Large Deviations Theory

一、 簡介

在機率論，大離差理論(large deviations theory)主要探討的是隨機序列機率分配尾部 (tails of random sequences of probability distributions)的漸進行為(asymptotic behavior)。大離差理論的一些基本概念早期可以追溯到 Laplace 的文獻，以及瑞典數學家 Harald Cramér (1938)和 Lundberg 的毀滅理論(ruin theory)。然而大離差理論正式被 Varadhan 提出是在 1966 年。

大離差理論提供了兩大方向的貢獻：

(1) 對 concentration of measures 提供啟發式的發展

(2) 機率測度收斂(convergence of probability measures) 的一般化

簡言之，大離差理論主要關心極端稀少或是尾部事件(extreme or tail events)的機率測度收斂問題。

關鍵字：
隨機向量(random vector)、大離差(large deviation)、熵(entropy)、Sanov 定理

二、 重要符號與假設

符號：

$X_1, X_2, \cdots$：獨立同態分配(independently and identically distributed)的隨機序列向量

V: 拓樸向量空間(topological vector space)

J: V 的 Borel 可測子集(measurable subset)，$J \subset V$

$$\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$$

$P(\bullet)$: 定義在機率測度空間的機率函數

μ: 由隨機向量 X 誘生在拓樸空間 V (induced on V by the random X)的機率測度

$$\mu_n(J) = P(\bar{X}_n \in J)$$

$$a_n(J) = n^{-1}\log \mu_n(J) = n^{-1}\log P(\bar{X}_n \in J)$$

$s(J) = \lim_{n\to\infty} a_n(J) = \lim_{n\to\infty} n^{-1}\log P(\bar{X}_n \in J)$：集合熵函數(set entropy function)

$s(v)=\inf\{s(J):v\in J, J \text{ is open convex}\}$：點熵函數(point entropy function)

對於 $m,n=1,2,\cdots$，設 $Y_{m,n}(\omega)=n^{-1}\sum_{i=m}^{m+n-1}X_i(\omega)$

Ω : a space of points ω

$\mathscr{P}$: a sigma-algebra of sets of Ω

$\mathscr{B}$: a sigma-algebra of Borel sets of V

Assumption 1.

(a) $Y_{m,n}$ 是一個 $\mathscr{P}-\mathscr{B}$ 可測轉換(measurable transformation) of Ω into V。

(b) 對於所有在 $\mathscr{B}$ 的 B_1 和 B_2，$P(Y_{1,m}\in B_1, Y_{m+1,n}\in B_2)=P(Y_{1,m}\in B_1)P(Y_{m+1,n}\in B_2)$

(c) 對於所有在 $\mathscr{B}$ 的 B，$P(Y_{m+1,n}\in B)=P(Y_{1,n}\in B)$

Assumption 2.

對於每一個 n，測度 μ_n 是正則(regular)。

Assumption 3.

對於每一個 n，測度 μ_n 是凸-正則(convex-regular)。

三、 Paper 重要定理與結果彙整

以下內容從四個方面彙整：

A. Lanford's entropy function 的存在性

B. 關於點熵 (point entropy)

C. 在連續線性的轉換下，Lanford's Theory 的不變性 (invariance)

D. Sanov 定理

A. Lanford's entropy function 的存在性

Lemma A1.

(a) 如果 $A\subset B$ 且 $s(A),s(B)$ 存在，則 $s(A)\le s(B)$。

(b) 對於 $i=1,2,\cdots,k$，如果 $s(A_i)$ 存在，則 $s\left(\bigcup_{i=1}^{k}A_i\right)$ 存在且等於 $\max\{s(A_i):1\le i\le k\}$。

Lemma A2.

如果 $J\subset V$ 是一個 Borel 可測凸集合，且對於所有 $n\ge k$，$\mu_n(J)>0$，則 $s(J)$ 存

在且

$$s(J)=\sup\{a_n(J):n\geq k\}$$

Lemma A3.

假設S是μ-測度的砥柱集合(support)；亦即$v\in S$若且唯若每一個v的 open neighborhood 有 positive μ-測度。
假設C是S的 closed convex hull。

集合S是封閉的(closed)，$\mu(S)=\mu(C)=1$，且每一個μ-測度 1 的封閉集合包含S。

Lemma A4.

