Edinburgh Research Explorer

Constraint-Driven Condensation in Large Fluctuations of Linear Statistics

Citation for published version:

Digital Object Identifier (DOI):
10.1103/PhysRevLett.112.020602

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Physical Review Letters

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Constraint driven condensation in large fluctuations of linear statistics

Juraj Szavits-Nossan† and Martin R. Evans‡

SUPA, School of Physics and Astronomy, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, United Kingdom

Satya N. Majumdar†

Laboratoire de Physique Théorique et Modèles Statistiques, UMR 8626, Université Paris Sud 11 and CNRS, Bât. 100, Orsay F-91405, France

Condensation is the phenomenon whereby one of a sum of random variables contributes a finite fraction to the sum. It is manifested as an aggregation phenomenon in diverse physical systems such as coalescence in granular media, jamming in traffic and gelation in networks. We show here that the same condensation scenario, which normally happens only if the underlying probability distribution has tails slower than exponential, can occur for light-tailed distributions in the presence of additional constraints. We demonstrate this phenomenon on the sample variance, whose probability distribution conditioned on the particular value of the sample mean undergoes a phase transition. The transition is manifested by a change in behavior of the large deviation rate function.

PACS numbers: 05.40.-a, 02.50.-r

In recent years the study of large deviations has come to the fore as a general framework in statistical physics [1]. This has proved especially useful in the domain of far-from-equilibrium phenomena where the form of stationary states is not immediately available. The standard large deviation theory describes the probability of events far away from the mean as $P(\Delta x) \propto \exp[-L I(\Delta x/L)]$ where $\Delta x \propto L$ is the deviation from the mean of some extensive observable $x$, $I$ is the rate function and $L$ is the system size. This contrasts with small deviations which are typically described by a Gaussian distribution. The fact that $P$ is dominated by the minimum of $I$ occurring at $\Delta x=0$ implies that large deviation rate functions may play the role of the free energy for nonequilibrium systems [2–4]. Similar large deviation principles also exist for time-extensive quantities such as the current or the activity in lattice gases [5, 7].

It is often observed that profound physical phenomena such as phase transitions have a striking manifestation in the large deviation theory. For example, the nonequivalence of microcanonical and canonical ensembles found in many systems with long-range interactions is manifested by non-convexity of the rate function [1]. Furthermore, nonanalytic behavior of $I$ signals a nonequilibrium phase transition, just as nonanalyticity of the free energy corresponds to an equilibrium phase transition. Recently there has been considerable interest in such nonequilibrium phase transitions [3, 10].

A simple, but significant, example is the condensation phase transition [11] which occurs, for example, in classical systems of interacting particles such as the zero-range process (ZRP) or models for the transfer of a continuous mass variable (for a comprehensive review see [12]). Remarkably, even though the dynamics contains interactions in these models, the steady-state probability of a microstate comprising masses $m_i$ at sites $i = 1, \ldots L$ takes a factorized form given by

$$P(m_1, \ldots, m_L) = \frac{1}{Z_L} \prod_{i=1}^L f(m_i) \delta(M - \sum_i m_i), \quad (1)$$

where $Z_L$ is the normalization constant given by

$$Z_L(M) = \prod_{i=1}^L \int dm_i f(m_i) \delta(M - \sum_i m_i). \quad (2)$$

Here correlations between sites are only generated by the global constraint imposed by the delta function that the total particle number is fixed to be $M$. When the underlying weight $f(m)$ is heavy-tailed in a sense that its tail decays slower than exponential, a condensation phase transition occurs as the global density of particles is raised above a critical value. In the condensed phase, a randomly selected site carries a macroscopic fraction of the particles which is referred to as the condensate. The condensation transition exhibited by the ZRP has thus served as a baseline model for studying aggregation phenomena in systems with more complex states such as coalescence in granular systems [13], jamming in traffic [14, 15], Ostwald ripening [16], gelation in networks [17], emulsification failure in polydisperse hard-sphere systems [18] etc. The dynamics of condensation has also proved of interest [19, 20] and recently generalizations to moving condensates have revealed curious dynamical effects [22, 23].

The connection between the simple factorized steady state of the ZRP and large deviations of a sum of random variables can be seen when one realizes that $P(m_1, \ldots, m_L)$ (Eq. (1)) is equivalent to the probability of picking $L$ independent and identically distributed random variables (iidrv) from a common distribution $f(m)$ (provided we normalize it properly), conditioned on the
fixed value of their total sum \[24, 25\]. In that context the partition function \(Z_L(M)\) itself becomes a probability distribution for the sum of \(L\) iidrv. By fixing \(M = \mu L\), where \(\mu \neq \langle m \rangle\), we condition random numbers on large deviations of their sum. Thus the constraint of fixed total mass, which is imposed by the dynamics of models such as ZRP, provides an elegant way of probing what would otherwise be rare events \[20\].

