Large fluctuations, weak convergence and all that

Razvan Teodorescu
T-13 and CNLS

June 14, 2007

Justification and Outline

Justification and Outline

- "Statistics" of non-equilibrium statistical physics papers (articles/day):

Justification and Outline

- "Statistics" of non-equilibrium statistical physics papers (articles/day):

2001-2003	50.68
2003-2005	58.86
2005-2007	72.19

Justification and Outline

- "Statistics" of non-equilibrium statistical physics papers (articles/day):

2001-2003	50.68
2003-2005	58.86
2005-2007	72.19

- Subjects represented at CNLS/T-DO:

Justification and Outline

- "Statistics" of non-equilibrium statistical physics papers (articles/day):

2001-2003	50.68
2003-2005	58.86
2005-2007	72.19

- Subjects represented at CNLS/T-DO:
  - ⋆ Fluctuations-dissipation in driven granular gases, 2D turbulence, . . .

Justification and Outline

- "Statistics" of non-equilibrium statistical physics papers (articles/day):

2001-2003	50.68
2003-2005	58.86
2005-2007	72.19

- Subjects represented at CNLS/T-DO:
  - ⋆ Fluctuations-dissipation in driven granular gases, 2D turbulence, . . .
  - ⋆ Large (extreme) deviations in non-linear PDE's

Justification and Outline

- "Statistics" of non-equilibrium statistical physics papers (articles/day):

2001-2003	50.68
2003-2005	58.86
2005-2007	72.19

- Subjects represented at CNLS/T-DO:
  - ⋆ Fluctuations-dissipation in driven granular gases, 2D turbulence, . . .
  - ⋆ Large (extreme) deviations in non-linear PDE's
  - ⋆ (Lax-Oleinik) entropy production in PDE's of hyperbolic type

Justification and Outline

- "Statistics" of non-equilibrium statistical physics papers (articles/day):

2001-2003	50.68
2003-2005	58.86
2005-2007	72.19

- Subjects represented at CNLS/T-DO:
  - ⋆ Fluctuations-dissipation in driven granular gases, 2D turbulence, . . .
  - ⋆ Large (extreme) deviations in non-linear PDE's
  - ⋆ (Lax-Oleinik) entropy production in PDE's of hyperbolic type
  - ⋆ Syncronization transition in stochastically coupled oscillators

Justification and Outline

- "Statistics" of non-equilibrium statistical physics papers (articles/day):

2001-2003	50.68
2003-2005	58.86
2005-2007	72.19

- Subjects represented at CNLS/T-DO:
  - ⋆ Fluctuations-dissipation in driven granular gases, 2D turbulence, . . .
  - ⋆ Large (extreme) deviations in non-linear PDE's
  - ⋆ (Lax-Oleinik) entropy production in PDE's of hyperbolic type
  - ⋆ Syncronization transition in stochastically coupled oscillators
  - ⋆ Kullback-Leibner divergence in quantum entanglement

Justification and Outline

- "Statistics" of non-equilibrium statistical physics papers (articles/day):

2001-2003	50.68
2003-2005	58.86
2005-2007	72.19

- Subjects represented at CNLS/T-DO:
  - ⋆ Fluctuations-dissipation in driven granular gases, 2D turbulence, . . .
  - ⋆ Large (extreme) deviations in non-linear PDE's
  - ⋆ (Lax-Oleinik) entropy production in PDE's of hyperbolic type
  - ⋆ Syncronization transition in stochastically coupled oscillators
  - ⋆ Kullback-Leibner divergence in quantum entanglement

A gentlemen's disagreement: large sample vs. strong beliefs

- An English affair . . .

A gentlemen's disagreement: large sample vs. strong beliefs

- An English affair . . .
  - Rev. Thomas Bayes (1702–1761): "all you need is a good prior (belief)"

A gentlemen's disagreement: large sample vs. strong beliefs

- An English affair . . .
  - ⋆ Rev. Thomas Bayes (1702–1761): "all you need is a good prior (belief)" – Bayesian inference

A gentlemen's disagreement: large sample vs. strong beliefs

- An English affair . . .
  - Rev. Thomas Bayes (1702–1761): "all you need is a good prior (belief)" – Bayesian inference
  - Sir Ronald Aylmer Fisher (1890–1962): "science starts with a large sample"

A gentlemen's disagreement: large sample vs. strong beliefs

- An English affair . . .
  - Rev. Thomas Bayes (1702–1761): "all you need is a good prior (belief)" – Bayesian inference
  - Sir Ronald Aylmer Fisher (1890–1962): "science starts with a large sample" – Fisherian inference

A gentlemen's disagreement: large sample vs. strong beliefs

- An English affair . . .
  - ⋆ Rev. Thomas Bayes (1702–1761): "all you need is a good prior (belief)" – Bayesian inference
  - ⋆ Sir Ronald Aylmer Fisher (1890–1962): "science starts with a large sample" – Fisherian inference

- . . . and not so English:

A gentlemen's disagreement: large sample vs. strong beliefs

- An English affair . . .
  - Rev. Thomas Bayes (1702–1761): "all you need is a good prior (belief)" – Bayesian inference
  - Sir Ronald Aylmer Fisher (1890–1962): "science starts with a large sample" – Fisherian inference

