14 juillet 2012
PROB #9

# 確率論 (*Probability Theory*) 第 9 週

w.yamamoto

## 1 幾つかの確率不等式

確率について、確率分布に拠らず成り立つ不等式が幾つかある。これらは、確率の近似的な評価に用いられたり、様々な定理の証明に用いられたりする。

### 1.1 マルコフの不等式

非負の確率変数 $X$ の期待値が有限の場合、任意の正数 $a > 0$ について

$$Pr[X \geq a] \leq \frac{1}{a} E[X] \tag{1}$$

が成り立つ。これを マルコフ (Markov) の不等式 という。[1]

証明は単純で、期待値を求める定積分の区間が $[0, \infty)$ と非負の領域であることと、その定積分よりも区間 $[a, \infty)$ での定積分の方が小さい、という事実を用いる。連続確率分布の場合、

$$\begin{aligned} E[X] &= \int_0^\infty x f_X(x)\, dx \\ &\geq \int_a^\infty x f_X(x)\, dx \;\; (\because \text{積分区間を狭めると非負の被積分関数の定積分は小さくなる}) \\ &\geq a \int_a^\infty f_X(x)\, dx \;\; (\because a \text{は区間} [a, \infty) \text{で最小値}) \\ &= a Pr[X \geq a] \end{aligned} \tag{2}$$

と証明できる。最後の不等式は、

$$\begin{aligned} \int_a^\infty x f_X(x)\, dx - a \int_a^\infty f_X(x)\, dx &= \int_a^\infty (x - a) f_X(x)\, dx \\ &\geq 0 \end{aligned} \tag{3}$$

と証明しても良い。

離散確率分布でも

$$\begin{aligned} E[X] &= \sum_{k=0}^\infty k p_X(k) \\ &\geq \sum_{k=a}^\infty k p_X(k) \;\; (\because \text{総和の区間を狭めると非負の関数の総和は小さくなる}) \\ &\geq a \sum_a^\infty f_X(x) \;\; (\because a \text{は区間} [a, \infty) \text{で最小値}) \\ &= a Pr[X \geq a] \end{aligned} \tag{4}$$

となる。

[1] この不等式は、参考書にはない。

マルコフの不等式から、累積分布関数 $F(x)$ について

$$
\begin{aligned}
F(x) &= Pr[X \le a] = 1 - Pr[X \ge a] \\
&\ge 1 - \frac{1}{x} E[X] \qquad (5)
\end{aligned}
$$

という関係が得られる。例えば期待値が $E[X] = 1$ のあらゆる非負の確率分布の累積分布関数は、図 1 の赤線より上になければならない、という強い不等式である。

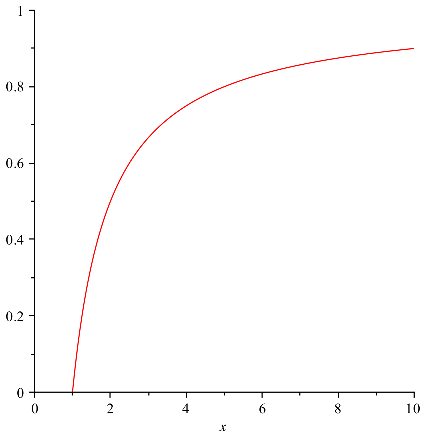


図 1 $E[X] = 1$ の場合のマルコフの不等式

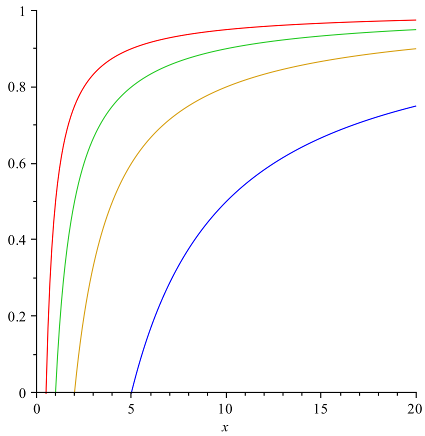


図 2 $E[X] = 0.5, 1, 2, 5$ の場合の下限

図 2 は、上から順に $E[X] = 1/2, 1, 2, 5$ の場合の下限である。いずれも期待値が変わると、累積分布関数の下限がそれに応じて変わり、これらの下限より下には累積分布関数は来ないことが分かる。

ところで図 1、図 2 からも分かるように、マルコフの不等式は $a > E[X]$ の範囲で、より精緻な区間を与える。$a \le E[X]$ では、確率は 0 以上 1 以下、という確率の公理のひとつの方がましである。そのことを確かめるために、マルコフの不等式の応用例を幾つか与えておく。

- 平均年収が 500 万円の世代の中で、年収が 5000 万円以上の人の割合は、1 割以下。

$$
Pr[X \ge 5,000] \le \frac{E[X]}{5,0000} = \frac{500}{5,000} = \frac{1}{10} \qquad (6)
$$

