

大偏差原理

大上由人

2024 年 11 月 8 日

1 大偏差原理

1.1 大偏差原理の定義

Def. 大偏差原理

A_n を、集合 B からとってきた確率変数として、その確率を $P(A_n)$ とする。このとき、大偏差原理が成立するとは、

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \ln P(A_n) = I_B \quad (1.1)$$

なる I_B が存在することをいう。このとき、 I_B を、集合 B に対するレート関数という。

すなわち、 $P(A_n)$ の支配的なふるまいが、 $n \rightarrow \infty$ で e^{-nI_B} に漸近することをいう。とくに、 I_B は、その減衰のスピードを特徴づける量となる。とくに定義の式は、

$$-\ln P(A_n) = nI_B + o(n) \quad (1.2)$$

と書ける。また、同じことではあるが、

$$-\frac{1}{n} \ln P(A_n) = I_B + o(1) \quad (1.3)$$

とも書ける。

大偏差原理に従わないような確率分布も存在する。例えば、 $P(A_n)$ が、 e^{-na} ($a > 0$) よりも素早く減衰する場合、超指数関数的に (Super-exponentially) 減衰するといい、 $I = \infty$ と書く。また、逆に、 $P(A_n)$ が、 e^{-na} ($a > 0$) よりも遅く減衰する場合、部分指数関数的に (Sub-exponentially) 減衰するといい、 $I = 0$ と書く。

また、上の定義は確率分布についてのものであるが、確率密度についても取り扱うことができる。これは、

$$P(S_n \in [s, s + ds]) = p(s) ds \quad (1.4)$$

を用いてあげること、

$$P(S_n \in [s, s + ds]) \sim e^{-nI_B(s)} ds \quad (1.5)$$

と書ける。とくに、これの対数をとることで、

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \ln P(S_n \in [s, s + ds]) = I_B(s) + \lim_{n \rightarrow \infty} \frac{1}{n} \ln ds = I_B(s) \quad (1.6)$$

となる。

1.2 数学的注意

上のような大偏差原理の特徴づけは、数学的にはすこし雑なものである。とくに、注意すべきケースについては、次のようなものがある。

1. 大偏差原理の定義式における極限が存在しない場合

このとき、一般には大偏差原理が成り立たない。しかし、 $P(A_n \in B)$ の上界や下界が存在する場合については、

$$e^{-nI_B^-} \leq P(A_n) \leq e^{-nI_B^+} \quad (1.7)$$

という関係が成り立つ。すなわち、レート関数 I_B の上下による大偏差原理が成り立つ。ただし、これ以降の議論では、基本的に確率分布の極限が存在することを仮定し、 $I_B^- = I_B^+ = I_B$ として議論を進める。

2. 離散変数と連続変数の取り扱い

大偏差原理は、離散変数に対しても連続変数に対しても成り立つ。ただし、離散変数を用いているときは、 $n \rightarrow \infty$ とすることで、連続変数として取り扱うことができる。

1.3 レート関数の計算

大偏差原理を実用するうえでは、以下の二点について考えなくてはならない。

1. 特定のランダム変数に対して、確率分布が大偏差原理を満たすことを確認する。
2. レート関数 I_B を計算する。

この二点目について、レート関数を具体的に求める方法は、大きく分けて以下の二つがある。

1. 確率分布を直接計算し、この分布に対してスターリングの公式などの漸近公式を利用する。
2. Gärtner-Ellis の定理を用いる。^{*1}

一つ目の方法では難しい場合があるため、二つ目の方法を用いることが多い。

^{*1} 初めて ä を打った。

Def. スケーリングキュムラント母関数

確率変数 A_n に対して、そのスケーリングキュムラント母関数を、

$$\lambda(k) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \langle e^{nkA_n} \rangle \quad (1.8)$$

と定義する。ただし、

$$\langle A_n \rangle = \int_{\mathbb{R}} e^{nka} P(A_n \in da) \quad (1.9)$$

である。

スケーリングキュムラント母関数は、確率変数 A_n の大偏差原理におけるレート関数 I_B と関係がある。具体的には、以下の定理が成り立つ。

Thm. Gärtner-Ellis の定理

確率変数 A_n に対して、そのスケーリングキュムラント母関数 $\lambda(k)$ が存在し、 $k \in \mathbb{R}$ で微分可能であるとする。このとき、 A_n は大偏差原理を満たし、そのレート関数は以下のルジャンドル変換で与えられる。

$$I(a) = \sup_k \{ka - \lambda(k)\} \quad (1.10)$$

が成り立つ。

すなわち、 $\lambda(k)$ が微分可能なとき、そのルジャンドル-フェンシエル変換によって、レート関数 I が求まる。

ただし、これですべてのレート関数が求まるわけでもないことに注意されたい。実際、 $\lambda(k)$ が存在するにもかかわらず、その $I(a)$ が求まらない場合も存在する (らしい、後で扱われるはず)。

