Gumbel Max-Stability

Motivation

The "family preservation" puzzle: When computing box intersections, we need to take maxima and minima of Gumbel-distributed coordinates. A crucial question arises: does the result remain Gumbel-distributed?

This is like asking: "If I take the maximum of two Gumbel-distributed numbers, do I get another Gumbel-distributed number, or do I get something else entirely?" If the answer were "something else", we'd be in trouble—we'd lose the mathematical structure that makes volume calculations tractable. We'd have to compute volumes for a different distribution, and then another, and another, creating an infinite regress.

Max-stability is the property that answers this question affirmatively. It ensures that the maximum of independent Gumbel random variables is itself Gumbel-distributed (after appropriate normalization). This **algebraic closure** property is what makes Gumbel boxes mathematically elegant: operations on Gumbel boxes produce Gumbel boxes, preserving the family throughout all computations. It's like a mathematical "closed system"—you can perform operations within the system without ever leaving it.

Historical context: Max-stability is a fundamental concept in extreme value theory. The Fisher-Tippett-Gnedenko theorem (Fisher & Tippett, 1928; Gnedenko, 1943) states that there are only three possible limiting distributions for normalized maxima: Gumbel (Type I), Fréchet (Type II), and Weibull (Type III). This theorem is the extreme value theory analog of the Central Limit Theorem: just as the CLT characterizes the limiting distribution of sums (the normal distribution), the Fisher-Tippett theorem characterizes the limiting distribution of maxima. The Gumbel distribution is the only one of these that is max-stable in the sense defined above—this is why it appears so naturally in problems involving extremes. When we take maxima of Gumbel random variables, we're operating within the "natural" family for extreme values, which is why the result remains Gumbel-distributed.

Connection to block maxima: In extreme value theory, the "block maxima" approach considers the maximum value in each block of observations. For example, if we have daily temperature data, we might consider the maximum temperature in each year. The Fisher-Tippett theorem tells us that, under appropriate conditions, these block maxima converge in distribution to one of the three extreme value distributions. The Gumbel distribution (Type I) arises when the underlying distribution has an exponential tail, which includes many common distributions like the normal, exponential, and gamma distributions. This universality property—that many different underlying distributions lead to the same limiting distribution for maxima—is what makes extreme value theory so powerful and why Gumbel distributions are so widely applicable.

Definition

A distribution G is **max-stable** if, for any $n \geq 1$, there exist constants $a_n > 0$ and b_n such that:

$$[G(a_n x + b_n)]^n = G(x)$$

Here, $[G(x)]^n$ denotes the n-th power of the CDF, which equals the CDF of $\max(X_1, \dots, X_n)$ where $X_1, \dots, X_n$ are independent random variables with distribution G. This means the maximum of n independent samples (after appropriate scaling and translation) has the same distribution as a single sample. In other words, the distribution is "closed" under the maximum operation.

Statement

Theorem (Gumbel Max-Stability). If $G_1, ..., G_k \sim \text{Gumbel}(\mu, \beta)$ are independent, then:

$$\max(G_1, ..., G_k) \sim \text{Gumbel}(\mu + \beta \ln k, \beta)$$

The location parameter shifts by $\beta \ln k$, preserving the Gumbel family.

Proof

The CDF of the maximum of k independent random variables is the product of their CDFs (since the maximum is $\leq x$ if and only if all k variables are $\leq x$):

$$P(\max(G_1, ..., G_k) \leq x) = \prod_{i=1}^{k} P(G_i \leq x) = [G(x)]^k$$

For Gumbel distribution $G(x) = e^{-e^{-\frac{x-\mu}{\beta}}}$:

$$[G(x)]^k = \left(e^{-e^{-\frac{x-\mu}{\beta}}}\right)^k = e^{-ke^{-\frac{x-\mu}{\beta}}}$$

The CDF of $\text{Gumbel}(\mu + \beta \ln k, \beta)$ is:

$$e^{-e^{-\frac{x-(\mu+\beta \ln k)}{\beta}}} = e^{-e^{-\frac{x-\mu}{\beta}+\ln k}} = e^{-ke^{-\frac{x-\mu}{\beta}}}$$

This matches exactly, proving max-stability.

Min-Stability

Corollary (Min-Stability). For MinGumbel, if $G_1, ..., G_k \sim \text{MinGumbel}(\mu, \beta)$ are independent:

$$\min(G_1, ..., G_k) \sim \text{MinGumbel}(\mu - \beta \ln k, \beta)$$

The location parameter shifts by $-\beta \ln k$, preserving the MinGumbel family.

