

### 3. Concentration of measure

Yoji Tomita

May 12, 2021

# Introduction

- 2章を前提として, この章では tail bound や concentration inequalities を求めるためのより上級の手法を紹介する.
- 3.1 : Concentration by entropic techniques
- 3.2 : A geometric perspective on concentration
- 3.3 : Wasserstein distances and information inequalities
- 3.4 : Tail bounds for empirical processes

## 3.1 Concentration by entropic techniques

- エントロピーと, 集中不等式導出のためのその関連テクニックに関する議論から始める.

### 3.1.1 Entropy and its properties

- 凸関数  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  と, 確率変数  $X \sim \mathbb{P}$  に対して,  $\phi$ -entropy を

$$\mathbb{H}_\phi(X) := \mathbb{E}[\phi(X)] - \phi(\mathbb{E}[X])$$

とする ( $X, \phi(X)$  の有限期待値は仮定).

- Jensen の不等式より,  $\phi$ -entropy は非負.
- これは  $X$  のばらつき加減を表す.
  - ▶ 極端な場合,  $X$  が a.s. で期待値と一致するなら,  $\mathbb{H}_\phi(X) = 0$ .

- 例 1 :  $\phi(u) = u^2$  なら  $\mathbb{H}_\phi(X)$  は分散.

$$\mathbb{H}_\phi(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \text{var}(X).$$

- 例 2 :  $\phi(u) = -\log u$ ,  $Z := e^{\lambda X}$  とすると,

$$\mathbb{H}_\phi(e^{\lambda X}) = -\lambda \mathbb{E}[X] + \log \mathbb{E}[e^{\lambda X}] = \log \mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}]$$

となり, centered cumulant generating function となる.

- この章では, 次の凸関数  $\phi : [0, \infty) \rightarrow \mathbb{R}$  に対する entropy を考える.

$$\phi(u) := \begin{cases} u \log u & \text{for } u > 0 \\ 0 & \text{for } u = 0 \end{cases}. \quad (3.1)$$

非負確率変数  $Z$  に対して,  $\phi$ -entropy は

$$\mathbb{H}(Z) = \mathbb{E}[Z \log Z] - \mathbb{E}[Z] \log \mathbb{E}[Z], \quad (3.2)$$

となる (ただし関連する期待値の存在は仮定).

- ▶ Shannon entropy や Kullback-Leibler divergence と関連がある (see Exercise 3.1).
  - ▶ 以後この entropy を考えるので,  $\mathbb{H}_\phi$  の subscript  $\phi$  は省略.
- $Z = e^{\lambda X}$  とすると,  $\mathbb{H}(e^{\lambda X})$  は  $X$  のモーメント母関数  $\varphi_X(\lambda) = \mathbb{E}[e^{\lambda X}]$  とその導関数  $\phi'_X(\lambda)$  で表せる.

$$\mathbb{H}(e^{\lambda X}) = \lambda \phi'_X(\lambda) - \varphi_X(\lambda) \log \varphi_X(\lambda). \quad (3.3)$$

### Example 3.1 (Entropy of a Gaussian random variable)

- $X$  は1次元正規分布  $X \sim \mathcal{N}(0, \sigma^2)$  とすると,  $\varphi_X(\lambda) = e^{\lambda^2 \sigma^2 / 2}$ ,  $\varphi'_X(\lambda) = \lambda \sigma^2 \varphi_X(\lambda)$  なので,

$$\mathbb{H}(E^{\lambda X}) = \lambda^2 \sigma^2 \varphi_X(\lambda) - \frac{1}{2} \lambda^2 \sigma^2 \varphi_X(\lambda) = \frac{1}{2} \lambda^2 \sigma^2 \varphi_X(\lambda). \quad (3.4)$$

- この節の残りで, このエントロピー (3.3) と tail bounds との関連性を説明していく.

## 3.1.2 Herbst argument and its extensions

- ある定数  $\sigma > 0$  が存在して、エントロピーが次の上限を満たすとする.

$$\mathbb{H}(e^{\lambda X}) \leq \frac{1}{2} \sigma^2 \lambda^2 \varphi_X(\lambda). \quad (3.5)$$

- ▶ Example 3.1 より, 正規分布  $X \sim \mathcal{N}(0, \sigma^2)$  は任意の  $\lambda \in \mathbb{R}$  に対し (3.5) をイコールで満たす.
- ▶ また任意の bounded な確率変数も (3.5) を満たす (Exercise 3.7).
- このとき, その確率変数は sub-Gaussian となる.