Gärtner-Ellis の定理の直感的な理解を試みる。 A_n に対して大偏差原理が成り立つと仮定すると、

$$P(A_n \in da) \asymp e^{-nI(a)} da \quad (1.11)$$

となる。このとき、

$$\langle e^{nkA_n} \rangle = \int_{\mathbb{R}} e^{nka} e^{-nI(a)} da \quad (1.12)$$

$$\sim \exp\left(n \sup_a \{ka - I(a)\}\right) \quad \because \text{鞍点法} \quad (1.13)$$

したがって、

$$\lambda(k) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \langle e^{nkA_n} \rangle = \sup_a \{ka - I(a)\} \quad (1.14)$$

となる。とくに、 $\lambda(k)$ が微分可能な時、逆ルジャンドル・フェンシエル変換によって

$$I(a) = \sup_k \{ka - \lambda(k)\} \quad (1.15)$$

が得られる。

1.4 クラメールの定理

i.i.d な確率変数 X_i について、考える。Gärtner-Ellis の定理を標本平均

$$S_n = \frac{1}{n} \sum_{i=1}^n X_i \quad (1.16)$$

に適用する。このとき、スケーリングキュムラント母関数は、

$$\lambda(k) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \langle e^{nkS_n} \rangle \quad (1.17)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \ln \langle e^{k \sum_{i=1}^n X_i} \rangle \quad (1.18)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \ln \prod_{i=1}^n \langle e^{kX_i} \rangle \quad (1.19)$$

$$= \ln \langle e^{kX} \rangle \quad (1.20)$$

この式が意味することは、一つの確率変数 X に対して、そのスケーリングキュムラント母関数を求めれば、標本平均のスケーリングキュムラント母関数が求まり、そのレート関数も求まるということである。

ex. Gauss 標本平均

hogehoge

ex. 指数標本平均

hogehoge

1.5 λ と I の関係

1.5.1 $k=0$ における値

Prop.

スケーリングキュムラント母関数 $\lambda(k)$ の $k=0$ における値は、以下の特徴を満たす。

$$\lambda(0) = 0 \quad (1.21)$$

$$\lambda'(0) = \lim_{n \rightarrow \infty} \langle A_n \rangle \quad (1.22)$$

$$\lambda''(0) = \lim_{n \rightarrow \infty} n \text{Var}(A_n) \quad (1.23)$$

Prf.

$\lambda(k)$ の定義より、

$$\lambda(0) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \langle e^{n \cdot 0 \cdot A_n} \rangle \quad (1.24)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \ln \langle 1 \rangle \quad (1.25)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \ln 1 \quad (1.26)$$

$$= 0 \quad (1.27)$$

次に、

$$\lambda'(k) = \lim_{n \rightarrow \infty} \frac{1}{n} \frac{\langle e^{nkA_n} n A_n \rangle}{\langle e^{nkA_n} \rangle} \quad (1.28)$$

$$= \lim_{n \rightarrow \infty} \frac{\langle A_n e^{nkA_n} \rangle}{\langle e^{nkA_n} \rangle} \quad (1.29)$$

$$\xrightarrow{k \rightarrow 0} \lim_{n \rightarrow \infty} \frac{\langle A_n \rangle}{1} \quad (1.30)$$

$$= \lim_{n \rightarrow \infty} \langle A_n \rangle \quad (1.31)$$

最後に、

$$\lambda''(k) = \frac{\partial}{\partial k} \lim_{n \rightarrow \infty} \frac{\langle A_n e^{nkA_n} \rangle}{\langle e^{nkA_n} \rangle} \quad (1.32)$$

$$= \lim_{n \rightarrow \infty} n \frac{\langle A_n^2 e^{nkA_n} \rangle \langle e^{nkA_n} \rangle - \langle A_n e^{nkA_n} \rangle^2}{\langle e^{nkA_n} \rangle^2} \quad (1.33)$$

$$\xrightarrow{k \rightarrow 0} \lim_{n \rightarrow \infty} n \frac{\langle A_n^2 \rangle - \langle A_n \rangle^2}{1} \quad (1.34)$$

$$= \lim_{n \rightarrow \infty} n \text{Var}(A_n) \quad (1.35)$$

よって、示された。 \square

1.5.2 凸性

$\lambda(k)$ が凸関数であることは、Hölder の不等式を用いて示すことができる。Hölder の不等式は、

$$\sum_i |y_i z_i| \leq \left(\sum_i |y_i|^{1/p} \right)^p \left(\sum_i |z_i|^{1/q} \right)^q, \quad (1.36)$$

