Exercises 1-3: Extreme Values

Valentina Ros, Alberto Rosso

Exercise 1: Gaussian tails and Gumbel scaling

In this exercise, we study the statistics of the minimum of M independent Gaussian random variables. Contrary to the bounded case, the distribution is unbounded from below, and the minimum is controlled by the asymptotic behavior of the left tail of the distribution.

Let $E_1, \ldots, E_M$ be independent, identically distributed random variables drawn from a Gaussian distribution with zero mean and variance σ^2. Throughout the exercise, we work in the limit of large M.

The cumulative distribution of a single variable is

$$P(E) = \int_{-\infty}^{E} \frac{dx}{\sqrt{2\pi\sigma^2}} \, e^{-x^2/2\sigma^2}. \tag{1}$$

Questions

1. Show that, for $E \to -\infty$, the cumulative distribution admits the asymptotic form

$$P(E) = \frac{\sigma}{\sqrt{2\pi}\,|E|} \, e^{-E^2/2\sigma^2} \left[1 + O\left(\frac{1}{E^2}\right)\right]. \tag{2}$$

   Tip: perform an integration by parts, or use the change of variable $t = x^2/2\sigma^2$.

2. Rewrite the asymptotic expression of $P(E)$ in exponential form,

$$P(E) = \exp\big(A(E)\big), \tag{3}$$

   and show that $E_{\min}$ satisfies the implicit equation

$$\log M \simeq \frac{E_{\min}^2}{2\sigma^2} + \log\left(\frac{\sqrt{2\pi}\,|E_{\min}|}{\sigma}\right). \tag{4}$$

   Tip: take the logarithm of the condition $M\,P(E_{\min}) \sim 1$.

3. Determine the leading behavior of the typical minimum for large M, and show that

$$E_{\min}^{(0)} \simeq -\sigma\sqrt{2\log M}. \tag{5}$$

   Tip: neglect subleading logarithmic terms.

4. Compute the first logarithmic correction to this result, and show that

$$E_{\min} \simeq -\sigma\sqrt{2\log M} + \sigma\,\frac{\operatorname{loglog} M + \log(4\pi)}{2\sqrt{2\log M}}. \tag{6}$$

   Tip: introduce the variable $y = |E_{\min}|/\sigma$ and write $y = \sqrt{2\log M} - \delta$ with $\delta \ll \sqrt{2\log M}$.

Exercise 2: Weakest-link statistics and the Weibull law

In this exercise, we consider a situation that is qualitatively different from the Gaussian case. The random variables are bounded from below, and the minimum is controlled by the behavior of the distribution close to the edge of its support. This will naturally lead to a different choice of scaling parameters a_L and b_L.

Consider a chain of length L subjected to a tensile force F. The chain breaks when its weakest link breaks. We denote by F_c the force required to break the chain.

Let $x_1, x_2, \dots, x_L$ be the breaking strengths of the individual links. They are assumed to be independent, identically distributed, and strictly positive random variables.

Throughout the exercise, we work in the limit of large L.

The strength of each link is drawn from a Gamma distribution with shape parameter $\alpha > 0$:

$$p(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-x}, \qquad x \geq 0. \tag{7}$$

Questions

1. Compute the typical breaking force F_c^{typ} of the chain and determine its dependence on L. *(Hint: use the condition $L\,P(x) = O(1)$, where $P(x) = \int_0^x p(t)\,dt$.)*

2. The breaking force of the chain is equal to the minimum of the L random variables. According to extreme value statistics, its distribution satisfies

$$Q_L(x) \equiv \text{Prob}(F_c > x) \sim \exp[-LP(x)]\,, \qquad P(x) = \int_0^x p(t)\,dt. \tag{8}$$

   Show that the appropriate scaling form is obtained by introducing

$$a_L = 0, \qquad b_L = F_c^{\text{typ}}, \tag{9}$$

   and defining the rescaled variable

$$z = \frac{x - a_L}{b_L}. \tag{10}$$

   Determine the limiting, L-independent distribution of z, and identify the corresponding universality class.

Exercise 3: Number of states close to the minimum

In this exercise, we study a further consequence of extreme value statistics. Rather than focusing only on the minimum energy, we characterize the number of states lying close to it. This quantity will play an important role in the discussion of the Random Energy Model.

Let $\{E_1, E_2, \dots, E_M\}$ be M independent, identically distributed random energies, and denote by

$$E_{\min} = \min\{E_1, \dots, E_M\}$$

their minimum.

Definition of $n(x)$. For a given realization of the disorder, define

$$n(x) = \#\{\, i \mid E_{\min} < E_i < E_{\min} + x \,\}, \tag{11}$$

that is, the number of states lying within an energy window x above the minimum. This is a random variable.

We are interested in its disorder average

$$\overline{n(x)} = \sum_{k=0}^{M-1} k \,\mathrm{Prob}[n(x) = k]. \tag{12}$$

Final goal. Show that, in the large-M limit, provided the minimum belongs to the *Gumbel universality class*,

$$\overline{n(x)} = e^{x/b_M} - 1, \tag{13}$$

where b_M is the scaling parameter introduced in the extreme value analysis.

Questions

1. Show that the probability that exactly k states lie in the interval $(E_{\min}, E_{\min} + x)$ can be written exactly as

$$\mathrm{Prob}[n(x) = k] = M\binom{M-1}{k}\int_{-\infty}^{\infty} dE\, p(E)\,[P(E+x)-P(E)]^k\,[1-P(E+x)]^{M-k-1}, \tag{14}$$

where $p(E)$ and $P(E)$ denote the probability density and cumulative distribution of the energies.

Tip: condition on the value of the minimum energy E and count the number of remaining variables in the interval.

2. Using the combinatorial identity

$$\sum_{k=0}^{M-1} k\binom{M-1}{k}(A-B)^k B^{M-1-k} = (M-1)(A-B)A^{M-2}, \tag{15}$$

show that the disorder average can be rewritten as

$$\overline{n(x)} = M(M-1)\int_{-\infty}^{\infty} dE\, p(E)\,[P(E+x) - P(E)]\,[1 - P(E)]^{M-2}. \tag{16}$$

3. Introducing

$$Q_{M-1}(E) = [1 - P(E)]^{M-1}, \tag{17}$$

show that $\overline{n(x)}$ can be written in the form

$$\overline{n(x)} = -M\int_{-\infty}^{\infty} dE\,[P(E+x) - P(E)]\frac{dQ_{M-1}(E)}{dE}. \tag{18}$$

Tip: use $p(E) = -\frac{d}{dE}[1 - P(E)]$.

4. Assume now that the extreme value statistics of the energies is described by the Gumbel distribution. Argue that for large M one may replace $Q_{M-1}(E) \simeq Q_M(E)$, and use the scaling form

$$-\frac{dQ_M(E)}{dE}\,dE \sim \exp\left(\frac{E-a_M}{b_M}\right)\exp\left[-\exp\left(\frac{E-a_M}{b_M}\right)\right]\frac{dE}{b_M}. \tag{19}$$

5. Introduce the scaling variable

$$z = \frac{E - a_M}{b_M}, \tag{20}$$

and argue that the dominant contribution to the integral comes from the region $E \simeq a_M$, where

$$P(E) \simeq \frac{1}{M} e^{(E-a_M)/b_M}. \tag{21}$$

Tip: recall that $Q_M(E)$ is sharply peaked around the typical minimum.

6. Compute the resulting integral and show that

$$\overline{n(x)} = \left(e^{x/b_M} - 1\right) \int_{-\infty}^{\infty} dz\, e^{2z - e^z} = e^{x/b_M} - 1. \tag{22}$$