Chapter Four : Chernoff and Hoeffding Bounds 発表レジュメ

當眞 ジェイソン翔

2018 年 4 月 23 日

この章では、確率変数の偏差 (平均 $\mathrm{E}[X]$ との差) を評価する不等式として Chernoff 上界と Hoeffding 上界を与える。これらは tail distribution (分布の裾) で起こる事象の確率を指数的に減少する関数で抑えることが出来る強力な不等式である。用いるのは、第 3 章で定義された Markov の不等式と、この章で説明するモーメント母関数である。

1 モーメント母関数の定義とその性質

定義 **1.1.** (モーメント母関数)

確率変数 X のモーメント母関数を $M_X(t) := \mathrm{E}[e^{tX}]$ で定義する。

定理 **1.2.**

X : 確率変数、$M_X(t)$: X のモーメント母関数とする。$^\forall n > 1$ で期待値と微分記号の交換が可能であると仮定すると、

$$\mathrm{E}[X^n] = M_X^{(n)}(0)$$

が成立する。

【証明】

仮定のもとで $M_X^{(n)}(t) = \dfrac{d^n}{dt^n}\mathrm{E}[e^{tX}] = \mathrm{E}[X^n e^{tX}]$。$t = 0$ を代入して与式を得る。■

注意 **1.3.**

モーメント母関数が $t = 0$ の近傍で存在すれば、期待値と微分記号の入れ替えが可能である。この本に出てくるすべての分布について、$t = 0$ の近傍でモーメント母関数が存在する。

例 **1.4.** (幾何分布)

$X \sim \mathrm{Ge}(p)$ のとき、$t < -\ln(1-p)$ であれば

$$\begin{aligned} M_X(t) &= \mathrm{E}[e^{tX}] \\ &= \sum_{k=1}^{\infty}(1-p)^{k-1}pe^{tk} \\ &= \frac{p}{1-p}\sum_{k=1}^{\infty}\{(1-p)e^t\}^k \\ &= \frac{p}{1-p}\cdot\frac{(1-p)e^t}{1-(1-p)e^t} = \frac{p}{1-p}\left(\frac{1}{1-(1-p)e^t} - 1\right) \end{aligned}$$

となる。

$M_X^{(1)}(t) = \dfrac{p}{1-p}\cdot\dfrac{(1-p)e^t}{\{1-(1-p)e^t\}^2}$ なので、$\mathrm{E}[X] = M_X^{(1)}(0) = \dfrac{p}{1-p}\cdot\dfrac{1-p}{p^2} = \dfrac{1}{p}$ である。

同様に計算することで、モーメント母関数を用いて高次のモーメントを計算することが可能である。

定理 1.5. (モーメント母関数の一意性)

X, Y：確率変数、$M_X(t), M_Y(t)$：X, Y のモーメント母関数とする。このとき、ある $\delta > 0$ が存在し、$^\forall t \in (-\delta, \delta)$ で $M_X(t) = M_Y(t)$ が成立するとき、この 2 つの確率変数が従う確率分布は一致する。

本には、証明は "beyond the scope of this book" と書かれているが、せっかくなので証明を与える。証明の方針としては、特性関数 $\varphi_X(t) = \mathrm{E}[e^{itX}]$ が存在し、任意の $t \in \mathbb{R}$ で $\varphi_X(t) = \varphi_Y(t)$ となることを言い、特性関数の一意性を用いるというものである。証明を書くにあたって、[2] を参考にした。

【証明】

まず、ある $r > 0$ が存在して、$^\forall t \in (-r, r)$ で $\dfrac{\mathrm{E}[|X|^k]}{k!} t^k \to 0 (k \to \infty) \cdots (*)$ が成立することを示す。

仮定より、$0 < s < \min(1, \delta)$ を満たす s のうち $\dfrac{\mathrm{E}[X^k]}{k!} s^k \to 0 (k \to \infty)$ を満たすものが存在する。このとき、r として $0 < r < s$ を満たす実数を選ぶ。すると、十分大きな k で $2kr^{2k-1} < s^{2k} \cdots (**)$ が成立する。また、任意の $x \in \mathbb{R}$ および $k \in \mathbb{N}$ について、不等式 $|x|^{2k-1} \leq 1 + |x|^{2k}$ が成立する。この不等式の両辺の期待値をとり、その後に不等式 $(**)$ を用いることで、十分大きな k について

$$
\begin{aligned}
\frac{\mathrm{E}[|X|^{2k-1}] r^{2k-1}}{(2k-1)!} &\leq \frac{r^{2k-1}}{(2k-1)!} + \frac{\mathrm{E}[|X|^{2k}] r^{2k-1}}{(2k-1)!} \\
&\leq \frac{r^{2k-1}}{(2k-1)!} + \frac{\mathrm{E}[|X|^{2k}] s^{2k}}{(2k)!}
\end{aligned}
$$