- . . . and not so English:
  - Cramér, Rao, Borel, Cantelli, Glivenko, Neyman

A gentlemen's disagreement: large sample vs. strong beliefs

- An English affair . . .
  - Rev. Thomas Bayes (1702–1761): "all you need is a good prior (belief)" – Bayesian inference
  - Sir Ronald Aylmer Fisher (1890–1962): "science starts with a large sample" – Fisherian inference
- . . . and not so English:
  - Cramér, Rao, Borel, Cantelli, Glivenko, Neyman
  - Blackwell, Ferguson, Lehman, Shannon, Kolmogorov

A gentlemen's disagreement: large sample vs. strong beliefs

- An English affair . . .
  - Rev. Thomas Bayes (1702–1761): "all you need is a good prior (belief)" – Bayesian inference
  - Sir Ronald Aylmer Fisher (1890–1962): "science starts with a large sample" – Fisherian inference
- . . . and not so English:
  - Cramér, Rao, Borel, Cantelli, Glivenko, Neyman
  - Blackwell, Ferguson, Lehman, Shannon, Kolmogorov
  - Basu, Levy, Dynkin, Wentzell, Freidlin, Varadhan

2007 Abel Prize - S.S.R. Varadhan (the "Nobel prize of mathematics")

" for his fundamental contributions to probability theory and in particular for creating a unified theory of large deviations. " - May 23, 2007.

Equilibrium statistics - a physicist's view

- Equilibrium = extremum of the appropriate thermodynamic potential (entropy, subject to constraints).

Equilibrium statistics - a physicist's view

- Equilibrium = extremum of the appropriate thermodynamic potential (entropy, subject to constraints).
- Why?

Equilibrium statistics - a physicist's view

- Equilibrium = extremum of the appropriate thermodynamic potential (entropy, subject to constraints).
- Why?
- Because Boltzmann said so;

Equilibrium statistics - a physicist's view

- Equilibrium = extremum of the appropriate thermodynamic potential (entropy, subject to constraints).
- Why?
- Because Boltzmann said so;
- Because Gibbs said so;

Equilibrium statistics - a physicist's view

- Equilibrium = extremum of the appropriate thermodynamic potential (entropy, subject to constraints).
- Why?
- Because Boltzmann said so;
- Because Gibbs said so;
- No,

Equilibrium statistics - a physicist's view

- Equilibrium = extremum of the appropriate thermodynamic potential (entropy, subject to constraints).
- Why?
- Because Boltzmann said so;
- Because Gibbs said so;
- No, it's because Shannon said so!

Equilibrium statistics - a physicist's view

- Equilibrium $=$ extremum of the appropriate thermodynamic potential (entropy, subject to constraints).
- Why?
- Because Boltzmann said so;
- Because Gibbs said so;
- No, it's because Shannon said so!

$$S(A \cup B) = S(A) + S(B) \Rightarrow S = -k_B \langle \log p \rangle = -k_B \sum_i p_i \log p_i.$$

The canonical ensemble - a refresher

- Maximize S subject to constraints $\sum_i p_i = 1$, $\sum_i p_i \epsilon_i = E$:

The canonical ensemble - a refresher

- Maximize S subject to constraints $\sum_i p_i = 1, \ \sum_i p_i \epsilon_i = E$:

The canonical ensemble - a refresher

- Maximize S subject to constraints $\sum_i p_i = 1, \ \sum_i p_i\epsilon_i = E$:
- $W(\{p_i\}) = -k_B \sum p_i \log p_i + \lambda_1(\sum_i p_i\epsilon_i - 1) + \lambda_2(\sum_i p_i - 1)$

The canonical ensemble - a refresher

- Maximize S subject to constraints $\sum_i p_i = 1, \ \sum_i p_i \epsilon_i = E$:
- $W(\{p_i\}) = -k_B \sum p_i \log p_i + \lambda_1 (\sum_i p_i \epsilon_i - 1) + \lambda_2 (\sum_i p_i - 1)$
- $\frac{\partial W}{\partial p_i} = 0 \rightarrow$

The canonical ensemble - a refresher

- Maximize S subject to constraints $\sum_i p_i = 1, \; \sum_i p_i \epsilon_i = E$:
- $W(\{p_i\}) = -k_B \sum p_i \log p_i + \lambda_1 (\sum_i p_i \epsilon_i - 1) + \lambda_2 (\sum_i p_i - 1)$
- $\frac{\partial W}{\partial p_i} = 0 \rightarrow \log p_i + 1 = \lambda_1 \epsilon_i + \lambda_2 \rightarrow$

The canonical ensemble - a refresher

- Maximize S subject to constraints $\sum_i p_i = 1, \ \sum_i p_i \epsilon_i = E$:
- $W(\{p_i\}) = -k_B \sum p_i \log p_i + \lambda_1 (\sum_i p_i \epsilon_i - 1) + \lambda_2 (\sum_i p_i - 1)$
- $\frac{\partial W}{\partial p_i} = 0 \rightarrow \ \log p_i + 1 = \lambda_1 \epsilon_i + \lambda_2 \rightarrow p_i = Z^{-1} e^{\lambda_1 \epsilon_i}$.

