Expansões de Edgeworth e Aproximações Pré-Assintóticas

Madalena Malva
CEAUL - Centro de Estatística e Aplicações da Universidade de Lisboa e ESTV - Escola Superior de Tecnologia de Viseu - malva@estv.ipv.pt

Sandra Mendonça
CEAUL - Centro de Estatística e Aplicações da Universidade de Lisboa e Universidade da Madeira - smendonca@uma.com

Dinis Pestana
CEAUL - Centro de Estatística e Aplicações da Universidade de Lisboa e Universidade de Lisboa - dinis.pestana@fc.ul.pt

Resumo: A ideia de aproximar uma função por outra é central em todas as áreas da Matemática, e a invenção dos polinómios ortogonais foi um avanço notável sobre a aproximação por polinómios de Taylor. As expansões de Edgeworth permitiram-nos construir o enquadramento adequado para repensar o teorema limite central clássico, e as expansões de Edgeworth diferidas *(tilted)* forneceram novos instrumentos para tratar grandes desvios de somas de variáveis aleatórias. Apresentamos alguns resultados similares para a aproximação de estatísticas ordinais, com uma breve incursão no tema de aproximações pré-assintóticas.

Palavras–chave: expansões assintóticas e pré-assintóticas, teoremas pré-limite.

Abstract: The idea of approximating a function by another function is a central one in all the areas of Mathematics, and the invention of the orthogonal polynomials was a remarkable advance on the approximations by Taylor polynomials. Edgeworth expansions allowed the construction of the right setup to rethink the classical central limit theorem, and the tilted Edgeworth expansions gave us new tools to handle large deviations in the sums of random variables. We present some similar results for the approximation of order statistics, with a brief incursion in the theme of pre-asymptotic approximations.

Keywords: asymptotic and pre-asymptotic expansions, pre-limits.

1 Introdução

Em paralelo com a teoria das somas de variáveis aleatórias, desenvolveu-se uma teoria de leis estáveis para máximos de sucessões de variáveis aleatórias (v.a.'s) independentes e identicamente distribuídas (iid). Desde o trabalho pioneiro de Fisher e Tippett [4] que estão identificadas três formas assintóticas, que Gnedenko [6] provou serem os únicos limites possíveis de máximos linearmente normalizados. A necessidade prática de trabalhar com variáveis dependentes alargou notavelmente a teoria de valores extremos, havendo inúmeros tipos de dependência fraca em que a forma assintótica da distribuição de máximos normalizados tem relação simples com a distribuição de extremos, que usamos na forma unificada de von Mises-Jenkinson: para x tal que $0 < H_\tau(x) < 1$,

$$H_\tau(x) = \exp\left[-(1+\tau x)^{-1/\tau}\right], \tag{1}$$

relembrando que para $\tau > 0$ é uma reparametrização da família de Fréchet$(1/\tau)$, para $\tau < 0$ da família de Weibull$(-1/\tau)$, e quando $\tau \to 0$ temos como limite a Gumbel padronizada.

Recentemente, Klebanov, Rachev e Szekely [10] contestaram a utilização de leis assintóticas, argumentando com a banalidade de que estamos sempre em situação de um número finito, aliás escasso, quase sempre, de observações. Terão porventura razão, pois desde Fisher e Tippett [4] se sabe que a convergência para a forma assintótica — a que chamam *ultimate* — pode ser muito lenta (exemplificam com a convergência do máximo de gaussianas para a Gumbel) e aproximações *penultimate* podem ser mais vantajosas. Gomes [7], usando expansões de Uzgören, deu uma explicação coerente do fenómeno, que mais tarde com Pestana [8] esclareceu ainda melhor mostrando que as aproximações pré-assintóticas serão melhores, por regra, quando uma lei está no domínio de atracção não-standard de uma lei estável de extremos.

No presente trabalho, exploramos duas vias para tentar clarificar os limites de duas abordagens que, do ponto de vista metodológico, se completam: uma tentativa de transpor as ideias exploradas na teoria de somas por Thiele, Gram, Charlier e Edgeworth, e que hoje constitui o campo sempre renovado das expansões de Edgeworth (veja-se o aproveitamento feito por Peter Hall [9] no que se refere à fundamentação do *bootstrap*, ou a explicitação dos resultados de Daniels [3] sobre aproximações com pontos de sela como aproveitamento de expansões de Edgeworth diferidas); e as expansões de Uzgören, tentando dilucidar as dificuldades de alargamento às Fréchet e, por dualidade, às Weibull referidas no final desse clássico.