が成立することが言える。よって、奇数次の項は $k \to \infty$ の極限で 0 に収束する。また、偶数次の項も $\mathrm{E}[|X|^{2k}] = \mathrm{E}[X^{2k}]$ であることから、0 に収束する。よって、このように定めた r について $(*)$ が成立する。

次に、不等式 $\left| e^{itx} \left(e^{ihx} - \displaystyle\sum_{k=0}^{n} \frac{(ihx)^k}{k!} \right) \right| \leq \dfrac{|hx|^{n+1}}{(n+1)!}$ を示す。

部分積分の公式より、

$$
\begin{aligned}
\int_0^x (x-s)^n e^{ihs} ds &= \left[-\frac{(x-s)^{n+1}}{n+1} e^{ihs} \right]_0^x + \frac{ih}{n+1} \int_0^x (x-s)^{n+1} e^{ihs} ds \\
&= \frac{x^{n+1}}{n+1} + \frac{ih}{n+1} \int_0^x (x-s)^{n+1} e^{ihs} ds
\end{aligned}
$$

が成立する。この関係を用いることで、帰納法より次式が成立することが言える。

$$
e^{ihx} = \sum_{k=0}^{n} \frac{(ihx)^k}{k!} + \frac{(ih)^{n+1}}{n!} \int_0^x (x-s)^n e^{is} ds
$$

よって、

$$
\begin{aligned}
\left| e^{itx} \left(e^{ihx} - \sum_{k=0}^{n} \frac{(ihx)^k}{k!} \right) \right| &= \left| e^{ihx} - \sum_{k=0}^{n} \frac{(ihx)^k}{k!} \right| \\
&= \left| \frac{(ih)^{n+1}}{n!} \int_0^x (x-s)^n e^{is} ds \right| \\
&\leq \frac{|h|^{n+1}}{n!} \int_0^x |x-s|^n ds = \frac{|hx|^{n+1}}{(n+1)!}
\end{aligned}
$$

となる。

この式の両辺の期待値を取ると、

$$\mathrm{E}\left[\left|e^{itX}\left(e^{ihX}-\sum_{k=0}^{n}\frac{(ihX)^k}{k!}\right)\right|\right]\leq\frac{|h|^{n+1}\mathrm{E}[|X|^{n+1}]}{(n+1)!}$$

となる。関数 $f(x)=|x|$ は凸関数なので、Jensen の不等式を用いることで

$$\left|\mathrm{E}\left[e^{itX}\left(e^{ihX}-\sum_{k=0}^{n}\frac{(ihX)^k}{k!}\right)\right]\right|\leq\frac{|h|^{n+1}\mathrm{E}[|X|^{n+1}]}{(n+1)!}$$

が導かれる。ここで X の特性関数を $\varphi_X(t)$ とすると、上の不等式は

$$\left|\mathrm{E}\left[e^{i(t+h)X}-\sum_{k=0}^{n}\frac{h^k}{k!}(iX)^k e^{itX}\right]\right|=\left|\varphi_X(t+h)-\sum_{k=0}^{n}\frac{\varphi_X^{(k)}(t)}{k!}h^k\right|\leq\frac{|h|^{n+1}\mathrm{E}[|X|^{n+1}]}{(n+1)!}$$

と整理される。最初に示したように、 $|h|\leq r\Longrightarrow\dfrac{|h|^{n+1}\mathrm{E}[|X|^{n+1}]}{(n+1)!}\to 0(n\to\infty)$ が成立するので、

$$\varphi_X(t+h)=\sum_{k=0}^{\infty}\frac{\varphi_X^{(k)}(t)}{k!}h^k,\ |h|\leq r$$

が成立する。これまでの議論を同様に適用することで、

$$\varphi_Y(t+h)=\sum_{k=0}^{\infty}\frac{\varphi_Y^{(k)}(t)}{k!}h^k,\ |h|\leq r$$

も成立する。

モーメント母関数 $M_X(t),M_Y(t)$ が ${}^\forall t\in(-r,r)$ で一致するので、任意の $k\in\mathbb{N}$ について $\mathrm{E}[X^k]=\mathrm{E}[Y^k]$ となることに注意する。

