Large Deviations

S.R.S. Varadhan
Courant Institute, NYU

Fields Institute
Toronto
April 13, 2011

- What is "Large Deviations" ?

- What is "Large Deviations" ?
- Stochastic model

- What is "Large Deviations" ?
- Stochastic model
- Mathematical Model with noise or randomness in it.

- What is "Large Deviations" ?
- Stochastic model
- Mathematical Model with noise or randomness in it.
- Questions do not have "yes" or "no" answers.

- What is "Large Deviations" ?
- Stochastic model
- Mathematical Model with noise or randomness in it.
- Questions do not have "yes" or "no" answers.
- Instead we have probabilities for "yes" and "no".

- Small probability is almost "no"

- Small probability is almost "no"
- The mathematical model tells us what the probability is.

- Small probability is almost "no"
- The mathematical model tells us what the probability is.
- Hard or nearly impossible to compute if the "size" is large.

- Small probability is almost "no"
- The mathematical model tells us what the probability is.
- Hard or nearly impossible to compute if the "size" is large.
- Asymptotic methods are needed.

- Laws of large numbers.

- Laws of large numbers.
- Certain questions have either very high or very low probabilities. Near certainty.

- Laws of large numbers.
- Certain questions have either very high or very low probabilities. Near certainty.
- Fluctuation theory. Small deviations are always possible, there are nontrivial probabilities associated with them.

- Laws of large numbers.
- Certain questions have either very high or very low probabilities. Near certainty.
- Fluctuation theory. Small deviations are always possible, there are nontrivial probabilities associated with them.
- Significant or "large" deviations are less likely and have only small probabilities of happening.

- Laws of large numbers.
- Certain questions have either very high or very low probabilities. Near certainty.
- Fluctuation theory. Small deviations are always possible, there are nontrivial probabilities associated with them.
- Significant or "large" deviations are less likely and have only small probabilities of happening.
- How small?

- Often exponentially small in the "size"

- Often exponentially small in the "size"
- The exponential constant is computable in many cases.

- Often exponentially small in the "size"
- The exponential constant is computable in many cases.
- There is a universal way of computing it and that is the main thrust of "Large Deviations" theory.

- Often exponentially small in the "size"
- The exponential constant is computable in many cases.
- There is a universal way of computing it and that is the main thrust of "Large Deviations" theory.
- Let us look at some examples.

- The simplest one is "flipping coins"

- The simplest one is "flipping coins"
- Toss a coin n times.

- The simplest one is "flipping coins"
- Toss a coin n times.
- Let ξ_n be the proportion of heads.

- The simplest one is "flipping coins"
- Toss a coin n times.
- Let ξ_n be the proportion of heads.
- If n is large $P[\xi_n \simeq \frac{1}{2}]$ is nearly 1.

- The simplest one is "flipping coins"
- Toss a coin n times.
- Let ξ_n be the proportion of heads.
- If n is large $P[\xi_n \simeq \frac{1}{2}]$ is nearly 1.
- $P[\xi_n \simeq p] \to 0$ if $p \neq \frac{1}{2}$

- The simplest one is "flipping coins"
- Toss a coin n times.
- Let ξ_n be the proportion of heads.
- If n is large $P[\xi_n \simeq \frac{1}{2}]$ is nearly 1.
- $P[\xi_n \simeq p] \longrightarrow 0$ if $p \neq \frac{1}{2}$
- At what rate?

- Can compute $P[\xi \simeq p] \simeq \frac{\binom{n}{np}}{2^n} = e^{-nH(p)+o(n)}$

- Can compute $P[\xi \simeq p] \simeq \frac{\binom{n}{np}}{2^n} = e^{-nH(p)+o(n)}$
- Stirling's formula. $n! = \exp[n \log n - n + o(n)]$

- Can compute $P[\xi \simeq p] \simeq \frac{\binom{n}{np}}{2^n} = e^{-nH(p)+o(n)}$
- Stirling's formula. $n! = \exp[n \log n - n + o(n)]$
- $H(p) = p \log(2p) + (1-p) \log(2(1-p))$

- Can compute $P[\xi \simeq p] \simeq \frac{\binom{n}{np}}{2^n} = e^{-nH(p)+o(n)}$
- Stirling's formula. $n! = \exp[n \log n - n + o(n)]$
- $H(p) = p \log(2p) + (1-p) \log(2(1-p))$
- $H(p) \geq 0$. $H(p) = 0 \Leftrightarrow p = \frac{1}{2}$

- Can compute $P[\xi \simeq p] \simeq \frac{\binom{n}{np}}{2^n} = e^{-nH(p)+o(n)}$
- Stirling's formula. $n! = \exp[n \log n - n + o(n)]$
- $H(p) = p \log(2p) + (1-p)\log(2(1-p))$
- $H(p) \geq 0$. $H(p) = 0 \Leftrightarrow p = \frac{1}{2}$
- Is there a recipe for computing $H(p)$?

