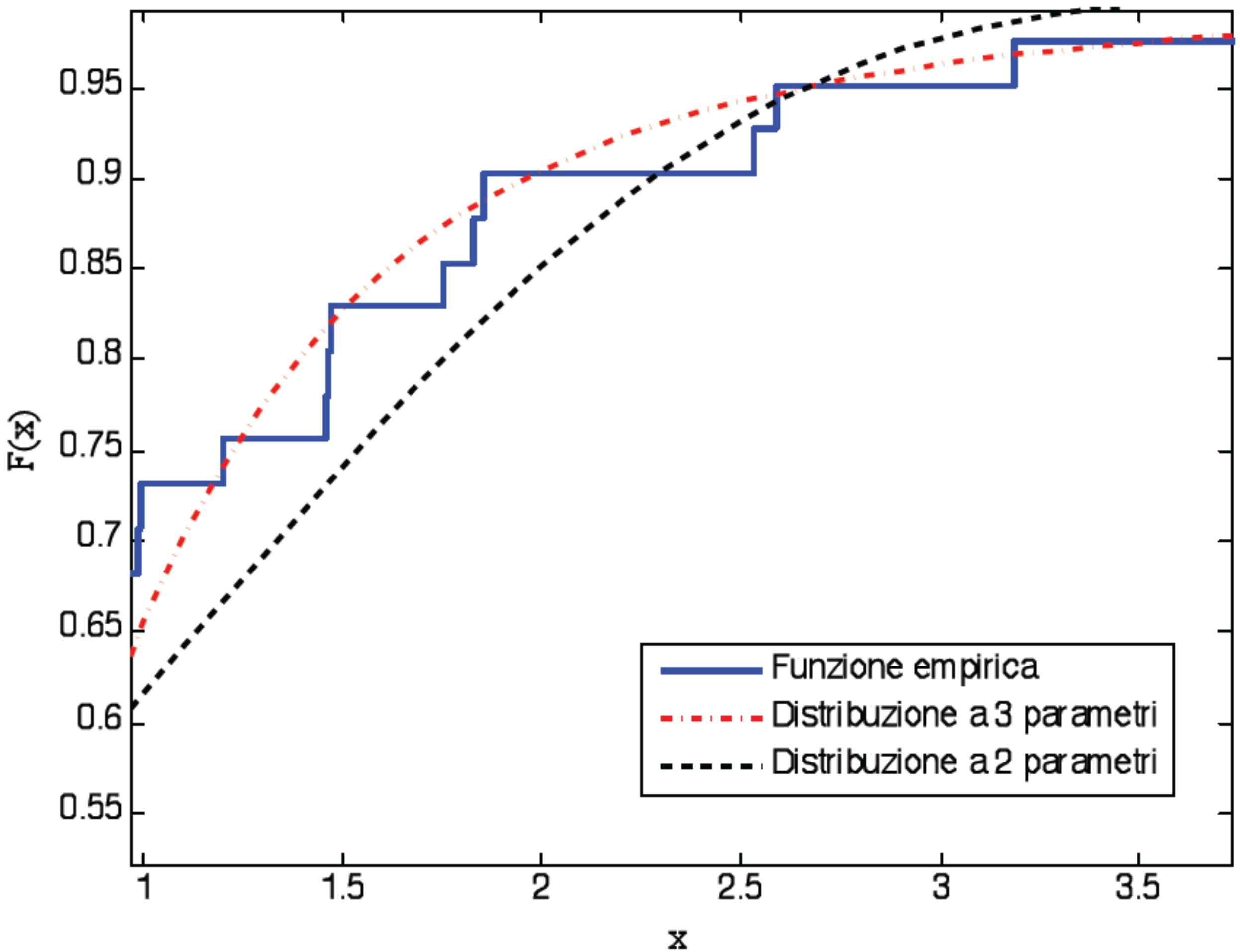
$$\widehat{\theta_3} \simeq 7.8590c + 2.9554c^2$$

$$c = \frac{2}{3+l_3/l_2} - \frac{\ln 2}{\ln 3}$$

$$\widehat{\theta_2} = \frac{l_2\widehat{\theta_3}}{\left(1-2^{-\widehat{\theta_3}}\right)\Gamma(1+\widehat{\theta_3})}$$

$$\widehat{\theta_1} = l_1 - \frac{\widehat{\theta_2}}{\widehat{\theta_3}}\left(1 - \Gamma(1+\widehat{\theta_3})\right)$$

Test di adattamento di Anderson-Darling



Distribuzione campionaria:

$$F_n(x)$$

$$
\begin{aligned}
F_n(x) &= 0, & x &< x_1 \\
F_n(x) &= \frac{i}{n}, & x_i &\le x < x_{i+1}, \\
F_n(x) &= 1, & x_n &\le x.
\end{aligned}
$$

Distribuzione teorica:

$$F_X(x)$$

dipende dalla distribuzione di probabilità utilizzata.

