Advanced Studies in Pure Mathematics 31, 2001
Taniguchi Conference on Mathematics Nara '98
pp. 265–286

Large Deviation and Hydrodynamic Scaling

Srinivasa R. S. Varadhan

§1. What are Large Deviations?

The theory of large deviations is, roughly speaking, a method of describing the rapidity with which probability distributions depending on a parameter approach the degenerate distribution at some point as the parameter becomes large.

Let us suppose that for each n there is a probabilty measure P_n on some space Ω_n defined on some σ-field Σ_n. There is a complete separable metric space X with its Borel sets $\mathcal{B}$, such that for each n there is a measurable map Φ_n of Ω_n into X. We denote the induced measure $P_n\Phi_n^{-1}$ on $(X, \mathcal{B})$ by Q_n. Actually it is the situation $(X, \mathcal{B}, Q_n)$ that will be of interest to us. As $n \to \infty$ the measures Q_n will converge weakly to a limit which will be degenrate at some point x_0 of X. This is usually a 'law of large numbers', statement. In particular for any closed set $A \subset X$, with $x_0 \notin A$,

$$\lim_{n\to\infty} Q_n(A) = 0. \tag{1.1}$$

If the parametrization has been chosen properly, the convergence in the limit (1.1) will often be exponentially fast and

$$\lim_{n\to\infty} \frac{1}{n} \log Q_n(A) = -\Psi(A) \tag{1.2}$$

will exist atleast for a large class nice sets. Since the exponential behavior of a sum is the same as that of the larger of the summands

$$\Psi(A \cup B) = \min\{\Psi(A), \Psi(B)\}$$

and one can expect $\Psi(A)$ to be given by a formula of the type

$$\Psi(A) = \inf_{x\in A} I(x)$$

Received December 2, 1999.

for some function $I(\cdot): X \to [0,\infty]$. The theory of large deviations is a large collection of interesting examples that fit this model. What makes the class of models interesting is the ability to identify the rate function $I(\cdot)$ in specific cases.

We say that the family Q_n on X satisfies the 'large deviation principle' with rate function $I(\cdot)$ if

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_n(A) \le -\inf_{x\in A} I(x) \quad \text{for closed sets } A \in X, \tag{1.3}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n(G) \ge -\inf_{x\in G} I(x) \quad \text{for open sets } G \in X. \tag{1.4}$$

Of course if $E \subset X$ is nice enough that

$$\inf_{x\in E^0} I(x) = \inf_{x\in E} I(x) = \inf_{x\in \overline{E}} I(x)$$

we get

$$\lim_{n\to\infty} \frac{1}{n} \log Q_n(E) = -\inf_{x\in E} I(x). \tag{1.5}$$

It is important that the function $I(\cdot)$ that can take the value $+\infty$ be lower semi-continuous and have compact level sets, i.e., for each $\ell < \infty$, the set

$$K_\ell = \{x : I(x) \le \ell\} \tag{1.6}$$

be a compact (closed and totally bounded) subset of X.

We will begin with some simple examples.

Example. Let α be a probability measure on R. Let P_n on $\Omega = R^n$ be the product measure $\alpha \times \alpha \times \cdots \times \alpha$. Let Φ_n be the map

$$\Phi_n(x_1, \ldots, x_n) = \frac{x_1 + \cdots + x_n}{n}.$$

The law of large numbers asserts that $Q_n \to \delta_a$ with $a = \int x\, d\alpha$. According to a theorem of Cramér [1] Q_n satisfies a large deviation principle with rate function

$$I(y) = \sup_\sigma [\sigma y - \log M(\sigma)] \tag{1.7}$$

where

$$M(\sigma) = \int e^{\sigma x} \alpha(dx). \tag{1.8}$$

Another example is the following.

Example. Let F be a finite alphabet $\mathcal{A}$ consisting of letters $\{a_1, \dots, a_k\}$. Let Ω_n consist of words $W = \{x_1, \dots, x_n\}$ of length n in $\mathcal{A}$. The probabilities $P_n(W)$ are all equal and since there are k^n words of length n

$$P_n(W) = \frac{1}{k^n}$$

for every word $W \in \Omega_n$. X is the space of probability distributions on $\{1, \dots, k\}$, i.e., $\{p_1, \dots, p_k : p_i \geq 0 \text{ and } \sum_i p_i = 1\}$. The map Φ_n is the empirical distibution

$$\Phi_n(x_1, \dots, x_n) = \left\{ \frac{\sum_j \delta_{a_i, x_j}}{n} \right\} = \left\{ \frac{n_i}{n} \right\} \tag{1.9}$$

where n_i is the number of times the letter a_i occurs in the word $W = \{x_1, \dots, x_n\}$. Again, by the law of large numbers, Q_n converges to $\delta_{\{1/k, \dots, 1/k\}}$.

$$\begin{aligned} Q_n[p_1, \dots, p_n] &\simeq \frac{n!}{(np_1)! \cdots (np_k)!} \frac{1}{k^n}, \\ I(p) &= -\lim_{n \to \infty} \frac{1}{n} \log Q_n[p_1, \dots, p_n] \\ &= \log k + \sum p_i \log p_i \\ &= \sum_i p_i \log \frac{p_i}{\frac{1}{k}}. \end{aligned}$$

A slightly more general form of the example is Sanov's theorem.

Example. If we define

$$P_n(W) = \frac{n!}{n_1! \cdots n_k!} \pi_1^{n_1} \cdots \pi_k^{n_k}$$

we get

$$I_\pi(p) = -\sum_i p_i \log \pi_i + \sum_i p_i \log p_i = \sum_i p_i \log \frac{p_i}{\pi_i}$$

or even more generally

Example. We take $\Omega_n = X \times X \times \cdots \times X$, $P_n = \alpha \times \alpha \times \cdots \times \alpha$ and Φ_n the map of Ω into the space $\mathcal{M}$ of all probability measures on X defined by

$$\Phi_n(x_1, \dots, x_n) = \frac{\delta_{x_1} + \cdots + \delta_{x_n}}{n}.$$

In this example $I_\alpha(\mu) < \infty$ only if $\mu \ll \alpha$ and

$$\frac{d\mu}{d\alpha} \log \frac{d\mu}{d\alpha} \in L_1(\alpha).$$

Then

$$I_\alpha(\mu) = \int_X \log \frac{d\mu}{d\alpha}\, d\mu = \int_X \frac{d\mu}{d\alpha} \log \frac{d\mu}{d\alpha}\, d\alpha. \tag{1.10}$$

Otherwise $I_\alpha(\mu) = +\infty$.

