Fluctuations of Maxima

Modelling Extremal Events: Chap 3

庄源

2024 年 4 月 16 日

南开大学精算学系

目录 i

章节框架介绍与示例

章节结构

- 主题：研究最大值的极限分布
- 3.1：累积分布函数 F 在满足什么条件时，能使这个极限分布存在？退一步讲，能否使中心化和标准化后的 M_n 依分布收敛于某个分布？
- 3.2：如果真的能收敛于一个简单的分布（不是 0 或 1 那样的常数)，那么这个分布是什么?
- 3.3：我怎样知道，对于给定的 F，它会收敛于极值分布中的哪一个？与中心极限定理相似，仿射变换的参数序列 c_n 和 d_n 到底是什么?
- 3.4：能否把三种极值分布归为一类？这些分布又跟广义帕累托分布（GPD）有什么关系?
- 3.5：有没有比依分布收敛更强的结论？(如几乎处处收敛)

凡例

定义、命题、定理、引理或推论

公式、定理编号均与原书相同。

图片的源代码均可在个人主页上找到。

评论

对应原书 Remark 部分。

例题

本章例题较多。

3.1 Limit Probabilities for Maxima

最大值随机变量的累积分布函数

- $X,\ X_1,\ X_2,\ \ldots$ 为 iid、非退化的随机变量，累积分布函数为 F；
- 定义样本最大值 M_n：

$$M_1 = X_1, \quad M_n = \max\left(X_1, \ldots, X_n\right), \quad n \geq 2$$

- 对于 iid 的 $X_1,\ X_2,\ \ldots$，有：

$$\begin{aligned} P\left(M_n \leq x\right) &= P\left(X_1 \leq x, \ldots, X_n \leq x\right) \\ &\overset{\text{iid}}{=} P\left(X_1 \leq x\right) \times \ldots \times P\left(X_n \leq x\right) = F^n(x), \quad x \in \mathbb{R}, \quad n \in \mathbb{N} \end{aligned}$$

最大值随机变量的收敛性

- 定义 x_F 为 F 的右端点，即 $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$；
- 若 $x < x_F$：

$$P\left(M_n \leq x\right) = F^n(x) \to 0, \quad n \to \infty$$

- 若 $x_F \leq \infty$，$x \geqslant x_F$：

$$P\left(M_n \leq x\right) = F^n(x) = 1$$

- M_n 是 n 的单调不减函数，则可以印证一个直觉性的结论：

$$M_n \xrightarrow{\text{a.s.}} x_F, \quad n \to \infty \tag{3.1}$$

寻找仿射变换下的收敛

- 是否存在两列数 $c_n > 0$，$d_n \in \mathbb{R}$，使得

$$c_n^{-1}\left(M_n - d_n\right) \xrightarrow{d} H, \quad n \to \infty \tag{3.2}$$

- 下式中，$u_n = c_n x + d_n$：

$$\begin{aligned} H(x) &= \mathrm{P}\left(c_n^{-1}\left(M_n - d_n\right) \leq x\right) \\ &= \mathrm{P}\left(M_n \leq c_n x + d_n\right) = \mathrm{P}\left(M_n \leq u_n\right) \end{aligned} \tag{3.3}$$

- F 满足什么性质时，$\mathrm{P}\left(M_n \leq u_n\right)$ 能够有收敛性，并不要收敛于一个简单的数?

泊松近似

Proposition 3.1.1 (Poisson Approximation)

对于给定 $\tau \in [0, \infty]$ 和一列实数 (u_n)，下面两式等价：

$$n\bar{F}\left(u_n\right) \rightarrow \tau, \tag{3.4}$$

$$\mathrm{P}\left(M_n \leq u_n\right) \rightarrow e^{-\tau}. \tag{3.5}$$

Page 116，证明见讲义

泊松近似（cont'd）

Remark：泊松近似的来源

在 $0 \leq \tau < \infty$ 时，令

$$B_n = \sum_{i=1}^{n} I_{\{X_i > u_n\}} \sim B\left(n, \bar{F}\left(u_n\right)\right)$$