上記のテイラー展開の式で $t=0$ とすると、$\varphi_X^{(k)}(0)=i^k\mathrm{E}[X^k]=i^k\mathrm{E}[Y^k]=\varphi_Y^{(k)}(0)$ となることから、$(-r,r)$ で $\varphi_X\equiv\varphi_Y$ となる。次に $t=-r+\varepsilon$ 及び $t=r-\varepsilon$ とすることで、$(-2r+\varepsilon,2r-\varepsilon)$ で $\varphi_X\equiv\varphi_Y$ になることが言える。$\varepsilon>0$ は任意なので、結局これは $(-2r,2r)$ で $\varphi_X\equiv\varphi_Y$ であることを示している。同様に $(-3r,3r)$ でも $\varphi_X\equiv\varphi_Y$ であることがいえ、結局任意の $t\in\mathbb{R}$ で $\varphi_X(t)=\varphi_Y(t)$ であることが言える。あとは特性関数の一意性[*1]より X と Y の従う確率分布が等しいことが従う。■

定理 1.6.

X,Y : 独立な確率変数のとき、$M_{X+Y}(t)=M_X(t)M_Y(t)$ が成立する。

【証明】

$$M_{X+Y}(t)=\mathrm{E}[e^{t(X+Y)}]=\mathrm{E}[e^{tX}\cdot e^{tY}]=\mathrm{E}[e^{tX}]\mathrm{E}[e^{tY}]=M_X(t)M_Y(t)$$

ここで、三つ目の等号で X と Y が独立であるとき e^{tX} と e^{tY} が独立であるということを用いた。■

[*1] 確率数理工学で扱われていたように、特性関数の一意性の証明には Levy の反転公式を使うが、実質的に行っているのは逆 Fourier 変換である。

2　Chernoff 上界

Chernoff 上界とは、Markov の不等式に e^{tX} を代入して得られる形の不等式である。Markov の不等式より、$t > 0$ のとき、関数 e^{tx} が単調増加になることから

$$\Pr(X \geq a) = \Pr(e^{tX} \geq e^{ta}) \leq \frac{\mathrm{E}[e^{tX}]}{e^{ta}}$$

が言える。これより

$$\Pr(X \geq a) \leq \min_{t>0} \frac{\mathrm{E}[e^{tX}]}{e^{ta}}$$

となる。同様に、$t < 0$ のとき、関数 e^{tx} が単調減少になることから

$$\Pr(X \leq a) = \Pr(e^{tX} \geq e^{ta}) \leq \frac{\mathrm{E}[e^{tX}]}{e^{ta}}$$

が言える。これより

$$\Pr(X \leq a) \leq \min_{t<0} \frac{\mathrm{E}[e^{tX}]}{e^{ta}}$$

が言える。

Chernoff 上界は、この形の不等式に、うまく t を選ぶことで導かれる。逆に、このような形で評価される上界はすべて Chernoff 上界と呼ばれる。

注意 **2.1.** 上記の Markov の不等式の右辺にはモーメント母関数 $M_X(t)$ が現れていることに注意する。

例 **2.2.** (Poisson 試行の和の Chernoff 上界)

n 個の独立な確率変数 X_i があり、各 X_i は確率 p_i で 1、確率 $1 - p_i$ で値 0 を取るものとする。このとき、X_i の和 $X = \displaystyle\sum_{i=1}^{n} X_i$ の tail distribution を評価したい。すなわち、$\mu = \mathrm{E}[X_i] = \displaystyle\sum_{i=1}^{n} p_i$ として、$\delta > 0$ に対して $\Pr(X \geq (1+\delta)\mu)$ 及び $\Pr(X \leq (1-\delta)\mu)$ を評価したい。これを Chernoff 上界を使って評価してみよう。

まず、モーメント母関数をそれぞれの X_i について求め、それを上から評価する。

$$\begin{aligned} M_{X_i}(t) = \mathrm{E}[e^{tX_i}] &= p_i e^{t\cdot 1} + (1-p_i)e^{t\cdot 0} \\ &= 1 + p_i(e^t - 1) \\ &\leq e^{p_i(e^t-1)} \end{aligned}$$

最後の変形では、${}^\forall x \in \mathbb{R}$ で成立する不等式 $1 + x \leq e^x$ を用いた。

次に、この評価を用いて X のモーメント母関数を上から評価する。各 X_i が独立であるから、

$$\begin{aligned} M_X(t) = \prod_{i=1}^{n} M_{X_i}(t) &\leq \prod_{i=1}^{n} e^{p_i(e^t-1)} \\ &= e^{\sum_{i=1}^{n} p_i(e^t-1)} = e^{(e^t-1)\mu} \end{aligned}$$

この X のモーメント母関数の評価をもとに、Poisson 試行の和の Chernoff 上界を次のように構成する。

定理 2.3.