である。^{*2}ここで $0 \leq p, q \leq 1$ かつ $p + q = 1$ である。この不等式を $\lambda(k)$ に適用すると、

$$\alpha \ln \langle e^{nk_1 A_n} \rangle + (1 - \alpha) \ln \langle e^{nk_2 A_n} \rangle \geq \ln \langle e^{n[\alpha k_1 + (1 - \alpha)k_2] A_n} \rangle \quad (1.37)$$

^{*2} ここではこの不等式の証明は行わない。初頭的な証明は以下で確認できる:<https://www.u-tokai.ac.jp/uploads/sites/12/2021/03/PP94-99.pdf>

となる。($p = \alpha, q = 1 - \alpha, y_i = e^{nk_1 A_n \alpha} p_i^\alpha, z_i = e^{nk_2 A_n (1-\alpha)} p_i^{1-\alpha}$ とおけばよい。)
したがって、

$$\alpha \lambda(k_1) + (1 - \alpha) \lambda(k_2) \geq \lambda(\alpha k_1 + (1 - \alpha) k_2). \quad (1.38)$$

を得る。これは、 $\lambda(k)$ が凸関数であることを示している。

1.5.3 ルジャンドル・フェンシエル双対性

特に、上の例で見たような、 $\lambda(k)$ が微分可能かつ凸関数である場合には、

$$I(a) = k(a)a - \lambda(k(a)) \quad (1.39)$$

$$a = \lambda'(k) \quad (1.40)$$

が成り立つ。

とくに、 $\lambda(k)$ が狭義凸関数である場合には、 $\lambda'(k)$ は単調増加関数であり、 $\lambda'(k(a)) = a$ を満たす関数 $k(a)$ が一意に存在する。すなわち、 $I'(a) = k(a)$ が成り立つ。また、このときは、 λ の傾きが I の傾きと一対一に対応する。($a = \lambda'(k), k = I'(a)$ は、その対応の式である。) この性質を、ルジャンドル双対性という。ただし、 $\lambda(k)$ が狭義凸関数でない場合にはそうなるとは限らない。

ex.

hoge

1.5.4 Varadhan の定理

実は、Gärtner-Ellis の定理の汎関数版が存在する。

Thm. Varadhan の定理

確率変数 A_n および A_n の関数 f に対して、そのスケーリングキュムラント母関数 $\lambda(f)$ が存在し、

$$\lambda(f) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \left\langle e^{nf(A_n)} \right\rangle = \sup_a \{f(a) - I(a)\} \quad (1.41)$$

が成り立つ。ただし、 $I(a)$ は、 A_n のレート関数である。

ここではこの証明は行わない。Varadhan の定理は、 $f(a) - I(a)$ が複数の極大値を持つときにも成り立つことが知られている。

1.6 レート関数の正值性

Prop.

スケーリングキュムラント母関数 $\lambda(k)$ が狭義凸関数であるとき、レート関数 $I(a)$ は、正值性を持つ。

Prf.

$$\lambda(0) = \sup_a \{0 - I(a)\} \quad (1.42)$$

$$= -\inf_a I(a) \quad (1.43)$$

$$= 0 \quad (1.44)$$

であるから、 $\inf_a I(a) = 0$ である。したがって、 $I(a)$ は正值性を持つ。 \square

1.6.1 レート関数の凸性

Prop.

$\lambda(k)$ が狭義凸関数かつ微分可能な時、レート関数 $I(a)$ は、凸関数である。

Prf.

ルジャンドル変換の一般論より、

$$I''(a)\lambda''(k) = 1 \quad (1.45)$$

が成り立つ。^{*3} $\lambda(k)$ が狭義凸関数であるとき、 $\lambda''(k) > 0$ であるから、 $I''(a) > 0$ となり、 $I(a)$ は凸関数である。 \square

とくに、i.i.d な確率変数 X_i の標本平均 S_n に対しては、

$$I''(a = \mu) = \frac{1}{\sigma^2} \quad (1.51)$$

が成り立つ。ただし、 μ は X_i の平均、 σ^2 は X_i の分散である。

1.6.2 大数の法則

$I(a)$ が最小値 0 をとる点 a^* が存在すると、

$$a^* = \lambda'(0) = \lim_{n \rightarrow \infty} \langle A_n \rangle \quad (1.52)$$

