Chebyshev の不等式と大数の法則

Th (Chebyshev's inequality)

X : r.v.

$E[X] = \mu$

$V[X] = \sigma^2 \ (\neq 0)$

$Z \equiv \dfrac{X-\mu}{\sigma}$

$\Rightarrow \forall \lambda > 0, \ P(|Z| \geqq \lambda) \leqq \dfrac{1}{\lambda^2}$

標準化された r.v.
Z について.

Proof (the case for X : conti.)

Let $\lambda > 0$.

It holds that

$$1 = V[Z] = E[Z^2]$$

$$= \int_{-\infty}^{\infty} z^2 f(z)\,dz \text{ , where } f \text{ is the p.d.f. for } Z$$

$$\geqq \int_{-\infty}^{-\lambda} z^2 f(z)\,dz + \int_{\lambda}^{\infty} z^2 f(z)\,dz$$

$$\geqq \int_{-\infty}^{-\lambda} \lambda^2 f(z)\,dz + \int_{\lambda}^{\infty} \lambda^2 f(z)\,dz$$

$$= \lambda^2 \left[\int_{-\infty}^{-\lambda} f(z)\,dz + \int_{\lambda}^{\infty} f(z)\,dz \right]$$

$$= \lambda^2 P(|Z| \geqq \lambda).$$

$$\therefore P(|Z| \geqq \lambda) \leqq \frac{1}{\lambda^2}.$$

//

- X はどのような分布をしていてもよい。

- $E[Z]=0$

- $P(|Z| \geqq \lambda) = P(Z \leqq -\lambda \text{ or } \lambda \leqq Z)$ は、

  Z が期待値である 0 から λ 以上ハズレる確率。

- ハズレ確率は 0 ではないが、$\dfrac{1}{\lambda^2}$ 以下に納まる。

  大きくハズレる確率は、かなり小さくなる。

  ($\lambda (>0)$ が大きいと、$1/\lambda^2$ は急速に小さくなる。)

- Cauchy 分布のように $E[X]$ や $V[X]$ が存在しない分布

  に対してはチェビシェフの不等式は適用できない。

  (この点については大数の法則や中心極限定理

  も同じ。)

Cor (Chebyshev's inequality)

X r.v.

$E[X] = \mu, \quad V[X] = \sigma^2 (\neq 0)$

$\Rightarrow \forall c > 0, \ P(|X-\mu| \geqq c) \leqq \dfrac{\sigma^2}{c^2}$

一般の r.v. について

Proof

Let $\begin{cases} c > 0 \\ \lambda = \dfrac{c}{\sigma} > 0 \end{cases}$

We know that

$$P\left(\left|\frac{X-\mu}{\sigma}\right| \geqq \lambda\right) \leqq \frac{1}{\lambda^2}.$$

$$\therefore P(|X-\mu| \geqq \lambda\sigma) \leqq \frac{1}{\lambda^2}$$

$$\therefore P(|X-\mu| \geqq c) \leqq \frac{\sigma^2}{c^2} \quad (\forall c > 0).$$

//

※ 大数の弱法則は、ここからすぐに導かれる。

$X_1, \cdots, X_n \sim iid(\mu, \sigma^2)$ のとき

$\overline{X}_n \equiv \dfrac{1}{n} \sum_{i=1}^{n} X_i$ とおくと

$E[\overline{X}_n] = \mu$

$V[\overline{X}_n] = \dfrac{\sigma^2}{n}$ である。

この $\overline{X}_n$ に対して、この Chebyshev's inequality を適用すればよい。

Cor

$X_1, \cdots, X_n \sim iid(\mu, \sigma^2)$ $(\sigma > 0)$

$S_n \equiv \sum_{i=1}^{n} X_i$

$\Rightarrow \forall c > 0,\ P(|S_n - n\mu| \geqq c) \leqq \dfrac{n\sigma^2}{c^2}$

