Entropy and Large Deviations

S.R.S. Varadhan
Courant Institute, NYU

Michigan State Universiy
East Lansing
March 31 , 2015

- Entropy comes up in many different contexts.

- Entropy comes up in many different contexts.
- in Physics, introduced by Rudolf Clausius in 1865

- Entropy comes up in many different contexts.
- in Physics, introduced by Rudolf Clausius in 1865
- In connection with heat transfer, relation between heat and work. Classical thermodynamics.

- Entropy comes up in many different contexts.
- in Physics, introduced by Rudolf Clausius in 1865
- In connection with heat transfer, relation between heat and work. Classical thermodynamics.
- Bulk Quantity

- Entropy comes up in many different contexts.
- in Physics, introduced by Rudolf Clausius in 1865
- In connection with heat transfer, relation between heat and work. Classical thermodynamics.
- Bulk Quantity
- Boltzmann around 1877 defined entropy as $c \log |\Omega|$, Ω is the set of micro states that correspond to a given macro state. $|\Omega|$ is its size.

- Shannon's Entropy. !948. Mathematical theory of communication. Modern information theory.

- Shannon's Entropy. !948. Mathematical theory of communication. Modern information theory.
- Given $\mathbf{p} = (p_1, \ldots, p_k)$,

$$H(\mathbf{p}) = -\sum_i p_i \log p_i$$

- Shannon's Entropy. !948. Mathematical theory of communication. Modern information theory.
- Given $\mathbf{p} = (p_1, \ldots, p_k)$,

$$H(\mathbf{p}) = -\sum_i p_i \log p_i$$

- Conditional Probability $P[X = i|\Sigma] = p_i(\omega)$, $E[p_i(\omega)] = p_i$

- Conditional Entropy

$$H(\omega) = -\sum_i p_i(\omega) \log p_i(\omega)$$

$$E[H(\omega)] = E[-\sum_i p_i(\omega) \log p_i(\omega)]$$

- Concavity

$$H = -\sum_i p_i \log p_i \geq E[-\sum_i p_i(\omega) \log p_i(\omega)]$$

- Concavity

$$H = -\sum_i p_i \log p_i \geq E[-\sum_i p_i(\omega) \log p_i(\omega)]$$

- X_i is a stationary stochastic process.
  $p_i(\omega) = P[X_1 = i | X_0, \dots X_{-n}, \dots]$.

$$H[P] = E^P\left[-\sum_i p_i(\omega) \log p_i(\omega)\right]$$

- Concavity

$$H = -\sum_i p_i \log p_i \geq E[-\sum_i p_i(\omega) \log p_i(\omega)]$$

- X_i is a stationary stochastic process.
  $p_i(\omega) = P[X_1 = i | X_0, \ldots X_{-n}, \ldots]$.

$$H[P] = E^P \left[-\sum_i p_i(\omega) \log p_i(\omega) \right]$$

- Conditioning is with respect to past history.

- Look at $p(x_1, \ldots, x_n)$

- Look at $p(x_1, \ldots, x_n)$
- $H_n(P) = -\sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n)$

- Look at $p(x_1, \ldots, x_n)$
- $H_n(P) = - \sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n)$
- $H_{n+m} \leq H_n + H_m$, H_n is $\uparrow$.

- Look at $p(x_1, \ldots, x_n)$
- $H_n(P) = -\sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n)$
- $H_{n+m} \leq H_n + H_m$, H_n is $\uparrow$.
- $H_{n+1} - H_n$ is $\downarrow$

- Look at $p(x_1, \dots, x_n)$
- $H_n(P) = -\sum_{x_1,\dots,x_n} p(x_1, \dots, x_n) \log p(x_1, \dots, x_n)$
- $H_{n+m} \le H_n + H_m$, H_n is $\uparrow$.
- $H_{n+1} - H_n$ is $\downarrow$
- $H(P) = \lim \frac{H_n}{n} = \lim [H_{n+1} - H_n]$

- Dynamical system. (X, Σ, T, P). $T : X \to X$ preserves P.