(a) 集合是S_n封閉的，且$\mu_n(S_n)=\mu_n(C_n)=1$。

(b) $S_n\supset S$且$C_n=C$。

(c) 如果J是一個開集合，則$J\cap C\neq\varnothing$若且唯若對於所有n夠大的情況下，$J\cap S_n$是非空集合(nonempty set)。

Lemma A5.

假設K是一個緊緻集合(compact set)。
如果$K\subset V$是一個緊緻集合，則

$$\limsup_{n\to\infty} a_n(K)\leq lan(K)$$

Lemma A6.

如果$K\subset V$是一個封閉凸集合(closed convex set)，則

$$\limsup_{n\to\infty} a_n(K)\leq lan(K)$$

Lemma A7.

如果J是一個開凸集合(open convex set)且$s(J)=\sup\{a_n(J):n\geq k\}>-\infty$，則

給定$\varepsilon>0$，存在一個緊緻凸集合$K\subset J$，使得$s(K)$存在且$s(K)\geq s(J)-\varepsilon$。

Theorem A1.

如果J是開凸集合(open convex set)的有限聯集(finite union)，則$s(J)$存在，

$-\infty\leq s(J)\leq 0$且$s(J)>-\infty$ 若且唯若 $J\cap C\neq\varnothing$。

Theorem A2.

函數 $s:V\to[-\infty,0]$ 是 concave 且 upper semicontinuous。

Theorem A3.

如果 J 是開凸集合(open convex set)的有限聯集(finite union)，則 $s(J)=lan(J)$。

Theorem A4.

假設 S 是 μ-測度的砥柱集合(support)；

亦即 $v\in S$ 若且唯若每一個 v 的 open neighborhood 有 positive μ-測度。

C 是 S 的 closed convex hull。

F 是所有 $v\in V$ 的集合，使得 $s(v)>-\infty$。

F° 是 F 的 interior (可能是 empty)

I 是 C 的 interior (可能是 empty)

(a) $F\subset C$

(b) 如果 F° 是非空集合，則 $F^\circ=I$ 且 $\overline{F}=C$。

(c) 如果 V 是有限維(finite-dimensional)且 I 是非空集合，則以下兩件事成立

(c-1) $F^\circ=I$ 且 $\overline{F}=C$

(c-2) s 在 I 上連續(continuous on I)

Corollary A1.

以下兩者其中之一會成立。

(1) 對於所有 v 在 V 裡，$s(v)=0$。

(2) $\sup\{s(v):v\in V\}=0$ 且 $\inf\{s(v):v\in V\}=-\infty$。

B. 關於點熵 (point entropy)

定義 $L_\theta(v,\varepsilon)=\{w:\theta(w)>\theta(v)-\varepsilon\}$

$H_\theta(v)=\{w:\theta(w)\geq\theta(v)\}$

$M(v)$ 是所有機率測度 v 的非空集合，使得 $E(X_1|\mu)=v$

$\sigma(v)=\sup\{-K(v):v\in M(v)\}$

Theorem B1.

對於每一個v在V裡，以下兩者皆會成立。

(1) $-c^*(v)=\inf\left\{s\left(H_\theta(v)\right):\theta\in V^*\right\}$。

(2) $-c^*(v)=\inf\left\{s\left(L_\theta(v,\varepsilon)\right):\theta\in V^*,0<\varepsilon<\infty\right\}$。

Theorem B2.

對於每一個v在V裡，$s(v)=-c^*(v)$。

Theorem B3.

(a) 函數$\sigma:V\to[-\infty,0]$是凹函數(concave)

(b) 對於所有v在V裡，$\sigma(v)\le s(v)$。

(c) 對於所有v在V裡，$\limsup_{w\to v}\sigma(w)=s(v)$。

(d) 對於每一個非空開集合J，$\sup\left\{\sigma(v):v\in J\right\}=lan(J)$

(e) 如果V是有限維(finite-dimensional)，則$\left\{v:\sigma(v)\neq s(v)\right\}$是一個$C$邊界的子集合

(f) 如果V是一維(one-dimensional)，對於所有v在V裡，$\sigma(v)=s(v)$。

Theorem B4.

