Large deviation theory:
From physics to mathematics and back

Hugo Touchette

National Institute for Theoretical Physics (NITheP)
Stellenbosch

44th Dutch Stochastics Meeting
Lunteren, The Netherlands
9 November 2015

Plan

Themes

- Typical states
- Fluctuations around typicality
- Many components

Outline

- A bit of history
- Basics of large deviations
- Equilibrium systems
- Nonequilibrium systems

⋮

Lewis (80s)
Graham (80s)

Lanford (1973)

Onsager (1953)

Einstein (1910)

Boltzmann (1877)

⋮

Ellis (1984)

Gärtner (1977)

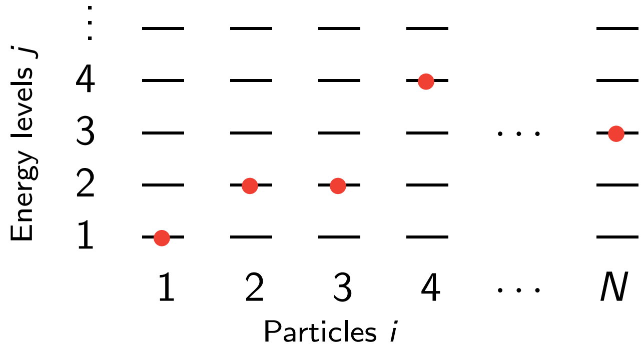
Freidlin-Wentzell (70s)

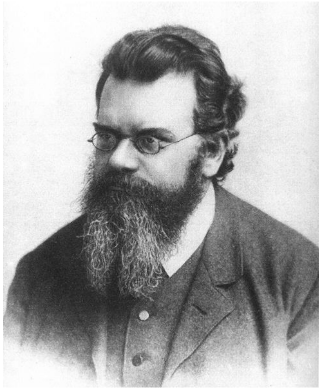
Donsker-Varadhan (70s)

Sanov (1957)

Cramér (1938)

Boltzmann (1877)





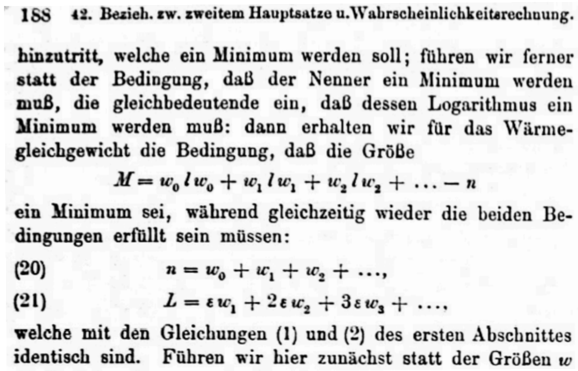
- Energy distribution:

$$w_j = \# \text{ particles in level } j$$

- Multinomial distribution:

$$\ln \frac{N!}{\prod_j w_j!} \approx -N \sum_j w_j \ln w_j = Ns(\mathbf{w})$$

- $P(\mathbf{w}) \approx e^{Ns(\mathbf{w})}$


188 42. Bezieh. zw. zweitem Hauptsatze u. Wahrscheinlichkeitsrechnung.

hinzutritt, welche ein Minimum werden soll; führen wir ferner statt der Bedingung, daß der Nenner ein Minimum werden muß, die gleichbedeutende ein, daß dessen Logarithmus ein Minimum werden muß: dann erhalten wir für das Wärmegleichgewicht die Bedingung, daß die Größe

$$M = w_0 \, l w_0 + w_1 \, l w_1 + w_2 \, l w_2 + \dots - n$$

ein Minimum sei, während gleichzeitig wieder die beiden Bedingungen erfüllt sein müssen:

(20) $n = w_0 + w_1 + w_2 + \dots,$

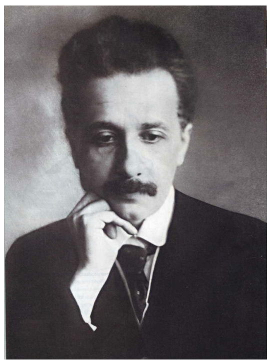
(21) $L = \varepsilon w_1 + 2\varepsilon w_2 + 3\varepsilon w_3 + \dots,$

welche mit den Gleichungen (1) und (2) des ersten Abschnittes identisch sind. Führen wir hier zunächst statt der Größen w

Einstein (1910)