Distribuzioni con più parametri sono più flessibili: si adattano meglio, ma con rischio di overfitting.

Test di adattamento di Anderson-Darling

$$Q^2 = n \int_{\text{all}x} [F_n(x) - F(x,\theta)]^2 \Psi(x) \mathrm{d}F(x)$$

Definizione teorica della statistica test (per test quadratici).

$\Psi(x)$ è una funzione peso che dipende dal test utilizzato.

$$\Psi(x) = [F(x, \theta)(1 - F(x, \theta))]^{-1}$$

Funzione peso per il test di Anderson-Darling: peso maggiore sulle code della distribuzione.

Definizione operativa della statistica test di Anderson-Darling:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} [(2i-1)\ln[F(x_i,\theta)] + (2n+1-2i)\ln[1 - F(x_i,\theta)]].$$

Lunghezza campione

Indice del campione ordinato

Valore della distribuzione teorica calcolato in corrispondenza dell'i-esimo valore campionario

Test di adattamento di Anderson-Darling

Il test è superato se $A^2 < A^2_{\text{limite}}(\alpha)$, dove $A^2_{\text{limite}}(\alpha)$ è riportato su tabelle specifiche per ogni distribuzione.

Per tenere conto delle diverse distribuzioni (e del diverso numero di parametri) si passa alla statistica test trasformata ω (test superato se $\omega < \omega_{\text{limite}}$):

$$\omega = 0.0403 + 0.116\left(\frac{A^2 - \xi_p}{\beta_p}\right)^{\frac{\eta_p}{0.851}} \qquad \text{se} \qquad 1.2\xi_p \le A^2$$

$$\omega = \left[0.0403 + 0.116\left(\frac{0.2\xi_p}{\beta_p}\right)^{\frac{\eta_p}{0.851}}\right]\frac{A^2 - 0.2\xi_p}{\xi_p} \qquad \text{se} \qquad 1.2\xi_p > A^2$$

α	ω_{limite}
0.10	0.347
0.05	0.461
0.01	0.743

Table 3. Coefficients to Be Set in Equation (11) for the Anderson-Darling Statistic, Asymptotic Case[a]

Distribution[b]	ξ_p	β_p	η_p
EV1 and EV2 (EV1 = Gumbel)	0.169	0.229	1.141
NORM and LN	0.167	0.229	1.147
GEV[c]	$0.147\,(1 + 0.13\,\hat\theta_3 + 0.21\,\hat\theta_3^2 + 0.09\,\hat\theta_3^3)$	$0.189\,(1 + 0.20\,\hat\theta_3 + 0.37\,\hat\theta_3^2 + 0.17\,\hat\theta_3^3)$	$1.186\,(1 - 0.04\,\hat\theta_3 - 0.04\,\hat\theta_3^2 - 0.01\,\hat\theta_3^3)$
GAM and LP3[d]	$0.145\,(1 + 0.17\,\hat\theta_3^{-1} + 0.33\,\hat\theta_3^{-2})$	$0.186\,(1 + 0.34\,\hat\theta_3^{-1} + 0.30\,\hat\theta_3^{-2})$	$1.194\,(1 - 0.04\,\hat\theta_3^{-1} - 0.12\,\hat\theta_3^{-2})$

[a]Here $\hat\theta_3$ is an asymptotic efficient estimator (usually maximum likelihood) of the shape parameter of the distribution.
[b]For tests of the EV2, LN, and LP3 distributions the data must be preliminarily log transformed.
[c]For the GEV distribution, if $\hat\theta_3 > 0.5$, $\hat\theta_3 = 0.5$ must be set in the regressions.
[d]For the GAM and LP3 distributions, if $\hat\theta_3 < 2$, $\hat\theta_3 = 2$ must be set in the regressions.