The canonical ensemble - a refresher

- Maximize S subject to constraints $\sum_i p_i = 1,\ \sum_i p_i \epsilon_i = E$:
- $W(\{p_i\}) = -k_B \sum p_i \log p_i + \lambda_1(\sum_i p_i \epsilon_i - 1) + \lambda_2(\sum_i p_i - 1)$
- $\frac{\partial W}{\partial p_i} = 0 \to \log p_i + 1 = \lambda_1 \epsilon_i + \lambda_2 \to p_i = Z^{-1} e^{\lambda_1 \epsilon_i}$.
- $S = -k_B \sum_i p_i \log p_i \to$

The canonical ensemble - a refresher

- Maximize S subject to constraints $\sum_i p_i = 1, \;\sum_i p_i \epsilon_i = E$:
- $W(\{p_i\}) = -k_B \sum p_i \log p_i + \lambda_1(\sum_i p_i \epsilon_i - 1) + \lambda_2(\sum_i p_i - 1)$
- $\frac{\partial W}{\partial p_i} = 0 \rightarrow \; \log p_i + 1 = \lambda_1 \epsilon_i + \lambda_2 \rightarrow p_i = Z^{-1} e^{\lambda_1 \epsilon_i}$.
- $S = -k_B \sum_i p_i \log p_i \rightarrow \lambda_1^{-1} \log Z = E + (k_B \lambda_1)^{-1} S$

The canonical ensemble - a refresher

- Maximize S subject to constraints $\sum_i p_i = 1, \; \sum_i p_i \epsilon_i = E$:
- $W(\{p_i\}) = -k_B \sum p_i \log p_i + \lambda_1(\sum_i p_i \epsilon_i - 1) + \lambda_2(\sum_i p_i - 1)$
- $\frac{\partial W}{\partial p_i} = 0 \rightarrow \; \log p_i + 1 = \lambda_1 \epsilon_i + \lambda_2 \rightarrow p_i = Z^{-1} e^{\lambda_1 \epsilon_i}$.
- $S = -k_B \sum_i p_i \log p_i \rightarrow \lambda_1^{-1} \log Z = E + (k_B \lambda_1)^{-1} S$
- Compare to $F = E - TS \rightarrow$

The canonical ensemble - a refresher

- Maximize S subject to constraints $\sum_i p_i = 1, \ \sum_i p_i \epsilon_i = E$:
- $W(\{p_i\}) = -k_B \sum p_i \log p_i + \lambda_1 (\sum_i p_i \epsilon_i - 1) + \lambda_2 (\sum_i p_i - 1)$
- $\frac{\partial W}{\partial p_i} = 0 \to \ \log p_i + 1 = \lambda_1 \epsilon_i + \lambda_2 \to p_i = Z^{-1} e^{\lambda_1 \epsilon_i}$.
- $S = -k_B \sum_i p_i \log p_i \to \lambda_1^{-1} \log Z = E + (k_B \lambda_1)^{-1} S$
- Compare to $F = E - TS \to \lambda_1 = -(k_B T)^{-1}$

The canonical ensemble - a refresher

- Maximize S subject to constraints $\sum_i p_i = 1, \ \sum_i p_i \epsilon_i = E$:
- $W(\{p_i\}) = -k_B \sum p_i \log p_i + \lambda_1 (\sum_i p_i \epsilon_i - 1) + \lambda_2 (\sum_i p_i - 1)$
- $\frac{\partial W}{\partial p_i} = 0 \rightarrow \ \log p_i + 1 = \lambda_1 \epsilon_i + \lambda_2 \rightarrow p_i = Z^{-1} e^{\lambda_1 \epsilon_i}$.
- $S = -k_B \sum_i p_i \log p_i \rightarrow \lambda_1^{-1} \log Z = E + (k_B \lambda_1)^{-1} S$
- Compare to $F = E - TS \rightarrow \lambda_1 = -(k_B T)^{-1}$
- $\boxed{p_i = Z^{-1} e^{-\epsilon_i / k_B T}, \ Z = e^{-F/k_B T}.}$

The canonical ensemble - a refresher

- Maximize S subject to constraints $\sum_i p_i = 1, \; \sum_i p_i \epsilon_i = E$:
- $W(\{p_i\}) = -k_B \sum p_i \log p_i + \lambda_1 (\sum_i p_i \epsilon_i - 1) + \lambda_2 (\sum_i p_i - 1)$
- $\frac{\partial W}{\partial p_i} = 0 \rightarrow \; \log p_i + 1 = \lambda_1 \epsilon_i + \lambda_2 \rightarrow p_i = Z^{-1} e^{\lambda_1 \epsilon_i}$.
- $S = -k_B \sum_i p_i \log p_i \rightarrow \lambda_1^{-1} \log Z = E + (k_B \lambda_1)^{-1} S$
- Compare to $F = E - TS \rightarrow \lambda_1 = -(k_B T)^{-1}$
- $\boxed{p_i = Z^{-1} e^{-\epsilon_i / k_B T},\; Z = e^{-F/k_B T}.}$

A closer look at entropy production - a Fisherian (frequentist) approach

- Why does entropy increase?

