Renormalization group approach to the statistics of extreme values and sums

Eric Bertin

Laboratoire de Physique, ENS Lyon

Collaboration with Géza Györgyi (Eötvös Univ., Budapest)

Itzykson conference, Saclay, 14-17 June 2011

Introduction and motivation

- Asymptotic distributions of extreme values of iid random variables known for long, but strong finite-size effects, not always easy to handle with standard probabilistic methods
- Idea: Use the renormalization language as a convenient tool to analyze fixed points and finite-size corrections, in spite of the absence of correlations
- Approach initiated by the Budapest group

  G. Györgyi, N. R. Moloney, K. Ozogány, and Z. Rácz, Phys. Rev. Lett. **100**, 210601 (2008).

  G. Györgyi, N. R. Moloney, K. Ozogány, Z. Rácz and M. Droz, Phys. Rev. E **81**, 041135 (2010).
- Aim of the present contribution: reformulate the results using a differential representation, which is more convenient

Outline

- Renormalization transform for extreme values of iid random variables
- Fixed points and linear stability
- Exact non-perturbative trajectories
- From extreme values to random sums

Outline

- Renormalization transform for extreme values of iid random variables
- Fixed points and linear stability
- Exact non-perturbative trajectories
- From extreme values to random sums

Basic renormalization idea

Extreme value statistics

- N iid random variables, distribution $\rho(x)$
- Integrated distribution $\mu(x) = \int_{-\infty}^{x} \rho(x')dx'$
- Integrated distribution for the maximum value

$$\mathrm{Prob}(\max(x_1, \ldots, x_N) < x) = \mu^N(x)$$

Decimation procedure

- Split the set of sufficiently large N random variables x_i into $N' = N/p$ blocks of p random variables each
- y_j the maximum value in the j^{th} block

$$\max(x_1, \ldots, x_N) = \max(y_1, \ldots, y_{N'})$$

- y_j are also i.i.d. random variables, with a distribution $\mu_p(y)$

$$\mu_p(y) = \mu^p(y)$$

Renormalization operation

Raising to a power and rescaling

$$[\hat{R}_p \mu](x) = \mu^p(a_p x + b_p)$$

- Necessity of scale and shift parameters a_p and b_p to lift degeneracy of the distribution
- Conditions to fix a_p and b_p to be specified later on

Parameterization of the flow

- p considered as continuous rather than discrete
- change of flow parameter $p = e^s$: distribution $\mu(x, s)$, parameters $a(s)$ and $b(s)$
- Parent distribution $\mu(x)$ obtained for $s = 0$

$$\mu(x, 0) = \mu(x)$$

Renormalization operation

Change of function

- double exponential form

$$\mu(x,s) = e^{-e^{-g(x,s)}}$$

- Link to the parent distribution: $g(x, s = 0) = g(x)$

Standardization conditions

- Conditions to fix the parameters $a(s)$ and $b(s)$

$$\mu(0,s) \equiv e^{-1}, \qquad \partial_x \mu(0,s) \equiv e^{-1}$$

- In terms of the function $g(x,s)$

$$g(0,s) \equiv 0, \qquad \partial_x g(0,s) \equiv 1$$

Renormalization operation

Renormalization of $\mu(x,s)$

$$\mu(x,s) \equiv [\hat{R}_s \mu](x) = \mu^{e^s}\big(a(s)x + b(s)\big)$$

Renormalization of $g(x,s) = -\ln[-\ln \mu(x,s)]$

$$g(x,s) = g\big(a(s)x + b(s)\big) - s.$$

Very simple transformation: linear change of variable in the argument and global additive shift.

However, one needs to determine $a(s)$ and $b(s)$.