- Generalize Boltzmann
- Macrostate: M_N
- Density of states (complexion):

$$W(m) = \# \text{ microstates with } M_N = m$$



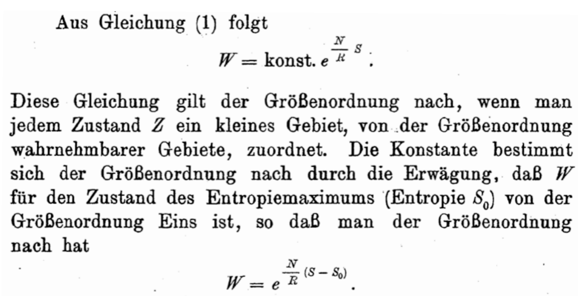
Einstein's postulate

$$W(m) = e^{Ns(m)}$$

- Probability:

$$P(m) = e^{N[s(m)-s(m^*)]}$$

- Equilibrium: $s(m^*)$ is max


Aus Gleichung (1) folgt

$$W = \text{konst.}\, e^{\frac{N}{R} S}.$$

Diese Gleichung gilt der Größenordnung nach, wenn man jedem Zustand Z ein kleines Gebiet, von der Größenordnung wahrnehmbarer Gebiete, zuordnet. Die Konstante bestimmt sich der Größenordnung nach durch die Erwägung, daß W für den Zustand des Entropiemaximums (Entropie S_0) von der Größenordnung Eins ist, so daß man der Größenordnung nach hat

$$W = e^{\frac{N}{R}(S - S_0)}.$$

Cramér (1938)

- Sample mean:

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad X_i \sim p(x) \text{ IID}$$

- Cumulant:

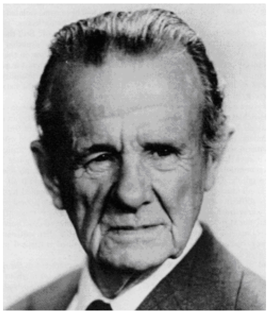
$$\lambda(k) = \ln E[e^{kX}] = \int_{\mathbb{R}} p(x)\, e^{kx}\, dx$$

- Probability density:

$$P(S_n = s) = e^{-nI(s)} \frac{1}{\sqrt{n}} \left(b_0 + \frac{b_1}{n} + \cdots \right)$$

- Rate function:

$$I(s) = \max_{k\in\mathbb{R}} \{ks - \lambda(k)\}$$



Harald Cramér (1893-1985)

On a d'ailleurs $b_0 = \frac{1}{h\bar{\sigma}\sqrt{2\pi}}$. En introduisant dans (21), on obtient donc

$$(28) \quad 1 - F_n\left(\frac{\bar{m}\sqrt{n}}{\sigma}\right) = \frac{1}{\sqrt{n}} e^{-(h\bar{m} - \log R)n} \left[b_0 + \frac{b_1}{n} + \cdots + \frac{b_{k-1}}{n^{k-1}} + O\left(\frac{1}{n^k}\right) \right].$$

Soit maintenant c un nombre donné tel que $0 < c < C_1$, et prenons h égal à la racine (unique) positive de l'équation (27). En introduisant cette valeur dans (28) et en posant

$$(29) \qquad \alpha = h\bar{m} - \log R$$

(où l'on voit facilement que α est toujours positif), on a le théorème suivant.

Sanov (1957)

- Sequence of IID RVs:

$$X_1, X_2, \ldots, X_n \qquad X_i \sim p(x)$$

- Empirical distribution:

$$L_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i,x}$$

$$P(L_n = \rho) \approx e^{-nD(\rho||p)}$$

- Relative entropy:

$$D(\rho||p) = \int dx\, \rho(x) \ln \frac{\rho(x)}{p(x)}$$

- Law of Large Numbers: $L_n \to \rho$

Ivan Nikolaevich Sanov (1919-1968)

Теорема 10. *Пусть $F(x)$ — функция распределения случайной величины ξ, $F_N(x)$ — эмпирическая функция распределения после N независимых наблюдений случайной величины ξ. Пусть $\Phi(x)$ — другая функция распределения, такая, что $\int_{-\infty}^{+\infty} \ln \frac{dF}{d\Phi} d\Phi$ существует. Пусть V_n — последовательность ε-окрестностей, содержащих $\Phi(x)$ и F-сходящихся к ней.*

