4.1. iid sample means

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

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

$$= \lim_{n\to\infty} \frac{1}{n} \ln \underbrace{E\, e^{k\Sigma_i X_i}}$$

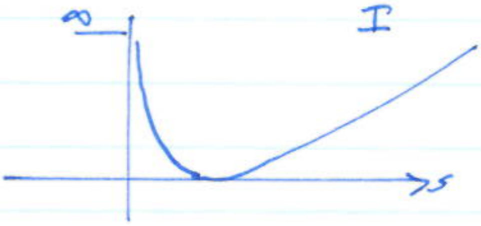
$$= E\,\Pi_i e^{kX_i}$$
$$= \Pi_i\, E e^{kX_i}$$
$$= E[e^{kX}]^n$$

$$= \ln E[e^{kX}] = \begin{cases} \ln \Sigma_i p_i e^{kX_i} \\ \ln \int p(x) e^{kx}\, dx \end{cases}$$ cumulant function

Example 1: $p(x) = m e^{-mx}$ $\qquad \mu = 1/m$

$$\lambda(k) = -\ln \frac{m-k}{m} = \ln \frac{m}{m-k} = -\ln(1-\mu k) \qquad k < 1/\mu$$

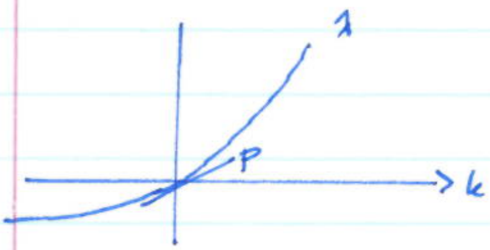
$$I(s) = k_s s - \lambda(k_s) = \frac{s}{\mu} - 1 - \ln\frac{s}{\mu} \qquad s > 0$$



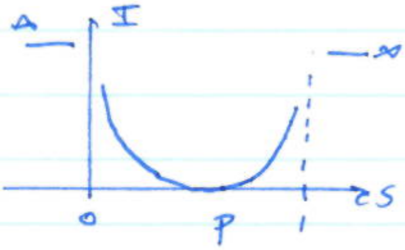
$$k_s : \quad \lambda'(k) = s$$

Example 2: Bernoulli $p(x)=\begin{cases} p & x=1 \\ 1-p & x=0 \end{cases}$

$\lambda(k) = \ln \sum_i p_i e^{kx_i} = \ln\left(pe^k + (1-p)\right)$



$I(s) = k_s s - \lambda(k_s)$
$= s \ln \frac{s}{p} + (1-s) \ln \frac{(1-s)}{1-p}$

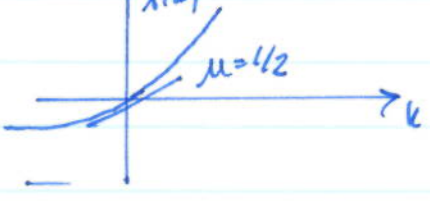


$k_s: \lambda'(k) = s$

Rem: $S_n = \frac{1}{n} \sum_{i=1}^{n} X_i \in \{0, 1/n, 2/n, \dots, \frac{n-1}{n}, 1\} \xrightarrow{n\to\infty} [0,1]$

Example 3: Uniform RV $p(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$

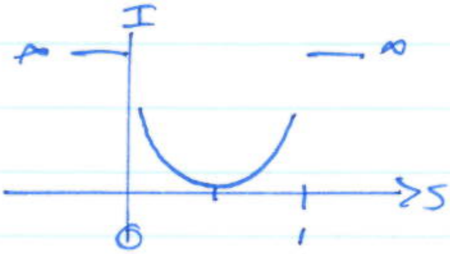
$\lambda(k) = \ln \int_0^1 p(x) e^{kx} dx = \ln \left. \frac{e^{kx}}{k} \right|_0^1 = \ln\left(\frac{e^k - 1}{k}\right)$



$k_s: \lambda'(k) = s$ transcendental eq.