There are some general principles in the theory which are relatively easy to establish. Here is one. A general property known as 'Contarction Principle' is the following:

Theorem 1.1. *Let P_n satisfy the large deviation property with rate function $I(\cdot)$ on X. Let $f : X \to Y$ be a continuous map into Y. Then $Q_n = P_n f^{-1}$ satifies a large deviation principle on Y with rate function $J(y) = \inf[I(x)\,;\, x : f(x) = y]$.*

We will illustrate the Contarction Principle by showing that Cramér's theorem can be obtained from Sanov's theorem. Consider the map $f : \mathcal{M} \to R$ defined by

$$f(\mu) = \int_R x\, d\mu. \tag{1.11}$$

Then the sample mean $(x_1 + \cdots + x_n)/n$ can be thought of as

$$\frac{x_1 + \cdots + x_n}{n} = f(\Phi_n(x_1, \ldots, x_n))$$

where Φ_n is the empirical distribution. A calculation shows that

$$I(y) = \inf_{\mu : \int x\, d\mu = y} I_\alpha(\mu)$$

which is the contraction principle. Actually Sanov's theorem can be sort of seen as a version of Cramér's theorem as well. We can repalce R by the locally convex topological vector space $\mathcal{M}(R)$ and replace α on R by the distribution β induced on $\mathcal{M}(R)$ by the map $x \to \delta_x$. The empirical distribution is just the sum of n independent $\mathcal{M}(R)$ valued random vectors with the common distribution β. The moment generating function is replaced by

$$M(V) = \int_{\mathcal{M}(R)} e^{\langle V, \mu \rangle}\, d\beta = \int_R e^{V(x)}\, d\alpha \tag{1.12}$$

and

$$H(\mu;\alpha) = I_\alpha(\mu) = \sup_{V(\cdot)} \left[\int_R V(x)\, d\mu - \log M(V) \right] \tag{1.13}$$
$$= \int_R \frac{d\mu}{d\alpha} \log \frac{d\mu}{d\alpha}\, d\alpha.$$

In particular

$$\int_R V(x)\, d\mu \le \log \int_R e^{V(x)}\, d\alpha + H(\mu;\alpha) \tag{1.14}$$

or for any $\sigma > 0$, replacing V by σV, we get

$$\int_R V(x)\, d\mu \le \frac{1}{\sigma} \log \int_R e^{\sigma V(x)}\, d\alpha + \frac{1}{\sigma} H(\mu;\alpha). \tag{1.15}$$

Another general principle is the following theorem on the exponential growth rate of integrals. It is basically a fancy version of the simple fact that for $a, b > 0$ we have

$$\lim_{n\to\infty} [a^n + b^n]^{1/n} = \max(a,b).$$

Theorem 1.2. *If the large deviation principle holds for some Q_n on X, with a rate function $I(\cdot)$, then for any real valued bounded continuous function $F(\cdot)$ on X*

$$\lim_{n\to\infty} \frac{1}{n} \log \int_X \exp[n\, F(x)]\, dQ_n = \sup_{x\in X} [F(x) - I(x)].$$

The book [2] is a good source for a discussion of these topics as well as for additional references.

§2. Hydrodynamic Scaling

The basic example of Hydrodynamic scaling is the derivation of Euler equations from the equations of classical mechanics. Let us start with a collection of $N \simeq \bar\rho \ell^3$ classical particles in a large periodic cube Λ_ℓ of side ℓ in R^3. The motion of the particles are governed by the equations of motion of a classical Hamiltonian dynamical system with energy given by

$$H(p,q) = \frac{1}{2} \sum_{i=1}^{N} \|p_i\|^2 + \frac{1}{2} \sum_{i\ne j} V(q_i - q_j). \tag{2.1}$$

Here, $q_i \in \Lambda_\ell$ is the position of the i-th particle and $p_i \in R^3$ is its velocity. The coordinates $k = 1, 2, 3$ refer to the three components of position or velocity. The repulsive potential $V \geq 0$ is an even function that is not identically zero and has compact support in R^3. The interaction in particular is short range. The classical equations of motion are

$$\frac{dq_i^k}{dt} = \frac{\partial H(p,q)}{\partial p_i^k} = p_i^k, \tag{2.2}$$

$$\frac{dp_i^k}{dt} = -\frac{\partial H(p,q)}{\partial q_i^k} = -\sum_{j=1}^{N} V_k(q_i - q_j), \tag{2.3}$$

where $V_k(q) = \partial V(q)/\partial q^k$ for $k = 1, 2, 3$ are the three components of the gradient of V. The dynamical system has five conserved quantities. The total number N of particles, the total momenta $\sum_{i=1}^N p_i^k$ for $k = 1, 2, 3$ and the total energy $H(p,q)$. The hydrodynamic scaling in this context consists of rescaling space and time by a factor of ℓ. The rescaled space is the unit torus $\mathbf{T}^3$ in 3-dimensions. The macroscopic quantities to be studied correspond to the five conserved quantities. The first one of these is the density, and is measured by a function $\rho(t,x)$ of t and x. For each $\ell < \infty$ it is approximated by $\rho_\ell(t,x)$, defined by

$$\int_{\mathbf{T}^3} J(x)\rho_\ell(t,x)\,dx = \frac{1}{\ell^3}\sum_{i=1}^{N} J\Big(\frac{q_i(\ell t)}{\ell}\Big). \tag{2.4}$$

