Teoria assintótica dos valores extremos

Nas aulas anteriores aprendemos resultados assintóticos de sequências de estatísticas obtidas de variáveis aleatórias independentes e identicamente distribuidas. Em particular estudamos a Lei Forte (Fraca) dos Grandes Números e o Teorema do Limite Central:

Sejam $X_1, X_2, \ldots, X_n \overset{iid}{\sim} X$, com $\mathbb{E}[X] = 0$ e $\mathrm{Var}(X) = 1$. Considere $\overline{X} = \sum_{i=1}^{n} X_i/n$.

Pela Lei Forte dos Grandes Números $\overline{X} \overset{q.c.}{\to} 0$ consequentemente $\overline{X} \overset{P}{\to} 0$ Como a convergência em probabilidade implica convergência em distribuição , temos

$$\overline{X} \overset{D}{\to} \begin{cases} 1 \text{ se } x \geq 0 \\ 0 \text{ se } x < 0 \end{cases}.$$

Contudo, se de certa forma, padronizamos a média amostral temos outro resultado $\sqrt{n} \cdot \overline{X} \overset{D}{\to} N(0,1)$
que é mais interessante.

Em resumo, procuramos $(a_n)_{n\geq 1}$, com $a_n > 0$ e $(b_n)_{n\geq 1}$ tais que

$$\frac{\sum_{i=1}^{n} X_i - b_n}{a_n} \overset{D}{\to} N(0,1).$$

Pelo Teorema do Limite Central

$$a_n = \sqrt{n\mathrm{Var}(X)} \text{ e } b_n = n\mathbb{E}[X]$$

são soluções.

Assim acontece com as propriedades assintóticas dos valores extremos.

Sejam $X_1, X_2, \ldots, X_n \overset{iid}{\sim} X$ tal que $X \sim U(0,1)$. Considere $X_{(n;1)} = \min(X_1, \ldots, X_n)$.

$$X_{(n;1)} \overset{P}{\to} 0 \text{ e } X_{(n;1)} \overset{D}{\to} \begin{cases} 1 \text{ se } x \geq 0 \\ 0 \text{ se } x < 0 \end{cases}.$$

Contudo

$$n \cdot X_{(1)} \overset{D}{\to} Exp(1).$$

As propriedades assintóticas dos valores extremos são importantes nas aplicações estatísticas. É desejável estimar o risco máximo que decorre das aplicações financeiras ao longo de horas, dias, meses, anos consecutivos. É importante estimar o número de componentes de um sistema em série para que funcione adequadamente por um longo período de tempo. Óbviamente, quando n converge para o infinito, $X_{(n;1)}$ converge quase certamente para o extremo inferior do suporte da função de distribuição F e converge em distribuição para este mesmo valor. Para evitar tais trivialidades padronizaremos $X_{(n;1)}$ por um parâmetro de escala a_n e um parâmetro de translação b_n e analisaremos o limite em distribuição da sequência padronizada

$$(\frac{X_{(n;1)} - b_n}{a_n})_{n \geq 1}.$$

Assumiremos, também para evitar trivialidades, que a distribuição limite é não degenerada.

O mesmo ocorre quando analisamos a sequência de máximos de variáveis aleatórias independentes e identicamente distribuidas . Observe qua $\max\{X_1, ..., X_n\} = -\min\{-X_1, ..., -X_n\}$ e que se $G(x)$ é o limite em distribuição de $X_{(n;1)}$, $\overline{G}(-x)$ é o limite em distribuição de $X_{(n;n)}$

As distribuições limites para a sequência de mínimos (máximos) de variáveis aleatórias independentes e identicamente distribuidas são de três tipos. Procuramos $(a_n)_{n\geq 1}$, com $a_n > 0$ e $(b_n)_{n\geq 1}$ tais que

$$\frac{X_{(1)} - b_n}{a_n} \overset{D}{\underset{?}{\rightarrow}} \begin{cases} W_1(x) = 1 - e^{-(x)^{\alpha}} \text{ se } x \geq 0,\ \alpha > 0 \\ W_2(x) = 1 - e^{-(-x)^{-\alpha}} \text{ se } x < 0,\ \alpha > 0 \\ \wedge(x) = 1 - e^{-e^{x}},\ \ -\infty < x < \infty \end{cases}.$$