- Dynamical system. (X, Σ, T, P). $T : X \to X$ preserves P.
- Shift T in a space of sequences. P any stationary process.

- Dynamical system. (X, Σ, T, P). $T : X \to X$ preserves P.
- Shift T in a space of sequences. P any stationary process.
- Isomorphism between (X, Σ, T, P) and (X', Σ', T', P')

$$S : X \to X'; ST = T'S, PS^{-1} = P'$$

- Dynamical system. (X, Σ, T, P). $T : X \to X$ preserves P.
- Shift T in a space of sequences. P any stationary process.
- Isomorphism between (X, Σ, T, P) and (X', Σ', T', P')

$$S : X \to X'; ST = T'S, PS^{-1} = P'$$

- Spectral Invariant. $(Uf)(x) = f(Ux)$ Unitary map in $L_2(X, \Sigma, P)$

- Dynamical system. (X, Σ, T, P). $T : X \to X$ preserves P.
- Shift T in a space of sequences. P any stationary process.
- Isomorphism between (X, Σ, T, P) and (X', Σ', T', P')

$$S : X \to X'; ST = T'S, PS^{-1} = P'$$

- Spectral Invariant. $(Uf)(x) = f(Ux)$ Unitary map in $L_2(X, \Sigma, P)$
- $VU = U'V$

- Entropy of a dynamical system.

$$(\Omega, \Sigma, P, T)$$

- Entropy of a dynamical system.

$$(\Omega, \Sigma, P, T)$$

- Each function f taking a finite set of values generates a stochastic process P_f, the distribution of $\{f(T^n x)\}$

- Entropy of a dynamical system.

$$(\Omega, \Sigma, P, T)$$

- Each function f taking a finite set of values generates a stochastic process P_f, the distribution of $\{f(T^n x)\}$
- $h(\Omega, \Sigma, P, T) = \sup_f h(P_f)$

- Entropy of a dynamical system.

$$(\Omega, \Sigma, P, T)$$

- Each function f taking a finite set of values generates a stochastic process P_f, the distribution of $\{f(T^n x)\}$
- $h(\Omega, \Sigma, P, T) = \sup_f h(P_f)$
- Invariant. Not spectral. $H(T^2, P) = 2H(T, P)$.

- Entropy of a dynamical system.

$$(\Omega, \Sigma, P, T)$$

- Each function f taking a finite set of values generates a stochastic process P_f, the distribution of $\{f(T^n x)\}$
- $h(\Omega, \Sigma, P, T) = \sup_f h(P_f)$
- Invariant. Not spectral. $H(T^2, P) = 2H(T, P)$.
- Computable. $P = \Pi \mathbf{p}$, $H(P) = h(\mathbf{p})$.

- Isomorphism Theorem of Ornstein

- Isomorphism Theorem of Ornstein
- If $h(\mathbf{p}) = h(\mathbf{q})$

- Isomorphism Theorem of Ornstein
- If $h(\mathbf{p}) = h(\mathbf{q})$
- The dynamical systems with product measures P, Q, i.e (F^∞, P) and (G^∞, Q) are isomorphic.

- Shannon's entropy and coding

- Shannon's entropy and coding
- Code in such a way that the best compression is achieved.

- Shannon's entropy and coding
- Code in such a way that the best compression is achieved.
- Incoming data stream has stationary statistics P.

- Shannon's entropy and coding
- Code in such a way that the best compression is achieved.
- Incoming data stream has stationary statistics P.
- Coding to be done into words in an alphabet of size r.

- Shannon's entropy and coding
- Code in such a way that the best compression is achieved.
- Incoming data stream has stationary statistics P.
- Coding to be done into words in an alphabet of size r.
- The compression factor is $c = \frac{H(P)}{\log r}$;

- The number of n tuples is k^n.