A straight forward differentiation with respect to t yields

$$\begin{aligned} \frac{d}{dt}\int_{\mathbf{T}^3} J(x)\rho_\ell(t,x)\,dx &= \frac{d}{dt}\frac{1}{\ell^3}\sum_{i=1}^{N} J\Big(\frac{q_i(\ell t)}{\ell}\Big) \\ &= \frac{1}{\ell^3}\sum_{i=1}^{N} (\nabla J)\Big(\frac{q_i(\ell t)}{\ell}\Big)\cdot p_i(\ell t) \\ &\simeq \int_{\mathbf{T}^3} (\nabla J)(x)\cdot \rho_\ell(t,x) u_\ell(t,x)\,dx \end{aligned} \tag{2.5}$$

where $u_\ell(t,x) = u_\ell^k(t,x)$, $k = 1, 2, 3$ are the components of the 'average' velocity of the fluid at the rescaled space time point x, t. This introduces three other macroscopic variables, which represent three coordinates of the momenta that are conserved. We can now write down the first of our five equations

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho u) = 0. \tag{2.6}$$

To derive the next three equations, using a test functions J, we differentiate for $k = 1, 2, 3$

(2.7)
$$\frac{d}{dt}\frac{1}{\ell^3}\sum_{i=1}^{N} J\Big(\frac{q_i(\ell t)}{\ell}\Big)p_i^k(\ell t) = \frac{1}{\ell^3}\sum_{i=1}^{N} p_i^k(\ell t)(\nabla J)\Big(\frac{q_i(\ell t)}{\ell}\Big)\cdot p_i(\ell t) - \frac{1}{\ell^2}\sum_{i=1}^{N}\sum_{j=1}^{N} J\Big(\frac{q_i(\ell t)}{\ell}\Big)V_k(q_i(\ell t) - q_j(\ell t)).$$

If we now use the skew-symmetry of $V_k = \partial V/\partial q_k$, we can rewrite the second term of the right hand side of equation (2.7) as

(2.8)
$$\begin{aligned}
&-\frac{1}{2\ell^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(J\Big(\frac{q_i(\ell t)}{\ell}\Big) - J\Big(\frac{q_j(\ell t)}{\ell}\Big)\Big)V_k(q_i(\ell t) - q_j(\ell t))\\
&\simeq -\frac{1}{2\ell^3}\sum_{i=1}^{N}\sum_{j=1}^{N} J_r\Big(\frac{q_i(\ell t)}{\ell}\Big)(q_i^r(\ell t) - q_j^r(\ell t))V_k(q_i(\ell t) - q_j(\ell t))\\
&= \frac{1}{\ell^3}\sum_{i=1}^{N}\sum_{j=1}^{N} J_r\Big(\frac{q_i(\ell t)}{\ell}\Big)\psi_k^r(q_i(\ell t) - q_j(\ell t))
\end{aligned}$$

with
$$\psi_k^r(q) = -\frac{1}{2}q^r V_k(q).$$

The next step is rather mysterious and requires considerable explanation. The quantities
$$\sum_{i=1}^{N} p_i^k p_i^r, \quad \sum_{i,j=1}^{N}\psi_k^r(q_i(t) - q_j(t))$$

are not conserved. They depend on combinations of individual velocities that are not conserved and on spacings between particles both of which change in the microscopic time scale and therefore do so rapidly in the macroscopic time scale. They should therefore be replaced by their space-time averages. By appealing to an 'Ergodic Theorem' they can be replaced by their averages with repect to their equilibrium distributions. The equilibrium 'ensemble' consists of an infinite collection of points $\{p_\alpha, q_\alpha\}$, in the phase space $R^3 \times R^3$. There is a natural five parameter family of measures $\mu_{\rho,u,T}$ that are invariant under spatial translations as well as the Hamiltonian dynamics. The points $\{p_\alpha\}$ are distributed according to a Gibbs Distribution with density ρ and formal interaction energy

$$\frac{1}{2T}\sum_{\alpha,\beta} V(q_\alpha - q_\beta).$$

In other words $\{q_\alpha\}$ is a point process obtained by taking infinite volume limit of $N = \ell^3\rho$ particles distributed in the cube of side ℓ in R^3 according to the joint density

$$\frac{1}{Z}\exp\left[-\frac{1}{2T}\sum_{1\le i\ne j\le N} V(q_i - q_j)\right]$$

where Z is the normalization constant. The velocities $\{p_\alpha\}$ are distributed independently of each other as well as of $\{q_\alpha\}$, having a common three dimensional Gaussian distribution with mean u and covariance TI. Assuming that the infinite volume limit exists in a reasonable sense it will be a point process defined as an infinite volume Gibbs measure $\mu_{\rho,T}$. The velocities $\{p_\alpha\}$ will be an independent Gaussian ensemble $\nu_{u,T}$. In the first term the quantities $p_i^k p_i^r$ are replaced by their expectations

$$u^k(t,x)u^r(t,x) + \delta_{k,r}T(t,x)$$

and in the second term $\psi_{k,r}$ are replaced by their expectations that involve the 'pressure' per unit volume in the Gibbs ensemble

$$\mathbf{P}_k^r(\rho,T) = \lim_{\ell\to\infty} E^{\mu_{\rho,T}}\left\{\frac{1}{\ell^3}\sum_{|q_\alpha|,|q_\beta|\le\ell} \psi_k^r(q_\alpha - q_\beta)\right\}.$$

This leads to the equation

$$\begin{aligned}(2.9)\qquad &\frac{d}{dt}\int_{\mathbf{T}^3} J(x)u^k(t,x)\,dx\\ &= \int_{\mathbf{T}^3}\sum_{r=1}^3 \frac{\partial J}{\partial x_r}(x)(u^k(t,x)u^r(t,x) + \delta_{k,r}T(t,x))\,dx\\ &+ \int_{\mathbf{T}^3}\sum_{r=1}^3 \frac{\partial J}{\partial x_r}(x)\mathbf{P}_k^r(\rho(t,x),T(t,x))\,dx.\end{aligned}$$