Тогда

$$P(F_N \in V_n) = e^{N\left[\int_{-\infty}^{+\infty} \ln \frac{dF}{d\Phi} d\Phi + \delta_n + O\left(\frac{n \ln N}{N}\right)\right]}, \qquad (45)$$

Large deviation theory

- Random variable: A_n
- Probability density: $P(A_n = a)$

Large deviation principle (LDP)

$$P(A_n = a) \approx e^{-nI(a)}$$

- Meaning of $\approx$:

$$\ln P(a) = -nI(a) + o(n)$$
$$\lim_{n\to\infty} -\frac{1}{n}\ln P(a) = I(a)$$

- Rate function: $I(a) \geq 0$

Goals of large deviation theory

1. Prove that a large deviation principle exists
2. Calculate the rate function

Varadhan's Theorem

- LDP:

$$P(A_n = a) \approx e^{-nI(a)}$$

- Exponential expectation:

$$E[e^{nf(A_n)}] = \int e^{nf(a)}\, P(A_n = a)\, da$$

- Limit functional:

$$\lambda(f) = \lim_{n\to\infty} \frac{1}{n}\ln E[e^{nf(A_n)}]$$

S. R. Srinivasa Varadhan
Abel Prize 2007

Theorem: Varadhan (1966)

$$\lambda(f) = \max_a\{f(a) - I(a)\}$$

Special case: $f(a) = ka$

$$\lambda(k) = \max_a\{ka - I(a)\}$$

Gärtner-Ellis Theorem

Scaled cumulant generating function (SCGF)

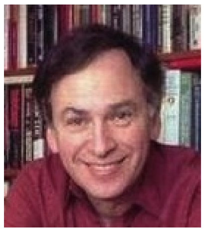
$$\lambda(k) = \lim_{n\to\infty} \frac{1}{n} \ln E[e^{nkA_n}], \qquad k \in \mathbb{R}$$

Theorem: Gärtner (1977), Ellis (1984)

If $\lambda(k)$ is differentiable, then

1. LDP:
$$P(A_n = a) \approx e^{-nI(a)}$$
2. Rate function:
$$I(a) = \max_k \{ka - \lambda(k)\}$$

- $I(a)$ is the Legendre transform of $\lambda(k)$



Richard S. Ellis

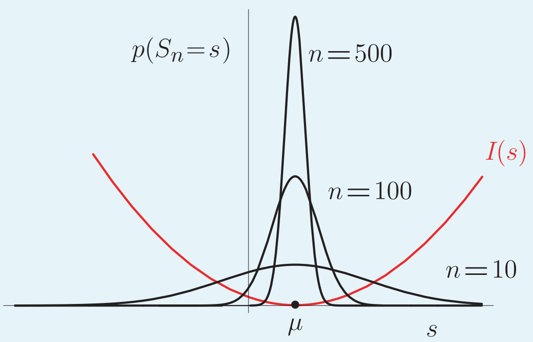
J. Gärtner

Cramer's Theorem

- Sample mean:
$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad X_i \sim p(x), \quad \text{IID}$$
- SCGF:
$$\lambda(k) = \lim_{n\to\infty} \frac{1}{n} \ln E[e^{nkS_n}] = \ln E[e^{kX}]$$

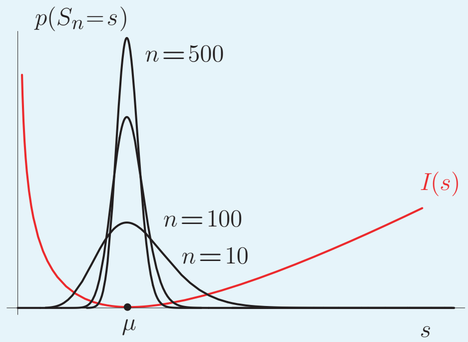
Gaussian

$$\lambda(k) = \mu k + \frac{\sigma^2}{2}k^2, \quad k \in \mathbb{R}$$
$$I(s) = \frac{1}{2\sigma^2}(s-\mu)^2, \quad s \in \mathbb{R}$$



Exponential

$$\lambda(k) = -\ln(1-\mu k), \quad k < \frac{1}{\mu}$$
$$I(s) = \frac{s}{\mu} - 1 - \ln\frac{s}{\mu}, \quad s > 0$$



Sanov's Theorem

- n IID random variables:

$$\omega = \omega_1, \omega_2, \ldots, \omega_n, \qquad P(\omega_i = j) = p_j$$

- Empirical frequencies:

$$L_{n,j} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\omega_i,j} = \frac{\#(\omega_i = j)}{n}, \qquad \mathbf{L}_n = (L_{n,1}, L_{n,2}, \ldots)$$

