ICFP M2 - Statistical physics 2
Homework n° 1

Grégory Schehr, Francesco Zamponi

In the TD 1 we have studied the distribution of the maximum M_n of a large number n of independent and identically distributed random variables $X_1, \dots, X_n$. One can investigate more detailed extremal properties of such large samples of random variables, for instance:

- what is the law of the second largest variable among $X_1, \dots, X_n$?
- what is the law of the k-th largest variable among $X_1, \dots, X_n$, for arbitrary k ?

To answer some of these questions we suggest the following approach. First, we recall a few results:

- If a_n and b_n are the series introduced in the TD that define the rescaling under which $(M_n - a_n)/b_n$ has a non-trivial limit, we have

$$F_X(a_n + b_n x) = 1 - \frac{\gamma(x)}{n} + o(1/n) \quad \text{i.e.} \quad \lim_{n\to\infty} F_X(a_n + b_n x)^n = G(x) = e^{-\gamma(x)} \ , \tag{1}$$

  where $G(x)$ is the cumulative distribution function of the rescaled variable $(M_n - a_n)/b_n$.

- A binomial distribution

$$p(k) = \mathrm{Binom}(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k} \tag{2}$$

  converges to a Poisson distribution when $n \to \infty$ with fixed $\lambda = pn$, i.e.

$$p(k) = \mathrm{Pois}(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda} \ . \tag{3}$$

- A multinomial distribution

$$\begin{aligned} p(k_1, \cdots, k_m) &= \mathrm{Multinom}(k_1, \cdots, k_m; n, p_1, \cdots, p_m) \\ &= \frac{n!}{k_1! \cdots k_m! (n - k_1 - \cdots - k_m)!} p_1^{k_1} \cdots p_m^{k_m} (1 - p_1 - \cdots - p_m)^{n - k_1 - \cdots - k_m} \end{aligned} \tag{4}$$

  converges to a product of independent Poisson distributions when $n \to \infty$ with fixed $\lambda_i = p_i n$, i.e.

$$p(k_1, \cdots, k_m) \quad \to \quad \mathrm{Pois}(k_1; \lambda_1) \cdots \mathrm{Pois}(k_m; \lambda_m) \ . \tag{5}$$

Keeping in mind these results:

- From the independent random variables $X_1, \dots, X_n$ define $\widehat{X}_1, \dots, \widehat{X}_n$ with $\widehat{X}_i = (X_i - a_n)/b_n$.
- Call $N_n([u, v])$ the (random) number of points $\widehat{X}_i$ among $\widehat{X}_1, \dots, \widehat{X}_n$ which falls in the interval $[u, v]$.
- Determine the probability distribution of $N_n([u, v])$, and of its limit $N([u, v])$ as $n \to \infty$.
- Characterize the joint law of $N_n([u_1, v_1]), \dots, N_n([u_p, v_p])$ when the intervals $[u_i, v_i]$ are disjoint, and then take the limit $n \to \infty$.
- Find back from this approach the distribution of the maximum derived in the TD.
  Hint: consider the probability that the maximum is smaller than x, and express it in terms of the variable $N([x, \infty[)$.
- Generalize this result to the k-th maximum.