Procuramos $(a_n)_{n\geq 1}$, com $a_n > 0$ e $(b_n)_{n\geq 1}$ tais que

$$\frac{X_{(n)} - b_n}{a_n} \overset{D}{\underset{?}{\rightarrow}} \begin{cases} W_1^*(x) = e^{-(-x)^{\alpha}} \text{ se } x \leq 0,\ \alpha > 0 \\ W_2^*(x) = e^{-x^{-\alpha}} \text{ se } x > 0,\ \alpha > 0 \\ \wedge^*(x) = e^{-e^{x}},\ \ -\infty < x < \infty \end{cases}.$$

A distribuição $W_1(x)$ é denominada distribuição de Weibull, na engenharia é utilizada para analisar dureza de certos materiais. A distribuição de Frechét, $W_2(x)$, é de pequeno interesse nas aplicações desde que é confinada a um suporte negativo e não pode originar-se como um limite em distribuição de variáveis aleatórias não negativas. Mesmo embora $\Lambda(x)$ permite valores negativos em seu domínio, é de algum interesse pois pode originar-se como o limite em distribuição de mínimos de tempos de vida, utilizados em seguros de vida e estudos de mortalidade. A distribuição $W_2^*(x)$ é de pequeno interesse nas aplicações desde que é confinada a um suporte negativo e não pode originar-se como um limite em distribuição de variáveis aleatórias não negativas.

A distribuição de Gumbel, $\Lambda^*(x)$ é apropriada para descrever medidas de valores extremos como altas temperaturas, velocidade do vento, $\Lambda(x)$ é utilizada como modelo para descrever o mínimo de temperaturas e pressão.

Note que se $X_1, ..., X_n$ são variáveis aleatórias i.i.d, com distribuição F,

$$P\left(X_{(n;1)} > x\right) = P\left(\min(X_1, \ldots, X_n) > x\right) \overset{iid}{=} \prod_{i=1}^{n} P\left(X_1 > x\right) = P\left(X_1 > x\right)^n$$

e

$$P\left(\frac{X_{(n;1)} - b_n}{a_n} > x\right) = P\left(X_{(n;1)} > a_n x + b_n\right) = \overline{F}(a_n x + b_n)^n,$$

onde $\overline{F}(x) = 1 - P(X \leq x)$.

Exemplo 0.1. Considere a distribuição de Weibull, $F(x) = 1 - e^{-x^\alpha}$, $\alpha > 0$ e $x \geq 0$

$$\overline{F}^n(a_n x + b_n) = e^{-n(a_n x + b_n)^\alpha}$$

Escolhemos $b_n = 0$ e $a_n = n^{-1/\alpha}$, $n = 1, 2, \ldots$ Segue que $\overline{F}^n(a_n x + b_n) = e^{-x^\alpha}$.

Concluímos que a distribuição de Weibull é fechada na formação de mínimos de V.A.(s).

Exemplo 0.2. Seja X uma variável aleatória uniformemente distribuida no intervalo $(0,1)$. $X \sim U(0,1)$, $F(x) = x$, $0 < x < 1$, $\overline{F}(x) = 1 - x$, $0 < x < 1$, $a_n = \frac{1}{n}$, $b_n = 0$, $n = 1, 2, \ldots$

$$\overline{F}^n(a_n x + b_n) = \left(1 - \frac{x}{n}\right)^n \to e^{-x}.$$

Definição 0.3. Duas funções de distribuições $H(\cdot)$ e $G(\cdot)$ são do mesmo tipo se existem constantes A e B, $A > 0$ tais que

$$G(Ax + B) = H(x) \forall\ x.$$

Se

$$X \sim N(\mu, \sigma^2) \text{ com distribuição } F$$
$$Z \sim N(0,1) \text{ com distribuição } \phi$$

$$F(x) = \phi\left(\frac{x - \mu}{\sigma}\right)$$

e as distribuições F e ϕ são do mesmo tipo.

Lema 0.4. *(Lema Fundamental) Seja $(F_n)_{n\geq 1}$ uma sequência de funções de distribuições tal que, para todo x valem a) e b):*

a) $F_n(a_n x + b_n) \xrightarrow{D} G(x)$
b) $F_n(a_n^* x + b_n^*) \xrightarrow{D} G^*(x)$,
onde $G(\cdot)$ e $G^*(\cdot)$ não são degeneradas. Então $G(\cdot)$ e $G^*(\cdot)$ são do mesmo tipo.

OBS: como consequência se obtemos $G(x)$ como o limite em distribuição, podemos obter as distribuições do mesmo tipo que $G(\cdot)$, modificando apropriadamente as sequências $(a_n)_{n\geq 1}$ e $(b_n)_{n\geq 1}$.