We now integrate by parts, remove the test function J and obtain from equation (2.9)

$$(2.10)\qquad \frac{d}{dt}(\rho u) + \nabla\cdot(\rho u\otimes u + \rho TI + \mathbf{P}(\rho,T)) = 0.$$

There is an equation of state that expresses the total energy per unit volume e as

$$e(\rho,u,T)=\frac{1}{2}\rho(|u|^2+3T)+f(\rho,T) \tag{2.11}$$

where $f(\rho,T)$, the potential energy per unit volume, is given by

$$f(\rho,T)=\lim_{\ell\to\infty}E^{\mu_{\rho,T}}\left\{\frac{1}{2\ell^3}\sum_{|q_\alpha|,|q_\beta|\le\ell}V(q_\alpha-q_\beta)\right\}.$$

Although we will not derive it, there is a similar equation for $e(t,x)$ that is obtained by differentiating

$$\frac{d}{dt}\frac{1}{2\ell^3}\sum_{i=1}^{N}J\Big(\frac{q_i(\ell t)}{\ell}\Big)\Big[|p_i(\ell t)|^2+\sum_{j=1}^{N}V(q_i(\ell t)-q_j(\ell t))\Big]$$

and proceeding in a similar fashion. It looks like

$$\frac{de}{dt}+\nabla\cdot\big[(e+T)u+\mathbf{P}(\rho,T)u\big]=0. \tag{2.12}$$

The five equations one for density given by equation (2.6), the three for velocities contained in equation (2.10) and finally the enrgy equation (2.12) constitute a first order system of non-linear hyperbolic conservation laws in the six variables $[\rho,u,T,e]$ with one relation between them given by equation (2.11). Given smooth initial data they have local solutions. Rigorous derivation of these equations does not exist.

We have made a basic assumption in the above derivation. If we take a small volume in space around the point (x,t) in macroscopic space-time and blow up the space by a factor of ℓ we will see a bunch of particles with velocities. The positions of these particles will form a point process in a big domain in R^3. The statistics of these points is assumed to be a Gibbs distribution $\mu_{\rho,T}$ corresponding to the density $\rho=\rho(t,x)$ and 'Temperature', $T=T(t,x)$. Given the positions, the velocities are assumed to be mutually independent and have a common Gaussian distribution with mean $u=u(t,x)$ and covariance $TI=T(t,x)I$. The five parameters (ρ,u,T) locally detemine a Gibbs-Gaussian equilibrium. The equations are derived under the assumption that this picture holds asymptotically for large ℓ. There is no known proof of this. While it is possible to prepare the initial state so that this property of loacl equlibrium holds at time $t=0$, there is no guarantee that this property persists at positive macroscopic times. If ρ, u, T are constants independent of

x at time 0, then we have a global equilibrium and that persists. But hydrodynamically this is the uninteresting case.

While the validity of the principle of local equlibrium is very hard to establish for the Hamiltonian system, it is not nearly so hard for stochastic systems of comparable type. Noise helps to establish local equilibria. This in many cases can be rigorously established and thence the corresponding hydrodynamical equations can be derived with full mathematical rigor.

We will consider a class of stochastic models that are called simple exclusion processes. They make sense on any finite or countable set X and for us it will be either the integer lattice $\mathbf{Z}^d$ in d-dimensions or $\mathbf{Z}^d_N$ obtained from it as a quotient by considering each coordinate modulo N. At any given time a subset of these lattice sites will be occupied by particles, with atmost one particle at each site. In other words some sites are empty while others are occupied with one particle. The particles move randomly. Each particle waits for an exponential random time and then tries to jump from its current site x to a new site y. The new site y is picked randomly according to a probability distribution $\pi(x, y)$. In particular $\sum_y \pi(x, y) = 1$ for every x. Of course a jump to y is not always possible. If the site is empty the jump is possible and is carried out. If the site already has a particle, the jump cannot be carried out and the particle forgets about it and waits for another chance, i.e., waits for a new exponential waiting time. If we normalize so that all waiting times have mean 1, the generator of the process can be written down as

$$(\mathcal{A}f)(\eta) = \sum_{x,y} \eta(x)(1 - \eta(y))\pi(x, y)[f(\eta^{x,y}) - f(\eta)] \tag{2.13}$$

where η represents the configuration with $\eta(x) = 1$ if there is a particle at x and $\eta(x) = 0$ otherwise. For each configuration η and a pair of sites x, y the new configuration $\eta^{x,y}$ is defined by

$$\eta^{x,y}(z) = \begin{cases} \eta(y), & \text{if } z = x, \\ \eta(x), & \text{if } z = y, \\ \eta(z), & \text{if } z \neq x, y. \end{cases} \tag{2.14}$$

We will mainly be concerned with the situation where the set X is $\mathbf{Z}^d$ or $\mathbf{Z}^d_N$, viewed naturally as an Abelian group with $\pi(x, y)$ being translation invariant and given by $\pi(x, y) = p(y - x)$ for some probability distribution p. It is convenient to assume that p has finite support. There are various possibilities. We will first consider the case $\sum_z zp(z) = m \neq 0$ that needs hyperbolic scaling and leads to Burgers equation with zero viscosity. In order to convey the idea it is sufficient to restrict ourselves to the case where $d = 1$, $p(1) = 1$ and $p(z) = 0$ for all $z \neq 1$. This is the totally asymmetric nearest neighbor simple exclusion model and of course $m = 1$. In this case the rescaling in time is done by a factor of N and the generator is

$$(\mathcal{A}_N f)(\eta) = N \sum_{x \in \mathbf{Z}_N} \eta(x)(1 - \eta(x+1))[f(\eta^{x,x+1}) - f(\eta)].$$

We can easily calculate

(2.16)

$$\begin{aligned} d&\left[\frac{1}{N} \sum_{x \in \mathbf{Z}_N} J\Big(\frac{x}{N}\Big)\eta_t(x)\right] \\ &= \left[\sum_{x \in \mathbf{Z}_N} \left[J\Big(\frac{x+1}{N}\Big) - J\Big(\frac{x}{N}\Big)\right]\eta_t(x)(1 - \eta_t(x+1))\right] dt + M_N(t) \\ &\simeq \frac{1}{N}\left[\sum_{x \in \mathbf{Z}_N} J'\Big(\frac{x}{N}\Big)\eta_t(x)(1 - \eta_t(x+1))\right] dt + o(1). \end{aligned}$$