- The number of n tuples is k^n.
- Shannon-Breiman-McMillan theorem: If P is stationary and ergodic, $P(E_n) \to 1$ where

$$E_n = \{|\frac{1}{n} \log p(x_1, \ldots, x_n) - H(P)| \leq \epsilon\}$$

- The number of n tuples is k^n.
- Shannon-Breiman-McMillan theorem: If P is stationary and ergodic, $P(E_n) \to 1$ where

$$E_n = \{|\frac{1}{n}\log p(x_1, \ldots, x_n) - H(P)| \le \epsilon\}$$

- Almost the entire probability under P is carried by nearly $e^{nH(\mathbf{P})}$ n tuples of more or less equal probability.

- The number of n tuples is k^n.
- Shannon-Breiman-McMillan theorem: If P is stationary and ergodic, $P(E_n) \to 1$ where

$$E_n = \{|\frac{1}{n}\log p(x_1, \ldots, x_n) - H(P)| \leq \epsilon\}$$

- Almost the entire probability under P is carried by nearly $e^{nH(\mathbf{P})}$ n tuples of more or less equal probability.
- $r^m = e^{nH(P)}$. $\quad c = \frac{m}{n} = \frac{H(\mathbf{P})}{\log r}$

- k^n sequences of length n from an alphabet of size k.

- k^n sequences of length n from an alphabet of size k.
- How many look like P?

- k^n sequences of length n from an alphabet of size k.
- How many look like P?
- $\exp[n(H(P)) + o(n)]$

- k^n sequences of length n from an alphabet of size k.
- How many look like P?
- $\exp[n(H(P)) + o(n)]$
- $H(P)$ is maximized when $P = P_0$, the uniform distribution on F^n

- k^n sequences of length n from an alphabet of size k.
- How many look like P?
- $\exp[n(H(P)) + o(n)]$
- $H(P)$ is maximized when $P = P_0$, the uniform distribution on F^n
- $H(P_0) = \log k$

- Kullback-Leibler information, Relative Entropy (1951)

$$H(\mathbf{q};\mathbf{p}) = \sum_i q_i \log \frac{q_i}{p_i}$$

- Kullback-Leibler information, Relative Entropy (1951)

$$H(\mathbf{q};\mathbf{p}) = \sum_i q_i \log \frac{q_i}{p_i}$$

- If $\mathbf{p}$ is the uniform distribution with mass $\frac{1}{k}$ at every point then

$$h(\mathbf{q}:\mathbf{p}) = \log k - h(\mathbf{q})$$

- Kullback-Leibler information, Relative Entropy (1951)

$$H(\mathbf{q};\mathbf{p}) = \sum_i q_i \log \frac{q_i}{p_i}$$

- If $\mathbf{p}$ is the uniform distribution with mass $\frac{1}{k}$ at every point then

$$h(\mathbf{q}:\mathbf{p}) = \log k - h(\mathbf{q})$$

- Two probability densities

$$H(g;f) = \int g(x) \log \frac{g(x)}{f(x)} dx$$

- Two probability measures

$$H(\mu, \lambda) = \int \log \frac{d\mu}{d\lambda} d\mu = \int \frac{d\mu}{d\lambda} \log \frac{d\mu}{d\lambda} d\lambda$$

- Two probability measures

$$H(\mu, \lambda) = \int \log \frac{d\mu}{d\lambda} d\mu = \int \frac{d\mu}{d\lambda} \log \frac{d\mu}{d\lambda} d\lambda$$

- Two stationary processes

$$H(Q, P) = E^Q[H(Q|\Sigma; P|\Sigma)]$$

- Two probability measures

$$H(\mu, \lambda) = \int \log \frac{d\mu}{d\lambda} d\mu = \int \frac{d\mu}{d\lambda} \log \frac{d\mu}{d\lambda} d\lambda$$

- Two stationary processes

$$H(Q, P) = E^Q[H(Q|\Sigma; P|\Sigma)]$$

- Has issues! OK if $P|\Sigma$ is globally defined.

- α, β are probability measures on X

- α, β are probability measures on X
- $P = \Pi\alpha$

- α, β are probability measures on X
- $P = \Pi\alpha$
- $$r_n(dx) = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$$

- α, β are probability measures on X
- $P = \Pi\alpha$
- $$r_n(dx) = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$$
- Sanov's Theorem.

$$P[r_n(dx) \simeq \beta] = \exp[-nH(\beta;\alpha) + o(n)]$$

- Large Deviations

- Large Deviations
- For closed sets C

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

- Large Deviations
- 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

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

- Large Deviations
- 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

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

- I is lower semi continuous and has compact level sets.