2 Expansões com polinómios ortogonais

Consideremos $\{X_i\}_{i\in\mathbf{N}_1}$ uma sucessão de v.a.'s iid com função de distribuição (fd) comum dada por F. Consideremos ainda $M_n = \max_{i=1,\ldots,n} X_i$.

Na teoria das somas, a expansão de Edgeworth dá-nos uma aproximação para a função de distribuição da soma de n v.a.'s iid devidamente normalizada, em torno de Φ, a fd de uma gaussiana padrão. Sabemos que, se a variância de X, σ_X^2 for finita, então a função de distribuição de $Y_n = \frac{\sum_{i=1}^n X_i - nE(X)}{\sigma_X\sqrt{n}}$, para n *grande*, pode ser aproximada pela distribuição gaussiana padrão, Φ.

Considerando para f_{Y_n} a expansão dada por

$$f_{Y_n} = \sum_{i=0}^{+\infty} c_i \varphi^{(i)},$$

tendo em conta que $\varphi^{(i)} = (-1)^i H_i \varphi$, onde H_i é o polinómio de Hermite de grau i e $c_i = \int_{\mathbf{R}} (-1)^i H_i(x) f_{Y_n}(x) dx$ obtém-se a seguinte expansão em polinómios ortogonais para f_{Y_n}:

$$f_{Y_n} \quad = \varphi - \tfrac{1}{3!} \tfrac{k_3'}{n^{1/2}} \varphi^{(3)} + \tfrac{1}{4!} \tfrac{k_4'}{n} \varphi^{(4)} + \tfrac{10!}{6!} \tfrac{(k_3')^2}{n} \varphi^{(6)} - \ldots$$

onde os valores k_ν' apelidados por Thiele [11] como semi-invariantes de $X - E(X)$, são os actualmente chamados de cumulantes de $X - E(X)$. Esta é de facto um expansão assintótica de f_{Y_n} em potências de $n^{1/2}$, com um resto que tem a mesma ordem do que o primeiro termo a ser desprezado.

Suponhamos que F tem função densidade de probabilidade (fdp) f e está no domínio de uma distribuição de extremos H. Para cada valor de $n \in \mathbf{N}_1$, vamos determinar uma sucessão de constantes $\{c_i\}_{i \in \mathbf{N}_0}$ e uma sucessão de polinómios $\{p_i\}_{i \in \mathbf{N}_0}$, ortogonais associados a H tais que

$$f_n(x) = \sum_{i \geq 0} c_i p_i(x) \frac{d}{dx} H(x) . \tag{2}$$

onde f_n é a fdp de Z_n. Designemos por Z a v.a. limite (em distribuição) da sucessão Z_n. Notemos que apenas a distribuição Gumbel tem momentos de todas as ordens. A de Fréchet(γ) e a de Weibull(γ) têm apenas momentos de ordem inferior a γ.

Multiplicando por $p_k(x)$ e integrando em $\mathbf{R}$ ambos os lados da igualdade (2) obtemos$\int_{\mathbf{R}} p_k(x) f_n(x) dx = c_k$. Tomando $p_k(x) = \sum_{i=0}^{k} a_{i,k} x^i$ resulta $c_k = \sum_{i=0}^{k} a_{i,k} E(Z^i)$ e

$$\int_{\mathbf{R}} p_k(x) p_m(x) h(x) dx \quad = \textstyle\sum_{i=0}^{k} \sum_{j=0}^{m} a_{i,k} a_{j,m} E\left(Z^{i+j}\right) = S(k,m) .$$

Os coeficientes dos polinómios podem ser obtidos através da solução das equações $S(k,m) = 0$ se $k \neq m$, e $S(k,m) = 1$ se $k = m$, para quaisquer naturais k e m. Por exemplo, para o primeiro conjunto de equações $S(0,1) = 0$ e $S(1,1) = 1$ obtemos

$$a_{0,1} = -\frac{E(Z)}{\sqrt{E(Z^2) - E^2(Z)}} \text{ e } a_{1,1} = \frac{1}{\sqrt{E(Z^2) - E^2(Z)}}.$$

Para o segundo conjunto de equações$S(0,2) = 0,\ S(1,2) = 0$ e $S(2,2) = 1$ obtemos

$$\begin{aligned} a_{0,2} &= \frac{E(Z)\,E\left(Z^3\right) - E^2\left(Z^2\right)}{\sqrt{var(Z)\,A(Z)}},\; a_{1,2} = \frac{E(Z)\,E\left(Z^2\right) - E\left(Z^3\right)}{\sqrt{var(Z)\,A(Z)}} \\ \text{e } a_{2,2} &= \frac{E\left(Z^2\right) - E^2(Z)}{\sqrt{var(Z)\,A(Z)}} \end{aligned}$$

onde

$$A(Z) = 2E(Z)\,E\left(Z^2\right)E\left(Z^3\right) - E^3\left(Z^2\right) - E^2\left(Z^3\right) - E^2(Z)\,E\left(Z^4\right) + E\left(Z^2\right)E\left(Z^4\right).$$