则

$$B_n \xrightarrow{d} \mathrm{Poi}(\tau) \iff \mathrm{E}(B_n) = n\bar{F}\left(u_n\right) \to \tau$$

于是有：

$$P\left(M_n \leq u_n\right) = \mathrm{P}\left(B_n = 0\right) \to \exp\{-\tau\}$$

在何时，$\mathrm{P}\left(M_n \leq u_n\right)$ 会收敛于简单的 0 或 1

Corollary 3.1.2

假设 $x_F < \infty$ 且

$$\bar{F}\left(x_F -\right) = F\left(x_F\right) - F\left(x_F -\right) > 0.$$

那么对于任意的序列 $\left(u_n\right)$ 都有：

$$P\left(M_n \leq u_n\right) \rightarrow \rho,$$

要么 $\rho = 0$，要么 $\rho = 1$。

Page 117，证明见讲义

在何时，$\mathrm{P}\left(M_n \leq u_n\right)$ 会收敛于简单的 0 或 1（cont'd）

Corollary 3.1.2

如果一个分布在很接近右端点（右端点有限大）处仍有跳，M_n 的非退化极限分布不存在。

Page 117，证明见讲义

无穷大右端点下的收敛条件

Theorem 3.1.3

令 F 为具有右端点 $x_F \leq \infty$ 的累积分布函数，再令 $\tau \in (0, \infty)$。存在满足 $n\bar{F}(u_n) \to \tau$ 的序列 (u_n)，当且仅当

$$\lim_{x \uparrow x_F} \frac{\bar{F}(x)}{\bar{F}(x-)} = 1 \tag{3.6}$$

当 F 为离散型分布时，$\bar{F}(x-) = \bar{F}(x-1)$。

几个无穷大右端点下离散分布的例子

例题 3.1.4–3.1.6

对于 Poisson 分布、几何分布和负二项分布，证明式 3.6 并不成立：

$$\lim_{x \uparrow x_F} \frac{\bar{F}(x)}{\bar{F}(x-)} = 1 \tag{3.6}$$

Page 118–119，证明见讲义

3.2 Weak Convergence of Maxima Under Affine Transformations

最大稳定分布

收敛到的分布是什么?

定义 3.2.1 最大稳定分布

如果对于 iid 的 X, $X_1, \ldots, X_n$，有合适的序列 $c_n > 0$，$d_n \in \mathbb{R}$ 和每个 $n > 2$，都有:

$$\max(X_1, \ldots, X_n) \stackrel{d}{=} c_n X + d_n \tag{3.7}$$

则称一个非退化的随机变量 X（对应的分布或累积分布函数）是最大稳定的。则有:

$$c_n^{-1}(M_n - d_n) \stackrel{d}{=} X \tag{3.8}$$

最大稳定分布和最大值极限分布的关系

定理 3.2.2 最大稳定分布和最大值极限分布的关系

最大稳定分布刚好就是任意可能的独立同分布随机变量最大值（经过中心化标准化）的极限分布。

证明方式：两种分布是“同一种”、“同一类”。

Page 121，证明见讲义

同分布（same distribution）和同类（same type）

定义

同分布（same distribution）为：

$$X \overset{d}{=} Y$$

同类（same type）为，存在 $a \in \mathbb{R}$ 和 $b > 0$，有：

$$X \overset{d}{=} bY + a$$

定理 A1.5：Convergence to Types Theorem

令 $A, B, A_1, A_2, \ldots$ 为随机变量，$b_n > 0, \beta_n > 0, a_n, \alpha_n \in \mathbb{R}$ 为常数，假设 $b_n^{-1}(A_n - a_n) \xrightarrow{d} A$，那么下面的关系：

$$\beta_n^{-1}(A_n - \alpha_n) \xrightarrow{d} B \tag{A.2}$$

成立，当且仅当

$$\lim_{n\to\infty} b_n/\beta_n = b \in [0, \infty), \quad \lim_{n\to\infty} (a_n - \alpha_n)/\beta_n = a \in \mathbb{R}. \tag{A.3}$$