Observação 0.5. Se $(X_n)_{n\geq 1}$ é uma sequência de variáveis aleatórias iid, com função de distribuição F, inversível, observe que

$$P\left(nF\left(X_{(n;1)}\right) > x\right) = P\left(F\left(X_{(n;1)}\right) > \frac{x}{n}\right) = P\left(X_{(n;1)} > F^{-1}\left(\frac{x}{n}\right)\right)$$
$$= \left(P\left(X_1 > F^{-1}\left(\frac{x}{n}\right)\right)\right)^n = \left(1 - P\left(X_1 \leq F^{-1}\left(\frac{x}{n}\right)\right)\right)^n = \left(1 - \frac{x}{n}\right)^n (*)$$

Exemplo 0.6. Seja $F(x) = 1 - x^{-\alpha},\ x \geq 1$

$$P\left(\frac{X_{(n;1)} - b_n}{a_n} > x\right) = P\left(X_{(n;1)} > a_n x + b_n\right) = P\left(F\left(X_{(n;1)}\right) > F\left(a_n x + b_n\right)\right)$$
$$= P\left(nF\left(X_{(n;1)}\right) > nF\left(a_n x + b_n\right)\right) \stackrel{*}{=} \left(1 - \frac{nF\left(a_n x + b_n\right)}{n}\right)^n$$
$$= \left[1 - \left(1 - (a_n x + b_n)^{-\alpha}\right)\right]^n = (a_n x + b_n)^{-n\alpha} \to e^{-x^{-\alpha}},$$

em que $a_n = -\frac{1}{n}$ e $b_n = 1$.

Definição 0.7. Uma função de distribuição $G(\cdot)$ é estável através do mínimo se para todo inteiro positivo k, existem constantes α_k e b_k tais que

$$\overline{G}^k(\alpha_k x + \beta_k) = \overline{G}(x)\ \forall\ x \in \mathbb{R}.$$

Exemplo 0.8. A) A distribuição de Gumbel é estável através do mínimo

$$\wedge(x) = 1 - e^{-e^x}, -\infty < x < \infty$$

Tomando $\alpha_k = 1$ e $\beta_k = -\ln(k)$ temos $e^{x-\ln(k)} = e^x e^{-\ln(k)} = e^x e^{\ln(1/k)} = \frac{1}{k}e^x$

$$\overline{\wedge}(\alpha_k x + \beta_k) = e^{-\frac{1}{k}e^x}$$

$$\overline{\wedge}^k(\alpha_k x + \beta_k) = \left(e^{-\frac{1}{k}e^x}\right)^k = e^{-e^x} = \overline{\wedge(x)}$$

B) A distribuição de Weibull é estável através do mínimo $W_1(x) = 1 - e^{-(x)^\alpha}$, $x \geq 0$. Tomando $\alpha_k = \frac{1}{k^{1/\alpha}}$ e $\beta_k = 0$ temos $\overline{W}_1(x) = e^{-(x)^\alpha}$ e

$$\overline{W}_1\left(\alpha_k x + \beta_k\right) = \overline{W}_1\left(\frac{x}{k^{1/\alpha}}\right) = e^{-\left(\frac{(x)^\alpha}{k}\right)}$$

$$\overline{W}_1^k\left(\alpha_k x + \beta_k\right) = e^{-\left(\frac{x^\alpha}{k}\right)k} = e^{-x^\alpha} = \overline{W}_1(x).$$

Teorema 0.9. *A função de distribuição $G(\cdot)$ é limite em distribuição de $X_{(n;1)}$ se, e somente se, é estável através do mínimo.*

Demonstração. A condição é necessária: Por hipótese temos: $\lim \overline{F}^n(a_n x + b_n) = \overline{G}(x)$.