Proof

Since $E[S_n] = n\mu$ and $V[S_n] = n\sigma^2$, ok. //

Cor

$X_1, \cdots, X_n \sim iid(\mu, \sigma^2)$ $(\sigma > 0)$

$S_n \equiv \sum_{i=1}^{n} X_i$

$\Rightarrow \forall c > 0,\ P\left(\left|\dfrac{S_n}{n} - \mu\right| \geqq c\right) \leqq \dfrac{\sigma^2/n}{c^2}$

Proof

Since $E\left[\dfrac{S_n}{n}\right] = \mu$ and $V\left[\dfrac{S_n}{n}\right] = \dfrac{\sigma^2}{n}$, ok. //

※ $\dfrac{S_n}{n} = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$.

例.（世論調査）

政党Aの
支持率 40%
国民

$$X_i = \begin{cases} 1 & \text{if } A\text{を支持} \\ 0 & \text{otherwise} \end{cases}$$

$$E[X_i] = 4/10$$

- $S_n \equiv \sum_{i=1}^{n} X_i$　n人に聞き取りしたときの政党Aの支持者数.
- $\dfrac{S_n}{n} \equiv \dfrac{1}{n}\sum_{i=1}^{n} X_i$　政党Aの支持率

$\boxed{\text{(1) } S_n \text{ と } \dfrac{S_n}{n} \text{ の期待値と分散を求めよ。}}$

$S_n \sim B\left(n, \frac{4}{10}\right)$ と考えられるので

$$E[S_n] = \frac{4}{10}n$$

$$V[S_n] = \frac{24}{100}n$$

従って.

$$E\left[\frac{S_n}{n}\right] = \frac{1}{n}E[S_n] = \frac{4}{10}$$

$$V\left[\frac{S_n}{n}\right] = \frac{1}{n^2}V[S_n] = \frac{24}{100}\cdot\frac{1}{n}$$

※ 聞き取り数 n を増やすと S_n の分散は大きくなるが.
政党Aの支持率 $\dfrac{S_n}{n}$ の分散は小さくなる。

(2) n人に聞き取りしたときの政党Aの支持率 $\frac{S_n}{n}$ が、その期待値 $4/10$ から $c\,(>0)$ 以上外れる確率の上限を(nに依存する形で)求めよ。

From Chebyshev's inequality,

$$P\left(\left|\frac{S_n}{n}-\frac{4}{10}\right|\geq c\right)\leq\frac{1}{c^2}V\left[\frac{S_n}{n}\right]=\frac{1}{c^2}\cdot\frac{24}{100}\cdot\frac{1}{n}.$$

(3) 400人に聞き取りしたときの政党Aの支持率 $\frac{S_{400}}{400}$ が、その期待値 $4/10$ から 5% 以上外れる確率は、せいぜい何%か?

$\left(\begin{array}{l} n=400 \\ c=5\%=\frac{1}{20} \end{array}\right.$ のケースなので、チェビシェフの不等式より

$$P\left(\left|\frac{S_{400}}{400}-\frac{4}{10}\right|\geq\frac{1}{20}\right)\leq\frac{1}{c^2}V\left(\frac{S_{400}}{400}\right)$$

$$=20^2\cdot\frac{24}{100}\cdot\frac{1}{400}=0.24.$$

よってせいぜい 24% である。

(4) 政党Aの支持率の調査結果が、その期待値 $4/10$ から 5%以上外れる確率を 10%以下にしたい。そのためには何人に聞き取りすればよいか？

n人に聞き取りするとすると、r.v. $\frac{S_n}{n}$ が

その期待値 $E\left[\frac{S_n}{n}\right] = \frac{4}{10}$ から 5%以上外れる

確率は Chebyshev's inequality

$$P\left(\left|\frac{S_n}{n} - \frac{4}{10}\right| \geq \frac{1}{20}\right) \leq \frac{1}{\left(\frac{1}{20}\right)^2} V\left[\frac{S_n}{n}\right]$$

$$= 400 \cdot \frac{24}{100} \cdot \frac{1}{n} = \frac{96}{n} \text{ より.}$$

$\frac{96}{n}$ 以下である。

この確率を 10%以下にするためには、

$$\frac{96}{n} \leq \frac{1}{10}$$

となるように n を決めればよい。

$$960 \leq n.$$

よって 960人以上に聞き取りすればよい。

例.