如果 (A.2) 成立，那么 $B \stackrel{d}{=} bA + a$，且 a, b 是唯一能让关系成立的常数。

向某类型收敛（Convergence to Types, cont’d）

定理 A1.5：Convergence to Types Theorem

令 $A, B, A_1, A_2, \ldots$ 为随机变量，$b_n > 0, \beta_n > 0, a_n, \alpha_n \in \mathbb{R}$ 为常数，假设 $b_n^{-1}(A_n - a_n) \xrightarrow{d} A$，那么下面的关系：

$$\beta_n^{-1}(A_n - \alpha_n) \xrightarrow{d} B \tag{A.2}$$

成立，当且仅当

$$\lim_{n\to\infty} b_n/\beta_n = b \in [0, \infty), \quad \lim_{n\to\infty} (a_n - \alpha_n)/\beta_n = a \in \mathbb{R}. \tag{A.3}$$

当 (A.2) 成立，A 非退化当且仅当 $b > 0$，A 和 B 属于同一类。

非高斯领域的中心极限定理：Fisher-Tippett theorem

Theorem 3.2.3: Fisher-Tippett theorem

让 (X_n) 是独立同分布随机变量序列。如果存在规范化参数 $c_n > 0$，$d_n \in \mathbb{R}$ 和一些非退化的累积分布函数 H 使得

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} H, \tag{3.9}$$

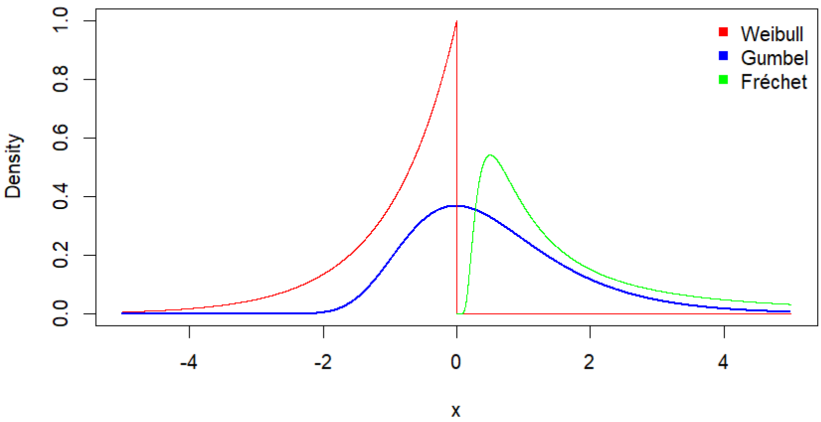
那么 H 属于以下三种分布的其中一种：

$$\text{Fréchet：}\quad \Phi_\alpha(x) = \begin{cases} 0, x \leq 0 \\ \exp\left\{-x^{-\alpha}\right\}, x > 0 \end{cases} \quad \alpha > 0.$$

$$\text{Weibull：}\quad \Psi_\alpha(x) = \begin{cases} \exp\left\{-(-x)^{\alpha}\right\}, x \leq 0 \\ 1, x > 0 \end{cases} \quad \alpha > 0.$$

$$\text{Gumbel：}\quad \Lambda(x) = \exp\left\{-e^{-x}\right\}, \quad x \in \mathbb{R}.$$

Fisher-Tippett 定理中的三种分布



Page 122，Fig 3.2.4，Fréchet 和 Weibull 分布的 α 都选为 1。

Fisher-Tippett theorem 的证明思路

对于所有 $t > 0$，有:

$$F^{[nt]}\left(c_{[nt]}x + d_{[nt]}\right) \to H(x), \quad x \in \mathbb{R}$$

也有:

$$F^{[nt]}\left(c_n x + d_n\right) = \left(F^n\left(c_n x + d_n\right)\right)^{[nt]/n} \to H^t(x)$$