A closer look at entropy production - a Fisherian (frequentist) approach

- Why does entropy increase?
- Because of Central Limit Theorem (R.A. Fisher)

A closer look at entropy production - a Fisherian (frequentist) approach

- Why does entropy increase?
- Because of Central Limit Theorem (R.A. Fisher)
- At given mean μ and variance σ^2, the normal distribution maximizes the Shannon entropy:

A closer look at entropy production - a Fisherian (frequentist) approach

- Why does entropy increase?
- Because of Central Limit Theorem (R.A. Fisher)
- At given mean μ and variance σ^2, the normal distribution maximizes the Shannon entropy:

$$S(X) \geq S(Y),\ X \sim N(\mu, \sigma^2),\ E(Y) = \mu,\ V(Y) = \sigma^2.$$

Entropy production in generalized sense

- Systems of PDE's with hyperbolic-type conservation laws

Entropy production in generalized sense

- Systems of PDE's with hyperbolic-type conservation laws

$$\partial_t \mathbf{u} + \nabla \cdot \mathbf{f}(\mathbf{u}) = 0, \quad \mathbf{u} \in \mathbb{R}^n, \quad \mathbf{x} \in \mathbb{R}^d$$

Entropy production in generalized sense

- Systems of PDE's with hyperbolic-type conservation laws

$$\partial_t \mathbf{u} + \nabla \cdot \mathbf{f}(\mathbf{u}) = 0, \quad \mathbf{u} \in \mathbb{R}^n, \quad \mathbf{x} \in \mathbb{R}^d$$

  Known to develop singularities (shocks, vortices) from smooth initial data

Entropy production in generalized sense

- Systems of PDE's with hyperbolic-type conservation laws

$$\partial_t \mathbf{u} + \nabla \cdot \mathbf{f}(\mathbf{u}) = 0, \quad \mathbf{u} \in \mathbb{R}^n, \quad \mathbf{x} \in \mathbb{R}^d$$

  Known to develop singularities (shocks, vortices) from smooth initial data

- How to continue the solution through a singularity? (many choices)

Entropy production in generalized sense

- Systems of PDE's with hyperbolic-type conservation laws

$$\partial_t \mathbf{u} + \nabla \cdot \mathbf{f}(\mathbf{u}) = 0, \quad \mathbf{u} \in \mathbb{R}^n, \quad \mathbf{x} \in \mathbb{R}^d$$

  Known to develop singularities (shocks, vortices) from smooth initial data

- How to continue the solution through a singularity? (many choices)

- Peter Lax (1953): physical solution increases the Lax-Oleinik entropy

Entropy production in generalized sense

- Systems of PDE's with hyperbolic-type conservation laws

$$\partial_t \mathbf{u} + \nabla \cdot \mathbf{f}(\mathbf{u}) = 0, \quad \mathbf{u} \in \mathbb{R}^n, \quad \mathbf{x} \in \mathbb{R}^d$$

  Known to develop singularities (shocks, vortices) from smooth initial data

- How to continue the solution through a singularity? (many choices)
- Peter Lax (1953): physical solution increases the Lax-Oleinik entropy
- Awarded the 2005 Abel prize

Entropy production in generalized sense

- Systems of PDE's with hyperbolic-type conservation laws

$$\partial_t \mathbf{u} + \nabla \cdot \mathbf{f}(\mathbf{u}) = 0, \quad \mathbf{u} \in \mathbb{R}^n, \quad \mathbf{x} \in \mathbb{R}^d$$

  Known to develop singularities (shocks, vortices) from smooth initial data

- How to continue the solution through a singularity? (many choices)

- Peter Lax (1953): physical solution increases the Lax-Oleinik entropy

- Awarded the 2005 Abel prize

Elements of large sample theory

- Sample of

Elements of large sample theory

- Sample of i.

Elements of large sample theory

- Sample of i.i.d.

Elements of large sample theory

- Sample of i.i.d. variables $X_1, X_2, \dots, X_n$

Elements of large sample theory

- Sample of i.i.d. variables $X_1, X_2, \ldots, X_n$
- p.d.f. $f_{X_n}(x)$, c.d.f. $F_{X_n}(x)$, m.g.f. $m_{X_n}(t) = E(e^{tX_n})$

Elements of large sample theory

- Sample of i.i.d. variables $X_1, X_2, \ldots, X_n$
- p.d.f. $f_{X_n}(x)$, c.d.f. $F_{X_n}(x)$, m.g.f. $m_{X_n}(t) = E(e^{tX_n})$
- $X_n \to Y$ weakly (in distribution) if

Elements of large sample theory

- Sample of i.i.d. variables $X_1, X_2, \ldots, X_n$
- p.d.f. $f_{X_n}(x)$, c.d.f. $F_{X_n}(x)$, m.g.f. $m_{X_n}(t) = E(e^{tX_n})$
- $X_n \to Y$ weakly (in distribution) if $F_{X_n}(x) \to F_Y(x)$

Elements of large sample theory

- Sample of i.i.d. variables $X_1, X_2, \ldots, X_n$
- p.d.f. $f_{X_n}(x)$, c.d.f. $F_{X_n}(x)$, m.g.f. $m_{X_n}(t) = E(e^{tX_n})$
- $X_n \to Y$ weakly (in distribution) if $F_{X_n}(x) \to F_Y(x)$
- $X_n \xrightarrow{d} Y$ iff. $E(e^{itX_n}) \to E(e^{itY})$ (characteristic functions)

Elements of large sample theory

- Sample of i.i.d. variables $X_1, X_2, \ldots, X_n$
- p.d.f. $f_{X_n}(x)$, c.d.f. $F_{X_n}(x)$, m.g.f. $m_{X_n}(t) = E(e^{tX_n})$
- $X_n \to Y$ weakly (in distribution) if $F_{X_n}(x) \to F_Y(x)$
- $X_n \xrightarrow{d} Y$ iff. $E(e^{itX_n}) \to E(e^{itY})$ (characteristic functions)
- Moments are not enough: $\frac{e^{-(\log x)^2/2}}{\sqrt{2\pi x}}$ and $\frac{e^{-(\log x)^2/2}}{\sqrt{2\pi x}}[1 + \sin(2\pi \log x)]$ have equal moments.