Suponhamos então que F está no domínio de $H_{3,0}$ e consideremos a aproximação dada pelas primeiras três parcelas da série de (2). Os primeiros momentos de uma v.a. $Z \frown$ Gumbel são dados por

$$\begin{array}{l} E(Z) = \gamma,\; E\left(Z^2\right) = \gamma^2 + \frac{\pi^2}{6},\; E\left(Z^3\right) = \gamma^3 + \gamma\frac{\pi^2}{2} + 2\zeta(3) \\ E\left(Z^4\right) = \gamma^4 + \gamma^2\,\pi^2 + \frac{3\pi^4}{20} + 8\gamma\zeta(3) \end{array}$$

onde γ é a constante de Euler ($\gamma \approx 0.5772$) e $\zeta(3)$ é a função zeta no ponto 3 ($\zeta(3) \approx 1.20206$). Como exemplos considerámos uma v.a. $Y \frown$ Exponencial(1), e uma v.a. $X \frown$ Gaussiana$(0,1)$, que sabemos pertencer ao domínio de $H_{3,0}$ e para as quais podemos tomar para constantes de atracção

$$a_n = \ln \text{ e } b_n = 1,\; a_n = (2\ln n)^{1/2} - \frac{1}{2}\frac{\ln\ln n + \ln(4\pi)}{(2\ln n)^{1/2}} \text{ e } b_n = (2\ln n)^{-1/2},$$

respectivamente. Infelizmente, os gráficos obtidos no programa Mathematica (Versão 5) não foram muito satisfatórios, principalmente no que se refere ao caso gaussiano. Nas Figuras 1 e 2 são apresentados três gráficos: a fdp do máximo normalizado de n ($n = 5$ na Figura 1 e $n = 15$ na Figura 2) v.a.'s i.i.d. com distribuição Exponencial(1), a fdp da distribuição limite (ou seja, da distribuição Gumbel) e a aproximação dada pela soma dos três primeiros termos da expansão acima referida. As Figuras 3 e 4 apresentam os gráficos correspondentes 'a distribuição gaussiana padrão.

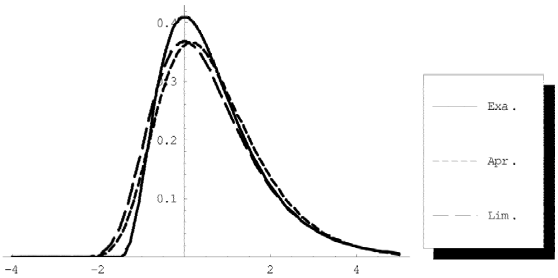

Figura 1: $X \frown$Exponencial(1), $n = 5$

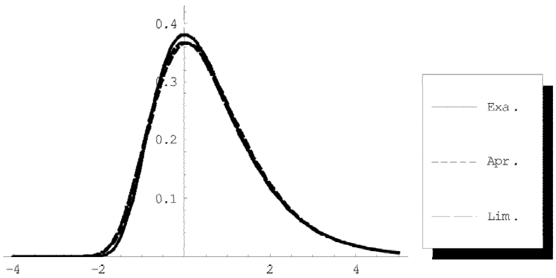

Figura 2: $X \frown$Exponencial(1), $n = 15$

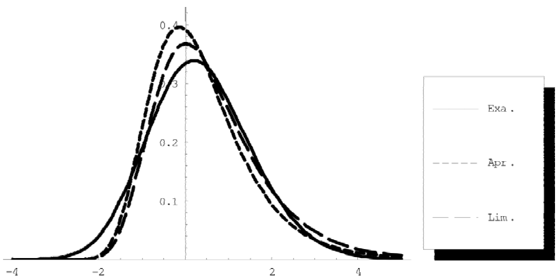

Figura 3: $X \frown$Gaussianal(0, 1), $n = 5$

3 Expansões de Uzgören

N. Uzgören [12] construiu uma expansão em série para $-\ln\left[-\ln F^n\left(a_n + b_n x\right)\right]$ para o caso particular de F ser uma fd com limite superior do suporte de F