↳ solve numerically using Maple/Mathematica

↳ $I(s) = k_s s - \lambda(k_s)$



$s \in [0,1]$

4.2. Parametric Legendre transform

~~Protokoll~~

Legendre transform:

$$I(s) = k_s \cdot s - \lambda(k_s) \qquad k_s : \lambda'(k) = s$$

① Fix s
② Solve $\lambda'(k) = s$ for $k \to k_s$
③ Get $I(s) = k_s \cdot s - \lambda(k_s)$
④ Repeat for $s \in$ Range S_n

Parametric Legendre transform:

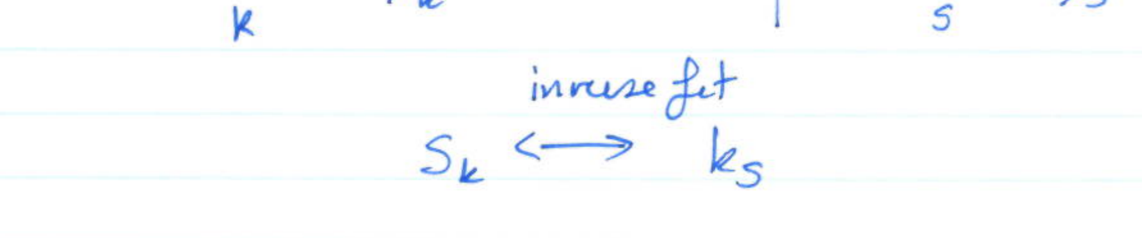
$$I(s_k) = k\, s_k - \lambda(k) \qquad s_k : \lambda'(k) = s_k$$

① Fix $k \in \mathbb{R}$
② Calculate $s_k = \lambda'(k)$
③ Get $I(s_k) = k \cdot s_k - \lambda(k)$
$= k \cdot \lambda'(k) - \lambda(k)$

Parametric plot: $\big(k, I(s_k)\big) = \big(k, k\lambda'(k) - \lambda(k)\big)$

④ Repeat $\forall k \in \operatorname{dom} \lambda$
↪ no solving needed!

Relation – duality



inverse fct

$$s_k \longleftrightarrow k_s$$

Rem: Exercise: uniform sample mean

4.3. Sanov's Theorem

Ref: Ivan N. Sanov, Math. Sib., 1957

· Sequence of iid RVs: $X = (X_1, X_2, \dots, X_n)$ $\quad X_i \sim p(x)$ iid

· Discrete RVs: $X_i \in \{1, 2, \dots, q\} = \Lambda$ $\quad$ q colors

$$p(j) = p_j \quad , \quad j \in \Lambda$$

· Empirical distribution:

$$L_{n,j} = \frac{\# X_i : X_i = j}{n} = \frac{\text{frequency of } j \text{ in } X}{n}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \underbrace{\delta_{X_i, j}}_{= \begin{cases} 1 & X_i = j \\ 0 & \text{otherwise} \end{cases}}$$

· Empirical vector:

$$L_n = (L_{n,1}, L_{n,2}, \dots, L_{n,q})$$

- $0 \leq L_{n,j} \leq 1$
- $\sum_{j \in \Lambda} L_{n,j} = 1$
- Looks like a probability distribution but not a " "

↳ empirical frequencies

Example: $\Omega = \{0,1\}$ $\quad P_0 = 1-p$ $\quad$ Bernoulli

$P_1 = p$

Possible sequence:

$$X = (0,0,1,0,1,1,0) \qquad n=7$$

$$L_{n,0} = \frac{\#\ 0\text{'s}}{7} = \frac{4}{7}$$

$$L_{n,1} = \frac{\#\ 1\text{'s}}{7} = \frac{3}{7}$$

$$L_n = \left(\frac{4}{7}, \frac{3}{7}\right) \qquad \text{normalized } \frac{3}{7}+\frac{4}{7}=1$$

Another sequence:

$$X = (0,0,0,0,0,0,0)$$

$$L_{n,0} = 1$$

$$L_{n,1} = 0$$

$$L_n = (1,0)$$

Rem: Suppose $p = 1/2$

- Most probable value of L_n as $n \to \infty$?
- Least likely L_n?
- $P_r(L_n)$?

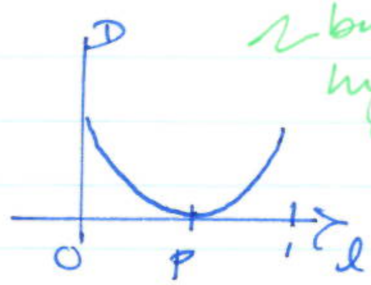
Theorem (Sanov)

- $X = (X_1, \ldots, X_n)$, $X_i \sim p(x)$ iid discrete or continuous

L_n satisfies the LDP
$$P(L_n = \ell) \asymp e^{-n D(\ell \| p)}$$
with rate function
$$D(\ell \| p) = \begin{cases} \sum_{j \in \Lambda} \ell_j \ln \frac{\ell_j}{p_j} & \text{discrete} \\ \int_\Lambda \ell(x) \ln \frac{\ell(x)}{p(x)} & \text{continuous} \end{cases}$$