In this work we go further and consider the effect of several global constraints on the simplest scenario of a factorized steady state. Our key observation is that condensation may be observed even when the underlying distribution \(f(m)\) is light-tailed. Our central object of study is the following partition function

\[
Z_L(M, V) = \int dm_1 \ldots \int dm_L \prod_{i=1}^{L} f(m_i) \times \delta(M - \sum_i m_i)\delta(V - \sum_i m_i^{1/p}). \tag{3}
\]

Generally we consider the linear statistic \(V = \sigma L = \sum_i m_i^{1/p}\), but we shall first focus on the case \(p = 1/2\) in which case \(V - M^2\) is the sample variance of \(L\) rvs drawn from \(f(m)\). In physical systems, the sample variance plays a crucial role in determining the strength of fluctuations and generally has a non-trivial probability distribution. The sample variance is of particular significance in models of traffic flow where \(m_i\) correspond to the headway between adjacent vehicles \(i\) and \(i+1\) and the sample variance gives a global measure of how bunched a set of \(L\) vehicles in a fixed spatial region is \[14\]. Another example is the much studied field of interface growth where \(m_i\) corresponds to the height at site \(i\) and \(V\) then corresponds to the roughness of the interface \[27, 29\]. Also we can mention volatility in financial models defined as the variance of random variables \(m_i\) taking the form of log returns \[30\]. Then \(i\) indexes the time interval and \(m_i\) is the log of the ratio of the price of a stock at time \(i\) and \(i-1\). In all these contexts a key question is how does a large deviation in the sample variance arise. The computation of \[30\] gives information on the conditioned probability density \(P(V|M)\) of finding \(V = \sum_i m_i^{1/p}\) given the value of the sum \(M = \sum_i m_i\) since one can write \(P(V|M)\) as

\[
P(V|M) = \frac{Z_L(M, V)}{Z_L(M)}. \tag{4}
\]

We shall show that asymptotic form of \(\ln Z_L(M, V)/L\), which forms the large deviation function, changes form according to the values of \(M\) and \(V\) thus indicating a condensation transition. We shall first demonstrate condensation transition for the exponential distribution \(f(m) = r \exp(-rm)\) which belongs to the class of light-tailed distributions which have exponentially bounded tails. After a straightforward generalization to other light-tailed cases, we shall conclude by analyzing the interesting case of heavy-tailed distributions, for which we find that the presence of two constraints can suppress the condensation that would otherwise happen if only one of the constraints was present.

Phase diagram. Our strategy for finding \(Z_L(M, V)\) is to calculate its Laplace transform \(\hat{Z}_L(s, \lambda)\) with respect to \(M\) and \(V\). The biggest advantage of working with iidrvs comes from the known form of \(\hat{Z}_L(s, \lambda)\), which has a factorizing property \(\hat{Z}_L(s, \lambda) = [g(s, \lambda)]^L\), where \(g(s, \lambda) = \int_0^\infty dm f(m)e^{-sm-\lambda m^{1/p}}\). The partition function \(Z_L(M, V)\) can be then found by the inversion formula

\[
Z_L(M, V) = \int_{c-i\infty}^{c+i\infty} \frac{ds}{2\pi i} \int_{d-i\infty}^{d+i\infty} \frac{d\lambda}{2\pi i} \times \exp\left(sM + \lambda V + L\ln(g(s, \lambda))\right). \tag{5}
\]

where the constants \(c\) and \(d\) are chosen to be right of any singularities. When \(L\) is large, it is natural to evaluate integrals in \(Z_L(M, V)\) using the saddle-point approximation, which amounts to solving the following saddle point equations for \(s\) and \(\lambda > 0\),

\[
\mu = \frac{\int_0^\infty mf(m)e^{-sm-\lambda m^{1/p}} \, dm}{g(s, \lambda)}, \tag{6a}
\]

\[
\sigma = \frac{\int_0^\infty m^{1/p}f(m)e^{-sm-\lambda m^{1/p}} \, dm}{g(s, \lambda)}. \tag{6b}
\]

For \(f(m) = r \exp(-rm)\) and \(p = 1/2\), the integrals in \[6a\] and \[6b\] can be expressed in terms of the complementary error function \(\text{erfc}(z)\),

\[
\mu = \frac{1 - \sqrt{\pi/e^2} \text{erfc}(z)}{\sqrt{\pi/e^2} \text{erfc}(z)} \equiv \frac{F_1(z)}{\sqrt{\lambda}}, \tag{7a}
\]

\[
\sigma = \frac{(1 + 2z^2/e^2) \text{erfc}(z) - 2z/\sqrt{\pi}}{2z\sqrt{\pi} \text{erfc}(z)} \equiv \frac{F_2(z)}{2\lambda}, \tag{7b}
\]

where we used shorter notation \(z = (s + r)/(2\sqrt{\lambda})\) and introduced functions \(F_1(z)\) and \(F_2(z)\). We can now combine \[7a\] and \[7b\] in a single equation for \(z\) in terms of \(F_1(z)\) and \(F_2(z)\),

\[
2\sigma \mu^2 = \frac{F_2(z)}{F_1^2(z)}. \tag{8}
\]