- 平均年収が 500 万円の世代の中で、年収が 200 万円以下の人の割合は、0 以上。

$$
Pr[X \ge 200] \ge 1 - \frac{E[X]}{200} = 1 - \frac{500}{200} = -\frac{3}{2} \qquad (7)
$$

- 1 企業あたりの平均従業員数が 2 千人とすると、10 万人以上の従業員を抱える企業の割合は、2% 以下。

$$
Pr[X \ge 2] \le \frac{E[X]}{100} = \frac{2}{100} \qquad (8)
$$

平均だけ分かれば、確率分布を知らずとも、上限や下限を与えられるのが、マルコフの不等式の実世界へのひとつの応用手段である。

マルコフの不等式は、すぐにチェビシェフの不等式の証明に使える。

## 1.2　チェビシェフの不等式

確率変数 $X$ が平均 $E[X] = \mu$、分散 $V[X] = \sigma^2$ を持つとき、任意の正数 $a > 0$ に対して

$$
Pr[|X - \mu| \ge a] \le \frac{\sigma^2}{a^2} \qquad (9)
$$

が成り立つ。これを チェビシェフ (Chevyshev) の不等式 という。[2]

この不等式は、

$$Pr\left[|X-\mu| \geq a\right] = Pr\left[|X-\mu|^2 \geq a^2\right] \tag{10}$$

と左辺を変形すれば、$|X-\mu|^2$ が非負の確率変数であることから、あとはマルコフの不等式を少し変形するだけで証明できる。

$$\begin{aligned} Pr\left[|X-\mu|^2 \geq a^2\right] &\geq \frac{E\left[|X-\mu|^2\right]}{a^2} && (11)\\ &= \frac{E\left[(X-\mu)^2\right]}{a^2} && (12)\\ &= \frac{\sigma^2}{a^2} && (13) \end{aligned}$$

この証明は、$X$ が連続確率変数でも離散確率変数でも、マルコフの不等式を証明してあることから、問題なく成り立つ。

チェビシェフの不等式は、マルコフの不当式とは異なり、確率変数 $X$ の定義域は非負でなくて良い。その代わりに、確率を評価する区間は、

$$X \geq a \text{ あるいは } X \leq b \tag{14}$$

といった片側区間ではなく、確率分布の平均 $\mu$ に対称な

$$\begin{aligned} |X-\mu| \geq a &\quad\leftrightarrow\quad X \leq \mu - a \text{ or } X \geq \mu + a\\ |X-\mu| \leq b &\quad\leftrightarrow\quad \mu - b \leq X \leq \mu + b \end{aligned}$$

などの区間でしか評価できない。

平均と分散が分かっている場合、チェビシェフの不等式、から

- 年収は非負とすれば、平均年収が 500 万円、標準偏差が 200 万円の世代の中で、年収が 5000 万円以上の人の割合は、0.004% 未満。

$$Pr\left[X - 500 \geq 4,500 \text{ or } X - 500 \leq -4,500\right] = Pr\left[X - 500 \geq 4,500\right] \leq \frac{200^2}{4,500^2} = 0.0000309... < 0.00004 \tag{15}$$

- 年収は非負とすれば、平均年収が 500 万円、標準偏差が 500 万円の世代の中で、年収が 5000 万円以上の人の割合は、1.3% 未満。

$$Pr\left[X - 500 \geq 4,500 \text{ or } X - 500 \leq -4,500\right] = Pr\left[X - 500 \geq 4,500\right] \leq \frac{500^2}{4,500^2} = \frac{1}{81} = 0.012345679... < q0.013 \tag{16}$$

- 1 企業あたりの平均従業員数が 5 百人、標準偏差を 2 百人とすると、10 万人以上の従業員を抱える企業の割合は、0.0005% 以下。

$$Pr\left[X - 5 \geq 995 \text{ or } X - 5 \leq -9995\right] = Pr\left[X \geq 1000\right] \leq \frac{2^2}{995^2} = 0.000004... < 0.000005 \tag{17}$$

など、マルコフの不等式よりは精緻な限界 (不等式) を得る。

チェビシェフの不等式は確率論の中では、確率の評価と大数の法則の証明など、幾つかの用法がある。

1. 確率の評価：上のような確率不等式による上限や下限を与えることができる
2. 大数の法則：証明に用いる

[2]参考書では p.24 で定理 2.3 として紹介されていて、p.53 の定理 4.1 の証明にも現れる。

## 1.3　イェンセンの不等式

確率変数 $X$ と、$X$ の定義域を含む区間上の凸関数 $h(x)$ を考える。$X$ の期待値と $h(X)$ の期待値がどちらも有限なとき、

$$E[h(X)] \quad \geq \quad h(E[X]) \tag{18}$$

が成り立つ。これを イェンセン (Jensen) の不等式 という。[3]