### Proposition 3.2 (Herbst argument)

エントロピー  $\mathbb{H}(e^{\lambda X})$  が (3.5) を任意の  $\lambda \in I$  (ただし  $I = [0, \infty)$  or  $\mathbb{R}$ ) について満たすとする. このとき,

$$\log \mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}] \leq \frac{1}{2} \lambda^2 \sigma^2 \quad \text{for all } \lambda \in I. \quad (3.6)$$



## Remarks:

- $I = \mathbb{R}$  なら, (3.6) は  $X - \mathbb{E}[X]$  がパラメータ  $\sigma$  の sub-Gaussian であることと同値.
- Chernoff argument より,  $I = [0, \infty)$  でも片側 tail-bound

$$\mathbb{P}[X \geq \mathbb{E}[X] + t] \leq e^{-\frac{t^2}{2\sigma^2}} \quad (3.7)$$

が得られ,  $I = \mathbb{R}$  なら両側 tail bounds

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq t] \leq 2e^{-\frac{t^2}{2\sigma^2}}$$

となる.

## Proof.

- $I = [0, \infty)$  の場合のみ示す ( $I = \mathbb{R}$  は演習とする).
- エントロピーのモーメント母関数による表現 (3.3) と仮定 (3.5) より,

$$\mathbb{H}(e^{\lambda X}) = \lambda \varphi'(\lambda) - \varphi(\lambda) \log \varphi(\lambda) \leq \frac{1}{2} \sigma^2 \lambda^2 \varphi(\lambda) \quad \forall \lambda \geq 0. \quad (3.8)$$

- 関数  $G$  を  $G(\lambda) := \frac{1}{\lambda} \log \varphi(\lambda)$  ( $\lambda \neq 0$ ) と定義し,  $\lambda = 0$  では連続性を満たすように

$$G(0) := \lim_{\lambda \rightarrow 0} G(\lambda) = \mathbb{E}[X] \quad (3.9)$$

とする.

- $G'(\lambda) = \frac{1}{\lambda} \frac{\varphi'(\lambda)}{\varphi(\lambda)} - \frac{1}{\lambda^2} \log \varphi(\lambda)$  より, (3.8) は  $G'(\lambda) \leq \frac{1}{2} \sigma^2$  となるので,  $\lambda_0 (> 0)$  から  $\lambda$  まですら両辺積分すると

$$G(\lambda) - G(\lambda_0) \leq \frac{1}{2} \sigma^2 (\lambda - \lambda_0).$$

- $\lambda_0 \downarrow 0$  とすると

$$G(\lambda) - \mathbb{E}[X] \leq \frac{1}{2} \sigma^2 \lambda$$

となり, これは (3.6) と同値である.

- 2章と同様に, 次は sub-exponential tail をもつ確率変数を考える.

### Proposition 3.3 (Bernsten entropy bound)

正整数  $b, \sigma$  が存在して, エントロピー  $\mathbb{H}(e^{\lambda X})$  は以下を満たすとする.

$$\mathbb{H}(e^{\lambda x}) \leq \lambda^2 \{b\varphi'_X(\lambda) + \varphi_X(\lambda)(\sigma^2 - b\mathbb{E}[X])\} \quad \text{for all } \lambda \in [0, 1/b). \quad (3.10)$$

このとき,

$$\log \mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}] \leq \sigma^2 \lambda^2 (1 - b\lambda)^{-1} \quad \text{for all } \lambda \in [0, 1/b). \quad (3.11)$$

Remarks:

- Chernoff argument より, Prop.3.3 は以下の sub-exponential tails をもつ変数の上側 Bernstein-type bound を含意する.

$$\mathbb{P}[X \geq \mathbb{E}[X] + \delta] \leq \exp \left( -\frac{\delta^2}{4\sigma^2 + 2b\delta} \right) \quad \text{for all } \delta \geq 0. \quad (3.12)$$

## Proof.

- 一般性を失わずに  $\mathbb{E}[X] = 0$  と  $b = 1$  を仮定できる (see Exercise 3.6).
- このとき (3.10) は次のように簡単化される.

$$\mathbb{H}(e^{\lambda X}) \leq \lambda^2 \{ \varphi'(\lambda) + \varphi(\lambda) \sigma^2 \} \quad \text{for all } \lambda \in [0, 1). \quad (3.13)$$

