Distance between the largest and second largest, and the density of states near the largest

EVS Lectures 3 and 4
(Dated: November 10, 2013)

I. INTRODUCTION

As we saw on the example of random energy model, extrem value statistics (EVS) gives us the distribution of the minimum energy E_{min}. The exponential distribution of E_{min} we obtained is an interesting results but to arrive at measurable quantities such as e.g. heat capacity at low temperatures, we would need the value of the first excited state as well or, in general, the density of states just above E_{min}.

In EVS the above problem is formulated as a question: How lonely is it at the top? Another version of the question: How crowded is it at the top?

The problem of crowding near extreme values comes up not only in physical problems such as the random energy model. Other examples are:

- Insurance: In addition to the value of largest losses, the knowledge about the number of near extreme losses is also crucial.
- Climate: The number of near extreme hurricanes are as important as the strength of the largest one.
- Astrophysics: The value of using the brightest galaxies as standard candles depends essentially on the brightness gap between the brightest and second brightest.
- Optimalization: Finding the exact optimum is impossible in many practical problems. One would like to know how many near optimal solutions exists and how close are they expected to be to the exact optimum.

In the following, first we shall calculate the gap between the largest and second largest for independent, identically distributed (i.i.d.) variables. Then we shall extend the calculation to the density of states near the maximum.

II. GAP BETWEEN THE LARGEST AND SECOND LARGEST: I.I.D. VARIABLES

A. General considerations

We have already calculated the distribution $P_1(x)$ of the largest value in scaled variable x which is related to the original variable z through

$$z = a_N x + b_N \tag{1}$$

where the coefficients a_N and b_N depend on the parent distribution and on the batch size N. Let's assume that we also know the distribution of the second larges $P_2(x)$ in the same scaling variable. Then the average distance between the largest and second largest $\Delta_{1,2}$ in the scaled variable is given by

$$\Delta_{1,2}(\gamma) = \langle x_1 \rangle - \langle x_2 \rangle = \int y P_1(y) dy - \int y P_2(y) dy \tag{2}$$

where γ is explicitely written in $\Delta_{1,2}(\gamma)$ in order to indicate the expected dependence on the parameter γ specifying the EVS limit distribution.

In the original variable z, we have from (1)

$$\langle z_1 \rangle - \langle z_2 \rangle = a_N(\langle x_1 \rangle - \langle x_2 \rangle) = a_N \Delta_{1,2}(\gamma) \,. \tag{3}$$

Since $\Delta_{1,2}(\gamma)$ is expected to be a finite number, we see that whether the gap increases, decreases, or stays finite in the $N \to \infty$ limit is determined by the factor a_N.

Thus, the gap is finite for an exponential parent distribution ($a_N \sim const$), it goes to zero for a Gaussian parent ($a_N \sim 1/\sqrt{\ln N}$), and it diverges for a power-law parent ($a_N \sim N^{1/\alpha}$).

B. Calculation of the second largest

Let the parent distribution in the original variable y be $p(y)$, and denote by $\mu(z) = \int_{-\infty}^{z} p(y)dy$ the integrated distribution giving the probability that $y < z$. Then the probability, $P_2(z)dz$, that the second largest is between z and $z + dz$ is given by

$$P_2(z)dz = N(N-1)\mu(z)^{N-2} p(z)dz[1-\mu(z)] \,. \tag{4}$$

Here $1 - \mu(z)$ is the probability that a selected measurement is above z, $p(z)dz$ is that one measurement is between z and $z + dz$ while $\mu(z)^{N-2}$ is the probability that the rest of the $N-2$ measurements is below z. The factor $N(N-1)$ is coming from combinatorics [note that the measurements are distinguishable and so, there is no factor 2 dividing $N(N-1)$].

