■ Summary of large deviations theory

Hugo Touchette Updated: March 19, 2013

A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications*, Springer, 1998.
R.S. Ellis, Entropy, *Large Deviations, and Statistical Mechanics*, Springer, 1985.

Notations.

$X^n = X_1 X_2 \dots X_n$	Sequence of random variables (RVs)
$x^n = x_1 x_2 \dots x_n$	Realization of X^n
$\mathcal{X} = \{x\}$	State or symbol space
$\mathcal{X}^n = \mathcal{X} \times \cdots \times \mathcal{X}$	Product space; space of $\{x^n\}$
$A_n : \mathcal{X}^n \mapsto \mathcal{A}$	"Observable"; $\{A_n\}$ is a sequence of RVs; $\mathcal{A}$ is the observable space

Examples.

$X^n = X_1 X_2 \dots X_n$, $X_i \sim p(x)$ IID	Independent and identically distributed RVs
$P(x^n) = p(x_1)p(x_2)\cdots p(x_n)$	Product measure for IID RVs
$L_n(x) = \frac{1}{n}\sum_{i=1}^n \delta_{x_i,x}$ or $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$	Empirical measure vector or type
$U_n(x^n) = \frac{1}{n}\sum_{i=1}^n u(x_i)$	Mean energy of n non-interacting particle (perfect gas)

Large deviation property.

$\{A_n\}_{n=1}^\infty$	Sequence of RVs
$\{P(A_n \in da)\}_{n=1}^\infty$	Sequence of probability measures
$P(A_n \in da) \asymp e^{-nI(a)} da$	Large deviation property (LDP)
$\lim_{n\to\infty} -\frac{1}{n} \ln \Pr(A_n \in da) = I(a)$	
$I(a)$	Rate function

Examples. (IID RVs $X_i \sim p(x)$)

$A_n = L_n$	Observable = type (vector of rational entries)
$P(L_n \in dl) \asymp e^{-nD(l\|\|p)} dl$	LDP for type; Sanov's theorem
$D(l\|\|p) = \sum_{x\in\mathcal{X}} l(x) \ln \frac{l(x)}{p(x)}$	Rate function = Kullback-Leibler distance
$\mathcal{X} = \{0,1\}$	Binary RVs
$R_n(x^n) = \frac{1}{n}\sum_{i=1}^n x_i$	Fraction of 1's in x^n
$P(R_n \in dr) \asymp e^{-nI(r)}$, $r \in [0,1]$	LDP for coin tossing; Cramer's theorem
$I(r) = \ln 2 + r \ln r + (1-r)\ln(1-r)$	

Principle of largest term and Laplace integral method.

$e^{na} + e^{nb} \asymp e^{n\max\{a,b\}}$	Principle of largest term (PLT)
$\sum_{i=1}^{m} e^{na_i} \asymp e^{n\max_i\{a_i\}}$	
$m = cte < \infty$ or $m = O(n^r)$, $r < \infty$	Finite of polynomial (in n) number of terms
$f_n = \int_{[\alpha,\beta]} g(x)e^{-n\phi(x)}dx \asymp g(x^*)e^{-n\phi(x^*)}$	Laplace approximation
1. $\phi'(x^*) = 0$, $x \in [\alpha, \beta]$	x^* is an extremal point
2. $\phi''(x^*) > 0$	x^* is a unique minimum of ϕ
3. $g(x^*) \neq 0$	

Varadhan's lemma.

$P(A_n \in da) \asymp e^{-nI(a)}da$ — LDP

$$\begin{aligned} E[e^{ng(A_n)}] &\asymp \int e^{ng(a)-nI(a)}da && \text{Exponential integral} \\ &\asymp e^{n\sup_a\{g(a)-I(a)\}} && \text{Laplace approximation} \\ \lambda[g] &= \lim_{n\to\infty} \frac{1}{n} \ln E[e^{ng(A_n)}] \\ &= \sup_a\{g(a) - I(a)\} && \text{Variational solution} \end{aligned}$$

Scaled cumulant generating function and the rate function.

$$\begin{aligned} &\{A_n\}_{n=1}^{\infty} && \text{Sequence of RVs} \\ P(A_n \in da) &\asymp e^{-nI(a)} \\ g(a) &= ka && \text{Linear test function} \\ \lambda(k) &= \lim_{n\to\infty} \frac{1}{n} \ln E[e^{nkA_n}] && \text{Scaled cumulant generating function (SCGF)} \\ &= \lim_{n\to\infty} \frac{1}{n} \ln \int e^{nkA_n(x^n)} P(X^n \in dx^n) \\ &= \lim_{n\to\infty} \frac{1}{n} \ln \int e^{nka} P(A_n \in da) \\ &= \lim_{n\to\infty} \frac{1}{n} \ln \int e^{n[ka-I(a)]}da \\ \lambda(k) &= \sup_a\{ka - I(a)\} && \text{Legendre-Fenchel transform of } I(a) \end{aligned}$$

Example. (IID RVs $X_i \sim p(x)$)

$$\begin{aligned} S_n(X^n) &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ \lambda(k) &= \lim_{n\to\infty} \frac{1}{n} \ln E[e^{knS_n}] && \text{Scaled moment generating function} \\ &= \lim_{n\to\infty} \frac{1}{n} \ln E[e^{kX_i}]^n \\ &= \ln E[e^{kX_i}] && \text{log-Laplace transform of } p(x) \end{aligned}$$

Gärtner-Ellis theorem.