$X_1, \cdots, X_n$: 独立な Poisson 試行に従う確率変数で、$\Pr(X_i = 1) = p_i$ とする。$X = \sum_{i=1}^{n} X_i,\ \mu = \mathrm{E}[X]$ とするとき、次の Chernoff 上界が成立する。

1. ${}^\forall\delta > 0$ に対し、
$$\Pr(X \geq (1+\delta)\mu) \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$$

2. $0 <^\forall \delta \leq 1$ に対し
$$\Pr(X \geq (1+\delta)\mu) \leq e^{-\mu\delta^2/3}$$

3. $R \geq 6\mu \Longrightarrow \Pr(X \geq R) \leq 2^{-R}$

【証明】

1. Markov の不等式より、任意の $t > 0$ に対し
$$\begin{aligned}\Pr(X \geq (1+\delta)\mu) &= \Pr(e^{tX} \geq e^{t(1+\delta)\mu}) \\ &\leq \frac{\mathrm{E}[e^{tX}]}{e^{t(1+\delta)\mu}} \\ &\leq \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} = \left(\frac{e^{(e^t-1)}}{e^{t(1+\delta)}}\right)^\mu\end{aligned}$$
${}^\forall\delta > 0$ に対し、$t = \ln(1+\delta) > 0$ とすることで
$$\Pr(X \geq (1+\delta)\mu) \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$$

2. $0 <^\forall \delta \leq 1$ に対し $\dfrac{e^\delta}{(1+\delta)^{(1+\delta)}} \leq e^{-\delta^2/3}$ を示せば良い。それは、関数
$$f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{3}$$
が $(0,1]$ で 0 以下であることを示せばただちに従う。
$$f'(\delta) = -\ln(1+\delta) + \frac{2}{3}\delta,\ f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3}$$
であり、$f'(\delta)$ が $0 < \delta < 1/2$ で単調減少、$\delta > 1/2$ で単調増加していることがわかる。$f'(0) = 0, f'(1) = -\ln 2 + \dfrac{2}{3} < 0$ なので、結局 $(0,1]$ で $f'(\delta) < 0$ である。 $f(0) = 0$ であるから、結局 $(0,1]$ で $f(\delta) < 0$ である。

3. $R = (1+\delta)\mu$ とおく。このとき、 $R \geq 6\mu \iff \delta = R/\mu - 1 \geq 5$。このとき、1. を用いれば
$$\begin{aligned}\Pr(X \geq (1+\delta)\mu) &\leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu \\ &\leq \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} \\ &\leq \left(\frac{e}{6}\right)^R \\ &\leq 2^{-R}\end{aligned}$$
二つ目の不等号では、$\mu \geq 0$ と $e^\mu \geq 1$ であることを用いた。■

反対向きの偏差についての Chernoff 上界も、同様に示される。

定理 **2.4.**

$X_1, \cdots, X_n$: 独立な Poisson 試行に従う確率変数で、$\Pr(X_i = 1) = p_i$ とする。$X = \sum_{i=1}^{n} X_i,\ \mu = \mathrm{E}[X]$ とするとき、$0 <^{\forall} \delta < 1$ に対し次の Chernoff 上界が成立する。

1.
$$\Pr(X \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$$

2.
$$\Pr(X \leq (1-\delta)\mu) \leq e^{-\mu\delta^2/2}$$

両方向からの評価を合わせることで、次の評価が得られる。

系 **2.5.**

$X_1, \cdots, X_n$: 独立な Poisson 試行に従う確率変数で、$\Pr(X_i = 1) = p_i$ とする。$X = \sum_{i=1}^{n} X_i,\ \mu = \mathrm{E}[X]$ とするとき、$0 <^{\forall} \delta < 1$ に対して次の Chernoff 上界が成立する。

$$\Pr(|X - \mu| \geq \delta\mu) \leq 2e^{-\mu\delta^2/3}$$

例 **2.6.** (コイン投げ)

表と裏がそれぞれ確率 1/2 で出るコインがあるとする。X を、このコインを n 回投げたときに表が出る回数を表す確率変数とする。それぞれのコイン投げの試行は独立であるとする。このとき、確率変数 X は、$p = 1/2$ の Bernoulli 試行を n 回繰り返したものと見ることが出来る。よって、系 2.5 を用いると

$$\Pr\left(\left|X - \frac{n}{2}\right| \geq \frac{1}{2}\sqrt{6n\ln n}\right) \leq 2\exp\left(-\frac{1}{3}\cdot\frac{n}{2}\cdot\frac{6\ln n}{n}\right) = \frac{2}{n}$$