Thus we have

$$P_2(z) = N(N-1)\mu(z)^{N-2} p(z)[1-\mu(z)] \,. \tag{5}$$

and we can also find the integrated probability distribution $M_2(z)$ for the second largest

$$M_2(z) = N\mu(z)^{N-1} - (N-1)\mu(z)^N \,. \tag{6}$$

In order to make the appropriate change of variables $z = a_N x + b_N$, it is convenient to write $M_2(z)$ as

$$M_2(z) = \mu(z)^N \left[1 + N\frac{1-\mu(z)}{\mu(z)}\right] . \quad (7)$$

We know that $\mu(z)^N = M_1(z)$ and, as a result of the change of variable $z = a_N x + b_N$, we shall have

$$\begin{aligned} \mu(a_N x + b_N)^N \to \mu(x)^N &= M_1(x) \\ &= \exp[-(1+\gamma x)^{-1/\gamma}] . \quad (8) \end{aligned}$$

The above form can also be used to write

$$\mu(x) = \exp[-(1+\gamma x)^{-1/\gamma}/N] . \quad (9)$$

Substituting the above expression into (7) and expanding in $1/N$ one obtains in the $N \to \infty$ limit

$$M_2(x) = M_1(x)[1 + (1+\gamma x)^{-1/\gamma}] . \quad (10)$$

Taking now the derivative by x gives us $P_2(x)$

$$P_2(x) = (1+\gamma x)^{-2/\gamma-1} \exp[-(1+\gamma x)^{-1/\gamma}] \quad (11)$$

The calculation of $\langle x_2 \rangle$ is simplified if we use another form of $P_2(x)$. Namely, let us write (10) in the following form

$$M_2(x) = M_1(x) + (1+\gamma x)P_1(x) . \quad (12)$$

Taking the derivative by x yields now

$$P_2(x) = (1+\gamma)P_1(x) + (1+\gamma x)P_1'(x) \quad (13)$$

and this form can be used to calculate $\langle x_2 \rangle$ as follows

$$\langle x_2 \rangle = \int_{-\infty}^{\infty} x P_2(x) dx = (1-\gamma)\langle x_1 \rangle - 1 . \quad (14)$$

Using the above form, we find the gap in the scaled variable

$$\Delta_{1,2}(\gamma) = \langle x_1 \rangle - \langle x_2 \rangle = \gamma \langle x_1 \rangle + 1 \quad (15)$$

where the γ-dependence is not only what we see explicitely, it is also through the γ-dependence of the average $\langle x_1 \rangle$.

For the original variable z, eq.(15) yields our final result

$$\langle z_1 \rangle - \langle z_2 \rangle = a_N(\gamma \langle x_1 \rangle + 1) . \quad (16)$$

C. Gaps for various γ-s

For $\gamma > 0$, we have the Frechet distribution and $a_N \sim N^\gamma$. Thus the distance between the largest and second largest diverges in the $N \to \infty$ limit.

The $\gamma = 0$ case (Fisher-Tippett-Gumber distribution) is special since the gap is just $\Delta_{1,2} = a_N$. Considering the frequently arising parent

$$p(y) \sim \exp(-y^\delta) \quad (17)$$

one has

$$a_N \sim [\ln N]^{-1+1/\delta} \quad (18)$$

and thus the limitig gap depends on δ

$$\langle z_1 \rangle - \langle z_2 \rangle = \begin{cases} \infty & \text{for } \delta < 1 , \\ \text{finite} \neq 0 & \text{for } \delta = 1 , \\ 0 & \text{for } \delta > 1 . \end{cases} \quad (19)$$

We note here that the above result remains valid for a more general class of parents frequently seen in physics: $p(y) \sim \exp(-y^\delta)/y^\alpha$.

Finally, we consider parents which have an upper threshold at a, and the parent approaches the threshold as $p(y) \sim (a-y)^\beta$. Then the limit distribution is the Weibull distribution, and the change of variable is given by $z = a_N x + a$ with $a_N \sim N^{-1/(\beta+1)}$. Thus, not very surprisingly, the gap goes to zero in the $N \to \infty$ limit.

III. DENSITY OF STATES NEAR THE MAXIMUM

The density of states near the minimum energy is needed e.g. if we would like to calculate the heat capacity of a system at low temperatures. The frequency (or density) of near maximum values is also important in general problems of EVS.