對於每一個v在V裡且$\theta\in\Theta$，$s_\theta(v)=s(v)+\theta(v)-c(\theta)$。

Corollary B1.

如果$E\left(X_1|\mu\right)=v$，則$s(v)=0$。

Corollary B2.

對於每一個v在V裡，$\sup\left\{s_\theta(v):\theta\in\Theta\right\}=\begin{cases}0, & \text{if } s(v)=-\infty\\ -\infty, & \text{if } s(v)>-\infty\end{cases}$

Corollary B3.

定義 W 是均數空間(space of means)。
假設 v 是在 W 裡的一點，則

(1) $s(v)=\sigma(v)>-\infty$

(2) 對於每一個 θ 在 $\Theta(v)$ 裡，$s(v)=-K(\mu_\theta)=-\theta(v)+c(\theta)$

如果 θ 和 δ 在 $\Theta(v)$ 裡，則在 V 上的 Borel sets，$\mu_\theta=\mu_\delta$。

C. 在連續線性的轉換下，Lanford's Theory 的不變性 (invariance)

定義 V^0 be equipped with a topology τ^0 in which it is a locally convex Hausdorff topological vector space

ξ be a continuous linear function on V into V^0.

定義 $X_n^0(\omega)=\xi(X_n(\omega))$，對於每一個 $n=1,2,\cdots$。

$\mu^0=\mu\xi^{-1}$

Transformed framework: V^0，τ^0，μ^0 和 $\{X_n^0\}$。

Theorem C1.

(1) V 是具有 locally convex Hausdroff topology τ 的拓樸向量空間(topological vector space)，$\{X_n\}$ 是一組序列在 $(\Omega,\mathscr{F},P)$ into V 上的可測轉換，使得 $\{X_n(\omega)\}$ 是 i.i.d 的過程且 $PX_n^{-1}=\mu$。

(2) **A. Lanford's entropy function 的存在性**

(3) **B. 關於點熵 (point entropy)**

上述(1)，(2)，(3) 的結論在 transformed framework 之下皆成立。

Theorem C2. 充分條件

(a) 函數 $s_0:V^0\to[-\infty,0]$ 是凹函數(concave)

(b) 對於所有 y 在 V^0 裡，$s_0(y)\le s^0(y)$。

(c) 對於所有 y 在 V^0 裡，$\limsup_{z\to y} s_0(z)=s^0(y)$。

(d) 對於每一個非空開集合 $B \subset V^0$，$\sup\{s_0(y): y \in B\} = \sup\{s^0(y): y \in B\}$

(e) 如果 V^0 是有限維(finite-dimensional)，則 $\{y: s_0(y) \neq s^0(y)\}$ is a subset of the boundary of the closed convex hull of the support of μ^0

(f) 如果 $V^0 = R^1$，對於所有 $y \in R^1$，$s_0(y) = s^0(y)$。

Theorem C3.

以下每一個條件皆是 $s_0(y) \equiv s^0(y)$ 的充分條件：

(a) C is τ_w-compact

(b) V is a Banach space and $\int_V \exp\left[t\|v\|\right] \mu(dv) < \infty,\ \forall t > 0$ holds

(c) V is a reflexive Banach space ($V = R^k = V^*$) and $\phi(\theta) < \infty$ in a neighborhood of $\theta = 0$

(d) $\phi(\theta)$ is τ_m^*-continuous at $\theta = 0$

(e) For each t, A_t is τ_w-compact.

(f) For each t, B_t is τ^0-closed.

Corollary C1.