Elements of large sample theory

- Sample of i.i.d. variables $X_1, X_2, \ldots, X_n$
- p.d.f. $f_{X_n}(x)$, c.d.f. $F_{X_n}(x)$, m.g.f. $m_{X_n}(t) = E(e^{tX_n})$
- $X_n \to Y$ weakly (in distribution) if $F_{X_n}(x) \to F_Y(x)$
- $X_n \xrightarrow{d} Y$ iff. $E(e^{itX_n}) \to E(e^{itY})$ (characteristic functions)
- Moments are not enough: $\frac{e^{-(\log x)^2/2}}{\sqrt{2\pi x}}$ and $\frac{e^{-(\log x)^2/2}}{\sqrt{2\pi x}}[1 + \sin(2\pi \log x)]$ have equal moments.

Weak convergence theorems

- Central Limit Theorem:

Weak convergence theorems

- Central Limit Theorem: if X_i i.i.d. with mean μ, variance σ^2, then

Weak convergence theorems

- Central Limit Theorem: if X_i i.i.d. with mean μ, variance σ^2, then

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \to N(\mu, \sigma^2/n)$$

Weak convergence theorems

- Central Limit Theorem: if X_i i.i.d. with mean μ, variance σ^2, then

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \rightarrow N(\mu, \sigma^2/n)$$

Generalization for non-identical, independent variables: Lindeberg-Feller theorem

Weak convergence theorems

- Central Limit Theorem: if X_i i.i.d. with mean μ, variance σ^2, then

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \to N(\mu, \sigma^2/n)$$

  Generalization for non-identical, independent variables: Lindeberg-Feller theorem

- Corrections to CLT – Edgeworth expansion:

Weak convergence theorems

- Central Limit Theorem: if X_i i.i.d. with mean μ, variance σ^2, then

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \rightarrow N(\mu, \sigma^2/n)$$

  Generalization for non-identical, independent variables: Lindeberg-Feller theorem

- Corrections to CLT – Edgeworth expansion: if $F_n(x)$ is the c.d.f. of $\sqrt{n}(\overline{X}_n - \mu)/\sigma$, and β_1 is the skewness coefficient, then

Weak convergence theorems

- Central Limit Theorem: if X_i i.i.d. with mean μ, variance σ^2, then

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \to N(\mu, \sigma^2/n)$$

  Generalization for non-identical, independent variables: Lindeberg-Feller theorem

- Corrections to CLT – Edgeworth expansion: if $F_n(x)$ is the c.d.f. of $\sqrt{n}(\overline{X}_n - \mu)/\sigma$, and β_1 is the skewness coefficient, then

$$F_n(x) = \mathrm{erf}(x) - \frac{\beta_1(x^2-1)}{6\sqrt{2n\pi}} e^{-x^2/2} + O(1/n)$$

Understanding the limits of CLT

- Deviations of order $O(\sqrt{n})$ are 'normal':

Understanding the limits of CLT

- Deviations of order $O(\sqrt{n})$ are 'normal':

$$P(n\overline{X}_n > n\mu + a\sqrt{n}) = P\left[\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} > a\right] \simeq \mathrm{erf}(a)$$

Understanding the limits of CLT

- Deviations of order $O(\sqrt{n})$ are 'normal':

$$P(n\overline{X}_n > n\mu + a\sqrt{n}) = P\left[\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} > a\right] \simeq \mathrm{erf}(a)$$

- Deviations of order $O(n)$ are 'large'

Understanding the limits of CLT

- Deviations of order $O(\sqrt{n})$ are 'normal':

$$P(n\overline{X}_n > n\mu + a\sqrt{n}) = P\left[\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} > a\right] \simeq \mathrm{erf}(a)$$

- Deviations of order $O(n)$ are 'large'

$$P(n\overline{X}_n > n\mu + na) = ?$$

- Everything in between – use asymptotics from Edgeworth expansions.

Understanding the limits of large sample theory

- A mice experiment ...

Understanding the limits of large sample theory

- A mice experiment ...

“ $33\frac{1}{3}$ % of the mice used in the experiment were cured by the test drug;

Understanding the limits of large sample theory

- A mice experiment ...

" $33\frac{1}{3}$ % of the mice used in the experiment were cured by the test drug; $33\frac{1}{3}$ % of the test population were unaffected by the drug and remained in a moribund condition;

Understanding the limits of large sample theory

- A mice experiment ...

" $33\frac{1}{3}$ % of the mice used in the experiment were cured by the test drug; $33\frac{1}{3}$ % of the test population were unaffected by the drug and remained in a moribund condition; the third mouse got away. "

(from R. A. Day, *AMS News*, vol. 41, no. 7, pp 486-494.)

Understanding the limits of large sample theory

- A mice experiment ...

  " $33\frac{1}{3}$ % of the mice used in the experiment were cured by the test drug; $33\frac{1}{3}$ % of the test population were unaffected by the drug and remained in a moribund condition; the third mouse got away. "

(from R. A. Day, *AMS News*, vol. 41, no. 7, pp 486-494.)

- Large sample test: small deviations must be gaussian.

Large deviations are macroscopic (how likely is this?)