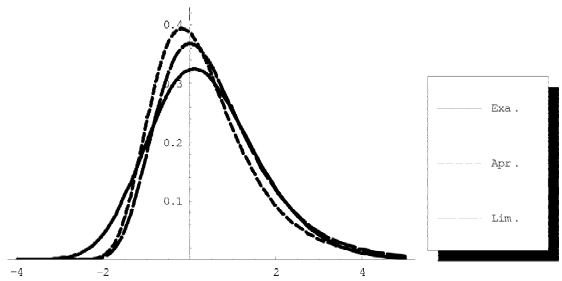

Figura 4: $X \frown$Gaussianal$(0,1)$, $n = 15$

(ω_F) igual a $+\infty$, contínua e diferenciável, com derivada $f(x)$ diferente de zero, para valore grandes de x. Supôs ainda que, tomando $g(x) = [1 - F(x)]/f(x)$, $g'(x) \xrightarrow[x \to +\infty]{} 0$. Sabendo que esta condição é suficiente para que F esteja no domínio de atracção de $H_{3,0}$, torna-se compreensível a escolha de Uzgören para as constantes de atracção: $a_n = x_n$ e $b_n = g(x_n)$ onde x_n é a solução da equação $1 - F(x_n) = 1/n$, $(n = 1, 2, \cdots)$. A expansão encontrada por N. Uzgören para as distribuições F que satisfazem a condição adicional $ng'(x_n) \xrightarrow[n \to +\infty]{} +\infty$ foi a seguinte

$$\begin{aligned} \ln[-\ln P^n(x_n + g(x_n)u)] &= -u - \tfrac{u^2}{2} \tfrac{d}{dx}\left(\tfrac{1}{g}\right)\Big|_{x=x_n} g^2(x_n) - \\ &- \tfrac{u^m}{m!} \tfrac{d^{m-1}}{dx^{m-1}}\left(\tfrac{1}{g}\right)\Big|_{x=x_n} g^m(x_n) - \ldots \end{aligned}$$

Consideremos F uma fd tal que $\omega_F = +\infty$ e para a qual existe um número real x_1 tal que, para $x \geq x_1$, $f(x) = F'(x)$ existe e é contínua. Suponhamos ainda que

$$\lim_{x \to +\infty} \frac{xf(x)}{1 - F(x)} = \lim_{x \to +\infty} \frac{x}{g(x)} = \gamma, \tag{3}$$

onde $\gamma \in (0, +\infty)$. Sabemos então que (cf., e.g., [5]) $F \in \mathcal{D}(H_{1,\gamma})$ e que as constantes de atracção, neste caso, podem ser dadas por $a_n = 0$ e $b_n = \inf\{x : 1 - F(x) \leq 1/n\}$. Supondo, tal como faz N. Uzgören [12], que $f(x)$ é diferente de zero, para valores suficientemente grandes de x, temos, para valores suficientemente grandes de n, $b_n = x_n = F^{-1}(1 - 1/n)$.

Seguindo os passos de Uzgören representemos F da seguinte forma:

$$F(x) = 1 - \exp[-h(x)]. \tag{4}$$

Da definição de g resulta $h = -\ln(gf)$ e de $1 - F(x_n) = f(x_n)g(x_n) = \frac{1}{n}$ resulta $h(x_n) = \ln n$. Elevando ambos os lados de (4) à potência n e tomando os simétricos dos logaritmos obtemos

$$\begin{aligned} -\ln F^n(x) &= -n \ln\{1-\exp[-h(x)]\} \quad = n\exp[-h(x)]\,\rho(x) \\ \Rightarrow \ln[-\ln F^n(x)] &= \ln n - h(x) + \ln\rho(x) \end{aligned}$$

onde $\rho(x) = \sum_{k=1}^{+\infty} \exp[-(k-1)\,h(x)]/k$.

Expandindo h em série de Taylor em torno de x_n, substituindo x por $x_n u$ e tendo em conta que $h'(x) = \frac{1}{g(x)}$, obtemos

$$h(x_n u) = \ln n \quad + \sum_{k=1}^{+\infty} \frac{1}{k!} \left.\frac{d^{k-1}}{dx^{k-1}} \frac{1}{g(x)}\right|_{x=x_n} x_n^k (u-1)^k .$$

Tendo em conta que se f for infinitamente diferenciável

$$\frac{d^k}{dx^k}[xf(x)] = k\frac{d^{k-1}}{dx^{k-1}}f(x) + x\frac{d^k}{dx^k}f(x), \tag{5}$$