Proof: The key insight is the relationship between MinGumbel and MaxGumbel via negation. If $G_i \sim \text{MinGumbel}(\mu, \beta)$, then $-G_i \sim \text{MaxGumbel}(-\mu, \beta)$. This follows from the definition: MinGumbel has CDF $F(x) = 1 - e^{-e^{\frac{x-\mu}{\beta}}}$, so $P(-G_i \leq x) = P(G_i \geq -x) = e^{-e^{\frac{-x-\mu}{\beta}}} = e^{-e^{-\frac{x+\mu}{\beta}}}$, which is the CDF of $\text{MaxGumbel}(-\mu, \beta)$.

Using the identity $\min(G_1, ..., G_k) = -\max(-G_1, ..., -G_k)$ and applying max-stability to the negated variables:

$$\max(-G_1, ..., -G_k) \sim \text{MaxGumbel}(-\mu + \beta \ln k, \beta)$$

Therefore:

$$\min(G_1, ..., G_k) = -\max(-G_1, ..., -G_k) \sim -\text{MaxGumbel}(-\mu + \beta \ln k, \beta) = \text{MinGumbel}(\mu - \beta \ln k, \beta)$$

This establishes min-stability: the minimum of independent MinGumbel random variables is itself MinGumbel-distributed, with the location parameter shifted by $-\beta \ln k$.

Why It Matters

Max-stability is not merely a theoretical curiosity—it is essential for the computational tractability of Gumbel box embeddings. When we compute box intersections (see the Log-Sum-Exp and Gumbel Intersection document):

- **Intersection minimum**: $z_{\{\text{cap}\}} = \max(\min_A, \min_B)$ remains MinGumbel-distributed (by max-stability applied to the minimum coordinates)
- **Intersection maximum**: $Z_{\{\text{cap}\}} = \min(\max_A, \max_B)$ remains MaxGumbel-distributed (by min-stability applied to the maximum coordinates)

This **algebraic closure** property means:
1. **Analytical tractability**: We can compute expected volumes using the Bessel function formula (see the Gumbel-Box Volume document) at every step, not just for initial boxes. The formula applies recursively to intersections of intersections.
2. **Consistent structure**: The mathematical framework remains consistent throughout all operations—intersections, unions, and compositions. Every operation produces Gumbel-distributed coordinates.
3. **Computational efficiency**: We avoid the need for numerical integration or Monte Carlo methods at each step. All volume calculations reduce to evaluating the Bessel function K_0.

Why the scale parameter stays the same: A crucial insight is that max-stability preserves the scale parameter β. When we take the maximum of k Gumbel random variables, only the location parameter shifts (by $\beta \ln k$), while β remains unchanged. This ensures that all boxes in a computation share the same scale parameter, maintaining consistency and enabling the use of a single temperature parameter throughout the model.

The logarithmic shift $\beta \ln k$ appears because the Gumbel CDF has the form $e^{-e^{-\frac{x-\mu}{\beta}}}$. When we raise this to the k-th power (for the maximum of k independent variables), we get $e^{-ke^{-\frac{x-\mu}{\beta}}}$. To match the standard Gumbel form, we need $ke^{-\frac{x-\mu}{\beta}} = e^{-\frac{x-\mu-\beta \ln k}{\beta}}$, which requires shifting the location parameter by $\beta \ln k$. This logarithmic dependence on k is characteristic of extreme value distributions and reflects the fact that taking maxima is a multiplicative operation in the exponential scale.

Intuition for the logarithmic shift: The shift $\beta \ln k$ can be understood intuitively: as we take the maximum of more independent Gumbel random variables, the expected value increases. However, the increase is sublinear (logarithmic) rather than linear. This is because extreme values grow slowly: doubling the number of samples doesn't double the expected maximum—it only increases it by $\beta \ln 2$. This logarithmic growth is a fundamental property of extreme value distributions and is why the Gumbel distribution is sometimes called the “double exponential” distribution: it involves exponentials of exponentials, leading to this logarithmic scaling behavior.

Without max-stability, each intersection would produce a different distribution family, making analytical calculations impossible and forcing us to resort to expensive numerical methods.

Example

Three independent Gumbel random variables:
- $G_1, G_2, G_3 \sim \text{Gumbel}(0, 1)$

Max-stability property:

$$\max(G_1, G_2, G_3) \sim \text{Gumbel}(\ln 3, 1) = \text{Gumbel}(1.099, 1)$$

Verification: The CDF of the maximum is:

$$P(\max(G_1, G_2, G_3) \leq x) = \left[e^{-e^{-x}}\right]^3 = e^{-3e^{-x}} = e^{-e^{-(x-\ln 3)}}$$

which is the CDF of $\text{Gumbel}(\ln 3, 1)$.