が成り立つ。^{*4} また、 $I(a)$ が a^* で微分可能であるならば、 $I'(a^*) = k(a^*) = 0$ が成り立つ。

^{*3}

$$I'(a) = k(a) \quad (1.46)$$

$$\lambda'(k) = a \quad (1.47)$$

であるから、

$$1 = \frac{\partial}{\partial a} \lambda'(k) \quad (1.48)$$

$$= \lambda''(k) \frac{\partial k(a)}{\partial a} \quad (1.49)$$

$$= \lambda''(k) I''(a) \quad (1.50)$$

となる。

^{*4} これは、 $I(a)$ が最小値 0 をとる点で微分可能な場合は自明。微分不可能な場合についても、 $I'(a)$ のグラフをひっくり返したものが $\lambda'(k)$ に対応することを考えれば、同様の結果が得られる。

$I(a)$ が 0 となる点は、 $n \rightarrow \infty$ としたときに確率分布が減衰しない点である。したがって、

$$P(A_n \in da^*) = \lim_{n \rightarrow \infty} P(A_n \in [a^*, a^* + da^*]) = 1 \quad (1.53)$$

が成り立つ。これは、大数の弱法則に対応し、 $A_n \rightarrow a^*$ を意味する。

ここで大事なこととして、大数の法則とは異なり、大偏差原理は、 A_n が a^* に収束する速度についての情報も持っていることである。正確に述べると、 $A_n \in B$ に対して、

$$\lim_{n \rightarrow \infty} P(A_n \in B) = \int_B P(A_n \in da) \asymp \int_B e^{-nI(a)} da \asymp e^{-n \inf_{a \in B} I(a)} \quad (1.54)$$

が成り立つ。したがって、 $a^* \notin B$ の場合には、 $P(A_n \in B) \rightarrow 0$ となり、 $a^* \in B$ の場合には、 $P(A_n \in B) \rightarrow 1$ となる。

1.6.3 中心極限定理

中心極限定理は、大偏差理論において、凸レート関数 $I(a)$ が唯一の最小値をもち、最小値 0 をとる点で二階微分可能な時に成り立つ。最小値まわりでのレート関数は、

$$I(a) \approx \frac{1}{2} I''(a^*) (a - a^*)^2 \quad (1.55)$$

となる。ただし、 a^* は $I(a)$ の最小値をとる点である。このとき、確率分布 $P(A_n)$ は、

$$P(A_n \in da) \approx e^{-nI(a)} \quad (1.56)$$

$$\approx \exp\left(-\frac{n}{2} I''(a^*) (a - a^*)^2\right) da \quad (1.57)$$

となる。したがって、たしかにガウス型になってくれる。とくに、i.i.d な確率変数 A_n の標本平均 S_n に対しては、 $I''(a^*) = \frac{1}{\sigma^2}$ であり、 $a^* = \mu$ であるから、

$$P(S_n \in ds) \approx \exp\left(-\frac{n}{2\sigma^2} (s - \mu)^2\right) ds \quad (1.58)$$

となる。これは、中心極限定理として知られているものである。

大偏差原理の優れている点は、レート関数がキュムラント母関数と結びついていることからわかるように、高次のゆらぎについても情報を持っていることである。

1.7 収束原理

他のレート関数の知識から新しいレート関数を求める方法として、収束原理がある。

Thm. 収束原理

ランダム変数 A_n のレート関数 $I_A(a)$ が与えられていたとする。このとき、 A_n の関数である $B_n = h(A_n)$ の確率分布 $P(B_n)$ を考える。ただし、 h は連続関数であるとする。このとき、 B_n のレート関数 $I_B(b)$ は、

$$I_B(b) = \inf_{a:h(a)=b} I_A(a) \quad (1.59)$$

となる。また、 $h(a) = b$ なる a が存在しない場合には、 $I_B(b) = \infty$ とする。

収束原理の直感的な説明を試みる。いま、

$$P(B_n \in db) = \int_{a:h(a)=b} P(A_n \in da) \quad (1.60)$$

$$\sim \int_{a:h(a)=b} e^{-nI_A(a)} da \quad (1.61)$$

$$\sim \exp\left(-n \inf_{a:h(a)=b} I_A(a)\right) db \quad (1.62)$$

となる。したがって、

$$I_B(b) = \inf_{a:h(a)=b} I_A(a) \quad (1.63)$$

が成り立つ。