この不等式の証明は、関数 $h(x)$ が下に凸であることを利用する。$h(x)$ が下に凸であれば、任意の点 $x_0$ に対して、ある直線 $g(x) = ax + b$ が存在して、$h(x_0) = g(x_0)$ と

$$h(x) \geq g(x), \ \forall x \Leftrightarrow h(x) - g(x) \geq 0, \ \forall x \tag{19}$$

とが成り立つ。$x_0 = E[X]$ とおけば、

$$\begin{aligned} E[h(X) - g(X)] &= \int_{x \in X} \{h(x) - g(x)\} f(x) \, dx \\ &\geq 0 \end{aligned} \tag{20}$$

は式 (19) に示した $h(x)$ の凸性から明らか。一方、

$$\begin{aligned} E[g(X)] &= E[aX + b] \\ &= aE[X] + b \\ &= g(E[X]) \end{aligned} \tag{21}$$

再び式 (19) に示した凸性と $h$ と $g$ の関係より、

$$g(E[X]) \quad = \quad h(E[X]) \tag{22}$$

以上より、

$$E[h(X)] \geq E[g(X)] \tag{23}$$

が示される。離散分布についても同様。

イェンセンの不等式から、$E\left[X^2\right]$ が有限の任意の確率分布について

$$\{E[X]\}^2 \leq E\left[X^2\right] \tag{24}$$

が分かる。

$E\left[X^2\right] < \infty$ を仮定できると

$$\begin{aligned} E[|X|] &= \int_{-\infty}^{\infty} |x| f_X(x) \, dx \\ &= \int_{-\infty}^{-1} |x| f_X(x) \, dx + \int_{-1}^{1} |x| f_X(x) \, dx + \int_{1}^{\infty} |x| f_X(x) \, dx \\ &\leq \int_{-\infty}^{-1} x^2 f_X(x) \, dx + \int_{-}^{1} f_X(x) \, dx + \int_{1}^{\infty} x^2 f_X(x) \, dx \\ &\leq \int_{-\infty}^{\infty} x^2 f_X(x) \, dx + \int_{-\infty}^{\infty} f_X(x) \, dx = E\left[X^2\right] + 1 \leq \infty \end{aligned} \tag{25}$$

が成り立つのはおまけ。

イェンセンの不等式から、次のことが示せる。

[3] この不等式も、参考書には無い。

- 算術平均 $(X_1+X_2+\cdots+X_n)/n$ と幾何平均 $(X_1X_2\cdots X_n)^{1/n}$ と調和平均 $n/\{1/X_1+1/X_2+\cdots 1/X_n\}$ の間に、次の大小関係が成り立つ。(通常、別のイェンセンの不等式からも示せるが、この確率版の不等式からも示せる)

$$\frac{1}{\frac{1}{n}\left(\frac{1}{X_1}+\cdots+\frac{1}{X_n}\right)} \leq (X_1X_2\cdots X_n)^{1/n} \leq (X_1+X_2+\cdots+X_n)/n \tag{26}$$

- 算術平均 $(X_1+X_2+\cdots+X_n)/n$ と幾何平均 $(X_1X_2\cdots X_n)^{1/n}$ と調和平均 $n/\{1/X_1+1/X_2+\cdots 1/X_n\}$ の間に、次の大小関係が成り立つ。

$$E_X\left[\frac{1}{\frac{1}{n}\left(\frac{1}{X_1}+\cdots+\frac{1}{X_n}\right)}\right] \leq E_X\left[(X_1X_2\cdots X_n)^{1/n}\right] \leq E_X\left[(X_1+X_2+\cdots+X_n)/n\right] \tag{27}$$

### 1.4　ヘルダーの不等式

２つの正の数 $p,q$ が $1/p+1/q=1$ を満たすとする。$E[|X|^p]$ が有限の確率変数 $X$ と、$E[|Y|^p]$ が有限の確率変数 $Y$ に対して、

$$E[|XY|] \quad \leq \quad \{E[|X|^p]\}^{1/p}\{E[|Y|^q]\}^{1/q} \tag{28}$$

が成り立つ。これを ヘルダー (Hölder) の不等式 という。[4]
ヘルダーの不等式で特に $p=2,q=2$ とした場合

$$E[|XY|] \quad \leq \quad \sqrt{E\left[|X|^2\right]}\sqrt{E\left[|Y|^2\right]} \tag{29}$$

を、コーシー・シュワルツ (Cauchy-Schwartz) の不等式 という。[5]