という評価ができる。これは、コイン投げで表が出る回数が $n/2$ の付近に集中していることを表している。つまり、ほとんどの場合、表が出る回数の $n/2$ に対する偏差が $O(\sqrt{n\ln n})$ であるということを示している。

Chernoff 上界が Chebyshev の不等式によって得られる上界に比べて、どれほど強い評価を与えているのかを確かめてみよう。そのために、コイン投げでの偏差が $n/4$ 以上となる確率

$$\Pr\left(\left|X - \frac{n}{2}\right| \geq \frac{n}{4}\right)$$

を両方の不等式で評価してみる。 Chebyshev の不等式を用いると

$$\Pr\left(\left|X - \frac{n}{2}\right| \geq \frac{n}{4}\right) \leq \frac{4}{n}$$

となる。一方、Chernoff 上界を用いると

$$\Pr\left(\left|X - \frac{n}{2}\right| \geq \frac{n}{4}\right) \leq 2\exp\left(-\frac{1}{3}\cdot\frac{n}{2}\cdot\frac{1}{2}\right) = 2e^{-n/12}$$

という評価ができる。よって、この場合は Chernoff 上界を用いたほうが Chebyshev の不等式を用いたときよりも良いオーダーの評価を与える。

例 **2.7.** (信頼区間の設定)

n 個の確率変数 X_i がベルヌーイ分布 Bernoulli(p) に独立に従っているとする。今、p が未知であるとして、X_i から p を推定したい。$\tilde{p} = \sum_{i=1}^{n} X_i/n$ とし、p を $\tilde{p}$ で推定することを考える。すると、$\tilde{p}$ は n を大きくすれば p に近づくことが期待される。これを信頼区間という概念を用いて定式化しよう。

定義 2.8. (信頼区間)

パラメータ p の $1-\gamma$ 信頼区間とは、ある実数 $\gamma > 0$ を用いて表せる閉区間 $[\tilde{p}-\delta, \tilde{p}+\delta]$ であって、

$$\Pr(p \in [\tilde{p}-\delta, \tilde{p}+\delta]) \geq 1-\gamma$$

を満たすものである。

推定という観点から、δ も γ も小さく取りたい。所望の δ, γ のためにどれくらいの標本データが必要かを検討する必要がある。そこで、δ, γ, n の間の関係式を導きたい。

$Y = \tilde{p}n$ とすると、$Y \sim \mathrm{B}(n,p)$ であることに注意する。ゆえに $\mathrm{E}[Y] = np$ である。このとき、

$$\begin{aligned}
\Pr(p \notin [\tilde{p}-\delta, \tilde{p}+\delta]) &= \Pr(np \notin [n(\tilde{p}-\delta), n(\tilde{p}+\delta)]) \\
&= \Pr\left(Y < np\left(1-\frac{\delta}{p}\right)\right) + \Pr\left(Y > np\left(1+\frac{\delta}{p}\right)\right) \\
&< \exp\left(-np\left(\frac{\delta}{p}\right)^2 \frac{1}{2}\right) + \exp\left(-np\left(\frac{\delta}{p}\right)^2 \frac{1}{3}\right) \\
&= e^{-n\delta^2/(2p)} + e^{-n\delta^2/(3p)} \\
&< e^{-n\delta^2/2} + e^{-n\delta^2/3}
\end{aligned}$$

最後の変形には $p \leq 1$ であることを用いた。これより、$\gamma = e^{-n\delta^2/2} + e^{-n\delta^2/3}$ とすることで、$[\tilde{p}-\delta, \tilde{p}+\delta]$ は p の $1-(e^{-n\delta^2/2} + e^{-n\delta^2/3})$ 信頼区間であることが導かれた。

対称性のある確率変数に対する Chernoff 上界

定理 2.9.

$X_1, \cdots, X_n$: 独立な確率変数で、

$$\Pr(X_i = 1) = \Pr(X_i = -1) = \frac{1}{2}$$

であるものとする。 $X = \sum_{i=1}^{n} X_i$ とするとき、任意の $a > 0$ に対して

$$\Pr(X \geq a) \leq e^{-a^2/(2n)}$$

【証明】

任意の $t > 0$ に対して、 $M_{X_i}(t) = \frac{1}{2}e^t + \frac{1}{2}e^{-t}$ である。 e^t と e^{-t} の Taylor 展開を足し合わせることで、

$$M_{X_i}(t) = \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} \leq \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!!} = \sum_{k=0}^{\infty} \frac{(t^2/2)^k}{k!} = e^{t^2/2}$$