Large deviations are macroscopic (how likely is this?)

Cramér’s legacy

Cramér's legacy

- Cramér-Rao lower bound, Cramér criterion, distribution of primes . . .

Cramér's legacy

- Cramér-Rao lower bound, Cramér criterion, distribution of primes . . .
- Let X_i be i.i.d. with m.g.f. $m(t) = E(e^{tX}) < \infty$, and $a > E(X)$. Then

Cramér's legacy

- Cramér-Rao lower bound, Cramér criterion, distribution of primes . . .
- Let X_i be i.i.d. with m.g.f. $m(t) = E(e^{tX}) < \infty$, and $a > E(X)$. Then

$$\boxed{P(\overline{X}_n > a) \to e^{-nQ(a)}, \quad Q(z) = \sup_t [zt - \log m(t)].}$$

- Example: (fair) coin tossing problem: H_n is number of heads in n trials,

Cramér's legacy

- Cramér-Rao lower bound, Cramér criterion, distribution of primes . . .
- Let X_i be i.i.d. with m.g.f. $m(t) = E(e^{tX}) < \infty$, and $a > E(X)$. Then

$$\boxed{P(\overline{X}_n > a) \to e^{-nQ(a)}, \quad Q(z) = \sup_t [zt - \log m(t)].}$$

- Example: (fair) coin tossing problem: H_n is number of heads in n trials,

$$P\left[H_n \geq n\frac{1+a}{2}\right] = ?$$

Cramér's legacy

- Cramér-Rao lower bound, Cramér criterion, distribution of primes . . .
- Let X_i be i.i.d. with m.g.f. $m(t) = E(e^{tX}) < \infty$, and $a > E(X)$. Then

$$\boxed{P(\overline{X}_n > a) \to e^{-nQ(a)}, \quad Q(z) = \sup_t [zt - \log m(t)].}$$

- Example: (fair) coin tossing problem: H_n is number of heads in n trials,

$$P\left[H_n \geq n\frac{1+a}{2}\right] =?$$

Bernoulli trials

Bernoulli trials

$$P_n = \sum_{k=n(1+a)/2}^{n} \frac{n!}{k!(n-k)!2^n}$$

Bernoulli trials

$$P_n = \sum_{k=n(1+a)/2}^{n} \frac{n!}{k!(n-k)!2^n} \simeq \left[\frac{n!}{k!(n-k)!2^n}\right]_{k=\frac{1+a}{2}n}$$

Bernoulli trials

$$P_n = \sum_{k=n(1+a)/2}^{n} \frac{n!}{k!(n-k)!2^n} \simeq \left[\frac{n!}{k!(n-k)!2^n}\right]_{k=\frac{1+a}{2}n}$$

$$P(H_n \geq n(1+a)/2) \simeq e^{-n[(1+a)\log(1+a)+(1-a)\log(1-a)]/2} = e^{-nQ[(1+a)/2]}$$

Bernoulli trials

$$P_n = \sum_{k=n(1+a)/2}^{n} \frac{n!}{k!(n-k)!2^n} \simeq \left[\frac{n!}{k!(n-k)!2^n}\right]_{k=\frac{1+a}{2}n}$$

$$P(H_n \geq n(1+a)/2) \simeq e^{-n[(1+a)\log(1+a)+(1-a)\log(1-a)]/2} = e^{-nQ[(1+a)/2]}$$

$$Q(z) = z\log z + (1-z)\log(1-z) + \log 2$$

Examples, properties of Cramér functions

Examples, properties of Cramér functions

- Gaussian $X \sim N(\mu, \sigma^2)$:

Examples, properties of Cramér functions

- Gaussian $X \sim N(\mu, \sigma^2)$: $Q(z) = \frac{(z-\mu)^2}{2\sigma^2}$

Examples, properties of Cramér functions

- Gaussian $X \sim N(\mu, \sigma^2)$: $Q(z) = \frac{(z-\mu)^2}{2\sigma^2}$
- Poisson $X \sim P(\lambda)$:

Examples, properties of Cramér functions

- Gaussian $X \sim N(\mu, \sigma^2)$: $Q(z) = \frac{(z-\mu)^2}{2\sigma^2}$
- Poisson $X \sim P(\lambda)$: $Q(z) = z \log(z/\lambda) + \lambda - z$

Examples, properties of Cramér functions

- Gaussian $X \sim N(\mu, \sigma^2)$: $Q(z) = \frac{(z-\mu)^2}{2\sigma^2}$
- Poisson $X \sim P(\lambda)$: $Q(z) = z \log(z/\lambda) + \lambda - z$
- Any Cramér function has a minimum at $z = \mu$: $Q(z) \geq Q(\mu) = 0$.