Gärtner-Ellis

- SCGF:

$$\lambda(\mathbf{k}) = \lim_{n\to\infty} \frac{1}{n} \ln E[e^{n\mathbf{k}\cdot\mathbf{L}_n}] = \ln \sum_{j=1}^{q} p_j\ e^{k_j}$$

- Rate function:

$$D(\boldsymbol{\mu}) = \inf_{\mathbf{k}}\{\mathbf{k}\cdot\boldsymbol{\mu} - \lambda(\mathbf{k})\} = \sum_{j=1}^{q} \mu_j \ln \frac{\mu_j}{p_j}$$

Beyond IID

Markov processes

$$\{X_t\}_{t=0}^{T}$$

$$A_T = \tfrac{1}{T}\int_0^T f(X_t)\,dt$$

- $P(A_T = a) \approx e^{-TI(a)}$
- Long time limit
- Donsker & Varadhan (1975)

SDEs

$$\dot{x}(t) = f(x(t)) + \sqrt{\epsilon}\,\xi(t)$$

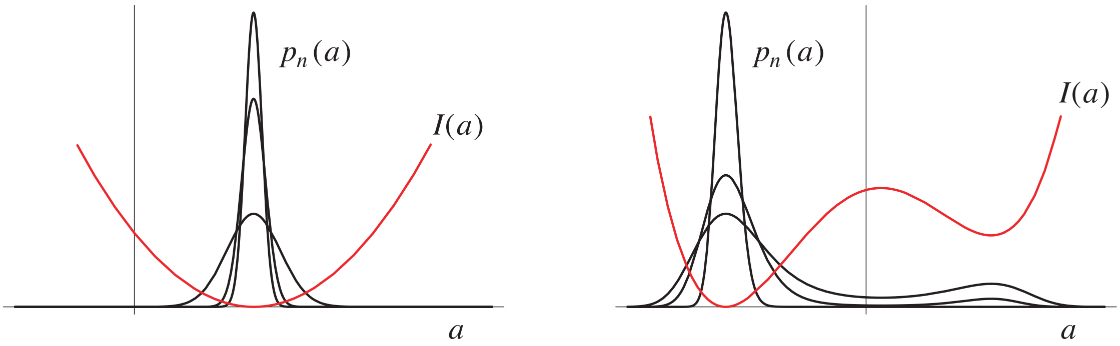
- $P[x] \approx e^{-I[x]/\epsilon}$
- Low noise limit
- Freidlin & Wentzell (1970s)
- Onsager & Machlup (1953)

Applications

- Noisy dynamical systems
- Interacting SDEs
- Stochastic PDEs
- Interacting particle systems
- RWs random environments
- Queueing theory
- Statistics, estimation
- Information theory

Summary

$$P(A_n = a) \approx e^{-nI(a)}$$

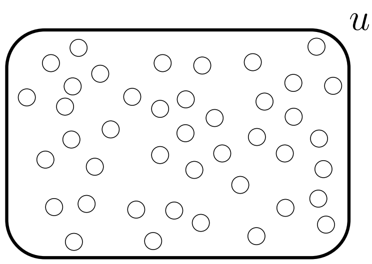


- Law of Large Numbers
  - Typical value = zeros of $I(a)$
- Central Limit Theorem
  - Quadratic minima = Gaussian fluctuations
  - Small deviations
- Large deviations
  - Fluctuations away from typical value

General theory of typical states and fluctuations

Equilibrium systems

- N particles
- Microstate: $\omega = \omega_1, \omega_2, \ldots, \omega_N$
- Statistical ensemble: $P(\omega)$
- Macrostate: $M_N(\omega)$
- Macrostate distribution:



$$P(M_N = m) = \sum_{\omega : M_N(\omega) = m} P(\omega)$$

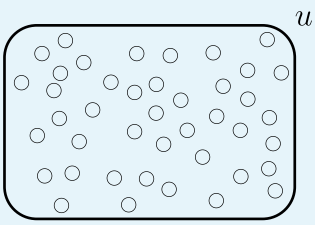
Problems

- Calculate $P(M_N = m)$
- Find most probable values of M_N (= equilibrium states)
- Study fluctuations around most probable values
- Thermodynamic limit $N \to \infty$