Entropy is King

- What is the secret recipe?

Entropy is King

- What is the secret recipe?
- Relative Entropy.

Entropy is King

- What is the secret recipe?
- Relative Entropy.
- Model introduces a probability measure μ on some space.

Entropy is King

- What is the secret recipe?
- Relative Entropy.
- Model introduces a probability measure μ on some space.
- The event we are looking for has small probability.

Entropy is King

- What is the secret recipe?
- Relative Entropy.
- Model introduces a probability measure μ on some space.
- The event we are looking for has small probability.
- Change the model

- In such a way that the probability is now nearly one as the size $n \to \infty$.

- In such a way that the probability is now nearly one as the size $n \to \infty$.
- Changes the measure $\mu_n \to \lambda_n$

- In such a way that the probability is now nearly one as the size $n \to \infty$.
- Changes the measure $\mu_n \to \lambda_n$
- $$H_n = \int \log \frac{d\lambda_n}{d\mu_n} d\lambda_n$$

- In such a way that the probability is now nearly one as the size $n \to \infty$.
- Changes the measure $\mu_n \to \lambda_n$
- $$H_n = \int \log \frac{d\lambda_n}{d\mu_n} d\lambda_n$$
- H_n is additive and grows linearly in n, i.e. $H_n \simeq nH$.

- In such a way that the probability is now nearly one as the size $n \to \infty$.
- Changes the measure $\mu_n \to \lambda_n$
- $$H_n = \int \log \frac{d\lambda_n}{d\mu_n} d\lambda_n$$
- H_n is additive and grows linearly in n, i.e. $H_n \simeq nH$.
- $P \simeq e^{-nH+o(n)}$

$$P(A) = \int_A e^{-\log \frac{dQ}{dP}} dQ$$

- $$P(A) = \int_A e^{-\log \frac{dQ}{dP}} dQ$$

- If $Q(A) \simeq 1$,

$$P(A) \geq Q(A) \exp[-\frac{1}{Q(A)} \int_A \frac{dQ}{dP} dQ] \simeq e^{-H(Q,P)}$$

- $$P(A) = \int_A e^{-\log \frac{dQ}{dP}} dQ$$
- If $Q(A) \simeq 1$,
$$P(A) \geq Q(A) \exp[-\frac{1}{Q(A)} \int_A \frac{dQ}{dP} dQ] \simeq e^{-H(Q,P)}$$
- For coin tossing, let the coin be biased with $P[\text{head}] = p$

- $$P(A) = \int_A e^{-\log \frac{dQ}{dP}} dQ$$
- If $Q(A) \simeq 1$,

$$P(A) \geq Q(A) \exp[-\frac{1}{Q(A)} \int_A \frac{dQ}{dP} dQ] \simeq e^{-H(Q,P)}$$

- For coin tossing, let the coin be biased with $P[\text{head}] = p$
- $\mu(k) = \binom{n}{k} 2^{-k}$

- $\lambda(k) = \binom{n}{k} p^k (1-p)^{n-k}$

- $\lambda(k) = \binom{n}{k} p^k (1-p)^{n-k}$
- $\sum_k \log \frac{\lambda(k)}{\mu(k)} \lambda(k) = nH(p)$

- $\lambda(k) = \binom{n}{k} p^k (1-p)^{n-k}$
- $\sum_k \log \frac{\lambda(k)}{\mu(k)} \lambda(k) = nH(p)$
- More generally if we add i.i.d.r.v's $S_n = \sum_{i=1}^{n} X_i$,

- $\lambda(k) = \binom{n}{k} p^k (1-p)^{n-k}$
- $\sum_k \log \frac{\lambda(k)}{\mu(k)} \lambda(k) = nH(p)$
- More generally if we add i.i.d.r.v's $S_n = \sum_{i=1}^n X_i$,
- Common distribution is α. $E_\alpha[X_i] = m$