Definition of density of observed values at a distance r from the maximum is as follows:

$$\rho(r, N) = \frac{1}{N} \sum_{i=1}^{N-1} \delta(r - (y_{max} - y_i)) \quad (20)$$

where N is the number of draws (batch size) from a parent distribution $p(y)$ and y_{max} is the maximum for a given batch of draws $y_{max} = \max\{y_1, y_2, ..., y_N\}$. Note that the maximum is not counted in the the sum (20), this is why it is only over $N-1$ values.

We shall be interested in the average density

$$\langle \rho(r, N) \rangle = \frac{1}{N} \sum_{i=1}^{N-1} \langle \delta(r - (y_{max} - y_i)) \rangle \quad (21)$$

where the averaging $\langle ... \rangle$ means that a large number of batches of N values of ys is prepared, the maximum y_{max} is determined in each of the batches, and r is measured from y_{max} in the given batch and, finally, the values $r = y_{max} - y_i$ are averaged over the batches.

Since the number of nonmaximal y values is $N-1$, the integral of $\langle \rho(r, N) \rangle$ is given by

$$\int_0^\infty dr \langle \rho(r, N) \rangle = 1 - \frac{1}{N} . \quad (22)$$

In order to calculate $\langle\rho(r,N)\rangle$, let us remember that the distribution of the maximum value of y_{max} is given through the parent distribution as

$$P(y_{max}=z,N)\equiv P_{max}(z,N)=Np(z)\mu(z)^{N-1} \quad (23)$$

where $\mu(z)$ is the integrated parent distribution

$$\mu(z)=\int_{-\infty}^{z} dyp(y)\,. \quad (24)$$

We can express $\langle\rho(r,N)\rangle$ through $P_{max}(z,N)$ and the conditional density of state $\langle\rho_c(r,z,N)\rangle$ defined as the density of state at a fixed z value of y_{max}. Indeed, $\langle\rho(r,N)\rangle$ is just an average over the position of the maximum

$$\langle\rho(r,N)\rangle=\int_{-\infty}^{\infty} dz\langle\rho_c(r,z,N)\rangle P_{max}(z,N)\,. \quad (25)$$

In order to calculate the conditional density $\langle\rho_c(r,z,N)\rangle$, we should find the conditional probability of $N-1$ draws all being smaller than z (the largest $y_1<z, y_2<z, ..., y_{N-1}<z$ in the batch). This conditional probabilty is given by

$$P_{cond}(y_1,y_2,...,y_{N-1})= \quad (26)$$
$$\frac{p(y_1)}{\int_{-\infty}^{z} dy_1 p(y_1)}\frac{p(y_2)}{\int_{-\infty}^{z} dy_2 p(y_2)}\cdots\frac{p(y_{N-1})}{\int_{-\infty}^{z} dy_{N-1} p(y_{N-1})}\,.$$

Now, the conditional density is obtained as

$$\langle\rho_c(r,z,N)\rangle= \quad (27)$$
$$\int_{-\infty}^{\infty} dy_1 ... \int_{-\infty}^{\infty} dy_{N-1}\rho(r,N)P_{cond}(y_1,...,y_{N-1})$$

and substituting the expressions (20) and (26) in the above integral, we obtain

$$\langle\rho_c(r,z,N)\rangle=\frac{N-1}{N}\frac{p(z-r)}{\int_{-\infty}^{z} dyp(y)}=\frac{N-1}{N}\frac{p(z-r)}{\mu(z)}\,. \quad (28)$$

We can now evaluate $\langle\rho(r,N)\rangle$ by substituting (23) and (28) into (25)

$$\langle\rho(r,N)\rangle=\int_{-\infty}^{\infty} dz\frac{N-1}{N}\frac{p(z-r)}{\mu(z)}Np(z)\mu(z)^{N-2} \quad (29)$$

and obtain our final result

$$\langle\rho(r,N)\rangle=\int_{-\infty}^{\infty} dzp(z-r)P_{max}(z,N-1) \quad (30)$$

where we have used the equality $(N-1)p(z)\mu(z)^{N-1}=P_{max}(z,N-1)$.