The martingale term is negligible and we need to do simple 'averaging'. The equlibria are the Bernoulli measures μ_ρ indexed by density. Since there is only a single invariant quantity, i.e., the number of particles, we can replace $\eta_t(x)(1-\eta_t(x+1))$ by its expected value $\rho(t,x)(1-\rho(t,x))$.

$$\frac{d}{dt}\int_{\mathbf{T}^1} J(\theta)\rho(t,\theta)\, d\theta = \int_{\mathbf{T}^1} J'(\theta)\rho(t,\theta)(1 - \rho(t,\theta))\, d\theta$$

or equivalently

$$\frac{\partial \rho}{\partial t} + \frac{\partial[\rho(1-\rho)]}{\partial \theta} = 0. \tag{2.17}$$

A different situation occurs when p is symmetric, i.e., $p(x) = p(-x)$. Let us look at the function

$$V_J(\eta) = \sum J(x)\eta(x)$$

and compute

$$(2.18)\qquad \begin{aligned}(\mathcal{A}V_J)(\eta) &= \sum_{x,y} \eta(x)(1-\eta(y))p(y-x)(J(y)-J(x))\\ &= \sum_{x,y} \eta(x)p(y-x)(J(y)-J(x))\\ &= \sum_{x,y} \eta(x)[(\mathbf{P}-I)J](x)\\ &= V_{(\mathbf{P}-I)J}(\eta).\end{aligned}$$

The space of linear functionals is left invariant by the generator. It is not difficult to see that

$$E_\eta\big[V_J(\eta(t))\big] = V_{J(t)}(\eta)$$

where

$$J(t) = \exp[t(\mathbf{P}-I)]J$$

is the solution of

$$\frac{d}{dt}J(t,x) = (\mathbf{P}-I)J(t,x).$$

It is almost as if the interaction had no effect and in fact for the calculation of expectations of 'one particle' functions it clearly does not. Let us start with a configuration on $\mathbf{Z}_N^d$ and scale space by N and time by N^2. The generator becomes $N^2\mathcal{A}$ and the particles can be visualized as moving on a lattice imbedded in the unit torus $\mathbf{T}^d$, with a spacing of $1/N$, that becomes dense as $N \to \infty$.

Let us consider the functional

$$\xi(t) = \frac{1}{N^d}\sum_x J\Big(\frac{x}{N}\Big)\eta_t(x).$$

We can write

$$\xi(t) - \xi(0) = \int_0^t V_N(\eta(s))\,ds + M_N(t)$$

where

$$V_N(\eta) = (N^2\mathcal{A}V_J)(\eta) = V_{J_N}(\eta)$$

with

$$\begin{aligned}(J_N)(x) &= N^2\sum\Big[J\Big(x+\frac{z}{N}\Big) - J(x)\Big]p(z)\\ &\simeq \frac{1}{2}(\Delta_C J)(x).\end{aligned}$$

Here Δ_C refers to the Laplacian

$$\sum_{i,j} C_{i,j} \frac{\partial^2}{\partial x_i \partial x_j}$$

with the covariance matrix C given by

$$C_{i,j} = \sum_x x_i x_j p(x).$$

$M_N(t)$ is a martingale and a very elementary calculation yields

$$E\{[M_N(t)]^2\} \le CtN^{-d}$$

essentially completing the proof in this case. Technically the empirical distribution $\nu_N(t)$ is viewed as a measure on $\mathbf{T}^d$ and $\nu_N(\,\cdot\,)$ is viewed as a stochastic process with values in the space $\mathcal{M}(\mathbf{T}^d)$ of nonnegative measures on $\mathbf{T}^d$. In the limit it lives on the set of weak solutions of the heat equation

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\Delta_C \rho \tag{2.19}$$

with the initial condition $\rho(0,x) = \rho_0(x)$ determined by

$$\int_{\mathbf{T}^d} J(x)\rho_0(x)\,dx = \lim_{n\to\infty} \frac{1}{N^d} \sum_{x \in \mathbf{Z}_N^d} J\Big(\frac{x}{N}\Big)\eta_0(x) \tag{2.20}$$

and the uniqueness of such weak soultions for given initial density establishes the validity of the scaling limit. We could have have computed two moments as in the noninteracting case. The expectation would have been no different from the noninteracting case since it involves only one particle functions. The variance involves two particle functions and would have involved slighlty more work, because the independence is not there. The martingale argument however is more general.

Let us now turn to the case where p has mean zero but is not symmetric. In this case

$$V_N(\eta) = N^{2-d} \sum_{x,y} \eta(x)(1-\eta(y))p(y-x)\Big[J\Big(\frac{y}{N}\Big) - J\Big(\frac{x}{N}\Big)\Big] \tag{2.21}$$

and we get stuck at this point. If p is symmetric, as we saw, we gain a factor of N^{-2}. Otherwise the gain is only a factor of N^{-1} which is not enough. We seem to end up with

$$\begin{aligned} & N^{-d}\sum_x\sum_y \eta(x) \\ & \qquad \times \left\langle \frac{1}{2}\Big[(\nabla J)\Big(\frac{x}{N}\Big) + (\nabla J)\Big(\frac{y}{N}\Big)\Big], N(1-\eta(y))(y-x)p(y-x)\right\rangle \\ & = \frac{1}{2N^d}\sum_x (\nabla J)\Big(\frac{x}{N}\Big) N\Psi_x \end{aligned}$$

where

$$\begin{aligned} \Psi_0 &= \frac{1}{2}\Big[\eta(0)\sum_z (1-\eta(z))zp(z) + (1-\eta(0))\sum_z \eta(-z)zp(z)\Big] \\ &= \frac{1}{2}\Big[-\eta(0)\sum_z \eta(z)zp(z) + (1-\eta(0))\sum_z \eta(-z)zp(z)\Big] \\ &= \frac{1}{2}\Big[\sum_z \eta(-z)zp(z) - \eta(0)\sum_z (\eta(z)+\eta(-z))zp(z)\Big]. \end{aligned}$$

The second sum is zero in the symmetric case and Ψ_0 can then be written as a 'gradient' $\Psi_0 = \sum_j \tau_{e_j}\xi_j - \xi_j$ where τ_{e_j} are shifts in the coordinate directions. This allows us to do summation by parts and gain a factor of N^{-1}. When this is not the case, we have a 'nongradient' model and the scaling limit can no longer be established by simple averaging.