(1) V 是具有 locally convex Hausdroff topology τ 的拓樸向量空間(topological vector space)，$\{X_n\}$ 是一組序列在 $(\Omega, \mathscr{F}, P)$ into V 上的可測轉換，使得 $\{X_n(\omega)\}$ 是 i.i.d 的過程且 $PX_n^{-1} = \mu$。

(2) **A. Lanford's entropy function 的存在性**

(3) **B. 關於點熵 (point entropy)**

上述(1)，(2)，(3) 的結論成立，並且將 τ 換成 τ_1。

在 τ_1 拓樸裡，s_1 是 V 上的點熵函數(point entropy function)，對於所有 v，$s_1(v) \geq s(v)$。

若在 Theorem C3. $s_0(y) \equiv s^0(y)$ 的六個充分條件皆滿足，則 $s_1(v) \equiv s(v)$。

D. Sanov 定理

Let $f:\{1,2,\cdots,r\}\to R$ be such that $\int fdP=0$. Define $\Gamma=\left\{Q\in P:\int fdQ\geq t\right\}$ for some $t>0$.

Let $X_1,X_2,\cdots$ be i.i.d. random variables with values in a finite set $\{1,2,\cdots,r\}$ and with distribution P (random variables in a continuous space will be considered at the end of this lecture). Denote by $\mathcal{P}$ the set of probabilities on $\{1,2,\cdots,r\}$.

Let $\hat{P}_n$ be the empirical distribution of X:

$$\hat{P}_n=\frac{1}{n}\sum_{k=1}^{n}\delta_{X_k}.$$

The law of large numbers states that $\hat{P}_n\to P$ a.s. To define a rare event, we fix $\Gamma\subseteq\mathcal{P}$ that does not contain P.

We are interested in behavior of probabilities of the form $\mathbf{P}[\hat{P}_n\in\Gamma]$, as $n\to\infty$

Consider $X_i\overset{i.i.d.}{\sim}Q$ for $i=1,2,\ldots,n$. Let $\mathcal{F}\subseteq\mathcal{P}(\mathcal{X})$ be a collection of probability distributions on the alphabet $\mathcal{X}$, denoting a rare event. Then,

$$\boxed{Q^n(\mathcal{F})\triangleq\mathbb{P}\{\mathbf{\Pi}_{X^n}\in\mathcal{F}\}\leq|\mathcal{P}_n|\,2^{-n\left(\inf_{P\in\mathcal{F}}D(P\|Q)\right)}}\,.$$

In addition, if Q and $\mathcal{F}$ satisfy certain regularity conditions, eg., there exists a sequence of types $\{P_n\in\mathcal{F}\cap\mathcal{P}_n\}$ such that

$$\limsup_{n\to\infty}D(P_n\|Q)=\inf_{P\in\mathcal{F}}D(P\|Q),$$

then the rate function can be completely characterized:

$$\boxed{\lim_{n\to\infty}\left\{-\frac{1}{n}\log Q^n(\mathcal{F})\right\}=\inf_{P\in\mathcal{F}}D(P\|Q)}\,.$$

四、 應用

Large deviation theory 應用在以下領域

(1) 保險理論的極端事件發生可能性的估計
(2) 訊息理論的 entropy 計算

從大數法則(Law of large number)得知，獨立同態隨機變量序列$\{X_1, X_2, \cdots, X_n\}$

的樣本平均(sample mean)

$$\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$$

會收斂到拓樸向量空間(topological vector space)裡隨機變量的均值 m。

在財務應用上，保險公司會對「隨機變量 $\bar{X}_n$ 遠離 m 的機率是多少」會感到興趣。因為此問題反映出「極端事件」發生的機率估計(probabilistic estimation)，是風險管理(risk management)的一個課題。

在風險管理中，常用變異數縮減(variance reduction)的技術：**重要抽樣法**(importance sampling)去估計違約機率(default probability)。可以利用

Cramér 定理：

令$\{X_1, X_2, \cdots, X_n\}$為實值 i.i.d 的隨機變量。在機率測度 P 之下，$E(X_1) < \infty$，

對於任何一個出象$x \geq E(X_1)$，

$$\lim_{n\to\infty} n^{-1} \ln P\left(n^{-1}\sum_{i=1}^{n} X_i \geq x\right) = -\inf_{y\geq x} \Gamma^*(y)$$

其中$\Gamma(\theta) = \ln E\left(e^{\theta X_i}\right)$是累積生成函數(cumulant generating function)

$\Gamma^*(\theta) = \sup_{\theta\in\Re}\left[\theta x - \Gamma(\theta)\right]$是 Fenchel-Legendre 轉換(Fenchel-Legendre transform)