由 Covergence to types theorem，存在函数 $\gamma(t) > 0, \delta(t) \in \mathbb{R}$ 满足

$$\lim_{n\to\infty} \frac{c_n}{c_{[nt]}} = \gamma(t), \quad \lim_{n\to\infty} \frac{d_n - d_{[nt]}}{c_{[nt]}} = \delta(t), \quad t > 0,$$

且

$$H^t(x) = H(\gamma(t)x + \delta(t)) \tag{3.10}$$

Fisher-Tippett theorem 的证明思路（cont’d）

由

$$H^t(x) = H(\gamma(t)x + \delta(t)) \tag{3.10}$$

有

$$\gamma(st) = \gamma(s)\gamma(t), \quad \delta(st) = \gamma(t)\delta(s) + \delta(t) \tag{3.11}$$

上述函数方程有三类解。

关于 Fisher-Tippett theorem 的几点注意事项：仿射变换

- 极限定理在仿射变换下也成立，只不过需要更换规范化参数。如果：

$$\lim_{n\to\infty} P\left(c_n^{-1}(M_n - d_n) \leq x\right) = H(cx+d)$$

那么，对于 $\tilde{c}_n = c_n/c$ 和 $\tilde{d}_n = d_n - dc_n/c$，有

$$\lim_{n\to\infty} P\left(\tilde{c}_n^{-1}\left(M_n - \tilde{d}_n\right) \leq x\right) = H(x)$$

关于 Fisher-Tippett theorem 的几点注意事项：Weibull 分布的定义

- 我们学过的 Weibull 分布定义域一般在 $[0, +\infty)$ 上：

$$F_\alpha(x) = 1 - e^{-x^\alpha}, \quad x \geq 0$$

- 书中的 Weibull 分布和我们学过的 Weibull 分布可以这样转换：

$$\Psi_\alpha(x) = 1 - F_\alpha(-x), \quad x < 0$$

- 部分精算教材中将 $\Psi_\alpha(x)$ 称为 “Weibull EV”，但在本书中，为了传统，还是使用 $\Psi_\alpha(x)$ 作为 Weibull 分布。

关于 Fisher-Tippett theorem 的几点注意事项：极值分布间的关系

- 虽然三种分布看起来十分不同，但它们有这样的转换关系：

$$X \text{ has df } \Phi_\alpha \iff \ln X^\alpha \text{ has df } \Lambda \iff -X^{-1} \text{ has df } \Psi_\alpha$$

Denition 3.2.6

在定理 3.2.3 中列出的 Φ_α, Ψ_α 和 Λ 叫作标准极值分布，相应的随机变量叫标准极值随机变量。Φ_α, Ψ_α 和 Λ 这一类的分布叫作极值分布，相应的随机变量叫极值随机变量。

标准极值分布的最大稳定特性

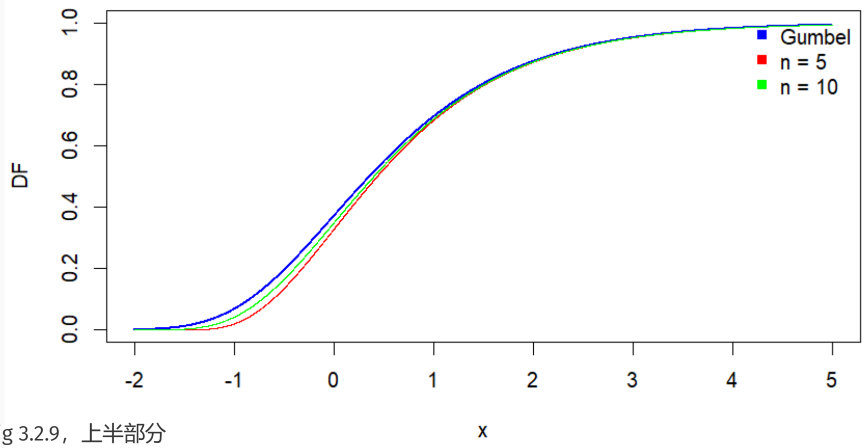
- Fréchet: $M_n \overset{d}{=} n^{1/\alpha} X$
- Weibull: $M_n \overset{d}{=} n^{-1/\alpha} X$
- Gumbel: $M_n \overset{d}{=} X + \ln n$