($k = 1, 2, ...$) e se F satisfizer a condição dada por (3), então, para $k = 1, 2, ...$ e x suficientemente grande,

$$\frac{d^k}{dx^k}f(x) \approx -\frac{\gamma+k}{x}\frac{d^{k-1}}{dx^{k-1}}f(x),$$

mostra-se que, para n suficientemente grande,

$$\frac{1}{k!}x_n^k(u-1)^k \left.\frac{d^{k-1}}{dx^{k-1}}\frac{1}{g(x)}\right|_{x=x_n} \approx -\gamma\frac{(1-u)^k}{k}, \tag{6}$$

o que conduz a

$$\begin{aligned} h(x_n u) &\approx \ln n + \sum_{k=1}^{+\infty} -\gamma\frac{(1-u)^k}{k} = \ln n - \gamma\ln u, \\ \Rightarrow \ln[-\ln F^n(x_n u)] &= \ln n - h(x_n u) + \ln\rho(x_n u) \\ &\approx -\gamma\ln u + \ln\rho(x_n u). \end{aligned}$$

Expandindo $\ln\rho$ temos

$$\begin{aligned} \ln\rho(x_n u) &= \sum_{i=1}^{+\infty}(-1)^{i+1} \\ &\sum_{k_1=2}^{+\infty}\cdots\sum_{k_i=2}^{+\infty} \frac{\left\{\exp\left[-\sum_{k=1}^{+\infty}\frac{x_n^k(u-1)^k}{k!}\frac{d^{k-1}}{dx^{k-1}}\left.\frac{f(x)}{1-F(x)}\right|_{x=x_n}\right]\right\}^{(k_1+\ldots+k_i-i)}}{n^{k_1+\ldots+k_i-i}\, i\, k_1\ldots k_i}. \end{aligned}$$

No trabalho de Uzgören é a diferença $\ln n - h(x_n u)$ que dá origem à expansão, sendo a parcela $\ln\rho(x_n u)$ negligenciada, por comparação aos termos da expansão obtida de $\ln n - h(x_n u)$. A expansão que Uzgören tem como parcela principal o termo referente à distribuição limite. Aqui o termo principal é $-\gamma\ln u$ e ficamos apenas dependentes de $\ln\rho(x_n u)$, se não tivermos em conta a série dada em (5), que tem um comportamento razoável apenas para valores de u no intervalo $(0,1)$, já que estamos a considerar a expansão $\ln u$. A velocidade de convergência de $\ln n - h(x_n u)$ para $-\gamma\ln u$ depende da velocidade de convergência do limite dado por (3).

4 Conclusões

Do exposto se conclui que as aproximações em teoria dos valores extremos oferecem ainda muitas dificuldades, mesmo no âmbito restrito de limites de máximos de observações iid. Por outro lado, os comentários de Klebanov, Rachev e Szekely não podem ser facilmente descartados, e a famosa frase de Tukey — mais vale uma solução aproximada de um problema exacto, do que uma solução exacta de um problema formulado com aproximações — incentiva-nos a manter todo o interesse nesse importante avanço da Matemática do século XIX, que foi a notável ideia de aproximar uma função pela expansão em série envolvendo outra função e as suas derivadas, nomeadamente quando tal nos leva a famílias decentes de polinómios ortogonais. As famílias indecentes, para já, deixamos para quem tenha mais fibra moral para lidar com elas.

Referências

[1] Cramér, H. (1925). *On some classes of series used in mathematical statistics.* Em *Proceedings of the Sixth Scandinavian Congress of Mathematicians*, 399-425. Copenhaga.

[2] Cramér, H. (1999). *Mathematical Methods of Statistics* (19th ed.). Princeton.

[3] Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.*, 25, 631-650.

[4] Fisher, R. A. e Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Cambridge Philosophical Society*, 24, 180-190.

[5] Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics* (2nd ed.). EUA: Robert E. Krieger.

[6] Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.*, 44, 423-453.

[7] Gomes, M. I. (1984). Penultimate limiting forms in extreme value theory. *Ann. Inst. Statist. Math.*, 36, 71-85.

[8] Gomes, M. I. e Pestana, D. D. (1987). *Non-standard domains of attraction and rates of convergence.* Em *New Perspectives in Theoretical and Applied Statistics*, 467-478. Wiley.

[9] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* New York: Springer.

[10] Klebanov, L. B., Rachev, S. e Szekely, G. (1999), Pre-limit theorems and their applications. *Acta Applicandae Mathematicae*, 58, 159-174.

[11] Thiele, T. N. (1903). *The theory of observations.* Londres: Charles & Edwin Layton (http://www.archive.org).

[12] Uzgören, N. T. (1954). *The asymptotic development of the distribution of the extreme values of a sample.* Em *Studies in Mathematics and Mechanics Presented to Richard von Mises by Friends, Colleagues and Pupils*, 346-353, New York: Academic Press.