Remarks:

- L_n is a vector - discrete case
  is a fct - continuous case
  ↳ Meaning of $P(L_n = \ell)$
- $D(\ell \| p) = \infty$ if ℓ not a distribution. (Values of L_n necessarily form a distribution.)
- $D(\ell \| p)$ = relative entropy of ℓ w/r to p
  = Kullback-Leibler distance between ℓ and p
  $D(\ell \| p) = 0 \Leftrightarrow \ell = p$
  = relative information
- $D(\ell \| p) = 0$ for $\ell = p$, $\ell_j = p_j$, $\ell(x) = p(x)$ unique zero
- $D(\ell \| p) > 0$ $\forall$ other ℓ
- $D(\ell \| p)$ convex in ℓ



~ but higher dim

Proof 1: Discrete case $\Lambda = \{1, 2, \dots, q\}$

Use multinomial distribution and Stirling's approximation

See exercises

→ Done by Boltzmann 1877.

→ Also in Sanov 1957

What about continuous RVs?

Proof 2: Gärtner - Ellis Theorem

$\Lambda = \{1, 2, \dots, q\}$, $X_1, \dots, X_n$, p_j

· Empirical frequency:

$$L_{n,j} = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\delta_{X_i, j}}_{= \begin{cases} 1 & X_i = j \\ 0 & X_i \neq j \end{cases}}$$

· Sample mean of iid Bernoulli RVs

· Empirical vector: $\vec{L}_n = (L_{n,1}, L_{n,2}, \dots, L_{n,q})$

· SCGF: $\vec{k} = (k_1, k_2, \dots, k_q)$

$$\lambda(\vec{k}) = \lim_{n \to \infty} \frac{1}{n} \ln E[e^{n \vec{k} \cdot \vec{L}_n}]$$

· iid sample mean: $\lambda(k) = \ln E[e^{kX}]$

$$\Rightarrow \lambda(\vec{k}) = \ln E[e^{\vec{k} \cdot \delta_{X_1, \cdot}}]$$

$$= \ln \sum_{x \in \Lambda} p(x)\, e^{\sum_j k_j \delta_{x,j}}$$

$$= \ln \sum_{x \in \Lambda} p(x)\, e^{k_x}$$

$$= \ln \sum_{j \in \Lambda} p_j\, e^{k_j}$$

Example: Bernoulli RVs $\Lambda = \{\underset{1-p}{0}, \underset{p}{1}\}$, $\vec{k} = (k_0, k_1)$, $\vec{L}_n = (L_{n,0}, L_{n,1})$

$$\lambda(\vec{k}) = (1-p)e^{k_0} + p e^{k_1}$$

GE Theorem :

- $\lambda(\vec{k}) < \infty \quad \forall \vec{k} \in \mathbb{R}^q$
- $\lambda(\vec{k})$ differentiable in vector sense $\forall \vec{k} \in \mathbb{R}^q$

$\Rightarrow$ ① LDP : $P(\vec{L}_n = \vec{\ell}) \asymp e^{-n I(\vec{\ell})}$

② Rate function:

$$I(\vec{\ell}) = \sup_{\vec{k} \in \mathbb{R}^d} \{\vec{k} \cdot \vec{\ell} - \lambda(\vec{k})\}$$

Vector Legendre-Fenchel transform

$$= \vec{k}_{\vec{\ell}} \cdot \vec{\ell} - \lambda(\vec{k}_{\vec{\ell}})$$

$$\vec{k}_{\vec{\ell}} : \nabla \lambda(\vec{k}) = \vec{\ell}$$

Solution: (~~with functional derivatives~~)

$$\lambda(\vec{k}) = \ln \sum_j p_j e^{k_j}$$

$$\frac{\delta}{\delta k_i} \lambda(\vec{k}) = \frac{1}{\sum_j p_j e^{k_j}} \sum_j p_j e^{k_j} \underbrace{\frac{\delta k_i}{\delta k_j}}_{= \delta_{i,j}}$$