In the following, we shall analyse the above expression in the large N limit ($N\to\infty$). We know that making the appropriate change of variables

$$z=a_N x+b_N\,, \quad (31)$$

the distribution of the maximum approaches a limit distribution

$$\lim_{N\to\infty} P_{max}(z=a_N x+b_N,N)dz\to P(x)dx \quad (32)$$

where $P(x)$ is a one-parameter family of distributions

$$P(x)=(1+\gamma x)^{-1/\gamma-1}\exp[-(1+\gamma x)^{-1/\gamma}] \quad (33)$$

with γ depending on the parent distribution, as discussed earlier.

The change of variables (31) suggests the same change of variables in the integral (30) thus yielding the following expression for $\langle\rho(r,N)\rangle$ in the the large N limit

$$\langle\rho(r,N)\rangle=\int_{-\infty}^{\infty} dx\, p(a_N x+b_N-r)\,P(x) \quad (34)$$

We have arrived to a form that is convenient to analyse in terms of some general properties of EVS. This is what we shall do in the rest of this lecture.

A. $a_N\to 0$ ($\gamma<0$)

In the Weibull class ($\gamma<0$), one has an upper limit ($y\le a$) for the parent distribution. The maximum converges to a and the change of variable (31) is of the form $z=a_N x-a$ where it is obvious from the convergence to a that $a_N\to 0$ for $N\to\infty$. Thus, the x-dependence in $p(a_N x+b_N-r)$ disappears and we have

$$\langle\rho(r,N)\rangle=p(a-r) \quad (35)$$

where the normalization of $P(x)$ ($\int_{-\infty}^{\infty} dxP(x)=1$) was used.

The $\langle\rho(r,N)\rangle=p(a-r)$ result is understandable. The threshold for the parent is the maximum value a and the density near the treshold is given by the density of draws near the maximum.

An interesting quantity to consider is the density of state in the limit of $r\to 0$, i.e. the density of states at the maximum. For the weibull class we obtain from (35)

$$\langle\rho(0,N)\rangle=p(a)\,. \quad (36)$$

Thus, the parent distribution at the treshold determines whether the density of state goes to zero, to a finite value, or to infinity.

B. $a_N\to 0$ ($\gamma=0$)

We have seen that $a_N\to 0$ occurs also in the Gumbel class ($\gamma=0$). E.g. $a_N\sim 1/\sqrt{\ln N}\to 0$ for a Gaussian parent and, in general, one has

$$a_N\sim(\ln N)^{-1+1/\delta} \quad \text{for a parent} \quad p(y)\sim\exp(-x^\delta)\,. \quad (37)$$

Thus $a_N \to 0$ for $\delta > 1$ for parents which have no threshold and decay to zero faster than an exponential (e^{-x}). The absence of threshold, however, yields a diverging b_N and, consequently, the simple Weibull result for the density of state becomes a bit more complicated

$$\langle \rho(r,N) \rangle = p(b_N - r) \,. \tag{38}$$

This result also means that the density of states at the maximum has a zero limit

$$\lim_{N\to\infty} \langle \rho(0,N) \rangle = p(b_N \to \infty) \to 0 \,. \tag{39}$$

We note that the above result also holds for more complicated parents [e.g. $\exp(-x^\delta)/x^\theta$ with $\delta > 1$] provided they decay to zero faster than an exponential.

C. $a_N \to \infty$ ($\gamma > 0$)

For parents decaying as a power law at large arguments, the limit distribution $P(x)$ is the Frechet distribution belonging to the $\gamma > 0$ EVS class. As we saw earlier, in this case, $a_N \sim N^\gamma \to \infty$ and $b_N = 0$, and equation (34) takes the form

$$\langle \rho(r,N) \rangle = \int_{-\infty}^{\infty} dx\, p(a_N x - r)\, P(x) \,. \tag{40}$$