## 2　確率分布同士の関係

これまでのところ、ベルヌーイ試行から出発して、以下の確率分布間の関係を得ている。

1. ベルヌーイ試行はすべての独立な離散試行の繰り返しの基本の確率分布
2. ベルヌーイ試行の独立な繰り返しを、$n$ 回連続して観測し、成功した回数を数えた時、その成功回数は二項分布に従う (ベルヌーイ分布に従う独立な確率変数の $n$ 個の和は二項分布に従う、とも)
3. ベルヌーイ試行の独立な繰り返しを、1 回の成功を得るまでの失敗回数を数えた時、その失敗回数は幾何分布に従う (ベルヌーイ分布に従う独立な確率変数の和が 1 となるまでの、失敗の観測回数は幾何分布に従う、とも)
4. ベルヌーイ試行の独立な繰り返しを、$k$ 回の成功を得るまでの失敗回数を数えた時、その失敗回数は負の二項分布に従う (ベルヌーイ分布に従う独立な確率変数の和が $k$ となるまでの、失敗の観測回数は幾何分布に従う、とも)
5. 微小時間ごとにベルヌーイ試行を独立に繰り返したとき、試行の成功間隔は指数分布に従う
6. ある繰り返し起こる事象の発生間隔が指数分布に従うとき、一定期間に起こる事象の回数はポアソン分布に従う
7. ある繰り返し起こる事象が、一定期間に起こる事象の回数がポアソン分布に従うとき、その事象の発生間隔は指数分布に従う

[4] この不等式も、参考書には記述がない。
[5] 参考書では、p.48 の問題 3.6 になっている。

8. 同一の指数分布に従う互いに独立な確率変数の和は、ガンマ分布に従う

この後、次の確率分布の関係は参考書に載っている。

1. どんな確率変数でも、互いに独立に同一の確率分布に従っていれば、確率変数の数が増えるにつれて、算術平均は正規分布に従うようになる
2. 正規分布に従う独立な確率変数の和は、正規分布に従う
3. t 分布
4. $\chi^2$ 分布
5. F 分布

載っていないのは、多項分布 (章末問題にはある)、超幾何分布、負の超幾何分布、一般超幾何分布、ベータ分布、ワイブル分布、非心 $\chi^2$ 分布、非心 $t$ 分布、非心 $F$ 分布、対数正規分布、逆正規分布、ロジスティック分布、グンベル分布 (二重指数分布)、コーシー分布、両側指数分布、パレート分布など。結構あるね ⋯ 。

# 3 離散確率変数の組

## 3.1 同時確率 (関数) と周辺確率 (関数)

まず表 1 のような確率を持つ、2 枚のコイン投げを考える。

表 1 2 枚のコイン投げの組

| | | コイン 2 | | |
|---|---|---|---|---|
| | | 表 | 裏 | 小計 |
| コイン 1 | 表 | 0.25 | 0.25 | 0.50 |
| | 裏 | 0.25 | 0.25 | 0.50 |
| | 小計 | 0.50 | 0.50 | 1.00 |

表 2 2 つのベルヌーイ試行の組

| | | $X_2$ | | |
|---|---|---|---|---|
| | | 1 | 0 | |
| $X_1$ | 1 | 0.25 | 0.25 | |
| | 0 | 0.25 | 0.25 | |
| | | | | 1.00 |

表 1 の確率事象は、確率変数を用いると、表 2 のように表せる。[6] 複数の事象を同時に考えることと、1 組の確率変数 $(X_1, X_2)$ を 同時に 扱うことは等しいので、以下では確率変数を用いて議論を進める。

まず、２つの確率変数 $X_1, X_2$ を 1 組の確率変数 $\boldsymbol{X} = (X_1, X_2)$ として扱う。この新しい二次元の確率変数は、同時確率変数、二次元確率変数、確率ベクトル、などと呼ばれる。表 2 の確率表を持つ同時確率変数 $\boldsymbol{X}$ の標本空間は、表を 1、裏を 0 として

$$\mathcal{X}_{X_1,X_2} = \{(0,0), (0,1), (1,0), (1,1)\} \tag{30}$$

の 4 つの点を要素に持つ離散集合である。[7]

$\boldsymbol{X} = (X_1, X_2)$ が $\mathcal{X}_{X_1,X_2}$ の各値を取る確率

$$Pr\,[X_1 = x_1 \ \text{ and } \ X_2 = x_2] \tag{31}$$

は、$X_1$ と $X_2$ が同時にとる値の組み合わせの確率であることから、同時確率関数 と呼ばれる。同時確率関数は

$$Pr\,[X_1 = x_1 \ \text{ and } \ X_2 = x_2] = p_{X_1,X_2}(x_1, x_2) = p_{12}(x_1, x_2) = p(x_1, x_2) \tag{32}$$