$$= \frac{1}{\sum_j p_j e^{k_j}} p_i e^{k_i} = \ell_i$$

normalized $\sum_i \ell_i = 1$

$$\Rightarrow k_i = \ln \ell_i - \ln p_i + \ln \sum_j p_j e^{k_j}$$

$$= \ln \frac{\ell_i}{p_i} + \ln \sum_j p_j e^{k_j} = \ln \frac{\ell_i}{p_i} + \lambda(\vec{k})$$

$$\Rightarrow I(\vec{\ell}) = \sum_i \ell_i \ln \frac{\ell_i}{p_i} + \sum_i \ell_i \lambda(\vec{k}) - \lambda(\vec{k})$$

$$= \sum_i \ell_i \ln \frac{\ell_i}{p_i} + \lambda(\vec{k}) - \lambda(\vec{k})$$

$$= D(\vec{\ell} \,||\, \vec{p})$$

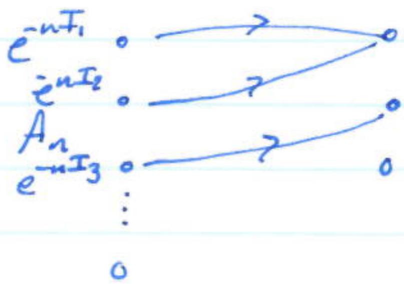
Also works in continuous case ☒

4.4. The contraction principle

Problem:

$$\left.\begin{array}{l} \cdot\ \text{LDP for } A_n \\ \cdot\ B_n = f(A_n) \end{array}\right\} \overset{?}{\Rightarrow} \text{LDP for } B_n$$

Example:



$p = e^{-nI_1} + e^{-nI_2} \asymp e^{-n \max\{I_1, I_2\}}$

B_n $\quad p = e^{-nI_3}$ ok.

Contraction principle: · A_n has LDP with rate function I_A.
· $B_n = f(A_n)$ with f continuous

Then

① LDP for B_n : $P(B_n = b) \asymp e^{-n I_B(b)}$

② Rate function:

$$I_B(b) = \underbrace{\inf_{a : f(a) = b}}_{\text{infimum = min}} I_A(a) = \inf_{a \in \underbrace{f^{-1}(b)}_{\text{pre-image}}} I_A(a)$$

"Proof": $P(B_n = b) = \int \delta(f(a) - b)\, P(A_n = a)\, da$

$$= \int_{f^{-1}(b)} |J|\ P(A_n = a)\, da$$

$$\asymp \int_{f^{-1}(b)} e^{-n I_A(a)}\, da$$

$$\asymp e^{-n \min I_A(a)}$$

$$\Rightarrow I_B(b) = \lim_{n\to\infty} -\frac{1}{n} \ln P(B_n = b) = \min_{a : f(a) = b} I_A(a)$$

■

Example: Contraction of Sanov LDP

$\downarrow$

Sample mean LDP

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$ Sample mean level 1

$$L_n(x) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i, x} & \text{discrete} \\ \frac{1}{n}\sum_{i=1}^{n} \delta(X_i - x) & \text{continuous} \end{cases}$$ Empirical vec level 2

$$S_n = f(\vec{L}_n) = \begin{cases} \sum_x L_n(x)\, x & \text{discrete} \\ \int dx\, L_n(x)\, x & \text{continuous} \end{cases}$$

$$\Rightarrow \quad I_1(s) = \min_{\vec{l} : f(\vec{l}) = s} I_2(\vec{l}) \qquad \text{iid or not}$$

- iid case: $I_2(\vec{l}) = D(\vec{l} \| p)$ Sanov LDP

$$\Rightarrow \quad I(s) = \min_{\vec{l} : f(\vec{l}) = s} D(\vec{l} \| p) = \min_{\vec{l} : \sum_x l(x) x = s} D(\vec{l} \| \vec{p})$$

- Explicit minimization:

$$l_s(x) = \frac{p(x)\, e^{\gamma_s x}}{E[e^{\gamma X}]} \qquad \text{normalized}$$

γ_s : Lagrange parameter fixed $\ni$ $\sum_x l_s(x) x = s$

$$\Rightarrow I(s) = D(\vec{l}_s \| \vec{p}) = \sum_x l_s(x) \ln \frac{l_s(x)}{p(x)}$$

$$= \gamma_s \underbrace{\sum_x l_s(x) x}_{=s} - \ln E[e^{\gamma_s X}]$$

$$= \gamma_s\, s - \lambda(\gamma_s)$$

$$= k_s\, s - \lambda(k_s) \qquad \text{Legendre transform } \gamma_s = k$$

Basis of Maximum entropy principle