Using the known limiting behaviors of complementary error function, \(\text{erfc}(z) \to \exp(-z^2)/(\sqrt{\pi}) \times (1 - 1/(2z^2))\) as \(z \to \infty\) and \(\text{erfc}(z) \to 2\) as \(z \to -\infty\), we can show that \[8\] admits a solution only when \(\mu^2 < \sigma < 2\mu^2\). The lower bound \(\mu^2 < \sigma\) is just Jensen’s inequality which states that \(V \leq M^2\), while the upper bound implies a breakdown of the saddle-point approximation for \(\sigma > 2\mu^2\) and signals a phase transition. We shall show below that the transition line \(\sigma_c(\mu) = 2\mu^2\), as presented in figure \[11\] separates a ‘fluid’ phase \((\mu^2 < \sigma < 2\mu^2)\) from the phase.
with a ‘condensate’ ($\sigma > 2\mu^2$). In the fluid phase all $L$ random variables contribute ‘cooperatively’ to accommodate atypical values of $M$ and $V$ by tuning $s$ and $\lambda$ that solve (7a) and (7b), respectively. On the contrary, large deviations in the condensed phase are typically realized by the square of one of the random variables contributing a macroscopic fraction of $V$.

\[
\Pi(L\sigma; s) \approx \ln(L\sigma - L\sigma_c(s); s). \tag{11}
\]

Again, this result states that the large deviation of $V$ arises through $L - 1$ random variables having typical values around $\sigma_c$ and one random variable taking the value $L(\sigma - \sigma_c(s))$. A comparison with exact $\Pi(V; s)$ obtained numerically for integer random variables, which is presented in figure 2 shows that (11) is a good approximation of $\Pi_L(V; s)$ even for lattices of size $L = 200$.

\[
\text{FIG. 2: (Color online) Expression (11) from large deviation theory (LDT, dashed line) compared to } \Pi(V; s) \text{ (ZRP, solid line). } \Pi_L(V; s) \text{ was obtained numerically for } L = 200 \text{ and } r + s = -\ln(0.9) \text{ via recursion relation } \Pi_L(V; s) = \sum_{L-1} w(v; s)\Pi_{L-1}(V - v; s) \text{ valid for integer random variables } \text{[12], where } w(v; s) \text{ is discrete version of Weibull distribution, } w(v; s) = \exp(-(r + s)/v) - \exp(-(r + s)/v + 1) \text{ and } v \text{ is an integer. For large } v, w(v; s) \text{ reduces to the continuous Weibull distribution defined earlier.}
\]

While the above analysis pertains to real $s$, to find $Z_L(M, V)$ by the inversion formula

\[
Z_L(M, V) = \int_{c-i\infty}^{c+i\infty} \frac{ds}{2\pi i} e^{sM + \text{Log}(s, 0)}\Pi(V; s), \tag{12}
\]

we have to allow complex $s$ for which (11) does not hold in general. However, we can calculate the integral in (12) using the saddle-point method by noting that for any complex $s$ with $s_0 = \text{Re}(s)$, the following inequality $|\Pi_L(V; s)| \leq \Pi_L(V; s_0)$ guarantees that $\Pi_L(V; s)$ decays slower than exponential in $L$. Calculating $Z_L(M, V)$ then amounts to solving $\mu = -\partial_s g(s, 0)/g(s, 0)$, which gives $s^* = 1/\mu - r$. A final step is to recognize that the integral we are then left with is in fact the saddle-point approximation of $Z_L(M)$, so that $Z_L(M, V) \approx$

\[
\text{FIG. 1: Phase diagram in the } \mu - \sigma \text{ plane for } f(m) = r e^{-rm} \text{ and } p = 1/2. \text{ Shaded area } \sigma < \mu^2 \text{ is forbidden owing to Jensen’s inequality.}
\]
$L \psi(L\sigma - L\sigma_c; s^*) \approx 2L$, where $\psi = c(s^*) = 2\mu^2$. The final result for $\sigma > 2\mu^2$ and large $L$ is

$$P(V = L\sigma| M = L\mu) \approx \frac{\sqrt{L}e^{-\mu^{-1}\sqrt{L(\sigma^2-2\mu^2)}}}{2\mu\sqrt{\sigma^2-2\mu^2}}. \quad (13)$$

This result is clearly different from what we find in the fluid phase, where $P(V| M)$ decays exponentially fast in $L$ with rate function $I(\mu, \sigma) = \sigma s^* + \sigma m^* + \ln g(s^*, \lambda^*) - \ln \mu - 1$, where $s^*$ and $\lambda^*$ solve (6a) and (6b). The condensation transition thus marks a change in the rate at which $P(V| M)$ decays for large $L$ that goes from exponential in the fluid phase to subexponential in the condensed phase. 