年間事故率 $\frac{1}{6}$
保険契約の母集団
互いに独立

$$X_i = \begin{cases} 1 & \text{if 1年以内に事故を起こす} \\ 0 & \text{if 無事故} \end{cases}$$

$$E[X_i] = \frac{1}{6}$$

$S_n = \sum_{i=1}^{n} X_i$　n件契約したときの
1年間の事故件数

$\frac{S_n}{n} \left(= \frac{1}{n} \sum_{i=1}^{n} X_i \right)$　事故契約の総契約数に対する比率

保険会社としてはこれらのr.v.に関心がある。

(1) S_nと$\frac{S_n}{n}$の期待値と分散は?

$S_n \sim B(n, \frac{1}{6})$と考えられるので

$$E[S_n] = \frac{n}{6}$$

$$V[S_n] = n \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36} n$$

従って

$$E\left[\frac{S_n}{n}\right] = \frac{1}{n} E[S_n] = \frac{1}{n} \cdot \frac{n}{6} = \frac{1}{6}$$

$$V\left[\frac{S_n}{n}\right] = \frac{1}{n^2} V[S_n] = \frac{1}{n^2} \frac{5}{36} n = \frac{5}{36} \cdot \frac{1}{n}$$

※ 契約件数nを増やすと、
事故件数の分散$V[S_n]$は大きくなるが、
事故契約比率の分散$V\left[\frac{S_n}{n}\right]$は小さくなる。

(2) 事故契約の（総契約数に対する）比率 $\frac{S_n}{n}$ が、その期待値 $\frac{1}{6}$ から $c\ (>0)$ 以上外れる確率の上限を求めよ。

From Chebyshev's inequality,

$$P\left(\left|\frac{S_n}{n}-\frac{1}{6}\right| \geqq c\right) \leqq \frac{\sigma^2}{c^2} = \frac{1}{c^2}\cdot\frac{5}{36}\cdot\frac{1}{n}$$

(3) $\frac{S_n}{n}$（事故契約比率）がその期待値 $\frac{1}{6}$ から5%以上外れる確率を、5%以下，3%以下にするために必要となる契約件数 n を求めよ。

- 5%以下にするためには、

$$P\left(\left|\frac{S_n}{n}-\frac{1}{6}\right| \geqq \frac{1}{20}\right) \leqq \frac{1}{20}$$

をみたすような n を求めればよい。

$c=\frac{1}{20}$ のケースで $\frac{1}{c^2}\frac{5}{36}\cdot\frac{1}{n} \leqq \frac{1}{20}$

が成り立つためには、

$$20^3\cdot\frac{5}{36} \leqq n$$

$$\therefore \frac{8000\times 5}{36} \leqq n \qquad \therefore 1111.11\cdots \leqq n$$

$\therefore n=$ 1112件以上.

- 3%以下にするためには

$$P\left(\left|\frac{S_n}{n}-\frac{1}{6}\right| \geqq \frac{1}{20}\right) \leqq (20)^2\cdot\frac{5}{36}\cdot\frac{1}{n} \leqq \frac{3}{100}$$

$\therefore 1851.85\cdots \leqq n$

$\therefore n=$ 1852件以上

(4) $\frac{S_n}{n}$ がその期待値 $\frac{1}{6}$ から 3% 以上外れる確率を 5% 以下，3% 以下にするために必要となる契約件数 n を求めよ。