- Prop.3.2 の証明と同様に  $G(\lambda) = \frac{1}{\lambda} \log \varphi(\lambda)$  を定義すると, (3.13) は  $G' \leq \sigma^2 + \frac{\varphi'}{\varphi}$  と同値になる.
- $\lambda_0 > 0$  を任意にとり  $\lambda_0$  から  $\lambda$  まで両辺積分すると,

$$G(\lambda) - G(\lambda_0) \leq \sigma^2(\lambda - \lambda_0) + \log \varphi(\lambda) - \log \varphi(\lambda_0).$$

- $\lambda_0 \downarrow 0$  とすると,  $\lim_{\lambda \downarrow 0} G(\lambda_0) = \mathbb{E}[X]$  と  $\log \varphi(0) = 0$  より,

$$G(\lambda) - \mathbb{E}[X] \leq \sigma^2 \lambda + \log \varphi(\lambda) \quad (3.14)$$

となる.

- (3.14) に  $G$  と  $\varphi$  を入れて書き換えると (3.11) が得られる.

□

### 3.1.3 Separately convex functions and the entropic method

- Entropy method は複数の確率変数からなる関数の concentration を考える時に強力.
- 関数  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  が *separately convex* であるとは, 各  $k \in \{1, \dots, n\}$  について 1 変数関数

$$y_k \mapsto f(x_1, \dots, x_{k-1}, y_k, x_{k+1}, \dots, x_n)$$

が任意の  $(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n) \in \mathbb{R}^{n-1}$  に対して凸であることをいう.

- また関数  $f$  がユークリッドノルムに対して  $L$ -Lipschitz であるとは,

$$|f(x) - f(x')| \leq L \|x - x'\| \quad \text{for all } x, x' \in \mathbb{R}^n.$$

が成り立つことをいう.

### Theorem 3.4

$\{X_i\}_{i=1}^n$  は独立な確率変数列でそれぞれのサポートは  $[a, b]$  に含まれるものとし,  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  は separately convex かつ  $L$ -Lipschitz であるとする. このとき, 任意の  $\delta > 0$  に対して

$$\mathbb{P}[f(X) \leq \mathbb{E}[f(X)] + \delta] \leq \exp\left(-\frac{\delta^2}{4L^2(b-a)^2}\right) \quad (3.16)$$

が成り立つ.

Remarks:

- この結果は Gaussian へ変数の Lipschitz 関数の upper tail bound を求めた Thm.2.26 の analogue だが, 独立かつ bounded な変数に対して適用できる.
- ただし, separately convexity の仮定は一般に取り除くことができない.
- $f$  が jointly convex の場合は lower tail bound の導出に他のテクニックが使える (cf Thm.3.24).

### Example 3.5 (Sharp bounds on Rademacher complexity)

- 有界部分集合  $\mathcal{A} \in \mathbb{R}^n$  は所与とし, 確率変数  $Z = \sup_{a \in \mathcal{A}} \sum_{k=1}^n a_k \epsilon_k$  を考える.
- $\epsilon_k \in \{-1, 1\}$  は i.i.d. な Rademacher variables.
- $Z$  は線形関数の  $\sup$  をとったものなので jointly (したがって separately) convex.
- 別の  $Z' = Z(\epsilon')$  に対し, 任意の  $a \in \mathcal{A}$  について

$$\langle a, \epsilon \rangle - Z' = \langle a, \epsilon \rangle - \sup_{a' \in \mathcal{A}} \langle a, \epsilon' \rangle \leq \langle a, \epsilon - \epsilon' \rangle \leq \|a\|_2 \|\epsilon - \epsilon'\|_2.$$

- $a \in \mathcal{A}$  について  $\sup$  をとると  $Z - Z' \leq \sup_{a \in \mathcal{A}} \|a\|_2 \|\epsilon - \epsilon'\|_2$ .
- よって,  $Z$  は  $\mathcal{W}(\mathcal{A}) := \sup_{a \in \mathcal{A}} \|a\|_2$ -Lipschitz.
- したがって, Theorem 3.4 より

$$\mathcal{P}[Z \leq \mathbb{E}[Z] + t] \leq \exp\left(-\frac{t^2}{16\mathcal{W}^2(\mathcal{A})}\right). \quad (3.17)$$

- 通常  $\mathcal{W}^2(\mathcal{A})$  は  $\sum_{k=1}^n \sup a_k^2$  よりずっと小さいので, Example 2.25 より強い bound となる.

## 3.1.4 Tensorization and separately convex functions

- 2つの lemma をもとに Theorem 3.4 を証明する.