- Sanov’s theorem is an example.

- Sanov's theorem is an example.
- P_n on $\mathcal{M}(X)$ with weak topology is the distribution of $r_n(dx)$ under $\Pi\alpha$.

- Sanov's theorem is an example.
- P_n on $\mathcal{M}(X)$ with weak topology is the distribution of $r_n(dx)$ under $\Pi\alpha$.
- $I(\beta) = h(\beta : \alpha)$.

- P is a stationary process, We have a string of n observations.

- P is a stationary process, We have a string of n observations.
- Ergodic theorem says most sequences resemble P.

- P is a stationary process, We have a string of n observations.
- Ergodic theorem says most sequences resemble P.
- What is the probability that they resemble Q?

- P is a stationary process, We have a string of n observations.
- Ergodic theorem says most sequences resemble P.
- What is the probability that they resemble Q?
- $\exp[-nH(Q;P)]$

- P is a stationary process, We have a string of n observations.
- Ergodic theorem says most sequences resemble P.
- What is the probability that they resemble Q?
- $\exp[-nH(Q;P)]$
- If $P = P_0$ then $H(Q;P) = \log k - H(Q)$.

- P is a stationary process, We have a string of n observations.
- Ergodic theorem says most sequences resemble P.
- What is the probability that they resemble Q?
- $\exp[-nH(Q;P)]$
- If $P = P_0$ then $H(Q;P) = \log k - H(Q)$.
- OK for nice P.

- Functional Analysis

- Functional Analysis
- $f \geq 0 \int f d\lambda = 1$

- Functional Analysis
- $f \geq 0 \int f d\lambda = 1$
-

$$\frac{d}{dp}\bigg|_{p=1} \left[\int f^p d\lambda \right] = \int f \log f d\lambda$$

- There is an analog of Holder's inequality in the limit as $p \to 1$.

- There is an analog of Holder's inequality in the limit as $p \to 1$.
- The duality

- There is an analog of Holder's inequality in the limit as $p \to 1$.
- The duality
- $x, y \in R$.
  $xy \leq \frac{|x|^p}{p} + \frac{|y|^q}{q}$,
  $\frac{|x|^p}{p} = \sup_y [xy - \frac{|y|^q}{q}], \frac{|y|^q}{q} = \sup_x [xy - \frac{|x|^p}{p}]$

- There is an analog of Holder's inequality in the limit as $p \to 1$.
- The duality
- $x, y \in R$.
  $xy \leq \frac{|x|^p}{p} + \frac{|y|^q}{q}$,
  $\frac{|x|^p}{p} = \sup_y [xy - \frac{|y|^q}{q}], \frac{|y|^q}{q} = \sup_x [xy - \frac{|x|^p}{p}]$
- Becomes, for $x > 0, y \in R$
  $x \log x - x = \sup_y [xy - e^y]$
  $e^y = \sup_{x>0} [xy - (x \log x - x)]$

- Hölder inequality becomes generalized Jensen's inequality.

$$\int g d\mu = \int f g d\lambda \leq \log \int e^g d\lambda + H(\mu, \lambda)$$

- Hölder inequality becomes generalized Jensen's inequality.

$$\int g d\mu = \int f g d\lambda \le \log \int e^g d\lambda + H(\mu, \lambda)$$

- $g = c\chi_A(x)$. Optimize over $c > 0$. $c = \log \frac{1}{\lambda(A)}$

- Hölder inequality becomes generalized Jensen's inequality.

$$\int g d\mu = \int f g d\lambda \leq \log \int e^g d\lambda + H(\mu, \lambda)$$

- $g = c\chi_A(x)$. Optimize over $c > 0$. $c = \log \frac{1}{\lambda(A)}$
- $e^g = \chi_{A^c} + \frac{1}{\lambda(A)}\chi_A$; $\int e^g d\lambda \leq 2$