・5% 以下にするためには，

$$P\left(\left|\frac{S_n}{n}-\frac{1}{6}\right| \geqq \frac{3}{100}\right)$$

$$\leqq \left(\frac{100}{3}\right)^2 \cdot \frac{5}{36} \cdot \frac{1}{n} \leqq \frac{1}{20}$$ より，

$$\left(\frac{100}{3}\right)^2 \cdot \frac{5}{36} \cdot 20 \leqq n.$$

$$\therefore 3086.4197\cdots \leqq n$$

$\therefore n = 3087$ 件以上

・3% 以下にするためには．

$$P\left(\left|\frac{S_n}{n}-\frac{1}{6}\right| \geqq \frac{3}{100}\right)$$

$$\leqq \left(\frac{100}{3}\right)^2 \cdot \frac{5}{36} \cdot \frac{1}{n} \leqq \frac{3}{100}$$ より．

$$\left(\frac{100}{3}\right)^2 \cdot \frac{5}{36} \cdot \frac{100}{3} \leqq n$$

$$\therefore 5144.032922\cdots \leqq n$$

$\therefore n = 5145$ 件以上

例.

潜在的
乗船客の
体重

$X_i \sim iid\ N(60, 10^2)$

$S_n = \sum_{i=1}^{n} X_i$ n人乗船したときの総体重

$\frac{S_n}{n} \left(= \frac{1}{n}\sum_{i=1}^{n} X_i\right)$ 〃 平均体重

(1) S_n と S_n/n の期待値と分散を求めよ.

正規分布の再生性より

$S_n \sim N(60n, 100n)$ と考えられるので

$E[S_n] = 60n$

$V[S_n] = 100n$

従って

$E\left[\frac{S_n}{n}\right] = 60$

$V\left[\frac{S_n}{n}\right] = \frac{100}{n}$

(2) n人乗船してきたとき乗船者の平均体重 $\frac{S_n}{n}$ が
その期待値 60kg から $c(>0)$ 以上外れる確率
の上限を（nに依存する形で）求めよ。

From Chebyshev's inequality,

$$P\left(\left|\frac{S_n}{n}-60\right| \geqq c\right) \leqq \frac{1}{c^2} V\left[\frac{S_n}{n}\right] = \frac{1}{c^2}\frac{100}{n}.$$

(3) 乗船者の平均体重 $\frac{S_n}{n}$ がその期待値である 60kg から
2kg 以上外れる確率を5%以下にするためには、
何人乗りの船を設計すればよいか？

$c = 2$kg のケースである。

$$\frac{1}{2^2}\frac{100}{n} \leqq \frac{5}{100}$$

つまり $500 \leqq n$ より

500人以上の船をつくればよい。

大数の法則 (law of large numbers)

Th (weak law of large numbers)

$X_1, \cdots, X_n \sim iid(\mu, \sigma^2)$ with $\sigma > 0$

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

$\Rightarrow \forall c > 0,\ P(|\bar{X}_n - \mu| \geqq c) \to 0$

Proof

We know that $E[\bar{X}_n] = \mu$ and $V[\bar{X}_n] = \frac{\sigma^2}{n}$.

Let $c > 0$.

From Chebyshev's inequality,

$$P(|\bar{X}_n - \mu| \geqq c) \leqq \frac{\sigma^2/n}{c^2} \to 0 \quad (n \to \infty)$$

$$(\forall c > 0).$$

//

※ このような収束を<u>確率収束</u>という。

※ 何種類かの拡張がある。

例（3割打者）

$$X_i = \begin{cases} 1 & \text{if ヒット} \\ 0 & \text{if 凡打} \end{cases}$$

$S_n = \sum_{i=1}^{n} X_i$　n 打席中のヒット数

$\dfrac{S_n}{n} \left(= \dfrac{1}{n}\sum_{i=1}^{n} X_i \right)$　n 打席のときの打率.