Since $p(y \to \infty) \to 0$, contribution to the above integral comes only from $x \approx r/a_N$ and we can write

$$\langle \rho(r,N) \rangle = A P(r/a_N) \,. \tag{41}$$

The constant A is determined from the normalization condition and one obtains

$$\langle \rho(r,N) \rangle = \frac{1}{a_N} P\left(\frac{r}{a_N}\right) = \tag{42}$$
$$\frac{1}{N^\gamma}\left(1+\frac{\gamma\, r}{N^\gamma}\right)^{-1-1/\gamma} \exp\left[-\left(1+\frac{\gamma\, r}{N^\gamma}\right)^{-1/\gamma}\right] .$$

We should note that the character of the result is different as compared to the case of $a_N \to 0$. The result there does not carry any universality, it depends on the parent distribution. Here, we obtained that the density of the state depends on the parent only in the sense that the asymptotics of the parent determines the parameter γ, otherwise the density of state is the same for a large class of parents.

D. $a_N \to \infty$ ($\gamma = 0$)

For the Gumbel class ($\gamma = 0$), we have $a_N \to \infty$ if the parent decay slower than a simple exponential

$$a_N \sim (\ln N)^{-1+1/\delta} \,, \quad [p(y) \sim \exp(-x^\delta) \,,\, \delta < 1] \,. \tag{43}$$

In this case, both a_N and b_N diverge and the only contribution to the integral

$$\langle \rho(r,N) \rangle = \int_{-\infty}^{\infty} dx\, p(a_N x + b_N - r)\, P(x) \tag{44}$$

comes from $x \approx (r - b_N)/a_N$. We should remember that we are in the $\gamma = 0$ class, thus $P(x)$ is the Gumbel distribution and so, we obtain

$$\langle \rho(r,N) \rangle = \frac{1}{a_N} P\left(\frac{r-b_N}{a_N}\right) = \tag{45}$$
$$\frac{1}{a_N} \exp\left[-\frac{r-b_N}{a_N} - \exp\left(-\frac{x-b_N}{a_N}\right)\right] .$$

The above results also hold for more complicated parents [e.g. $\exp(-x^\delta)/x^\theta$ with $\delta < 1$] provided they decay to zero than than an exponential but faster than any power law.

E. $a_N \to$ finite ($\gamma = 0$)

The borderline case between $a_N \to 0$ and $a_N \to \infty$ is the case of exponential decay of the parent. In this case, we have $a_N \to a =$ finite and $b_N \to \infty$ and the x dependence cannot be eliminated from $p(ax + b_N - r)$. The result is a convolution of the parent with the Gumbel distribution

$$\langle \rho(r,N) \rangle = \int_{-\infty}^{\infty} dx\, p(ax + b_N - r)\, P(x) = R(r - b_N) \tag{46}$$

where

$$R(u) = \int_{-\infty}^{\infty} dx\, p(ax - u)\, \exp\left(-x - e^{-x}\right) \,. \tag{47}$$

F. General remarks for the density of states

Not very surprisingly, the structure of the results for the density of states parallel that of the results for the gap between the largest and the second largest. For Weibull and for Gumbel with $\delta > 1$, the gap closes in the large N limit, and the density of state is determined by the parent distribution near the largest. For Frechet and for Gumbel with $\delta < 1$, on the other hand, the gap between the largest and the second largest diverges, only the large scale features of the parent matter, and a universal density distribution emerges. The most complicated case is the simple exponential parent which is at the borderline between the above cases and the density distribution picks up some features of both the parent and the limit distribution.

IV. HOMEWORKS

Homework 4: Consider a mol of H_2 gas at room temperature. Estimate the expected maximum velocity in this gas. Estimate how accurately is the maximum velocity given. Estimate the difference between the largest and second largest velocities in the gas.

Homework 5: Distribution of the maximum daily temperatures at the Amistad Dam. See link **Homework No.5** on the homepage of the course.

Homework 6: Determine the average gap between the maximum and second maximum of daily temperatures at the Amistad Dam. For temperature data see link **Homework No.5** on the homepage of the course.

[1] S. Sabhapandit and S. N. Majumdar, *Density of near-extreme events*, Phys. Rev. Lett. **98** 140201 (2007).