[6] 表を 1、裏を 0 としたのは $X_1$ 及び $X_2$ の値に、それぞれのコインの 表が出る回数、という意味を持たせるためのみ。

[7] 表 2 に合わせるなら、$(1,1), (1,0), \ldots$ のように降順で記す方が分かりやすい。しかし、確率変数の値を数値で表す時はこのように昇順で表すことが多いことを記しておく。表記だけの問題で、どう書いても同じ集合である。

のように、関数の引数が複数あることは共通しているが、$p$ の部分は割と自由に書かれてしまう。

全ての確率変数が同時に何れかの値の組み合わせを取る確率を同時確率と呼ぶのに対し、どれか一部の確率変数のみが同時に何れかの値の組み合わせを取る確率を 周辺確率 と呼ぶ。例えば、$X_1$ のみを観測するときの確率を $X_1$ の周辺確率という。事象「$X_1 = x_1 \text{ and } X_2 = 0$」と事象「$X_1 = x_1 \text{ and } X_2 = 1$」は互いに疎なので、それらの和集合

$$\{X_1 = x_1 \text{ and } X_2 = 0\} \cup \{X_1 = x_1 \text{ and } X_2 = 1\} \tag{33}$$

は

$$\{X_1 = x_1\} \text{ and } \{X_2 = 0 \text{ or } X_2 = 1\} \tag{34}$$

となる。もう 1 歩進めると、

$$\{X_1 = x_1\} \text{ and } \{X_2\text{の値は何でもいい}\} \tag{35}$$

であり、周辺確率は

$$Pr\left[X_1 = x_1\right] = \sum_{k=0}^{1} Pr\left[X_1 = x_1 \text{ and } X_2 = k\right] \tag{36}$$

と、$X_1 = x_1$ は固定し、それ以外の確率変数のすべての組み合わせの確率を足し合わせて求める。表 2 で表される確率事象の場合、$X_1$ の周辺確率は

$$\begin{aligned} Pr\left[X_1 = x_1\right] &= Pr\left[X_1 = x_1 \text{ and } X_2 = 0\right] + Pr\left[X_1 = x_1 \text{ and } X_2 = 1\right] \\ &= p\left(x_1, 0\right) + p\left(x_1, 1\right) \\ &= 0.25 + 0.25 = 0.5 \end{aligned} \tag{37}$$

となる。この値は、次の表 3 や表 4 の「周辺」の列と記してある。$X_2$ についても、$X_2$ のみを観測するときの確率は $X_2$ の周辺確率であり、同様に記してある。

表 3 互いに独立な 2 枚のコイン投げの同時確率表

| | | コイン 2 | | コイン 1 |
|---|---|---|---|---|
| | | 表 | 裏 | 周辺 |
| コイン 1 | 表 | 0.25 | 0.25 | 表 0.50 |
| | 裏 | 0.25 | 0.25 | 裏 0.50 |
| コイン 2 周辺 | | 表 | 裏 | |
| | | 0.50 | 0.50 | 1.00 |

表 4 互いに独立な 2 つのベルヌーイ試行の同時確率表

| | | $X_2$ | | $X_1$ |
|---|---|---|---|---|
| | | 1 | 0 | 周辺 |
| $X_1$ | 1 | 0.25 | 0.25 | 0.50 |
| | 0 | 0.25 | 0.25 | 0.50 |
| $X_2$ 周辺 | | 0.50 | 0.50 | 1.00 |

より一般に、$n$ 個の確率変数 $X_1, X_2, \ldots, X_n$ があるとき、同時確率関数を

$$Pr\left[X_1 = x_1, \ \ldots, \ X_n = x_n\right] \quad = \quad p\left(x_1, \ldots, x_n\right) \tag{38}$$

と書くと、$i$ 番目の確率変数 $X_i$ の周辺確率関数は

$$\begin{aligned} Pr\left[X_i = x_i\right] &= p_i\left(x_i\right) \\ &= \sum_{k_1} \sum_{k_2} \cdots \sum_{k_{i-1}} \sum_{k_{i+1}} \cdots \sum_{k_n} p\left(k_1, k_2, \ldots, k_{i-1}, x_i, k_{i+1}, \ldots, k_n\right) \end{aligned} \tag{39}$$

と求める。これは $X_i = x_i$ 以外、$X_1, X_2, \ldots, X_{i-1}, X_{i+1} \ldots, X_n$ については、あらゆる可能性をすべて足し合わせてしまった確率、とも言える。

## 3.2　独立性と従属性

表 4 の確率表から、確率変数 $X_1$ しか観測できない場合の、周辺確率分布は

$$\begin{aligned} Pr[X_1 = 1] &= Pr[X_1 = 1 \text{ and } X_2 = 0] + Pr[X_1 = 1 \text{ and } X_2 = 1] \\ &= 0.25 + 0.25 = 0.5 \qquad (40) \\ Pr[X_1 = 0] &= Pr[X_1 = 0 \text{ and } X_2 = 0] + Pr[X_1 = 0 \text{ and } X_2 = 1] \\ &= 0.25 + 0.25 = 0.5 \qquad (41) \end{aligned}$$