Differential representation

Iteration of the RG transformation

$$g(x, s + \Delta s) = [\hat{R}_{\Delta s} g](x, s)$$

Infinitesimal transformation $\Delta s = ds$

$$g(x, s + ds) = [\hat{R}_{ds} g](x, s)$$

- More explicitly, with $a(ds) = 1 + \gamma(s) ds$ and $b(ds) = \eta(s) ds$:

$$g(x, s + ds) = g\Big(\big(1 + \gamma(s) ds\big) x + \eta(s) ds, s\Big) - ds$$

  where the functions $\gamma(s)$ and $\eta(s)$ are to be specified
- Linearizing with respect to ds, we get

$$\partial_s g(x, s) = \big(\gamma(s) x + \eta(s)\big) \partial_x g(x, s) - 1$$

Determination of $\gamma(s)$ and $\eta(s)$

- Standardiz. conditions $g(0,s) \equiv 0$ and $\partial_x g(0,s) \equiv 1$ yield

$$\eta(s) \equiv 1$$

$$\gamma(s) = -\partial_x^2 g(0,s)$$

Partial differential equation of the flow

$$\partial_s g(x,s) = (1+\gamma(s)x)\partial_x g(x,s) - 1$$

Outline

- Renormalization transform for extreme values of iid random variables
- Fixed points and linear stability analysis
- Exact non-perturbative trajectories
- From extreme values to random sums

Fixed points of the flow

- Stationary solution $g(x,s) = f(x)$:

$$0 = (1 + \gamma x) f'(x) - 1$$

with $\gamma = -f''(0)$

- Using the standardization condition $f(0) = 0$

$$f(x;\gamma) = \int_0^x (1+\gamma y)^{-1} dy = \frac{1}{\gamma} \ln(1+\gamma x)$$

- Fixed point integrated distribution

$$M(x;\gamma) = e^{-e^{-f(x;\gamma)}} = e^{-(1+\gamma x)^{-1/\gamma}}$$

Easy way to recover the well-known generalized extreme value distributions, obtained here as a fixed line of the RG transformation

Linear perturbations

- Perturbation $\phi(x,s)$ introduced through

$$g(x,s) = f(x) + f'(x)\,\phi(x,s)$$

- Linearized partial differential equation

$$\partial_s \phi(x,s) = (1+\gamma x)\,\partial_x \phi(x,s) - \gamma\,\phi(x,s) - x\,\partial_x^2 \phi(0,s)$$

- Convergence properties to the fixed point distribution are obtained from the analysis of this PDE

Definition

- Perturbations of the form

$$\phi(x,s) = \epsilon(s)\,\psi(x)$$

- Standardiz. conditions for $\psi(x)$: $\quad \psi(0)=0, \quad \psi'(0)=0$
- To lift the ambiguity of the factorization $\epsilon(s)\,\psi(x)$, we impose $\psi''(0)=-1$, which sets the scale of ψ.
- The condition $\gamma(s) = -\partial_x^2 g(0,s)$ translates into

$$\epsilon(s) = \gamma(s) - \gamma$$

Equation for $\psi(x)$ (notation $\dot{\epsilon} \equiv \frac{d}{ds}$)

$$\dot{\epsilon}(s)\psi(x) = \epsilon(s)\big((1+\gamma x)\psi'(x) - \gamma\psi(x) + x\big)$$

Eigenfunctions

Solvability condition

- Equation can be solved only if

$$\frac{\dot{\epsilon}(s)}{\epsilon(s)} = \gamma'$$

  which implies

$$\epsilon(s) \propto e^{\gamma' s}$$

- Differential equation for $\psi(x)$:

$$(1+\gamma x)\psi'(x) = (\gamma+\gamma')\psi(x) - x$$

Solution for the Weibull and Fréchet cases ($\gamma \neq 0$)

$$\psi(x;\gamma,\gamma') = \frac{1+(\gamma'+\gamma)x-(1+\gamma x)^{\gamma'/\gamma+1}}{\gamma'(\gamma'+\gamma)}$$

in the range of x such that $1+\gamma x > 0$.