$\{A_n\}_{n=1}^{\infty}$	Sequence of RVs
$\lambda(k) = \lim_{n\to\infty} \frac{1}{n} \ln E[e^{nkA_n}]$	SGMF

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

$P(A_n \in da) \asymp e^{-nI(a)}da$

$$I(a) = \sup_k \{ka - \lambda(k)\} \qquad \text{Gärtner-Ellis}$$

Examples. (Sums of IID RVs $X_i \sim p(x)$)

$$S_n(X^n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Gaussian	$X_i \sim \mathcal{N}(\mu, \sigma^2)$	$x \in \mathbb{R}$
$\lambda(k) =$	$\mu k + \frac{1}{2}\sigma^2 k^2$	$k \in \mathbb{R}$
$I(s) =$	$\dfrac{(s-\mu)^2}{2\sigma^2}$	$s \in \mathbb{R}$

Exponential	$p(x) = me^{-mx}$	$x \in \mathbb{R}_+$
$\lambda(k) =$	$-\ln[(m-k)/m]$	$k < m$
$I(s) =$	$ms - 1 - \ln ms$	$s \in \mathbb{R}_+$

Binary $\{\pm 1\}$	$p(\pm 1) = 1/2$	
$\lambda(k) =$	$\ln\cosh k$	$k \in \mathbb{R}$
$I(s) =$	$\dfrac{(1+s)}{2}\ln(1+s) + \dfrac{(1-s)}{2}\ln(1-s)$	$s \in [0,1]$

Three levels of description.

$A_n(x^n) : \mathcal{X}^n \mapsto \mathbb{R}$	Scalar observable
$P(A_n \in da) \asymp e^{-nI(a)}da$	Level-1 LDP
L_n	Empirical measure
$P(L_n \in dl) \asymp e^{-nI(l)}dl$	Level-2 LDP
$M_n^{(n)}$	Joint-n empirical measure
$P(M_n^{(n)} \in dm) \asymp e^{-nI(m)}dm$	Level-3 LDP

Contraction principle.

$\{A_n\}_{n=1}^{\infty}$	$\xrightarrow{f:\mathcal{A}\mapsto\mathcal{B}}$	$\{B_n\}_{n=1}^{\infty};\ B_n = f(A_n)$	Contraction
$a \in \mathcal{A}$		$b \in \mathcal{B}$	
$P(A_n \in da) \asymp e^{-nI(a)}da$		$P(B_n \in db) \asymp e^{-nI(b)}db$	
		$I_B(b) = \min_{a\in\mathcal{A}:b=f(a)} I_A(a)$	Largest term estimate

Markov chains and the pair empirical measure.

$X^n = X_1X_2\dots X_n,\ X_i \in \mathcal{X}$	Sequence of RVs
$X_{n+1} = X_1,\ X_0 = X_n$	Periodic boundary conditions
$P(x^n) = \prod_{i=1}^{n} P(x_{i+1}\|x_i)$	Probability measure under Markov condition
$P(x_{i+1}\|x_i) = P(x'\|x)$	Transition probability matrix
$L_n^2(x,y) = \frac{1}{n}\sum_{i=1}^{n}\delta_{(x_i,x_{i+1}),(x,y)} = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i,x}\delta_{x_{i+1},y}$	Pair empirical measure
$P(L_n^2 \in dl) \asymp e^{-nD(l\|\|P)}dl$	LDP for the pair empirical measure
$D(l\|\|P) = \sum_{x\in\mathcal{X},y\in\mathcal{X}} l(x,y)\ln\frac{l(x,y)}{l(x)P(y\|x)}$	Extended information divergence
$L_n(x) = \sum_{y\in\mathcal{X}} L_n(x,y)$	Type projection (marginal empirical measure)

Large deviation property (rigorous definition).

$\{P(A_n \in da)\}_{n=1}^{\infty}$ satisfies a LDP with rate n and rate function I if

(1) I is a rate function (cf. below)	
(2) $\lim_{n\to\infty}\sup\frac{1}{n}\ln P(A_n \in [A]) \le -I([A])$	$[A]$ closed subset of $\mathcal{A}$
(3) $\lim_{n\to\infty}\inf\frac{1}{n}\ln P(A_n \in]A[) \ge -I(]A[)$	$]A[$ open subset of $\mathcal{A}$
$I(A) = \inf_{a\in A} I(a)$	LDP rate function