より、成功確率が $p = 0.5$ のベルヌーイ試行そのものである。$X_2$ も同様に、

$$\begin{aligned} Pr[X_2 = 1] &= Pr[X_1 = 0 \text{ and } X_2 = 1] + Pr[X_1 = 1 \text{ and } X_2 = 1] \\ &= 0.25 + 0.25 = 0.5 \qquad (42) \\ Pr[X_2 = 0] &= Pr[X_1 = 0 \text{ and } X_2 = 0] + Pr[X_1 = 1 \text{ and } X_2 = 0] \\ &= 0.25 + 0.25 = 0.5 \qquad (43) \end{aligned}$$

と、成功確率が $p = 0.5$ のベルヌーイ試行となる。このことは表 3 および表 4 と共通する。

次に <u>独立性</u> の定義のひとつ

$$X \perp\!\!\!\perp Y \Leftrightarrow Pr[X = x \text{ and } \ Y = y] = Pr[X = x] \times Pr[Y = y] \qquad (44)$$

を思い出せば、直ちに、

$$\begin{aligned} &Pr[\text{「コイン 1 で表が出る」 and 「コイン 2 で表が出る」}] \\ &\quad = 0.25 = 0.50 \times 0.50 \\ &\quad = Pr[\text{「コイン 1 で表が出る」}] \times Pr[\text{「コイン 2 で表が出る」}] \qquad (45) \end{aligned}$$

の関係が

$$(\text{コイン 1, コイン 2}) \in \{(\text{表, 表}), (\text{表, 裏}), (\text{裏, 表}), (\text{裏, 裏})\} \qquad (46)$$

の全ての組み合わせについて確認でき、コイン 1 とコイン 2、それぞれの試行は互いに独立と分かる。このことから、この同時確率分布は、周辺分布がベルヌーイ試行であり、しかも同時分布も互いに独立なベルヌーイ試行の組み合わせになっている、と結論づけられる。このことが、表 3 と表 4 のタイトルには、「互いに独立な … 」と記した根拠である。

では、次の表 5 と表 6 はどうだろうか。

表 5 互いに従属な 2 枚のコイン投げの同時確率表

| | | コイン 2 | | |
|---|---|---|---|---|
| | | 表 | 裏 | 周辺 |
| コイン 1 | 表 | 0.30 | 0.20 | 0.50 |
| | 裏 | 0.20 | 0.30 | 0.50 |
| 周辺 | | 0.50 | 0.50 | 1.00 |

表 6 互いに従属なベルヌーイ試行の同時確率表

| | | $X_2$ | | |
|---|---|---|---|---|
| | | 1 | 0 | 周辺 |
| $X_1$ | 1 | 0.30 | 0.20 | 0.50 |
| | 0 | 0.20 | 0.30 | 0.50 |
| 周辺 | | 0.50 | 0.50 | 1.00 |

表 5 はすぐに、確率変数を用いて表 6 のように書き直せるので、ここでも確率変数を用いて議論する。表 6 の確率表から、確率変数 $X_1$ しか観測できない場合の、周辺確率分布は

$$\begin{aligned} Pr[X_1 = 1] &= Pr[X_1 = 1 \text{ and } X_2 = 0] + Pr[X_1 = 1 \text{ and } X_2 = 1] \\ &= 0.2 + 0.3 = 0.5 \qquad (47) \\ Pr[X_1 = 0] &= Pr[X_1 = 0 \text{ and } X_2 = 0] + Pr[X_1 = 0 \text{ and } X_2 = 1] \\ &= 0.3 + 0.2 = 0.5 \qquad (48) \end{aligned}$$

より、成功確率が $p = 0.5$ のベルヌーイ試行そのものである。$X_2$ も同様に、

$$
\begin{aligned}
Pr[X_2 = 1] &= Pr[X_1 = 0 \text{ and } X_2 = 1] + Pr[X_1 = 1 \text{ and } X_2 = 1] \\
&= 0.2 + 0.3 = 0.5 \qquad (49) \\
Pr[X_2 = 0] &= Pr[X_1 = 0 \text{ and } X_2 = 0] + Pr[X_1 = 1 \text{ and } X_2 = 0] \\
&= 0.3 + 0.2 = 0.5 \qquad (50)
\end{aligned}
$$

と、成功確率が $p = 0.5$ のベルヌーイ試行となる。このことは表 3 および表 4 と共通する。

しかし同時確率は、周辺確率の積にはならない。

$$
\begin{aligned}
&Pr[\text{「コイン 1 で表が出る」 and 「コイン 2 で表が出る」}] \\
&= 0.3 \neq 0.50 \times 0.50 \\
&\neq Pr[\text{「コイン 1 で表が出る」}] \times Pr[\text{「コイン 2 で表が出る」}] \qquad (51)
\end{aligned}
$$