Exactly the same situation arises in the symmetric case if we make the probabilities of jumps $p(x) = 1/2d$ for the $2d$ nearest neighbors and 0 otherwise, but change the rates so that the generator reads

$$(\mathcal{A}f)(\eta) = \sum_{|x-y|=1} a_{x,y}(\eta)[f(\eta^{x,y}) - f(\eta)] \tag{2.22}$$

where $a_{x,y}(\eta)$ are translation invariant and satisfy the 'detailed balance', conditions relative to the Bernoulli measures.

There are several good sources for this and related material. In particular the book [4], the monograph [10] and the notes [3] contain all of this material as well as more references.

§3. Large Deviation Methods in Hydrodynamic Scaling

A rigorous proof of the validity of the hydrodynamic scaling limit depends on establishing some sort of a local ergodic theorem. There are several ways of carrying this out depending on the circumstances.

But the methods that are fairly general use ideas from large deiviations in some form. We will consider the example of the totally asymmetric simple exclusion process in one dimension. Suppose we are give a smooth function $\rho(t,x)$ on $[0,T] \times \mathbf{T}$ that satisfies $0 < c \leq \rho(t,x) \leq 1-c < 1$ and solves Burgers equation (2.17)

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}[\rho(t,x)(1-\rho(t,x)] = 0.$$

For any smooth $\rho(x)$ from $\mathbf{T} \to (0,1)$ we can associate a local or slowly varying equlibrium state

$$f^N_{\rho(\,\cdot\,)}(\eta) = \prod_{x=1}^{N} \rho\Big(\frac{x}{N}\Big)^{\eta_x}\Big(1-\rho\Big(\frac{x}{N}\Big)\Big)^{1-\eta_x}. \tag{3.1}$$

We could guess that the state at time t is more or less

$$f_N(t,\eta) = \prod_{x=1}^{N} \rho\Big(t,\frac{x}{N}\Big)^{\eta_x}\Big(1-\rho\Big(t,\frac{x}{N}\Big)\Big)^{1-\eta_x}. \tag{3.2}$$

Even if we start at time 0 with initial distribution $f_N(0,\eta)$ the true state at time t is the solution $g_N(t,\eta)$ of the Kolmogorov forward equation

$$\frac{\partial g_N}{\partial t} = \mathcal{L}^*_N g_N \tag{3.3}$$

with the initial condition

$$g_N(0,\eta) = f_N(0,\eta). \tag{3.4}$$

We wish to compare the true solution g_N to our guess f_N. They match at $t=0$. What about $t>0$?

The comparison is done by

$$H_N(t) = H(g_N(t,\,\cdot\,); f_N(t,\,\cdot\,)). \tag{3.5}$$

It is controlled by establishing a Gronwall type of inequality

$$\frac{dH_N(t)}{dt} \leq C H_N(t) + \text{“}\mathit{error}\text{”} \tag{3.6}$$

that leads to

Theorem 3.1. *We have*

$$\lim_{N\to\infty}\sup_{0\le t\le T}\frac{1}{N}H_N(t)=0 \tag{3.7}$$

which in turn implies the validity of the hydrodynamic limit

$$\lim_{N\to\infty}\frac{1}{N}\sum_{x\in\mathbf{Z}_N}J\Big(\frac{x}{N}\Big)\eta_t(x)=\int_{\mathbf{T}}J(x)\rho(t,x)\,dx.$$

One key ingredient is the validity of the local equilibrium principle which needs to be established. This takes the following form. Let ℓ be an intermediate scale, i.e., $1\ll\ell\ll N$. For any local function $g(\eta)$ let us define

$$\hat{g}(\rho)=E^{\mu_\rho}[g(\eta)] \tag{3.8}$$

where μ_ρ is the Bernoulli measure with density ρ. We look at the difference

(3.9)
$$D_{\ell,N,g}(\eta)=\frac{1}{N}\sum_x\left|\frac{1}{2\ell+1}\sum_{y:|y-x|\le\ell}g(\tau_y\eta)-\hat{g}\Big(\frac{1}{2\ell+1}\sum_{y:|y-x|\le\ell}\eta(y)\Big)\right|.$$

Establishing hydrodynamic limit requires

$$\lim_{\epsilon\to 0}\limsup_{N\to\infty}E^{P_N}\left[\int_0^T D_{\epsilon N,N,g}(\eta_t)\,dt\right]=0 \tag{3.10}$$

where P_N is the process starting from an arbitrary initial configuration. The relative entropy considerations reduce this to proving estimates of the form

$$\lim_{\ell\to\infty}\limsup_{N\to\infty}E^{P_N}[D_{\ell,N,g}(g)]=0. \tag{3.11}$$

This can be reduced to proving a much stronger estimate for the process Q_N in equilibrium.

$$\lim_{\ell\to\infty}\limsup_{N\to\infty}\frac{1}{N}\log E^{Q_N}\left[\exp N\int_0^T D_{\epsilon N,N,g}(\eta_t)\,dt\right]=0 \tag{3.12}$$

for every g.

This in turn can be estimated in terms of Feynman-Kac representation and variational formulae involving Dirichlet forms.

In the case of diffusive scaling this approach is often powerful enough to yield equation (3.10).