Note que

$$\overline{F}^{nk}(a_{nk}x + b_{nk}) \xrightarrow{D} \overline{G}(x)$$

$$\left[\overline{F}^n(a_{nk}x + b_{nk})\right]^k \xrightarrow{D} \overline{G}(x),$$

que pode ser escrito como $\overline{F}^n(a_{nk}x + b_{nk}) \xrightarrow{D} \overline{G}(x)^{1/k}$. Segue do Lema Fundamental que $\overline{G}(x)$ e $\overline{G}^{1/k}(x)$ são do mesmo tipo e portanto existem constantes α_k e β_k tais que

$$\overline{G}^k(\alpha_k x + \beta_k) = \overline{G}(x).$$

Para provar que a condição é suficiente, basta notar que $G(\cdot)$ é o limite em distribuição de $X(1)$ com $X_1, \ldots, X_n$ sendo iid a $G(\cdot)$:

$$P\left(\frac{X_{(n;1)} - b_n}{a_n} > x\right) = \overline{G}^n(a_n x + b_n).$$

Tomando $a_n = \alpha_n$ e $b_n = \beta_n$, temos

$$\lim_{n\to\infty} P\left(\frac{X_{(n;1)} - b_n}{a_n} > x\right) = \lim_{n\to\infty} \overline{G}^n(\alpha_n x + \beta_n) =$$

$$\lim_{n\to\infty} \overline{G}(x) = \overline{G}(x).$$

□

Seja $G(\cdot)$ estável através do mínimo, isto é, $\exists\ \alpha_k > 0$ e β_k tal que

$$\overline{G}^k(\alpha_k x + \beta_k) = \overline{G}(x)\ \forall\ x \qquad (**)$$

No que segue, as provas omitidas são encontradas no livro Statistical theory of reliability and lifetesting: probability models, Barlow and Proschan, 1981.

Lema 0.10. *(Lema 1) Sejam α_k e β_k constantes satisfazendo $(**)$ para uma distribuição $G(\cdot)$. Então, para todo positivo inteiro j e k temos*

$$\alpha_{jk} = \alpha_j \alpha_k,$$
$$\beta_{jk} = \beta_k + \alpha_k \beta_j = \beta_j + \alpha_j \beta_k.$$

(Prova omitida)

Lema 0.11. *(Lema 2) Se em $(**)$ $\alpha_j = 1$ para algum $j > 1$, então $\alpha_j = 1$ para todo j.*

(Prova omitida)

Exemplo 0.12. A distribuição de Gumbel, $\wedge(x) = 1 - e^{-e^x}$ é estável através do mínimo. $\alpha_k = 1$ e $\beta_k = -\ln(k)$.

Lema 0.13. *(Lema 3) Se em* $(**)$ $\alpha_j < 1$ *para algum* $j > 1$*, então*

a) *Existe* x_0 *tal que* $G(x_0) = 0$ *e* $G(x) > 0$ *para todo* $x > x_0$.
b) $\beta_k/(1-\alpha_k) = x_0$ *para todo* $k > 1$.

(Prova omitida)

Exemplo 0.14. A distribuição de Weibull, $W_1(x) = 1 - e^{-x^\alpha}$ é estável através do mínimo. $\alpha_k = \frac{1}{k^{1/\alpha}} < 1$ e $\beta_k = 0$.

Lema 0.15. *(Lema 4) Se em* $(**)$ $\alpha_j > 1$ *para algum* $j > 1$*, então*

a) *Existe* x_0 *tal que* $G(x_0) = 1$ *e* $G(x) < 1$ *para todo* $x < x_0$.
b) $\beta_k/(1-\alpha_k) = x_0$ *para todo* $k > 1$.

(Prova omitida)

Lema 0.16. *(Lema 5) Sejam* $\alpha_j,\ j = 1, 2, \ldots$ *as constantes em* $(**)$. *Então:*

a) $\alpha_j < 1$ *para todo* $j > 1$*, ou*
b) $\alpha_j = 1$ *para todo* $j \geq 1$*, ou*
c) $\alpha_j > 1$ *para todo* $j > 1$

(Prova omitida)

No lema 3 e lema 4 podemos, sem perda de generalidade, tomar $x_0 = 0$. Assim o problema de encontrar uma distribuição estável através do mínimo se reduz a encontar solução para

$$\overline{G}^n(\alpha_n x) = \overline{G}(x) \text{ com } G(x) = 0 \text{ para } x \leq 0 \ (\alpha_n < 1)$$

ou

$$\overline{G}^n(\alpha_n x) = \overline{G}(x) \text{ com } G(x) = 1 \text{ para } x \geq 0 \ (\alpha_n > 1)$$

ou

$$\overline{G}^n(x + \beta_n) = \overline{G}(x)(\alpha_n = 1).$$

E-mail address: `bueno@@ime.usp.br`

Departamento de Estatística, Instituto de Matemática e Estatística, Universidade de São Paulo, Caixa Postal 66281, CEP 05311-970, São Paulo, Brazil