- $\lambda(k) = \binom{n}{k} p^k (1-p)^{n-k}$
- $\sum_k \log \frac{\lambda(k)}{\mu(k)} \lambda(k) = nH(p)$
- More generally if we add i.i.d.r.v's $S_n = \sum_{i=1}^{n} X_i$,
- Common distribution is α. $E_\alpha[X_i] = m$
- $P[\frac{S_n}{n} = x] \simeq e^{-nI(x)}$

- $\lambda(k) = \binom{n}{k} p^k (1-p)^{n-k}$
- $\sum_k \log \frac{\lambda(k)}{\mu(k)} \lambda(k) = nH(p)$
- More generally if we add i.i.d.r.v's $S_n = \sum_{i=1}^n X_i$,
- Common distribution is α. $E_\alpha[X_i] = m$
- $P[\frac{S_n}{n} = x] \simeq e^{-nI(x)}$
- Change α to β so that $E_\beta[X] = x$

- $\lambda(k) = \binom{n}{k} p^k (1-p)^{n-k}$
- $\sum_k \log \frac{\lambda(k)}{\mu(k)} \lambda(k) = nH(p)$
- More generally if we add i.i.d.r.v's $S_n = \sum_{i=1}^n X_i$,
- Common distribution is α. $E_\alpha[X_i] = m$
- $P[\frac{S_n}{n} = x] \simeq e^{-nI(x)}$
- Change α to β so that $E_\beta[X] = x$
- $H(\beta) = \int \log \frac{d\beta}{d\alpha} d\beta$

- β is not unique.

- β is not unique.
- $I(x) = \inf_{\beta : E_\beta[X]=x} H(\beta)$

- β is not unique.
- $I(x) = \inf_{\beta : E_\beta[X]=x} H(\beta)$
- Change the model, but do it frugally!

General Formulation

- X is a Polish space.

General Formulation

- X is a Polish space.
- $P_n \to \delta_x$

General Formulation

- X is a Polish space.
- $P_n \longrightarrow \delta_x$
- $P_n(A) = \exp[-c(A)n + o(n)]$

General Formulation

- X is a Polish space.
- $P_n \longrightarrow \delta_x$
- $P_n(A) = \exp[-c(A)n + o(n)]$
- $c(A \cup B) = \inf\{c(A), c(B)\}$

General Formulation

- X is a Polish space.
- $P_n \longrightarrow \delta_x$
- $P_n(A) = \exp[-c(A)n + o(n)]$
- $c(A \cup B) = \inf\{c(A), c(B)\}$
- $c(A) = \inf_{x \in A} I(x)$

General Formulation

- For closed sets C,

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(C) \leq -\inf_{x\in C} I(x)$$

General Formulation

- For closed sets C,

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(C) \leq -\inf_{x\in C} I(x)$$

- For open sets G,

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(G) \geq -\inf_{x\in G} I(x)$$

General Formulation

- For closed sets C,

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(C) \le -\inf_{x\in C} I(x)$$

- For open sets G,

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(G) \ge -\inf_{x\in G} I(x)$$

- If $\inf_{x\in A^o} I(x) = \inf_{x\in \bar{A}} I(x)$

$$\lim_{n\to\infty} \frac{1}{n} \log P_n(A) = -\inf_{x\in A} I(x)$$

Some simple facts

- $F : X \to Y$

Some simple facts

- $F : X \to Y$
- $Q_n = P_n F^{-1}$

Some simple facts

- $F : X \to Y$
- $Q_n = P_n F^{-1}$
- $$J(y) = \inf_{x \in F^{-1}(y)} I(x)$$

Some simple facts

- $F : X \to Y$
- $Q_n = P_n F^{-1}$
- $$J(y) = \inf_{x \in F^{-1}(y)} I(x)$$
- $$\int \exp[nF(x)]dP_n = \exp[n \sup_x [F(x) - I(x)] + o(n)]$$

Ventcell-Freidlin Theory

- $$dx(t) = b(x(t))dt, x(0) = x$$

Ventcell-Freidlin Theory

- $$dx(t) = b(x(t))dt, x(0) = x$$
- Perturb by Brownian Motion

Ventcell-Freidlin Theory

- $$dx(t) = b(x(t))dt, x(0) = x$$
- Perturb by Brownian Motion
- $$dx(t) = b(x(t))dt + \sqrt{\epsilon}d\beta(t)$$

Ventcell-Freidlin Theory

- $$dx(t) = b(x(t))dt, x(0) = x$$
- Perturb by Brownian Motion
- $$dx(t) = b(x(t))dt + \sqrt{\epsilon} d\beta(t)$$
- $$P_\epsilon \longrightarrow \delta_{x(\cdot)}$$

Ventcell-Freidlin Theory

- $$dx(t) = b(x(t))dt, x(0) = x$$
- Perturb by Brownian Motion
- $$dx(t) = b(x(t))dt + \sqrt{\epsilon} d\beta(t)$$
- $$P_\epsilon \rightarrow \delta_{x(\cdot)}$$
- Exponential

$$\exp[-\frac{1}{\epsilon} I(f(\cdot))]$$

- What is $I(f)$?