在不同的 F 下，各自最大值的极限分布是什么？

例 3.2.7 指数分布最大值的极限分布

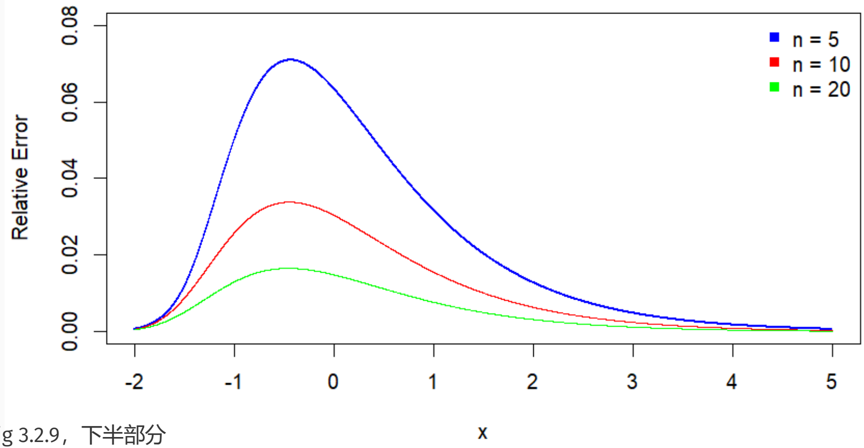
$$\begin{aligned} P\left(M_n - \ln n \leq x\right) &= \left(P(X \leq x + \ln n)\right)^n \\ &= \left(1 - n^{-1} e^{-x}\right)^n \\ &\to \exp\left\{-e^{-x}\right\} = \Lambda(x), \quad x \in \mathbb{R} \end{aligned}$$

从图像查看最大值的依分布收敛特性：指数 VS Gumbel



Page 126，Fig 3.2.9，上半部分

使用 Gumbel 分布近似多个指数变量最大值的相对误差



Page 126，Fig 3.2.9，下半部分

在不同的 F 下，各自最大值的极限分布是什么？(cont'd)

例 3.2.8 柯西分布最大值的极限分布

柯西分布的 df 为：

$$f(x) = \left(\pi\left(1 + x^2\right)\right)^{-1}, \quad x \in \mathbb{R}$$

所以有：

$$\lim_{x\to\infty} \frac{\bar{F}(x)}{(\pi x)^{-1}} = \lim_{x\to\infty} \frac{f(x)}{\pi^{-1}x^{-2}} = \lim_{x\to\infty} \frac{\pi x^2}{\pi\left(1 + x^2\right)} = 1$$

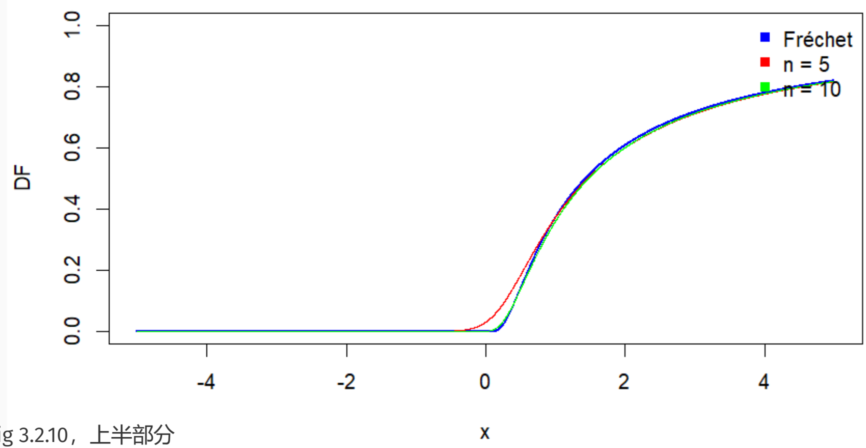
也就是 $\bar{F}(x) \sim (\pi x)^{-1}$。

在不同的 F 下，各自最大值的极限分布是什么？(cont'd)