### Lemma 3.7

$X, Y \sim \mathbb{P}$ , i.i.d. とすると, 任意の関数  $g : \mathbb{R} \rightarrow \mathbb{R}$  に対し以下が成り立つ.

$$\mathbb{H}(e^{\lambda g(X)}) \leq \lambda^2 \mathbb{E} \left[ (g(X) - g(Y))^2 e^{\lambda g(X)} \mathbb{I}[g(X) \geq g(Y)] \right] \quad \text{for all } \lambda > 0. \quad (3.20a)$$

さらに  $X$  のサポートが  $[a, b]$  に含まれ,  $g$  が凸かつ Lipschitz なら,

$$\mathbb{H}(e^{\lambda g(X)}) \leq \lambda^2 (b - a)^2 \mathbb{E} \left[ (g'(X))^2 e^{\lambda g(X)} \right] \quad \text{for all } \lambda > 0, \quad (3.20b)$$

が成り立つ ( $g'$  は  $g$  の導関数).



- Lemma 3.7 は, 凸かつ Lipschitz な関数はほとんどいたるところ微分可能であるという事実を使っている (Rademacher's Theorem).
- また,  $g$  が  $L$ -Lipschitz なら  $\|g'\|_\infty \leq L^1$  なので, (3.20b) は以下を含意する.

$$\mathbb{H}(e^{\lambda g(X)}) \leq \lambda^2 L^2 (b-a)^2 \mathbb{E}[e^{\lambda g(X)}] \quad \text{for all } \lambda > 0.$$

- したがって Proposition 3.2 より

$$\mathbb{P}[g(X) \geq \mathbb{E}[g(X)] + \delta] \leq e^{-\frac{\delta^2}{4L^2(b-a)^2}}$$

となるので, Lemma 3.7 は Theorem 3.4 の 1 変数バージョンをただちに導く.

- しかし (3.20b) では  $L$  でなく  $g'$  とより強い bound となっており, これが Theorem 3.4 の導出において重要となる.

---

<sup>1</sup>関数  $f$  に対して  $\|f\|_\infty$  は  $f$  の本質的上限  $\|f\|_\infty := \inf\{C \geq 0 : |f(x)| \leq C \text{ almost every } x.\}$ .

## Proof of Lemma 3.7

- エントロピーの定義より,

$$\begin{aligned}\mathbb{H}(e^{\lambda g(X)}) &= \mathbb{E}_X[\lambda g(X)e^{\lambda g(X)}] - \mathbb{E}_X[e^{\lambda g(X)}] \log(\mathbb{E}_Y[e^{\lambda g(Y)}]) \\ &\leq \mathbb{E}_X[\lambda g(X)e^{\lambda g(X)}] - \mathbb{E}_{X,Y}[e^{\lambda g(X)} \lambda g(Y)] \quad (\text{Jensen's inequality}) \\ &= \frac{1}{2} \mathbb{E}_{X,Y} \left[ \lambda \{g(X) - g(Y)\} \{e^{\lambda g(X)} - e^{\lambda g(Y)}\} \right] \\ &= \lambda \mathbb{E} \left[ \{g(X) - g(Y)\} \{e^{\lambda g(X)} - e^{\lambda g(Y)}\} \mathbb{I}[g(X) \geq g(Y)] \right]. \quad (3.22)\end{aligned}$$

となる (ただし最後の等式は  $X, Y$  の対称性より).

- 指数関数の凸性より, 任意の  $s \geq t$  に対し  $e^s - e^t \leq e^s(s - t)$  なので,

$$(s - t)(e^s - e^t)\mathbb{I}[s \geq t] \leq (s - t)^2 e^s \mathbb{I}[s \geq t]$$

- これを (3.22) に適用すると, (3.20a)

$$\mathbb{H}(e^{\lambda g(X)}) \leq \lambda^2 \mathbb{E}[(g(X) - g(Y))^2 e^{\lambda g(X)} \mathbb{I}[g(X) \geq g(Y)]]. \quad (3.23)$$

が得られる.

- さらに  $g$  が凸で  $x, y \in [a, b]$  とすると,  $g(x) - g(y) \leq g'(x)(x - y)$  より,

$$(g(x) - g(y))^2 \mathbb{I}[g(x) \geq g(y)] \leq (g'(x))^2 (x - y)^2 \leq (g'(x))^2 (b - a)^2$$

となるので, これを (3.23) に適用すれば (3.20a) が得られる. □