Solution for the Gumbel case ($\gamma = 0$)

$$\psi(x;\gamma') = \frac{1}{\gamma'^2}\left(1+\gamma' x - e^{\gamma' x}\right)$$

Empirical interpretation

- N variables in the block $\Rightarrow s = \ln N$
- Convergence $g(x, s = \ln N) \to f(x)$
- Corrections proportional to $e^{\gamma' s} \propto N^{\gamma'}$
  (if $\gamma' = 0$: logarithmic convergence in N).
- Interpretation of $\gamma' > 0$? Are there unstable solutions?

  $\Rightarrow$ Can we look at non-perturbative solutions?

Outline

- Renormalization transform for extreme values of iid random variables
- Fixed points and linear stability
- Exact non-perturbative trajectories
- From extreme values to random sums

Motivation

Unstable solutions around the fixed point may seem counterintuitive: can we find an example of full RG trajectory starting from an unstable direction?

Back to the equations: the Gumbel case

- Equation to be solved

$$\partial_s g(x,s) = (1+\gamma(s)x)\partial_x g(x,s) - 1$$

- Ansatz for the solution starting from $f(x) = x$

$$g(x,s) = x + \epsilon(s)\psi(x;\gamma'(s))$$

- Same as linear perturbation, except that γ' depends on s

Non-perturbative solutions

- Equation of the flow

$$\dot{\epsilon}\,\psi + \epsilon\,\dot{\gamma}'\;\partial_{\gamma'}\psi - (1+\epsilon x)\,\epsilon\,\partial_x\psi - \epsilon\,x = 0$$

- Specific properties of $\psi(x)$, resulting from the knowledge of its explicit form

$$\partial_{\gamma'}\psi = -\frac{2}{\gamma'}\psi + \frac{x}{\gamma'}\partial_x\psi$$

$$\partial_x\psi = \gamma'\psi - x$$

- This results in

$$\left(\dot{\epsilon} - 2\epsilon\frac{\dot{\gamma}'}{\gamma'} - \epsilon\gamma'\right)\psi + \left(\epsilon\frac{\dot{\gamma}'}{\gamma'} - \epsilon^2\right)x\partial_x\psi = 0$$

Equations for $\epsilon(s)$ and $\gamma'(s)$

$$\dot{\epsilon} = 2\epsilon^2 + \epsilon\gamma'$$

$$\dot{\gamma}' = \epsilon\gamma'.$$

- PDE for the RG flow $\Rightarrow$ system of two coupled nonlinear ordinary differential equations
- Only a restricted family of solutions, not the full flow
- Visualization in a two-dimensional parameter space (ϵ, γ').

Solution for ϵ

- Look for a parametric solution $\epsilon(\gamma')$
- One finds

$$\frac{d\epsilon}{d\gamma'} = \frac{2\epsilon}{\gamma'} + 1$$

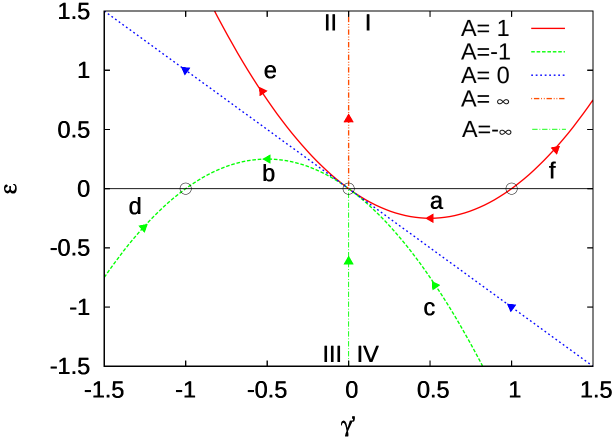
- Solution: parabola

$$\epsilon = A\gamma'^2 - \gamma' \,, \qquad A = \frac{\epsilon_0 + \gamma'_0}{\gamma'^2_0}$$