- What is $I(f)$?
- Perturb the ODE to

$$dx(t) = f'(t)dt + \sqrt{\epsilon}d\beta(t)$$

- What is $I(f)$?
- Perturb the ODE to

$$dx(t) = f'(t)dt + \sqrt{\epsilon}d\beta(t)$$

- $$Q_\epsilon \to \delta_{f(\cdot)}$$

- What is $I(f)$?
- Perturb the ODE to

$$dx(t) = f'(t)dt + \sqrt{\epsilon}d\beta(t)$$

-

$$Q_\epsilon \longrightarrow \delta_{f(\cdot)}$$

-

$$\frac{dQ_\epsilon}{dP_\epsilon} = \exp[\frac{1}{\epsilon}\int_0^T (f'(t) - b(x(t))(dx(t) - b(x(t)dt) \\ - \frac{1}{2\epsilon}\int_0^T \|f'(t) - b(x(t))\|^2 dt]$$

Exit problem

- $$I(f(\cdot)) = \frac{1}{2}\int_0^T \|f'(t) - b(x(t)\|^2 dt$$

Exit problem

- $$I(f(\cdot)) = \frac{1}{2}\int_0^T \|f'(t) - b(x(t)\|^2 dt$$
- $$V(x) = \inf_{\substack{T, f(0)=x_0 \\ f(T)=x\in\delta G}} I_T(f)$$

Exit problem

- $$I(f(\cdot)) = \frac{1}{2}\int_0^T \|f'(t) - b(x(t)\|^2 dt$$

- $$V(x) = \inf_{\substack{T, f(0)=x_0 \\ f(T)=x\in\delta G}} I_T(f)$$

- $$\inf_{x\in\delta G} V(x)$$

Feynamn-Kac and eigen-values

- $\pi(x, y)$ is a Markov Chain.

Feynamn-Kac and eigen-values

- $\pi(x, y)$ is a Markov Chain.
- Consider

$$\pi(x, y)e^{V(y)}$$

Feynamn-Kac and eigen-values

- $\pi(x, y)$ is a Markov Chain.
- Consider

$$\pi(x, y)e^{V(y)}$$

- Spectral radius is $\rho(V)$ and $\lambda(V) = \log \rho(V)$

Feynamn-Kac and eigen-values

- $\pi(x, y)$ is a Markov Chain.
- Consider

$$\pi(x, y)e^{V(y)}$$

- Spectral radius is $\rho(V)$ and $\lambda(V) = \log \rho(V)$
- $$\lambda(V) = \lim_{n\to\infty} \frac{1}{n} \log E_x[\exp[V(X_1) + \cdots + V(X_n)]]$$

- $Y = \mathcal{M}(X)$

- $Y = \mathcal{M}(X)$
- $(X_1, X_2, \ldots, X_n) \rightarrow \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} = \nu_n(\omega)$

- $Y = \mathcal{M}(X)$
- $(X_1, X_2, \ldots, X_n) \to \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} = \nu_n(\omega)$
- $\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \to \lambda$

- $Y = \mathcal{M}(X)$
- $(X_1, X_2, \ldots, X_n) \to \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} = \nu_n(\omega)$
- $\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \to \lambda$
- $P[\nu_n(\omega) \simeq \mu] = \exp[-nI(\mu) + o(n)]$

- $Y = \mathcal{M}(X)$
- $(X_1, X_2, \ldots, X_n) \to \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} = \nu_n(\omega)$
- $\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \to \lambda$
- $P[\nu_n(\omega) \simeq \mu] = \exp[-nI(\mu) + o(n)]$
- $\lambda(V) = \sup_\mu[\int V(x)\mu(dx) - I(\mu)]$