In the last example alluded to in the previous section, i.e., nongradient models, the analysis involves writing the 'current', in the form

$$W_{x,x+1} = c(\eta(x) - \eta(x+1)) + \text{'\textit{negligible terms}'} \tag{3.13}$$

corresponding to a projection in $L_2(P_\rho)$. The left handside can be thought of as a closed one form while the negligible terms are the exact ones. One has to prove that the codimension of the negligible terms is one and can be represented by the density gradient term. Because the analysis is carried out seperately in each equilibrium this determines $c = c(\rho)$ and one ends up with an equation of the form

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\frac{\partial}{\partial x}\left[c(\rho)\frac{\partial \rho}{\partial x}\right]. \tag{3.14}$$

The proofs agian involve establishing superexponetial estimates in equlibrium and use Jensen type inequality (1.15) to go from equilibrium to nonequilibrium.

§4. Large Deviations in Hydrodynamic Scaling

Let us consider $k_N \simeq N^d$ independent random walks on the lattice $\mathbf{Z}_N^d$ of $\mathbf{Z}^d$ modulo N. If we denote their trajectories by $\{x_1(\cdot), \ldots, x_{k_N}(\cdot)\}$ and rescale them as $\frac{1}{N}x_i(N^2 t) = y_i(t)$ we have k_N noninteracting rescaled random walks and the empirical process

$$R_{N,\omega} = \frac{1}{N^d}\sum_{i=1}^{k_N} \delta_{y_i(\cdot)} \tag{4.1}$$

will converge on the Skorohod space of trajectories $D[[0,T];\mathbf{T}^d]$ to a Brownian motion with covariance

$$\langle Ca, b\rangle = \sum_z p(z)\langle z, a\rangle\langle z, b\rangle \tag{4.2}$$

where $p(\cdot)$ is the probability distribution of a single step and it is assumed to be symmetric. Of course we need to assume that the initial distribution

$$\nu_{N,\omega} = \frac{1}{N^d}\sum_{i=1}^{k_N} \delta_{y_i(0)} \tag{4.3}$$

has a limit μ and if we take Q to be the Brownian motion with Covariance C and initial distribution μ then

$$\lim_{N\to\infty} R_{N,\omega} = Q \tag{4.4}$$

in probability. Because of the way we have normalized, the total mass of Q, which is the same as total mass of μ, is given by

$$\bar{\rho} = \lim_{N\to\infty} \frac{k_N}{N^d}.$$

We can ask about the probabilities of large deviations in this context. It is a minor variation of Sanov's theorem and the rate function with normalization by N^d is given by the relative entropy

$$\mathcal{I}(R) = H(R; Q). \tag{4.5}$$

We would like to see how this changes if we go from the context of independent random walks to an interacting model like simple exclusion. We will keep the jump distribution as the same $p(\,\cdot\,)$.

We saw before that the hydrodynamic limt in this case was still given by equation (2.19). However the behavior of $R_{N,\omega}$ is more complex. For that we have to understand how a tagged particle will behave in equlibrium as well as nonequlibrium. It is known, (see [6]) that a tagged particle in equlibrium will diffuse like a Brownian Motion with some covariane $S(\rho)$ that depends on the density ρ. This is to be expected, because in low density there is very little interaction and one expects $S(\rho) \to C$ as $\rho \to 0$. On the other hand at high density, i.e., when $\rho \to 1$ there is gridlock and one should expect $S(\rho) \to 0$. (The one dimensional nearest neighbor case is different due to blocking and $S(\rho) \equiv 0$ in that one case.)

One expects therefore that if the initial condition is a random configuration chosen from equilibrium with density ρ then, in probability,

$$\lim_{N\to\infty} R_{N,\omega} = Q_\rho \tag{4.6}$$

where Q_ρ is the Browninan motion (with total mass ρ) having covariance $S(\rho)$ and initial density ρ.

In nonequlibrium the situation is a lot more complicated. First of all the density itself is given by the solution $\rho(t,x)$ of the heat equation (2.19) with initial condition $\rho(x)$ which is determined as the limit of the empirical distribution of the initial configuration in the sense of equation (3.4). The tagged particle will only see its immediate neighborhood and will behave as if it is in equilibrium at density $\rho(t,x)$ if it finds itself at time t at the point x. It is reasonable then to expect it to behave like a diffusion with the second order or diffusion coefficients equal to $S(\rho(t,x))$. It could have an additional first order or drift term. It is more convenient to write the expected backward generator in divergence form as

$$\mathcal{L}_t = \frac{1}{2}\nabla S(\rho(t,x))\nabla + c(t,x)\nabla. \tag{4.7}$$

Of course we can tag any are all of the particles and the empirical process is the same whether they or tagged or not. Therefore the solution $\rho(t,x)$ of the heat equation (2.19) must also be a solution of the forward equation corresponding to (4.7), i.e.,

$$\frac{\partial \rho(t,x)}{\partial t} = \frac{1}{2}\nabla S(\rho(t,x))\nabla\rho(t,x) - \nabla\cdot c(t,x)\rho(t,x). \tag{4.8}$$

This means

$$\frac{1}{2}\nabla C\nabla\rho(t,x) = \frac{1}{2}\nabla S(\rho(t,x))\nabla\rho(t,x) - \nabla\cdot c(t,x)\rho(t,x).$$

One can guess (with fingers crossed) that

$$c(t,x) = [S(\rho(t,x) - C]\frac{\nabla\rho(t,x)}{2\rho(t,x)}. \tag{4.9}$$

That this is indeed true is a result in [9] which is based on results of [7].

We now turn to large deviations. To simplify the presentation we assume that the initial configuration is deterministic. Otherwise we have to factor in the large deviation behavior of the initial profile and this adds an extra term to all the rate functions.

Large deviations are invariably obtained by perturbing the dynamics in such a way that the modification produces the needed deviation. The modified process will have, after suitable normalization, some entropy relative to the original process. This can be thought of as 'cost' of the modification. It is conceivable that there are lots of modifications with different costs that produce the same desired deviation. The rate function is always the minimum of such costs. If one can run through a large class of modifications one gets a large deviation lower bound which is the minimum of the costs over that class of modifications. One tries then to match it with an upper bound by some other method.