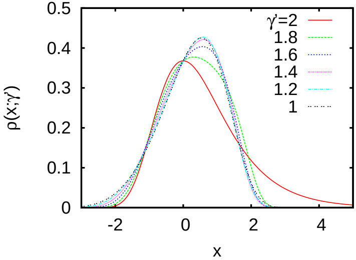
Implicit solution for s

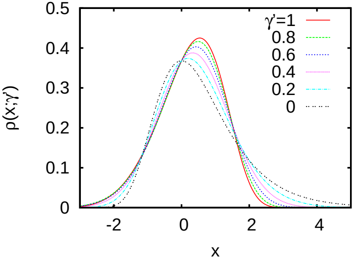
$$s(\gamma') = \frac{1}{\gamma'} - \frac{1}{\gamma'_0} + \frac{\epsilon_0 + \gamma'_0}{\gamma'^2_0} \ln\left(1 + \frac{\gamma'_0}{\epsilon_0} - \frac{\gamma'^2_0}{\epsilon_0 \gamma'}\right)$$

Parameter space (ϵ, γ')



Starting close to the Gumbel distribution ($\gamma' = 2$)... and coming back to it (at $\gamma' = 0$) after an excursion





Flow around the Fréchet distribution ($\gamma > 0$)

- Ansatz

$$g(x,s) = f\Big(x + \epsilon(s)\psi\big(x;\bar{\gamma}(s), B\bar{\gamma}(s)\big);\gamma_0\Big)$$

with

$$f(x;\gamma_0) = \frac{1}{\gamma_0}\ln(1+\gamma_0 x)$$

and

$$\psi(x;\gamma,\gamma') = \frac{1 + (\gamma'+\gamma)x - (1+\gamma x)^{\gamma'/\gamma+1}}{\gamma'(\gamma'+\gamma)}$$

Flow around the Fréchet distribution ($\gamma > 0$)

- Ansatz

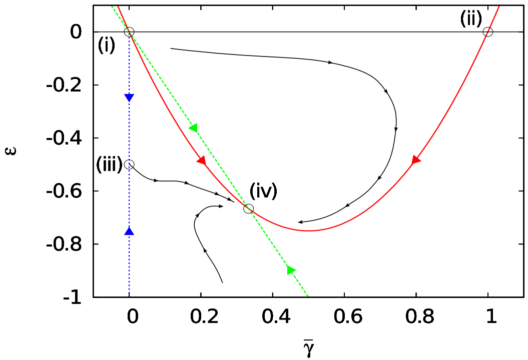
$$g(x,s) = f\left(x + \epsilon(s)\psi(x;\bar{\gamma}(s), B\bar{\gamma}(s)); \gamma_0\right)$$

- B constant parameter
- $\epsilon(s)$ and $\bar{\gamma}(s)$ two functions satisfying

$$\dot{\epsilon} = 2\epsilon^2 + \epsilon(\gamma_0 + (B+1)\bar{\gamma})$$

$$\dot{\bar{\gamma}} = \bar{\gamma}(\epsilon + \gamma_0 - \bar{\gamma})$$

Diagram of the flow (Fréchet)



Starting from a Fréchet distribution of parameter γ_0 [fixed points (i) and (ii)], one ends up at another Fréchet distribution of parameter $\gamma_1 \neq \gamma_0$ [fixed points (iii) and (iv)]

E. Bertin, G. Györgyi, J. Stat. Mech. P08022 (2010)

Outline

- Renormalization transform for extreme values of iid random variables
- Fixed points and linear stability
- Exact non-perturbative trajectories
- From extreme values to random sums

Formal analogy between sums and extremes

Extreme value statistics for iid random variables

- Relevant mathematical object: integrated distribution $\mu(x)$
- Integrated distribution of the maximum of N iid random variables

$$\mu_N(x) = \mu(x)^N$$

- Linear rescaling of x to preserve the standardiz. conditions

Statistics of sums of iid random variables

- Relevant mathematical object: characteristic function $\Phi(q)$
- Characteristic function for the sum of N iid random variables

$$\Phi_N(q) = \Phi(q)^N$$

- Linear rescaling

Same formal structure, only the objects differ

Renormalization transform for sums

- Random sum

$$Z = \frac{1}{a_N} \sum_{i=1}^{N} z_i$$

with z_i i.i.d. numbers each with density $P(z)$

- a_N scaling factor ensuring a non-degenerate limit distribution
- Characteristic (or moment generating) function

$$\Phi(q) = \int_{-\infty}^{\infty} dz\, e^{iqz}\, P(z)$$

- Restriction to even distributions $P(z) = P(-z)$ so that $\Phi(q)$ is real (makes RG calculations easier)
- One also has $\Phi(q) = \Phi(-q)$