- Hölder inequality becomes generalized Jensen's inequality.

$$\int g d\mu = \int f g d\lambda \leq \log \int e^g d\lambda + H(\mu, \lambda)$$

- $g = c\chi_A(x)$. Optimize over $c > 0$. $c = \log \frac{1}{\lambda(A)}$
- $e^g = \chi_{A^c} + \frac{1}{\lambda(A)}\chi_A$; $\int e^g d\lambda \leq 2$
- $$\mu(A) \leq \frac{H(\mu, \lambda) + \log 2}{\log \frac{1}{\lambda(A)}}$$

- Entropy grows linearly but the Lp norms grow exponentially fast.

- Entropy grows linearly but the Lp norms grow exponentially fast.
- Useful inequality for interacting particle systems

- Statistical Mechanics. Probability distributions are often defined through an energy functional.

- Statistical Mechanics. Probability distributions are often defined through an energy functional.
- $$p_n(x_1, x_2, \ldots, x_n)$$
$$= [c(n)]^{-1} \exp[-\sum_{i=1}^{n-k+1} F(x_i, x_{i+1}, \ldots, x_{i+k-1})]$$

- $$c_n = \sum_{x_1,\ldots,x_n} \exp[-\sum_{i=1}^{n-k+1} F(x_i, x_{i+1}, \ldots, x_{i+k-1})]$$

- $$c_n = \sum_{x_1,\ldots,x_n} \exp[-\sum_{i=1}^{n-k+1} F(x_i, x_{i+1}, \ldots, x_{i+k-1})]$$
- $\exp[nH(P)]$, P typical sequences.

- $$c_n = \sum_{x_1,\ldots,x_n} \exp[-\sum_{i=1}^{n-k+1} F(x_i, x_{i+1}, \ldots, x_{i+k-1})]$$
- $\exp[nH(P)]$, P typical sequences.
- $\exp[-n \int F dP]$

- $$c_n = \sum_{x_1,\ldots,x_n} \exp[-\sum_{i=1}^{n-k+1} F(x_i, x_{i+1}, \ldots, x_{i+k-1})]$$
- $\exp[nH(P)]$, P typical sequences.
- $\exp[-n \int F dP]$
- The sum therefore is $e^{-n(\int F dP - H(P))}$

- Then by Laplace asymptotics

$$\lim_{n\to\infty} \frac{\log c_n}{n} = \sup_P [-\int F\, dP + H(P)]$$

- Then by Laplace asymptotics

$$\lim_{n\to\infty} \frac{\log c_n}{n} = \sup_P[-\int F\,dP + H(P)]$$

- If the sup is attained at a unique P then

$$P \simeq c^{-1} \exp[-\sum_{i=-\infty}^{\infty} F(x_i, x_{i+1}, \ldots, x_{i+k-1})]$$

- Then by Laplace asymptotics

$$\lim_{n\to\infty} \frac{\log c_n}{n} = \sup_P [-\int F\, dP + H(P)]$$

- If the sup is attained at a unique P then

$$P \simeq c^{-1} \exp[-\sum_{i=-\infty}^{\infty} F(x_i, x_{i+1}, \ldots, x_{i+k-1})]$$

- The averages $\frac{1}{n}\sum_{i=1}^{n-k+1} G(x_i, x_{i+1}, \ldots, x_{i+k-1})$ converge in probability under p_n to $E^P[G(x_1, \ldots, x_k)]$.

- Log Sobolev inequality.

- Log Sobolev inequality.
- T_t is a Markov semigroup, on (X, Σ)
  $T_t f \geq 0$ if $f \geq 0$ and $T_t 1 = 1$.