が確かめられる。他の値の組み合わせについても同様である。

ここで確率論では、独立の否定語は非独立ではなく、<u>従属</u>であったことを思い起こして欲しい。独立でなければ従属であるといい、式 (44) の対偶

$$X \not\perp\!\!\!\perp Y \Leftrightarrow Pr[X = x \text{ and } Y = y] \neq Pr[X = x] \times Pr[Y = y] \qquad (52)$$

が成り立つ。

よって確率変数 $X_1$ と $X_2$ は周辺分布はどちらも $p = 0.5$ のベルヌーイ試行だが、互いに従属であることが分かる。

表 4 と表 6 は人工的な例だが、確率を計算する上で、複数の確率変数が互いに独立か否か、また検討する事象の候補が互いに疎か否か、は計算結果に大きな影響を与える。実際には独立ではない変数同士に独立性を仮定することは、その影響をしっかり検討してからにするのが良い。

## 3.3 互いに独立な確率変数についての補足

1 組の確率変数が互いに独立であるとき、$(X_1, X_2)$ を得る試行 (あるいは実験) は行ったが、$X_1$ の結果のみ知らされたとする。この時は条件付き確率の公式

$$Pr[X_2 = x_2 | X_1 = x_1] = \frac{Pr[X_1 = x_1 \text{ and } X_2 = x_2]}{Pr[X_1 = x_1]} \qquad (53)$$

を用いれば、表 2 のような条件付き確率表を得ることができる。

表 2 $X_1$ の結果が 1 もしくは 0 と知らされた時の $X_2$ の条件付き確率 $Pr[X_2|X_1]$

| | | $X_2$ 未観測 | | |
|---|---|---|---|---|
| | | 1 | 0 | $Pr[X_2 = 0 \text{ or } 1 \mid X_1]$ |
| $X_1$ | 1 | 0.50 | 0.50 | 1.0 |
| 条件 | 0 | 0.50 | 0.50 | 1.0 |

表 2 から、$X_1$ の値のみを知らされたとき、0, 1 いずれの値であっても $X_2$ の条件付き分布は変わらないことが分かる。このように、<u>一部の値を知らされても残りの部分に関して全く得をしない</u>のが、独立性の 1 つの意味である。

確率計算上は、考えるべき確率変数が互いに独立であれば、同時確率が周辺確率の積で得られることのメリットはとても大きい。例えば 100 個の確率変数 $X_1, X_2, \ldots, X_{100}$ の同時確率が 100 次元の表で

$$Pr[X_1 = x_1 \text{ and } X_2 = x_2 \text{ and } \cdots X_{100} = x_{100}] = p(x_1, x_2, \ldots, x_{100}) \qquad (54)$$

と定義されるべきところを、$X_1, X_2, \ldots, X_{100}$ が互いに独立であれば、

$$Pr\left[X_1 = x_1 \text{ and } X_2 = x_2 \text{ and } \cdots X_{100} = x_{100}\right] = p\left(x_1, x_2, \ldots, x_1 00\right) = p_1\left(x_1\right) \times p_2\left(x_2\right) \times \cdots \times p_{100}\left(x_{100}\right) \tag{55}$$

と、1 次元の確率を 100 個、で定義が完了する。

表 3 独立試行と従属試行での確率関数の定義に必要な組み合わせの違い

| 複数同時試行の例 | 互いに独立で同一 | 互いに独立 | 互いに従属 |
|---|---|---|---|
| $n$ 回の同時ベルヌーイ試行 | 2 | $2n$ | $2^n$ |
| $n$ 個のさいころの同時投げ | 6 | $6n$ | $6^n$ |

例えば互いに独立なベルヌーイ試行を $n = 100$ 回行う場合、126 穣[8]の組み合わせを検討しなければならない。さいころならば、1 無量大数にさらに 10 億を掛けた数の組み合わせについて検討しなければならない。一般に、定義しなければならない確率の数が多くなると、そのまま確率分布のパラメータの数も多くなる。

ところが互いに独立で同一な確率分布に従う試行であるならば、ベルヌーイ試行の場合は何回繰り返そうが、2 つの値 $X = 0, 1$ についての (周辺) 確率を定めれば、同時確率は自動的に計算できる。さいころ投げの場合も 6 つの値 $X = 1, 2, 3, 4, 5, 6$ の確率だけを定めれば、同時確率は自動的に計算できる。つまり、少ないパラメータの数で全ての事象の起こる確率を定めることができる。