例 3.2.8 柯西分布最大值的极限分布

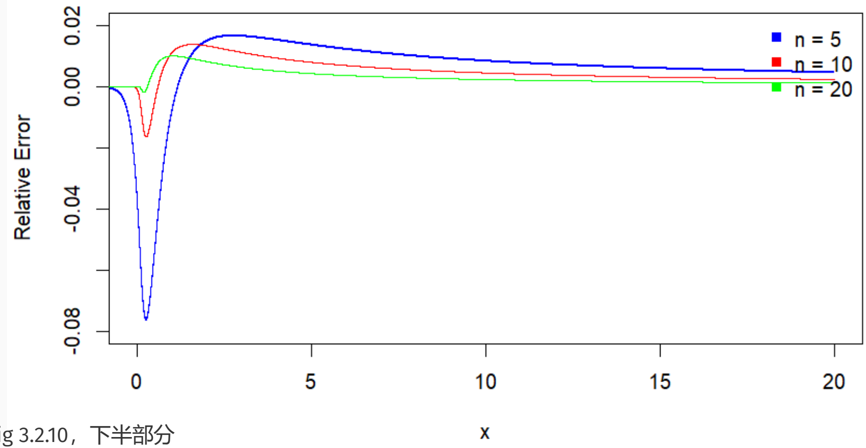
$$\begin{aligned} P\left(M_n \leq \frac{nx}{\pi}\right) &= \left(1 - \bar{F}\left(\frac{nx}{\pi}\right)\right)^n \\ &= \left(1 - \frac{1}{nx} + o(1)\right)^n \\ &\to \exp\left\{-x^{-1}\right\} = \Phi_1(x), \quad x > 0 \end{aligned}$$

从图像查看最大值的依分布收敛特性：柯西 VS Fréchet



Page 127，Fig 3.2.10，上半部分

使用 Fréchet 分布近似多个柯西变量最大值的相对误差



Page 127，Fig 3.2.10，下半部分

3.3 Maximum Domains of Attraction and Norming Constants

动机

- c_n 和 d_n 应该怎样选取？（其实在上一节中已经告诉了我们）
- 如何很快就知道，分布 F 下最大值的极限分布是 Fréchet、Weibull 和 Gumbel 中的哪一个？（最大吸引域和冯·米塞斯条件）

最大吸引域

Definition 3.3.1

如果存在常数 $c_n > 0, d_n \in \mathbb{R}$ 满足 $c_n^{-1}(M_n - d_n) \xrightarrow{d} H$，就说随机变量 X（X 的累积分布函数 F，X 的分布）属于极值分布 H 的最大吸引域。该关系可写为 $X \in \mathrm{MDA}(H)$ $(F \in \mathrm{MDA}(H))$。

上述条件也相当于在说：

$$\lim_{n\to\infty} P\left(M_n \le c_n x + d_n\right) = \lim_{n\to\infty} F^n\left(c_n x + d_n\right) = H(x), \quad x \in \mathbb{R}$$

最大吸引域的表征（后续证明最大吸引域的方式）

Proposition 3.3.2

F 属于极值分布 H 的最大吸引域，具有规范化参数 $c_n > 0, d_n \in \mathbb{R}$ 当且仅当

$$\lim_{n\to\infty} n\bar{F}\left(c_n x + d_n\right) = -\ln H(x), \quad x \in \mathbb{R}.$$

$H(x) = 0$ 时，极限视作 ∞。

Page 128，证明直接使用 Poisson 近似

其它已讨论过的基础知识

- 正则变化相关知识（附录 A3.1）
- 尾等价（Definition 3.3.3）

Definition 3.3.5：分位数函数

$$F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}, \quad 0 < t < 1$$

定义了 F 的 t 分位数。

Thanks!