- $Y = \mathcal{M}(X)$
- $(X_1, X_2, \ldots, X_n) \to \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} = \nu_n(\omega)$
- $\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \to \lambda$
- $P[\nu_n(\omega) \simeq \mu] = \exp[-nI(\mu) + o(n)]$
- $\lambda(V) = \sup_\mu [\int V(x)\mu(dx) - I(\mu)]$
- $I(\mu) = \sup_V [\int V(x)\mu(dx) - \lambda(V)]$

- Change $\pi(x, y) \longrightarrow \pi'(x, y)$

- Change $\pi(x, y) \longrightarrow \pi'(x, y)$
- $\mu\pi' = \mu$

- Change $\pi(x, y) \longrightarrow \pi'(x, y)$
- $\mu\pi' = \mu$
- Now $\nu_n \longrightarrow \mu$

- Change $\pi(x,y) \longrightarrow \pi'(x,y)$
- $\mu\pi' = \mu$
- Now $\nu_n \longrightarrow \mu$
- $$\frac{1}{n}\log\frac{dQ_n}{dP_n} = \frac{1}{n}\sum_i \log\frac{\pi'(X_i, X_{i+1})}{\pi(X_i, X_{i+1})}$$

- Change $\pi(x,y) \longrightarrow \pi'(x,y)$
- $\mu\pi' = \mu$
- Now $\nu_n \longrightarrow \mu$
- $$\frac{1}{n}\log\frac{dQ_n}{dP_n} = \frac{1}{n}\sum_i \log\frac{\pi'(X_i, X_{i+1})}{\pi(X_i, X_{i+1})}$$
- $$H(\pi') = \sum_{x,y} \mu(x)\pi'(x,y)\log\frac{\pi'(x,y)}{\pi(x,y)}$$

- Change $\pi(x,y) \longrightarrow \pi'(x,y)$
- $\mu\pi' = \mu$
- Now $\nu_n \longrightarrow \mu$
- $$\frac{1}{n}\log\frac{dQ_n}{dP_n} = \frac{1}{n}\sum_i \log\frac{\pi'(X_i, X_{i+1})}{\pi(X_i, X_{i+1})}$$
- $$H(\pi') = \sum_{x,y}\mu(x)\pi'(x,y)\log\frac{\pi'(x,y)}{\pi(x,y)}$$
- $$I(\mu) = \inf_{\pi':\mu\pi'=\mu} H(\pi')$$

Continuous time

- Generator $\mathcal{A}$

Continuous time

- Generator $\mathcal{A}$
- $$\lambda(V) = \lim_{t\to\infty} \frac{1}{t} \log E_x[\exp[\int_0^t V(x(s))ds]]$$

Continuous time

- Generator $\mathcal{A}$
- $$\lambda(V) = \lim_{t\to\infty} \frac{1}{t} \log E_x[\exp[\int_0^t V(x(s))ds]]$$
- $$\lambda(V) \Leftrightarrow I(\mu)$$

Continuous time

- Generator $\mathcal{A}$
- $$\lambda(V) = \lim_{t\to\infty} \frac{1}{t} \log E_x[\exp[\int_0^t V(x(s))ds]]$$
- $$\lambda(V) \Leftrightarrow I(\mu)$$
- $$I(\mu) = \sup_{u>0} - \int \frac{\mathcal{A}u}{u} d\mu$$

Reversible case

- $\mathcal{A}$ is self adjoint with respect to α.

Reversible case

- $\mathcal{A}$ is self adjoint with respect to α.
- $$I(\mu) = \mathcal{D}(\sqrt{\frac{d\mu}{d\alpha}})$$

Reversible case

- $\mathcal{A}$ is self adjoint with respect to α.
- $$I(\mu) = \mathcal{D}(\sqrt{\frac{d\mu}{d\alpha}})$$
- $$\lambda(V) = \sup_{g:\|g\|_{2,\alpha}=1} \left[\int V g^2 d\alpha - \mathcal{D}(g) \right]$$

Reversible case

- Topology of X.

Reversible case

- Topology of X.
- Smoothing.

Reversible case

- Topology of X.
- Smoothing.
- $d = 1$, Brownian motion, L_1 or similar.

Future

- Scaling limits of Interacting particles

Future

- Scaling limits of Interacting particles
- Random walks in a random environment

Future

- Scaling limits of Interacting particles
- Random walks in a random environment
- Homogenization of Hamilton Jacobi equations

Future

- Scaling limits of Interacting particles
- Random walks in a random environment
- Homogenization of Hamilton Jacobi equations
- Random graphs, Random matrices

THE END