Equilibrium large deviations

Microcanonical

Einstein (1910)



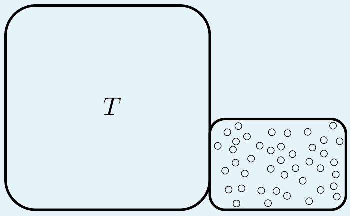
$$P_u(M_N = m) = e^{S(u,m)/k_B}$$

- Extensivity: $S \sim N$
- LDP:

$$P_u(M_N = m) \approx e^{-NI_u(m)}$$

Canonical

Landau (1937)



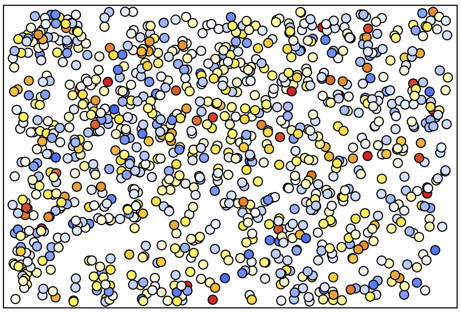
$$P_\beta(M_N = m) = e^{-F(\beta,m)}$$

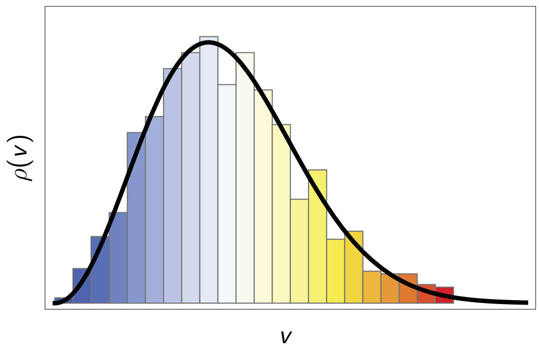
- Extensivity: $F \sim N$
- LDP:

$$P_\beta(M_N = m) \approx e^{-NI_\beta(m)}$$

- Exponential concentration of probability
- Equilibrium states = minima and zeros of I

Maxwell distribution





- Velocity distribution:

$$L_N(v) = \frac{\#\text{ particles with } v_i \in [v, v+\Delta v]}{N\Delta v}$$

Sanov's Theorem

$$P_u(L_N = \rho) \approx e^{-NI_u(\rho)}$$

- Equilibrium distribution:

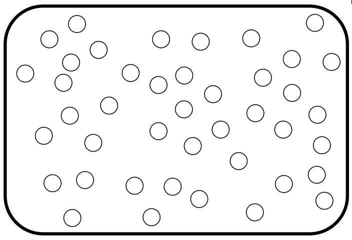
$$\rho^*(v) = c\, v^2 e^{-\frac{mv^2}{2k_B T}}$$

Entropy and free energy

- Density of states:

$$\Omega(u) = \#\ \omega \text{ with } U/N = u$$

- Large deviation form: $\Omega(u) \approx e^{Ns(u)}$



Gärtner-Ellis Theorem

$$s(u) = \min_\beta \{\beta u - \varphi(\beta)\}$$

- Free energy:

$$\varphi(\beta) = \lim_{N\to\infty} -\frac{1}{N} \ln Z(\beta), \qquad Z(\beta) = \int d\omega\ e^{-\beta U(\omega)}$$

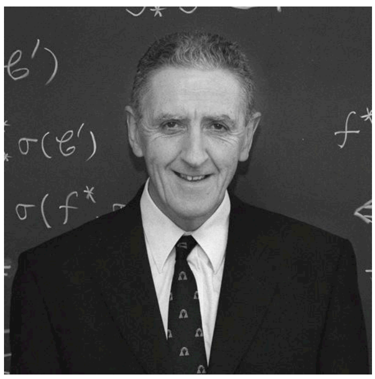
- $Z(\beta)$ = partition function = generating function
- $\varphi(\beta)$ = free energy = SCGF
- Basis of Legendre transform in thermodynamics

Sources and applications

- Finite-range systems
  Lanford (1973)
- Spin systems
  Ellis (1980s)
- Bose condensation
  Lewis (1980s)
- 2D turbulence
- Long-range systems
- Quantum systems
  Lenci, Lebowitz (2000)
- Spin glasses