In our example the possible perturbations are of the jump rates $N^2p(z)$ of the speeded up dynamics. If the magnitude of the perturbation is $\lambda_N \ll N^2$ then the magnitude of the entropy 'cost' is $\lambda_N^2 N^{-2}$

per particle. This suggests a perurbation of order N to obtain a total entropy 'cost' of order N^d. We therefore consider a perturbed generator of the form

$$(\mathcal{A}_{t,N,q(\cdot,\cdot,\cdot)}f)(\eta) = \sum_{x,y} \eta(x)(1-\eta(y)) \times \Big[N^2 p(y-x) + Nq\Big(t, \frac{x}{N}, \cdot\Big)\Big][f(\eta^{x,y}) - f(\eta)]$$

where $q(t,x,z)$ is a nice function of t, x and z. If we denote by

$$b(t,x) = \sum zq(t,x,z) \tag{4.11}$$

the effect of the perturbation is to produce a solution of the following modified equation as the hydrodynamic limit.

$$\frac{\partial \rho(t,x)}{\partial t} = \frac{1}{2}\nabla C \nabla \rho(t,x) - \nabla \cdot [b(t,x)\rho(t,x)(1-\rho(t,x))] \tag{4.12}$$

with the same initial condition given by (2.20). Given $\rho(\cdot,\cdot)$ we view (4.12) as an equation for $b(\cdot,\cdot)$ and denote the set of solutions by $\mathcal{B}_{\rho(\cdot,\cdot)}$. For a given $b(\cdot,\cdot)$ the set of $q(\cdot,\cdot,\cdot)$ that satisfy (4.11) is denoted by $\mathcal{Q}_{b(\cdot,\cdot)}$. The entropy cost when divided by N^d converges to

$$\frac{1}{2}\int_0^T \int_{T^d} \Big[\sum_z \frac{q(t,x,z)^2}{p(z)}\Big] \rho(t,x)(1-\rho(t,x))\, dxdt. \tag{4.13}$$

Minimizing (4.13) over $\mathcal{Q}_{b(\cdot,\cdot)}$ yields
(4.14)

$$\mathcal{E}(b(\cdot,\cdot)) = \frac{1}{2}\int_0^T \int_{T^d} \langle b(t,x), C^{-1} b(t,x)\rangle \rho(t,x)(1-\rho(t,x))\, dxdt$$

Minimizing $\mathcal{E}(b(\cdot,\cdot))$ over $\mathcal{B}_{\rho(\cdot,\cdot)}$ gives us the rate function for the large deviation of the empirical density which is the family of one dimensionaal marginals of $R_{N,\omega}$. This was done in [5]. Next, we need to consider the effect of the perturbation on the motion of the tagged particle. This will produce for $R_{N,\omega}$, a weak limit Q_b in probability, with the same initial distribution but with the new backward generator

$$\mathcal{L}_b = \frac{1}{2}\nabla S(\rho(t,x)\nabla + c(t,x) \cdot \nabla + b(t,x)(1-\rho(t,x)) \cdot \nabla. \tag{4.15}$$

Finally we can write down the rate function $I(R)$ for the large deviations of $R_{N,\omega}$. From R in addition to its one dimensional marginals

$\rho(\,\cdot\,,\,\cdot\,)$ we can consider the 'currents'

$$\Lambda_R(f) = E^R\left[\int_0^T \langle f(s,x(s)), dx(s)\rangle\right]. \tag{4.16}$$

From $\mathcal{B}_{\rho(\,\cdot\,,\,\cdot\,)}$ we look for a $\bar{b}(\,\cdot\,,\,\cdot\,)$ such that

$$\Lambda_R(f) \equiv \Lambda_{Q_{\bar{b}}}(f)$$

and call it $\overline{R}$. The rate function turns out to be

$$I(R) = \mathcal{E}(\bar{b}(\,\cdot\,,\,\cdot\,)\,) + H(R;\overline{R}). \tag{4.17}$$

If the marginal of R does not match the initial density or if we have trouble defining (4.17) at any stage then $I(R)$ is $+\infty$. It turns out that $\bar{b}(\,\cdot\,,\,\cdot\,)$ if it exists is unique. Details of these results can be found in [8].

References

[1] H. Cramér, Sur un nouveau théeorème-limite de la théorie des probabilités, Actualités Scientifiques et Industrielles, number **736**, Hermann, Paris (1938), 5–23.

[2] J.-D. Dueschel and W. D. Stroock, "Large Deviations", Academic Press, Boston, 1989.

[3] L. Jensen and H.-T. Yau, "Hydrodynamic Scaling Limits of Simple Exclusion Models, Probability Theory and Applications", (P. H. Elton and S. R. S Varadhan, eds.), IAS-Park City Mathematics Series, Vol. **6**, American Mathematical Society, Providence, Rhode Island, 1999, pp. 167–225.

[4] C. Kipnis and C. Landim, "Scaling limits of interacting particle systems", Springer-Varlag, Berlin, 1999.

[5] C. Kipnis, S. Olla and S. R. S. Varadhan, Hydrodynamics and large deviations for simple exclusion processes, Comm. Pure Appl. Math., **42** (1989), 115–137.

[6] C. Kipnis and S. R. S. Varadhan, Central Limit Theorem for additive functionals of reversible Markov processes and applications to simple exclusion, Comm. Math. Phys., **104** (1986), 1–19.

[7] J. Quastel, Diffusion of color in the simple exclusion process, Comm. Pure Appl. Math., **45** (1992), 623–679.

[8] J. Quastel, F. Rezakhanlou and S. R. S. Varadhan, Large deviations for the symmetric simple exclusion process in dimensions $d \geq 3$, Probab. Theory Related Fields, **113,** no. 1 (1999), 1–84.

[9] F. Rezakhanlou, Propagation of chaos for symmetric simple exclusions, Comm. Pure Appl. Math., **47,** no. 7 (1994), 943–957.

[10] H. Spohn, "Large Scale Dynamics of Interacting Particles", Springer-Verlag, New York, 1991.

Courant Institute of Mathematical Sciences
251 Mercer Street
New York, NY 10012
U.S.A.
varadhan@cims.nyu.edu