このように、確率分布を定義する際に実現値の組み合わせを考える必要がなくなるのが、独立性の仮定の最大のメリットといえる。

### 3.4 互いに従属な確率変数についての補足

確率論の講義ではまず、互いに独立な確率変数についての計算を多く紹介する。分散の加法性しかり、モーメント母関数の性質しかり、和の分布しかり、である。しかし周辺分布が同じことと、互いに独立なことは、異なる仮定である。周辺確率分布は同じでも、互いに独立か、あるいは互いに従属かで、同時分布は異なる。同時分布が変われば、条件付き分布も変わる。よって、誤った独立性の仮定の下で、条件付き確率や同時確率を計算するなど、楽観的な確率計算に基づいてリスクを評価するのは危険であり、正しい確率の計算には正しいモデルの設定が欠かせない。

## 4 講義と参考書の共通部分と差異

共通部分：

- 1 章は同じ
- 2 章は一部同じ
- 3 章は、離散確率分布と連続確率分布を併記するようにした
- 4 章は後半に回す
- 5 章は 3 章の説明とともに説明した
- 6 章も 3 章の説明とともに説明した

差異：

- 講義では測度論の利用は避けてみた

[8] じょう、と読む。$10^2 8$ の単位のこと。一、万、億、兆、京 (けい)、垓 (がい)、𥝱 (じょ)、穣 (じょう)、溝 (こう)、澗 (かん)、正 (せい)、載 (さい)、極 (ごく)、恒河沙 (ごうがしゃ)、阿僧祇 (あそうぎ)、那由多 (なゆた)、不可思議 (ふかしぎ)、無量大数 (むりょうたいすう) と続く、日本の単位の 7 つ目。参考までに英語は 126 穣回は、1.26 nonillion times と言う。

- 講義ではスティルチェス積分の利用も避けてみた
- 講義ではガンマ分布を加えた
- 講義では多項分布も教えた
- 講義では正規分布を一番最後に回した

## 5 レポート略解

**#8-1** 典型的な計算問題。
**#8-2** 典型的な計算問題。
**#8-3** 証明をフォローする、は、証明を追いかける、意味なので、証明をしっかり理解した上で、転記しながら、自分にとって不明な点がないように補えば良い。
**#8-4** マルコフの不等式は、チェビシェフの不等式に並んで紹介されるときは、チェビシェフの不等式と同等の証明でいい。Wikipedia を写した人がいるようだが、確率論における証明はともかく、一般的証明の方は、理解していますか？と確認の意味で、×にしてあるそう。折角、頑張ってこの講義では避けて回っている測度論をがっちり使っている。可測集合の概念も、ルベーグ積分 $\int d\mu$ も、この講義では避けて来たけど、理解できている人がいるのなら、こちらもスティルチェス積分を全面に押し出して講義を行うなど、内容をもう少し高度にしても良いかもしれない、という感想を抱いた。

## 6 レポート課題

**#9-1** 気象庁の波浪観測情報[9]を参考に、どこか適当な場所や地域について、適当な長さの期間の波の高さの平均を調べ、$10m$ 以上の高さの波が来る確率を評価しなさい。(地震性の津波のための計算方式ではないことに注意)
**#9-2** 分散も計算して、同様に $10m$ 以上の高さの波が来る確率を評価しなさい。
**#9-3** (参考書を参考に) コーシー・シュワルツの不等式を証明しなさい。

## 7 レポート提出要領

下記の要領でレポートを作成し、提出すること。

| | |
|---|---|
| 課題番号 | #9 (2012.06.14 出題) |
| 提出期限 | 2012 年 6 月 25 日 午後 4 時 30 分 |
| 提出場所 | 西 5 号館 3 階総合情報学科事務室前の集合ポストの「確率論」とある投函口 |
| 様　　式 | Ａ４もしくはＢ５（ルーズリーフ可、両面可） |
| そ の 他 | 丸写しは採点していて飽きるし、剽窃は自分のためにならない<br>各自が自力で取り組むことを、切に願う<br>成書を参考にするなとは言わないが、参考にした書籍があれば、著者への礼儀として必ず記すこと<br>表紙はつけないこと<br>１ページ目の上部に、「講義名」「レポート番号」「学籍番号」「氏名」「投函日」を記すこと |

[9]http://www.jma.go.jp/jp/wave/ で地図上のどこかをクリックすると、波の高さのデータに出会える。

下記は見本である。

| | |
|---|---|
| | 提出日:2012/06/11 |
| 確率論レポート #8 | |
| | 学籍番号:0000000 |
| | 氏　　名:電通　大 |
| 課題#8-1 | |

## 参考文献


[1] 微分積分学の教科書.

[2] 永田靖 (2005) 「統計学のための数学入門 30 講」朝倉書店.

[3] 宮川雅巳 (1988) 「統計技法」 共立出版.

[4] 稲垣宣生 (2003) 「数理統計学」改訂版, 裳華房.