- Large deviation structure
- Typical states and fluctuations

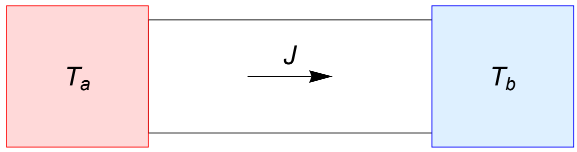
Oscar Lanford III (1940-2013)



John T. Lewis (1932-2004)

Nonequilibrium systems

- Process: X_t
  - One or many particles
  - Markov process
  - External forces
  - Boundary reservoirs
- Trajectory: $\{x_t\}_{t=0}^T$
- Path distribution: $P[x]$
- Observable: $A_{N,T}[x]$



Problems

- Calculate $P(A_{N,T} = a)$
- Find most probable values of $A_{N,T}$ (= stationary states)
- Study fluctuations around typical values
- Scaling limits:

$$N \to \infty \qquad T \to \infty \qquad \text{noise} \to 0$$

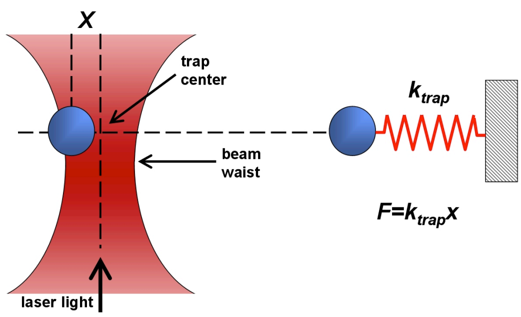
Example: Pulled Brownian particle

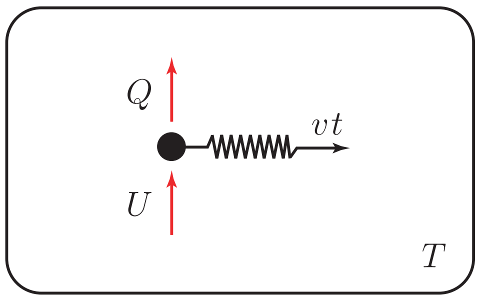
- Glass bead in water
- Laser tweezers
- Langevin dynamics:

$$m\ddot{x}(t) = \underbrace{-\alpha\dot{x}}_{\text{drag}} \underbrace{-k[x(t)-vt]}_{\text{spring force}} + \underbrace{\xi(t)}_{\text{noise}}$$

- Fluctuating work:

$$\underbrace{W_T}_{\text{work}} = \underbrace{\Delta U}_{\text{potential}} + \underbrace{Q_T}_{\text{heat}}$$





LDP

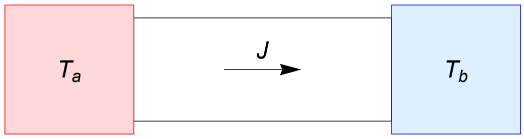
$$P(W_T = w) \approx e^{-TI(w)}$$

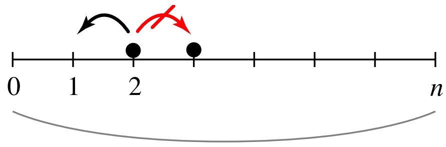
Fluctuation relation

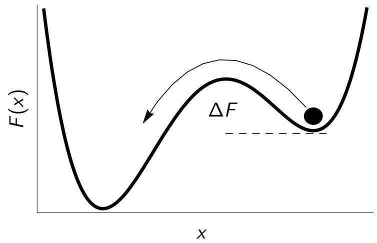
$$\frac{P(W_T = w)}{P(W_T = -w)} = e^{Tcw}$$

Applications

- Driven nonequilibrium systems
- Interacting particle models
  - Current, density fluctuations
  - Macroscopic, hydrodynamic limit
- Thermal activation
  - Kramers escape problem
- Disorded systems
- Multifractals
- Chaotic systems
- Quantum systems







- Exponentially rare fluctuations
- Exponential concentration of typical states
- Same theory for equilibrium and nonequilibrium systems

Summary

- Random variables — ensembles — stochastic systems
- Most probable values — equilibrium states — typical states
- Fluctuations — rare events
- Rate function = entropy
- Cumulant function = free energy
- Scaling limit: $N \to \infty$, $T \to \infty$, $\epsilon \to 0$
- Unified language for statistical mechanics

H. Touchette
The large deviation approach to statistical mechanics
Physics Reports **478**, 1-69, 2009

`www.physics.sun.ac.za/~htouchette`

Next talk

- Markov processes conditioned on large deviations
- When a fluctuation happens, how does it happen?