- Log Sobolev inequality.
- T_t is a Markov semigroup, on (X, Σ)
  $T_t f \geq 0$ if $f \geq 0$ and $T_t 1 = 1$.
- Symmetric with respect to μ

- Log Sobolev inequality.
- T_t is a Markov semigroup, on (X, Σ)
  $T_t f \geq 0$ if $f \geq 0$ and $T_t 1 = 1$.
- Symmetric with respect to μ
- $f \geq 0, \int f d\mu = 1$

$$\int f \log f d\mu \leq cD(\sqrt{f})$$

- Log Sobolev inequality.
- T_t is a Markov semigroup, on (X, Σ)
  $T_t f \geq 0$ if $f \geq 0$ and $T_t 1 = 1$.
- Symmetric with respect to μ
- $f \geq 0, \int f d\mu = 1$

$$\int f \log f d\mu \leq cD(\sqrt{f})$$

-

$$\frac{d}{dt} \int f_t \log f_t d\mu \leq -\frac{1}{2} D(\sqrt{f}) \leq -c' \int f_t \log f_t d\mu$$

- Exponential decay, spectral gap etc.

- Exponential decay, spectral gap etc.
- Infinitesimal Version.

- Exponential decay, spectral gap etc.
- Infinitesimal Version.
- $\{f(x,\theta)\}$

- Exponential decay, spectral gap etc.
- Infinitesimal Version.
- $\{f(x,\theta)\}$
- $$H(\theta,\theta_0) = \int f(x,\theta)\log\frac{f(x,\theta)}{f(x,\theta_0)}d\mu$$

- Exponential decay, spectral gap etc.
- Infinitesimal Version.
- $\{f(x,\theta)\}$
- $$H(\theta,\theta_0) = \int f(x,\theta) \log \frac{f(x,\theta)}{f(x,\theta_0)} d\mu$$
- $H(\theta,\theta_0)$ has a minimum at θ_0.

- Exponential decay, spectral gap etc.
- Infinitesimal Version.
- $\{f(x,\theta)\}$
- $$H(\theta,\theta_0) = \int f(x,\theta) \log \frac{f(x,\theta)}{f(x,\theta_0)} d\mu$$
- $H(\theta,\theta_0)$ has a minimum at θ_0.
- $$I(\theta(0)) = \frac{d^2 H}{d\theta^2}\Big|_{\theta=\theta_0}$$

- Optimizing Entropy.

- Optimizing Entropy.
- Toss a coin n times.

- Optimizing Entropy.
- Toss a coin n times.
- Had $\frac{3n}{4}$ heads.

- Optimizing Entropy.
- Toss a coin n times.
- Had $\frac{3n}{4}$ heads.
- Was it a steady lead or burst of heads?

- Optimizing Entropy.
- Toss a coin n times.
- Had $\frac{3n}{4}$ heads.
- Was it a steady lead or burst of heads?
- $P[\frac{1}{n}S(nt) \simeq f(t)]$

- $f'(t) = p(t)$

- $f'(t) = p(t)$
- $e^{-n \int_0^1 h(p(t))dt}$

- $f'(t) = p(t)$
- $e^{-n \int_0^1 h(p(t))dt}$
- Minimize $\int_0^1 h(p(t))dt$ subject to $\int_0^1 p(s)ds = \frac{3}{4}$

- $f'(t) = p(t)$
- $e^{-n\int_0^1 h(p(t))dt}$
- Minimize $\int_0^1 h(p(t))dt$ subject to $\int_0^1 p(s)ds = \frac{3}{4}$
- $h(p) = \log 2 + p \log p + (1-p)\log(1-p)$

- $f'(t) = p(t)$
- $e^{-n \int_0^1 h(p(t))dt}$
- Minimize $\int_0^1 h(p(t))dt$ subject to $\int_0^1 p(s)ds = \frac{3}{4}$
- $h(p) = \log 2 + p \log p + (1-p)\log(1-p)$
- Convexity. $p(s) = \frac{3s}{4}$.

- General Philosophy

- General Philosophy
- $P(A)$ is to be estimated.

- General Philosophy
- $P(A)$ is to be estimated.
- $P(A)$ is small.