假設$\{X_1, X_2, \cdots, X_n\}$為實值 i.i.d 的$N(0,1)$隨機變量，且$1 \ll c$，從 **Cramér 定理**

以及動差母函數$E\left(e^{\theta X_i}\right) = \exp\left(\frac{\theta^2}{2}\right)$，可以得知

$$\lim_{n\to\infty} \frac{1}{n} \ln P\left(n^{-1}\sum_{i=1}^{n} X_i \geq x\right) = -\frac{x^2}{2}$$

或是

$$P\left(n^{-1}\sum_{i=1}^{n}X_i \ge x\right) \approx \exp\left(-\frac{nx^2}{2}\right)。$$

$\{X_1, X_2, \cdots, X_n\}$ 為 i.i.d 的 $N(0,1)$，則 $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i \sim N(0,1)$，又 $X \sim N(0,1)$，則違約機率

$$P\left(X \ge c := \sqrt{n}x\right) = P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i \ge \sqrt{n}x\right)。$$

由 **Cramér 定理**可以得到

$$P(X \ge c) \approx \exp\left(-\frac{c^2}{2}\right)，1 \ll c。$$

透過 $\mu = c$ 的選擇，二階動差

$$E_\mu\left[I(x > c)e^{\mu^2}e^{-2\mu x}\right] \approx \exp\left(-c^2\right)$$

亦可以用 **Cramér 定理**推出。這是大離差理論在 Importance sampling 的重要應用。

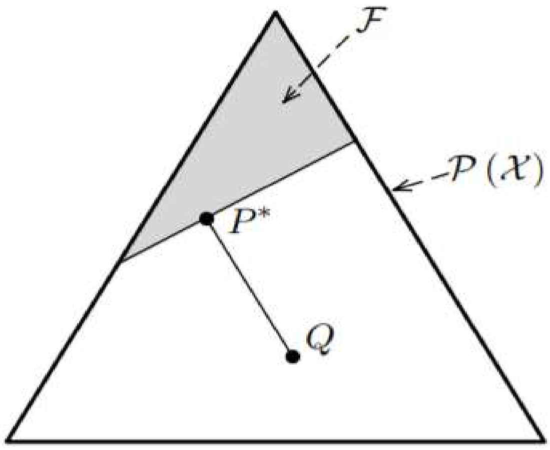
在訊息理論(information theory)，Sanov 定理提供了在某機率分配之下，觀察到非典型樣本，計算機率邊界(probability bound)。以大離差理論的術語來說，Sanov 定理在 Kullback-Liebler divergence 下去計算出獨立同態隨機變數序列的 rate function。

The Large Deviation Principle (LDP) characterizes the limiting behavior of the probability of certain rare events.

Since the probability of rare events are usually exponentially decaying, the rate function (rate of convergence), $\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{P}\{\text{rare event}\}$, is the main object of interest.

For example:

Ensemble	Rare Event	Rate Function
$X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$	$\left\{\left\lvert\frac{1}{n}\sum_{i=1}^{n}X_i\right\rvert\right\} \ge \varepsilon$	at least $\frac{\log e}{2}\varepsilon^2$
$X_i \overset{\text{i.i.d.}}{\sim} \text{Ber}(p)$	$\left\{\left\lvert\frac{1}{n}\sum_{i=1}^{n}X_i - p\right\rvert\right\} \ge \varepsilon$	at least $\frac{\log e}{3p}\varepsilon^2$



Sanov's Theorem tells us the probability that, after getting n i.i.d. samples of a distribution, the empirical distribution is still "far" away from the actual distribution, is exponentially small.

"How far" is measured by KL divergence.

五、 參考資料

[1] Online lecture notes from Princeton University
https://blogs.princeton.edu/sas/2013/10/10/lecture-3-sanovs-theorem/

[2] Online lecture notes from National Taiwan University (臺大電機系 王奕翔教授)
http://homepage.ntu.edu.tw/~ihwang/Teaching/Fa15/Slides/IT_Lecture_08_P1_v1.pdf

[3] 韓傳祥（2011）。*金融隨機計算*。新竹：新陸書局。