Transformation of the characteristic function

- $\Phi_N(q)$ the characteristic function of the sum Z

$$\Phi_N(q) = \Phi^N(a_N q)$$

- Renormalization transform, with $s = \ln N$

$$\Phi(q, s) = [\hat{R}_s \Phi](q) = \Phi^{e^s}\big(a(s)\, q\big)$$

- Standardization conditions

$$\Phi(1, s) = \Phi(-1, s) \equiv e^{-1}$$

Remark: a single rescaling parameter $a(s)$ because $\langle z \rangle$ due to the parity of $P(z)$, no need for an additive shift

Double exponential form

- The function $h(q,s)$ is introduced as

$$\Phi(q,s) = e^{-e^{-h(q,s)}}$$

- Parity $h(q,s) = h(-q,s)$: considering only a half-axis in q is enough
- To prepare for the analogy to EVS, we shall consider the negative semi-axis, i.e. $q < 0$

Renormalization transform for $h(q,s)$

$$h(q,s) = h\big(a(s)\,q\big) - s$$

with standardiz. conditions $h(0,s) = +\infty$ and $h(-1,s) = 0$

Partial differential equation of the flow

- Along the same lines as for EVS, one can consider an infinitesimal renormalization transformation, eventually yielding

$$\partial_s h(q,s) = q\,\gamma(s)\,\partial_q h(q,s) - 1$$

- $\gamma(s)$ is given by

$$\gamma(s) = -\frac{1}{\partial_q h(-1,s)}$$

- Very similar to the PDE obtained for extreme values
- Differences between the PDEs arise from the different choices made for the standardization conditions

Fixed point solution

- One looks for a solution $h(q,s) = f(q)$
- Ordinary differential equation for $f(q)$

$$q\,\gamma\, f'(q) = 1, \qquad q < 0$$

- Solution satisfying $f(-1) = 0$

$$f(q;\gamma) = \frac{1}{\gamma} \ln(-q)$$

Result for the characteristic function

$$\Phi(q;\gamma) = e^{-|q|^{-\frac{1}{\gamma}}}$$

Characteristic function of the symmetric Lévy distribution, of parameter $\alpha = -1/\gamma$.

Also precisely corresponds to the original form of the Weibull distribution ($\gamma < 0$) obtained by Fisher and Tippett (1928).

Here, one restriction: $\gamma \leq -\frac{1}{2}$, equivalent to $0 < \alpha \leq 2$

Further comments

- Linear stability analysis (eigenfunctions, ...) can be performed in the same way as for extreme value statistics

  E. Bertin, G. Györgyi, J. Stat. Mech. P08022 (2010)

- Exact non-perturbative solutions describing the crossover from one fixed point to another can be given.

- Full analysis in the general case of an arbitrary distribution $P(z)$, without symmetry assumption

  E. Bertin, G. Györgyi, Z. Simon, in preparation

Conclusion

On the present work

- Renormalization is a convenient tool to analyze fixed points and finite size corrections
- Analysis of finite size corrections made easy by the use of eigenfunctions
- Emphasis put here on function space aspects, but the approach also allows one to determine the parameters γ and γ' from the parent distribution

What remains to be done?

- Can be applied to variants of the present problems, for instance, statistics of $\max(x_1^{q(n)}, \ldots, x_i^{q(n)})$
- Is renormalization without correlation really renormalization? Extension to correlated variables welcome... but yet unclear