- General Philosophy
- $P(A)$ is to be estimated.
- $P(A)$ is small.
- Find Q such that $Q(A) \simeq 1$

- General Philosophy
- $P(A)$ is to be estimated.
- $P(A)$ is small.
- Find Q such that $Q(A) \simeq 1$
- $$P(A) = Q(A)\frac{1}{Q(A)}\int_A e^{-\log\frac{dQ}{dP}}dQ$$

- $$P(A) \geq Q(A) \exp[-\frac{1}{Q(A)} \int_A \log \frac{dQ}{dP} dQ]$$

- $$P(A) \geq Q(A) \exp[-\frac{1}{Q(A)} \int_A \log \frac{dQ}{dP} dQ]$$

- Asymptotically more or less

$$P(A) \geq \exp[-h(Q : P)]$$

- $$P(A) \geq Q(A) \exp[-\frac{1}{Q(A)} \int_A \log \frac{dQ}{dP} dQ]$$
- Asymptotically more or less
  $$P(A) \geq \exp[-h(Q : P)]$$
- Optimize over Q.

- Example. Markov Chain

- Example. Markov Chain
- $\{p_{i,j}\}$. Invariant distribution p

- Example. Markov Chain
- $\{p_{i,j}\}$. Invariant distribution $\mathbf{p}$
- What is the probability that the empirical r_n is close to $\mathbf{q}$.

- Example. Markov Chain
- $\{p_{i,j}\}$. Invariant distribution $\mathbf{p}$
- What is the probability that the empirical r_n is close to $\mathbf{q}$.
- Find $\{q_{i,j}\}$ such that the invariant distribution is $\mathbf{q}$

- Example. Markov Chain
- $\{p_{i,j}\}$. Invariant distribution $\mathbf{p}$
- What is the probability that the empirical r_n is close to $\mathbf{q}$.
- Find $\{q_{i,j}\}$ such that the invariant distribution is $\mathbf{q}$
- $h(Q : P) = \sum_{i,j} q_i q_{i,j} \log \frac{q_{i,j}}{p_{i,j}}$

- Example. Markov Chain
- $\{p_{i,j}\}$. Invariant distribution $\mathbf{p}$
- What is the probability that the empirical r_n is close to $\mathbf{q}$.
- Find $\{q_{i,j}\}$ such that the invariant distribution is $\mathbf{q}$
- $h(Q : P) = \sum_{i,j} q_i q_{i,j} \log \frac{q_{i,j}}{p_{i,j}}$
- Optimize over $\{q_{i,j}\}$ with $\mathbf{q}$ as invariant distribution.

- Example. Markov Chain
- $\{p_{i,j}\}$. Invariant distribution $\mathbf{p}$
- What is the probability that the empirical r_n is close to $\mathbf{q}$.
- Find $\{q_{i,j}\}$ such that the invariant distribution is $\mathbf{q}$
- $h(Q : P) = \sum_{i,j} q_i q_{i,j} \log \frac{q_{i,j}}{p_{i,j}}$
- Optimize over $\{q_{i,j}\}$ with $\mathbf{q}$ as invariant distribution.
- Best possible lower bound is the upper bound.

- Random graphs

- Random graphs
- p is the probability of edge.

- Random graphs
- p is the probability of edge.
- Need more triangles $\binom{N}{3}\tau$, $\tau > p^3$.

- Random graphs
- p is the probability of edge.
- Need more triangles $\binom{N}{3}\tau$, $\tau > p^3$.
- Fix $\int f(x,y)f(y,z)f(z,x)dxdydz = \tau$

- Random graphs
- p is the probability of edge.
- Need more triangles $\binom{N}{3}\tau$, $\tau > p^3$.
- Fix $\int f(x,y)f(y,z)f(z,x)dxdydz = \tau$
- Minimize $H(f) =$
$\int [f(x,y)\log\frac{f(x,y)}{p} + (1 - f(x,y))\log\frac{(1-f(x,y))}{1-p}]dxdy$

- Random graphs
- p is the probability of edge.
- Need more triangles $\binom{N}{3}\tau$, $\tau > p^3$.
- Fix $\int f(x,y)f(y,z)f(z,x)dxdydz = \tau$
- Minimize $H(f) =$
  $\int [f(x,y)\log\frac{f(x,y)}{p} + (1-f(x,y))\log\frac{(1-f(x,y))}{1-p}]dxdy$
- $f = \tau^{\frac{1}{3}}$?

THANK YOU