Examples, properties of Cramér functions

- Gaussian $X \sim N(\mu, \sigma^2)$: $Q(z) = \frac{(z-\mu)^2}{2\sigma^2}$
- Poisson $X \sim P(\lambda)$: $Q(z) = z\log(z/\lambda) + \lambda - z$
- Any Cramér function has a minimum at $z = \mu$: $Q(z) \geq Q(\mu) = 0$.
- For some distributions, $Q(z) - Q(-z) = \mathrm{const} \cdot z$,

Examples, properties of Cramér functions

- Gaussian $X \sim N(\mu, \sigma^2)$: $Q(z) = \frac{(z-\mu)^2}{2\sigma^2}$
- Poisson $X \sim P(\lambda)$: $Q(z) = z\log(z/\lambda) + \lambda - z$
- Any Cramér function has a minimum at $z = \mu$: $Q(z) \geq Q(\mu) = 0$.
- For some distributions, $Q(z) - Q(-z) = \mathrm{const} \cdot z$, fluctuation theorem

Examples, properties of Cramér functions

- Gaussian $X \sim N(\mu, \sigma^2)$: $Q(z) = \frac{(z-\mu)^2}{2\sigma^2}$
- Poisson $X \sim P(\lambda)$: $Q(z) = z \log(z/\lambda) + \lambda - z$
- Any Cramér function has a minimum at $z = \mu$: $Q(z) \geq Q(\mu) = 0$.
- For some distributions, $Q(z) - Q(-z) = \mathrm{const} \cdot z$, fluctuation theorem
- Cramér function may be seen as generalized thermodynamic potentials (see $\sup$ condition)

Generalizations of Cramér's theorem

Generalizations of Cramér's theorem

- Cramér theorem: large deviation principle for sample mean

Generalizations of Cramér's theorem

- Cramér theorem: large deviation principle for sample mean
- Sanov theorem: equivalent question for empirical measure (large deviation principle in the space of measures)

Generalizations of Cramér's theorem

- Cramér theorem: large deviation principle for sample mean
- Sanov theorem: equivalent question for empirical measure (large deviation principle in the space of measures)
- Varadhan lemma: a Laplace transform of Sanov theorem (large deviation theory for functions of the measure)

Generalizations of Cramér's theorem

- Cramér theorem: large deviation principle for sample mean
- Sanov theorem: equivalent question for empirical measure (large deviation principle in the space of measures)
- Varadhan lemma: a Laplace transform of Sanov theorem (large deviation theory for functions of the measure)

Sanov's theorem and Kullback-Leibner divergence

- Sample of i.i.d. $X_{1,2}, \ldots, X_n$

Sanov's theorem and Kullback-Leibner divergence

- Sample of i.i.d. $X_{1,2}, \ldots, X_n$ empirical measure $l_n = \frac{1}{n} \sum_{i=1}^{n} \delta(X_i)$

Sanov's theorem and Kullback-Leibner divergence

- Sample of i.i.d. $X_{1,2}, \ldots, X_n$ empirical measure $l_n = \frac{1}{n}\sum_{i=1}^{n} \delta(X_i)$
- Distance function between measures $d(f,g) = \int |f-g| dx$

Sanov's theorem and Kullback-Leibner divergence

- Sample of i.i.d. $X_{1,2}, \ldots, X_n$ empirical measure $l_n = \frac{1}{n}\sum_{i=1}^{n} \delta(X_i)$
- Distance function between measures $d(f,g) = \int |f-g| dx$
- Limit (equilibrium) distribution $X_n \to Y$, p.d.f. f_Y

Sanov's theorem and Kullback-Leibner divergence

- Sample of i.i.d. $X_{1,2}, \ldots, X_n$ empirical measure $l_n = \frac{1}{n}\sum_{i=1}^{n} \delta(X_i)$
- Distance function between measures $d(f,g) = \int |f-g|dx$
- Limit (equilibrium) distribution $X_n \to Y$, p.d.f. f_Y
- Sanov's theorem:

Sanov's theorem and Kullback-Leibner divergence

- Sample of i.i.d. $X_{1,2}, \dots, X_n$ empirical measure $l_n = \frac{1}{n}\sum_{i=1}^{n} \delta(X_i)$
- Distance function between measures $d(f,g) = \int |f - g| dx$
- Limit (equilibrium) distribution $X_n \to Y$, p.d.f. f_Y
- Sanov's theorem: $\boxed{P[d(l_n, f_Y) > a] \to \exp[-n \inf_g Q_Y(g)]}$

Sanov's theorem and Kullback-Leibner divergence

- Sample of i.i.d. $X_{1,2}, \ldots, X_n$ empirical measure $l_n = \frac{1}{n}\sum_{i=1}^{n} \delta(X_i)$
- Distance function between measures $d(f,g) = \int |f - g| dx$
- Limit (equilibrium) distribution $X_n \to Y$, p.d.f. f_Y
- Sanov's theorem: $\boxed{P[d(l_n, f_Y) > a] \to \exp[-n \inf_g Q_Y(g)]}$
- $Q_Y(g) = \int g(x) \log g(x)/f_Y(x) dx = K(g, f_Y)$, the Kullback-Leibner divergence, $K(g,f) \geq K(g,g) = 0$ (Shannon-Kolmogorov lemma)

Sanov's theorem and Kullback-Leibner divergence

- Sample of i.i.d. $X_{1,2}, \ldots, X_n$ empirical measure $l_n = \frac{1}{n}\sum_{i=1}^{n} \delta(X_i)$
- Distance function between measures $d(f,g) = \int |f-g|dx$
- Limit (equilibrium) distribution $X_n \to Y$, p.d.f. f_Y
- Sanov's theorem: $\boxed{P[d(l_n, f_Y) > a] \to \exp[-n \inf_g Q_Y(g)]}$
- $Q_Y(g) = \int g(x) \log g(x)/f_Y(x)dx = K(g, f_Y)$, the Kullback-Leibner divergence, $K(g,f) \geq K(g,g) = 0$ (Shannon-Kolmogorov lemma)