よって、 $M_X(t) = \prod_{i=1}^{n} \mathrm{E}[e^{tX_i}] \leq e^{t^2 n/2}$。

そこで Markov の不等式を用いると、

$$\Pr(X \geq a) = \frac{\mathrm{E}[e^{tX}]}{e^{ta}} \leq e^{t^2 n/2 - ta}$$

という評価を得る。$t = a/n$ とすることで、

$$\Pr(X \geq a) \leq e^{-a^2/(2n)}$$

を得る。■

対称性より $\Pr(X \leq -a) \leq e^{-a^2/(2n)}$ も従う。この 2 つの評価を合わせることで次の系が従う。

系 **2.10.**

$X_1, \cdots, X_n$: 独立な確率変数で、

$$\Pr(X_i = 1) = \Pr(X_i = -1) = \frac{1}{2}$$

であるものとする。 $X = \sum_{i=1}^{n} X_i$ とするとき、任意の $a > 0$ に対して

$$\Pr(|X| \geq a) \leq 2e^{-a^2/(2n)}$$

$Y_i = (X_i + 1)/2$ という線形変換を施すことで次の系が従う。

系 **2.11.**

$Y_1, \cdots, Y_n$: 独立な確率変数で、

$$\Pr(Y_i = 1) = \Pr(X_i = 0) = \frac{1}{2}$$

であるものとする。 $Y = \sum_{i=1}^{n} Y_i, \mu = \mathrm{E}[Y] = n/2$ とするとき、

1. 任意の $a > 0$ に対して

$$\Pr(Y \geq \mu + a) \leq e^{-2a^2/n}$$

2. 任意の $\delta > 0$ に対して

$$\Pr(Y \geq (1 + \delta)\mu) \leq e^{-\delta^2 \mu}$$

【証明】

1. 定理 2.9 と同じ表記を用いると、$Y = \sum_{i=1}^{n} Y_i = \frac{1}{2}\left(\sum_{i=1}^{n} X_i\right) + \frac{n}{2} = \frac{1}{2}X + \mu$ となる。定理 2.9 より

$$\Pr(Y \geq \mu + a) = \Pr(X \geq 2a) \leq e^{-4a^2/(2n)} = e^{-2a^2/n}$$

2. 1. で $a = \delta\mu$ とすればよい。■

ここで、系 4.10 の 2. の評価は、定理 2.3 の 2. の評価の指数部の 1/3 の部分が 1 になっているため、より強い評価であることに注意せよ。

逆側についても同様に評価することが出来る。

系 **2.12.**

$Y_1, \cdots, Y_n$: 独立な確率変数で、

$$\Pr(Y_i = 1) = \Pr(X_i = 0) = \frac{1}{2}$$

であるものとする。 $Y = \sum_{i=1}^{n} Y_i, \mu = \mathrm{E}[Y] = n/2$ とするとき、

1. 任意の $0 < a < \mu$ に対して

$$\Pr(Y \leq \mu - a) \leq e^{-2a^2/n}$$

2. 任意の $0 < \delta < 1$ に対して

$$\Pr(Y \leq (1 - \delta)\mu) \leq e^{-\delta^2 \mu}$$

例 **2.13.** (Set Balancing Problem)

$A=(a_{ij})$ を、すべての成分が $\{0,1\}$ のいずれかである $n\times m$ 行列とする。いま、

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

とする。$\boldsymbol{b}=(b_i)$ のすべての成分が $\{-1,1\}$ のいずれかであるときに、

$$\|A\boldsymbol{b}\|_\infty = \max_{i=1,\cdots,n} |c_i|$$

を最小化する $\boldsymbol{b}$ を求めたい。

驚くべきことに、$\boldsymbol{b}$ のそれぞれの要素を、$\Pr(b_i=-1)=\Pr(b_i=1)=1/2$ となるように独立に値を割り当てることで、ほとんどの場合 $\|A\boldsymbol{b}\|_\infty = O(\sqrt{m\ln n})$ となる。

定理 **2.14.**

$\boldsymbol{b}$ を、それぞれの成分が独立に $\{-1,1\}$ から等しい確率で選ばれたランダムベクトルとする。このとき、

$$\Pr(\|A\boldsymbol{b}\|_\infty \geq \sqrt{4m\ln n}) \leq \frac{2}{n}$$

【証明】

A の i 行目を $\boldsymbol{A}_i$ と表記する。$k=\displaystyle\sum_{j=1}^{m} A_{ij}$ とする。

もし $k\leq\sqrt{4m\ln n}$ であれば、明らかに $|\boldsymbol{A}_i\cdot\boldsymbol{b}|=|c_i|\leq\sqrt{4m\ln n}$ である。もし $k>\sqrt{4m\ln n}$ であれば、

$$Z_i=\sum_{j=1}^{m} a_{ij}b_j$$

の k 個の非ゼロ要素が独立かつ確率 $1/2$ で値として 1 か -1 をとる確率変数であることに注意する。

そこで系 2.10 を用いると、

$$\Pr(|Z_i|>\sqrt{4m\ln n}) \leq 2e^{-4m\ln n/(2k)} \leq 2e^{-2\ln n} = \frac{2}{n^2}$$