In a finite system, this change of scale however does not happen exactly for $\sigma = \sigma_c$. Close to the transition point where $|\sigma - \sigma_c| \sim O(L^{-1/2})$, one can show that $P(V| M)$ is smooth and follows Gaussian distribution with the standard deviation $2\mu^2/\sqrt{L}$. The transition line is thus shifted by the amount $4\mu^2/3$ [32].

**Generalizations.** So far we have considered the exponential distribution $f(m) = \exp(-rm)$ and $p = 1/2$. Most generally, we have to distinguish between distributions that decay exponentially or faster (light-tailed) or slower than exponential (heavy-tailed). For simplicity, we will consider here only distributions that either have strictly exponential tails, $f(m) \sim \exp(-rm)$ for large $m$, or tails that are not exponentially bounded (heavy-tailed distributions). We also have to distinguish between cases $p < 1$ and $p > 1$. The fundamental difference in going from $p < 1$ to $p > 1$ is that for $p > 1$ the condensation transition happens whereby one of the $m_i$'s takes a macroscopic fraction of $M$ rather than $V$. Here, we consider only the $p < 1$ case and leave $p > 1$ for a subsequent publication [32].

For a general distribution $f(m)$ and $p \neq 1/2$ we can no longer perform the integrals in (6a) and (6b) explicitly. However, by considering the r.h.s. of (6a) and (6b) as functions of $\mu(s, \lambda)$ and $\sigma(s, \lambda)$ respectively, we can show that (a) $\mu = \mu(s, \lambda)$ has a unique solution $s_\mu(\lambda)$ for any given $\mu$ and any $\lambda > 0$, and (b) $\sigma = \sigma(s_\mu(\lambda), \lambda)$ has a unique solution provided $\sigma < \sigma(s_\mu(0), 0)$. The limiting value $\sigma_c(\mu) = \sigma(s_\mu(0), 0)$ arises because $\lambda$ must always stay non-negative, otherwise the integrals in (6a) and (6b) will diverge. Proofs of (a) and (b) are straightforward [32].

The question then arises as to what happens for $\sigma > \sigma_c(\mu)$? Notice that so far we have not distinguished between light-tailed and heavy-tailed distributions, which becomes important when considering the limit $\lim_{\lambda \to 0} \sigma(s_{\mu}(\lambda), \lambda)$. If this limit is finite for any $\mu$, then $\sigma_c(\mu) = \sigma(s_{\mu}(0), 0)$ becomes the transition line separating fluid from condensed phase and the corresponding phase diagram will be as in figure [1]. This type of behavior will happen for any light-tailed distribution with a tail $\propto \exp(-rm)$, in which case $s_{\mu}(0) > -r$ can solve (6a) for any $\mu$. On the other hand, if $f(m)$ has a tail that decays slower than exponential, then $s_{\mu}(0)$ does not exist for $\mu > \mu_c$, nor does the transition line $\sigma_c(s_{\mu}(0), 0)$. Instead, both saddle point equations can be solved with some positive $\lambda$ and negative $s_{\mu}(\lambda)$. The corresponding phase diagram, presented in figure [2] for a Pareto distribution, has a transition line that follows $\sigma_c(\mu) = \sigma(s_{\mu}(0), 0)$ for $\mu < \mu_c$ and then turns into straight line $\mu = \mu_c$ for $\sigma > \sigma_c(\mu_c)$. For $\mu > \mu_c$, it is thus no longer possible to go to the condensed phase by increasing $\sigma$. For $\mu > \mu_c$ there is also an interesting effect that fixing $V = \sigma L$ has suppressed condensation that would have appeared in $M$ had there been no constraint on $V$.

![Phase diagram](image-url)
Note Added. After we submitted this work, we received a manuscript [35] which deals with a related problem and recovers figure 3. Also we became aware of a close connection of condensation in our work to breathers (localized solutions) in the discrete nonlinear Schrödinger equation and related models [36]. Finally, the $p = 1/2$ case with a uniform underlying distribution has been studied using probabilistic techniques in [37].

JSN and MRE would like to acknowledge funding from EPSRC under grant number EP/J007404/1. SNM acknowledges support by ANR grant 2011-BS04-013-01 WALKMAT.

---

[11] This form of condensation bears a close relation to Bose-Einstein condensation in the ideal Bose gas but is distinct from the liquid-gas transition which is also sometimes referred to as condensation.
[30] Whether the stock prices are true iidrvs is still much debated today.