$S_n \sim B\left(n, \frac{3}{10}\right)$ なので

$$E[S_n] = \frac{3}{10}n$$

$$V[S_n] = n \cdot \frac{3}{10} \cdot \frac{7}{10} = \frac{21}{100}n$$

$$E\left[\frac{S_n}{n}\right] = \frac{1}{n}E[S_n] = \frac{3}{10}$$

$$V\left[\frac{S_n}{n}\right] = \frac{1}{n^2}V[S_n] = \frac{21}{100} \cdot \frac{1}{n}$$

"3割打者"の実際の打率 $\dfrac{S_n}{n}$ が3割から

$c\,(>0)$ 以上外れる確率は. Chebyshev's inequality より

$$P\left(\left|\frac{S_n}{n} - \frac{3}{10}\right| \geq c\right) \leq \frac{1}{c^2}V\left[\frac{S_n}{n}\right]$$

$$= \frac{1}{c^2} \cdot \frac{21}{100} \cdot \frac{1}{n}$$

よって $P\left(\left|\dfrac{S_n}{n} - \dfrac{3}{10}\right| \geq c\right) \to 0 \ (n \to \infty) \ (^\forall c > 0)$.

つまり. 打席数 n を増やせば. 真の3割打者の

実際の打率 $\dfrac{S_n}{n}$ が3割から たとえわずかでも

外れる確率は 0に収束する。

大数の法則

問題1.（標準化確率変数についてのチェビシェフの不等式）

確率変数Xの期待値を$E[X] = \mu$、分散を$V[X] = \sigma^2(\neq 0)$とする。また、$Z = \frac{X-\mu}{\sigma}$と定義する。このとき、任意の正の数$\lambda > 0$に対して、

$$P(|Z| \geq \lambda) \leq \frac{1}{\lambda^2}$$

が成り立つことを、Xが連続確率変数の場合について示せ。

問題2.（一般の確率変数についてのチェビシェフの不等式）

確率変数Xの期待値を$E[X] = \mu$、分散を$V[X] = \sigma^2(\neq 0)$とする。このとき、任意の正の数$c > 0$に対して、

$$P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

が成り立つことを、問題1で扱った標準化確率変数についてのチェビシェフの不等式を用いて示せ。

問題3. 大数の弱法則の定理のステートメントを記述し、それを証明せよ。

問題4.（サイコロ投げ）

サイコロをn回投げる第i回目を考える。1の目が出た場合$X_i = 1$、他の目が出た場合$X_i = 0$と定めると、X_iは

$$X_i = \begin{cases} 1 & \text{with prob. } 1/6 \\ 0 & \text{with prob. } 5/6 \end{cases}$$

という（離散）確率変数である。さらに、$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$とおくと、これは$n$回投げて1の目が出た回数の比率を表す確率変数である。

（1）$\bar{X}_n$の期待値$E[\bar{X}_n]$と分散$V[\bar{X}_n]$を求めよ。

（2）チェビシェフの不等式を用いて、$\bar{X}_n$が（1）で求めた期待値から$c > 0$以上外れる確率の上限を求めよ。

（3）大数の弱法則が成り立つことを確認せよ。

問題5.（保険契約）

損害保険契約1件につき、その契約者が一年間に事故を起こす確率は1%である。

（1）保険加入者の事故率がその期待値である1%と5%以上食い違う確率の上限を、契約件数nに依存する形で求めなさい。

（2）事故を起こした契約の（総契約数に対する）比率が、その期待値から5%以上外れる確率を3%以下にするためには、契約件数を何件以上にすればよいか？

（3）社長が交代し、事故を起こした契約の（総契約数に対する）比率が、その期待値から3%以上外れる確率を5%以下にする方針になった。契約件数を何件以上にすればよいか？

問題6.（世論調査）

政党Aの支持率は60%と言われている。今回、選挙を前にして世論調査を行うことになった。政党Aの支持率が60%というのが正しいとして、この調査結果としての政党Aの支持率が60%から5%以上外れる確率が1%以下にしたい。そのためには、何人に聞き取り調査を行えばいいか。

問題7. レジュメの例などを参考にして、チェビシェフの不等式や大数の法則を用いる例を自分で考え、それに解答を与えなさい。