という評価が得られる。最後の不等号には $m\geq k$ であるということを用いた。

よって、

$$\begin{aligned} \Pr(\|A\boldsymbol{b}\|_\infty \geq \sqrt{4m\ln n}) &= \Pr\left(\bigcup_{i=1}^{n} |Z_i| \geq \sqrt{4m\ln n}\right) \\ &\leq \sum_{i=1}^{n} \Pr(|Z_i| \geq \sqrt{4m\ln n}) \\ &\leq n\cdot\frac{2}{n^2} = \frac{2}{n} \end{aligned}$$

が成立する。■

3 Hoeffding の不等式

Hoeffding の不等式は、Chernoff 上界による評価を、値域が有限な確率変数について与えるものである。証明は確率数理工学の授業で行っているため、今回はそこで紹介されていた方法で証明を行う。

定理 3.1. (Hoeffding の不等式)
$X_1, \cdots, X_n$: 独立な確率変数で、 $\mathrm{E}[X_i] = \mu_i,\ \Pr(a_i \leq X_i \leq b_i) = 1$ であるとする。このとき、

$$\Pr\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i \geq \varepsilon\right) \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

【証明】

$Z_i = X_i - \mu_i, \alpha_i = a_i - \mu_i, \beta_i = b_i - \mu_i$ とする。すると、任意の $t > 0$ に対して、Markov の不等式より

$$\begin{aligned}
\Pr\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i \geq \varepsilon\right) &= \Pr\left(\sum_{i=1}^n Z_i \geq \varepsilon\right) \\
&= \Pr(e^{t\sum_{i=1}^n Z_i} \geq e^{t\varepsilon}) \\
&\leq \frac{\mathrm{E}[e^{t\sum_{i=1}^n Z_i}]}{e^{t\varepsilon}} \\
&\leq \left(\prod_{i=1}^n \mathrm{E}[e^{tZ_i}]\right) \cdot e^{-t\varepsilon}
\end{aligned}$$

$\mathrm{E}[e^{tZ_i}]$ を評価していく。
$\alpha_i \leq Z_i \leq \beta_i$ より、 $\lambda = \dfrac{Z_i - \alpha_i}{\beta_i - \alpha_i}$ とおくと、$0 \leq \lambda \leq 1$ であって、$Z_i = \lambda\beta_i + (1-\lambda)\alpha_i$ と書ける。
e^{tx} の凸性より

$$e^{tZ_i} \leq \lambda e^{t\beta_i} + (1-\lambda)e^{t\alpha_i}$$

が成立し、両辺の期待値をとると、

$$\begin{aligned}
\mathrm{E}[e^{tZ_i}] &\leq \mathrm{E}[\lambda]e^{t\beta_i} + \mathrm{E}[1-\lambda]e^{t\alpha_i} \\
&= -\frac{\alpha_i}{\beta_i - \alpha_i}e^{t\beta_i} + \frac{\beta_i}{\beta_i - \alpha_i}e^{t\alpha_i} \\
&= pe^{t\alpha_i} + (1-p)e^{t\beta_i}
\end{aligned}$$

最後の式で $p = \beta_i/(\beta_i - \alpha_i)$ と置いた。
ここで

$$\psi(t) = \ln \mathrm{E}[e^{tX_i}] = \ln(pe^{t\alpha_i} + (1-p)e^{t\beta_i})$$

とすると、

$$\begin{aligned}
\psi(0) &= \log(p + (1-p)) = 0 \\
\psi'(0) &= \left.\frac{p\alpha_i e^{t\alpha_i} + (1-p)\beta_i e^{t\beta_i}}{pe^{t\alpha_i} + (1-p)e^{t\beta_i}}\right|_{t=0} \\
&= 0(p \text{ の定義より}) \\
\psi''(t) &= \frac{p\alpha_i^2 e^{t\alpha_i} + (1-p)\beta_i^2 e^{t\beta_i}}{pe^{t\alpha_i} + (1-p)e^{t\beta_i}} - \left(\frac{p\alpha_i e^{t\alpha_i} + (1-p)\beta_i e^{t\beta_i}}{pe^{t\alpha_i} + (1-p)e^{t\beta_i}}\right)^2 \\
&= q\alpha_i^2 + (1-q)\beta_i^2 - (q\alpha_i + (1-q)\beta_i)^2 \\
&= \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 = \mathrm{V}[Y]
\end{aligned}$$