Varadhan's lemma and Laplace transforms

- $\Omega =$ space of deviations, $P_n(X_n > a)$ satisfies LDP with Cramér function $Q(a)$, then

Varadhan's lemma and Laplace transforms

- $\Omega =$ space of deviations, $P_n(X_n > a)$ satisfies LDP with Cramér function $Q(a)$, then

- Varadhan lemma: $\boxed{\int e^{nF(a)} dP_n(a) \to \exp\{n \sup_{\Omega}[F(a) - Q(a)]\}}$

Varadhan's lemma and Laplace transforms

- $\Omega =$ space of deviations, $P_n(X_n > a)$ satisfies LDP with Cramér function $Q(a)$, then

- Varadhan lemma: $\boxed{\int e^{nF(a)} dP_n(a) \to \exp\{n \sup_\Omega [F(a) - Q(a)]\}}$

- It's an integrated version of LDP

Varadhan's lemma and Laplace transforms

- $\Omega =$ space of deviations, $P_n(X_n > a)$ satisfies LDP with Cramér function $Q(a)$, then

- Varadhan lemma: $\boxed{\int e^{nF(a)} dP_n(a) \to \exp\{n \sup_\Omega [F(a) - Q(a)]\}}$

- It's an integrated version of LDP

- In the exponent, it effectively does a Lagrange transform of thermodynamic potentials:

Varadhan's lemma and Laplace transforms

- $\Omega =$ space of deviations, $P_n(X_n > a)$ satisfies LDP with Cramér function $Q(a)$, then
- Varadhan lemma: $\boxed{\int e^{nF(a)} dP_n(a) \to \exp\{n \sup_\Omega [F(a) - Q(a)]\}}$
- It's an integrated version of LDP
- In the exponent, it effectively does a Lagrange transform of thermodynamic potentials: $Q \to Q - F$

Large deviations in steady-states

- System in steady-state, change external, intensive parameter fast

Large deviations in steady-states

- System in steady-state, change external, intensive parameter fast
- Observe deviations in conjugate (extensive) quantity Σ:

Large deviations in steady-states

- System in steady-state, change external, intensive parameter fast
- Observe deviations in conjugate (extensive) quantity Σ: $\frac{P(\Sigma)}{P(-\Sigma)} = e^{\mathrm{const}\cdot\Sigma}$

Large deviations in steady-states

- System in steady-state, change external, intensive parameter fast
- Observe deviations in conjugate (extensive) quantity Σ: $\frac{P(\Sigma)}{P(-\Sigma)} = e^{\mathrm{const}\cdot\Sigma}$
- Follows from LDP $P(\Sigma) \simeq e^{Q(\Sigma)}$ and fluctuation theorem $Q(\Sigma) - Q(-\Sigma) = \mathrm{const} \cdot \Sigma$

Large deviations in steady-states

- System in steady-state, change external, intensive parameter fast
- Observe deviations in conjugate (extensive) quantity Σ: $\frac{P(\Sigma)}{P(-\Sigma)} = e^{\mathrm{const}\cdot\Sigma}$
- Follows from LDP $P(\Sigma) \simeq e^{Q(\Sigma)}$ and fluctuation theorem $Q(\Sigma) - Q(-\Sigma) = \mathrm{const} \cdot \Sigma$
- Derived for Langevin processes in V. Y. Chernyak, M. Chertkov, C. Jarzynski, cond-mat/0605471

Large deviations in steady-states

- System in steady-state, change external, intensive parameter fast
- Observe deviations in conjugate (extensive) quantity Σ: $\frac{P(\Sigma)}{P(-\Sigma)} = e^{\mathrm{const}\cdot\Sigma}$
- Follows from LDP $P(\Sigma) \simeq e^{Q(\Sigma)}$ and fluctuation theorem $Q(\Sigma) - Q(-\Sigma) = \mathrm{const}\cdot\Sigma$
- Derived for Langevin processes in V. Y. Chernyak, M. Chertkov, C. Jarzynski, cond-mat/0605471

Integrated large deviations

- Average the large deviation result $\langle e^{-\mathrm{const}\cdot\Sigma}\rangle$ by a Laplace-transform (Varadhan) formula:

$$\langle e^{-\mathrm{const}\cdot\Sigma}\rangle = \int_{\Sigma(-\lambda)}^{\Sigma(\lambda)} \rho(\Sigma) e^{-\mathrm{const}\cdot\Sigma} d\Sigma \sim \exp\left[\log\frac{\rho_+}{\rho_-} + Q(\Sigma_-) - Q(\Sigma_+)\right]$$

- For most systems, it becomes thermodynamic identity (free energy)

Integrated large deviations

- Average the large deviation result $\langle e^{-\mathrm{const}\cdot\Sigma}\rangle$ by a Laplace-transform (Varadhan) formula:

$$\langle e^{-\mathrm{const}\cdot\Sigma}\rangle = \int_{\Sigma(-\lambda)}^{\Sigma(\lambda)} \rho(\Sigma) e^{-\mathrm{const}\cdot\Sigma} d\Sigma \sim \exp\left[\log\frac{\rho_+}{\rho_-} + Q(\Sigma_-) - Q(\Sigma_+)\right]$$

- For most systems, it becomes thermodynamic identity (free energy)

$$\langle e^{-\Sigma}\rangle = e^{F(-\lambda)-F(\lambda)}.$$

- Derived for Langevin processes in V. Y. Chernyak et. al., op. cit.