ただし $q = pe^{t\alpha_i}/(pe^{t\alpha_i} + (1-p)e^{t\beta_i})$、$Z$ は $\Pr(Y = \alpha_i) = q, \Pr(Y = \beta_i) = 1 - q$ である確率変数である。

$\alpha_i \le Y \le \beta_i$ より

$$\mathrm{V}[Y] = \mathrm{E}[(Y - \mathrm{E}[Y])^2] \le \mathrm{E}\left[\left(Y - \frac{\alpha_i + \beta_i}{2}\right)^2\right] = \frac{(\beta_i - \alpha_i)^2}{4} = \frac{(b_i - a_i)^2}{4}$$

が成立する。よって、Taylor の定理を用いることで、

$$\psi(t) = \psi(0) + t\psi'(0) + \frac{t^2}{2}\psi''(\tilde{t})$$

となる $0 \le \tilde{t} \le t$ の存在が言える。上記の不等式と合わせて、

$$\psi(t) \le \frac{(b_i - a_i)^2}{8}t^2$$

が言える。よって、

$$\mathrm{E}[e^{tZ_i}] = e^{\psi(t)} \le \exp\left(\frac{(b_i - a_i)^2}{8}t^2\right)$$

が成立し、最初の Markov の不等式での評価と組み合わせることで

$$\Pr\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i \ge \varepsilon\right) \le \exp\left(\frac{\sum_{i=1}^{n}(b_i - a_i)^2}{8}t^2 - t\varepsilon\right)$$

という評価を得る。

$$f(t) = \frac{\sum_{i=1}^{n}(b_i - a_i)^2}{8}t^2 - t\varepsilon$$

を考えると、$f(t)$ は $t = \dfrac{4\varepsilon}{\sum_{i=1}^{n}(b_i - a_i)^2}$ で最小値をとり、その値は $-\dfrac{2\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}$ である。

上不等式で $t = \dfrac{4\varepsilon}{\sum_{i=1}^{n}(b_i - a_i)^2}$ とすることで

$$\Pr\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i \ge \varepsilon\right) \le \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

を得る。■

上記の証明と全く同じ議論で

$$\Pr\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i \le -\varepsilon\right) \le \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

となることも示される。この 2 つの不等式の評価を足し合わせることで、

$$\Pr\left(\left|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right| \ge \varepsilon\right) \le 2\exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

が言える。

例 **3.2.**

1. n 個の確率変数 X_i が独立に一様分布 $\mathrm{U}[0, l]$ に従っているとする。このとき、$\mu = \mathrm{E}[X_i] = l/2$ で、Hoeffding の不等式より

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{l}{2}\right| \ge \varepsilon\right) \le 2e^{-2n\varepsilon^2/l^2}$$

   となる。$\varepsilon = \delta\mu$ とすると、以下が成立する。

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{l}{2}\right| \ge \delta\mu\right) \le 2e^{-n\delta^2/2}$$

2. n 個の確率変数 Y_i が独立に一様分布 $\mathrm{U}[0,i]$ に従っているとする。また、$Y = \sum_{i=1}^{n} Y_i$ とする。このとき、$\mathrm{E}[Y_i] = i/2$ で $\mu = \mathrm{E}[Y] = n(n+1)/4$ である。Hoeffding の不等式より

$$\begin{aligned}\Pr\left(\left|Y - \frac{n(n+1)}{4}\right| \geq \varepsilon\right) \leq 2e^{-2\varepsilon^2/\sum_{i=1}^{n} i^2} &= 2e^{-2\varepsilon^2/(n(n+1)(2n+1)/6)} \\ &= 2e^{-12\varepsilon^2/(n(n+1)(2n+1))}\end{aligned}$$

よって、$\varepsilon = \delta\mu$ として以下が成立する。

$$\Pr(|Y - \mu| \geq \delta\mu) \leq 2e^{-12\delta^2 n^2(n+1)^2/(16n(n+1)(2n+1))} \leq 2e^{-3n\delta^2/8}$$

参考文献

[1] Michael Mitzenmacher and Eli Upfal, “Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis”, Cambridge University Press, 2nd ed., 2017.

[2] Patrick Billingsley, “Probability and Measure”, Wiley-Interscience publication, anniversary ed., 2012.