28 juillet 2012
PROB #11

確率論 (*Probability Theory*) 第 11 週

w.yamamoto

1　今後の予定

今回を含めて講義はあと 5 回なので、そろそろまとめを意識し始める。今後の講義内容は、次のように予定している。

- 同時分布と条件付き分布と周辺分布 (概念整理, 多項分布と分割表を例に)
- 正規分布 (1 次元, 様々な準備)
- 同時分布と条件付き分布と周辺分布 (正規分布 (2 次元) を例に)
- 確率分布に関する収束定理 1: 大数の法則 (マルコフの定理やチェビシェフの定理の続き)
- 確率分布に関する収束定理 2: 中心極限定理 (大数の法則の続き)
- マルコフ過程 (OR に送るかも)

2　離散確率変数の組

2.1　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	コイン 2 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	X_2 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。[1] 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

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

の 4 つの点を要素に持つ離散集合である。[2]

[1] 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。
[2] 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。

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

$$Pr\,[X_1 = x_1 \ \text{ and } \ X_2 = x_2] \tag{2}$$

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

$$Pr\,[X_1 = x_1 \ \text{ and } \ X_2 = x_2] = p_{X_1,X_2}\,(x_1, x_2) = p_{12}\,(x_1, x_2) = p\,(x_1, x_2) \tag{3}$$

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「$X_1 = x_1$ and $X_2 = 0$」と事象「$X_1 = x_1$ and $X_2 = 1$」は互いに疎なので、それらの和集合

$$\{X_1 = x_1 \ \text{ and } \ X_2 = 0\} \cup \{X_1 = x_1 \ \text{ and } \ X_2 = 1\} \tag{4}$$

は

$$\{X_1 = x_1\} \ \text{and} \ \{X_2 = 0 \ \text{or} \ X_2 = 1\} \tag{5}$$

となる。もう 1 歩進めると、

$$\{X_1 = x_1\} \ \text{and} \ \{X_2\text{の値は何でもいい}\} \tag{6}$$

であり、周辺確率は

$$Pr\,[X_1 = x_1] = \sum_{k=0}^{1} Pr\,[X_1 = x_1 \ \text{ and } \ X_2 = k] \tag{7}$$

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

$$\begin{aligned} Pr\,[X_1 = x_1] &= Pr\,[X_1 = x_1 \ \text{ and } \ X_2 = 0] + Pr\,[X_1 = x_1 \ \text{ and } \ X_2 = 1] \\ &= p\,(x_1, 0) + p\,(x_1, 1) \\ &= 0.25 + 0.25 = 0.5 \end{aligned} \tag{8}$$

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 表	裏	コイン 1 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 0.50	裏 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 1	0	X_1 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

$$Pr\,[X_1 = x_1, \ \ldots, \ X_n = x_n] \ = \ p\,(x_1, \ldots, x_n) \tag{9}$$

と書くと、i 番目の確率変数 X_i の周辺確率関数は

$$\begin{aligned} Pr\,[X_i = x_i] &= p_i\,(x_i) \\ &= \sum_{k_1}\sum_{k_2}\cdots\sum_{k_{i-1}}\sum_{k_{i+1}}\cdots\sum_{k_n} p\,(k_1, k_2, \ldots, k_{i-1}, x_i, k_{i+1}, \ldots, k_n) \end{aligned} \tag{10}$$

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。

2.2　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

$$\begin{aligned} Pr[X_1 = 1] &= Pr[X_1 = 1 \text{ and } X_2 = 0] + Pr[X_1 = 1 \text{ and } X_2 = 1] \\ &= 0.25 + 0.25 = 0.5 \end{aligned} \tag{11}$$

$$\begin{aligned} Pr[X_1 = 0] &= Pr[X_1 = 0 \text{ and } X_2 = 0] + Pr[X_1 = 0 \text{ and } X_2 = 1] \\ &= 0.25 + 0.25 = 0.5 \end{aligned} \tag{12}$$

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、

$$\begin{aligned} Pr[X_2 = 1] &= Pr[X_1 = 0 \text{ and } X_2 = 1] + Pr[X_1 = 1 \text{ and } X_2 = 1] \\ &= 0.25 + 0.25 = 0.5 \end{aligned} \tag{13}$$

$$\begin{aligned} Pr[X_2 = 0] &= Pr[X_1 = 0 \text{ and } X_2 = 0] + Pr[X_1 = 1 \text{ and } X_2 = 0] \\ &= 0.25 + 0.25 = 0.5 \end{aligned} \tag{14}$$

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 独立性 の定義のひとつ

$$X \perp\!\!\!\perp Y \Leftrightarrow Pr[X = x \text{ and } Y = y] = Pr[X = x] \times Pr[Y = y] \tag{15}$$

を思い出せば、直ちに、

$$\begin{aligned} & Pr[\text{「コイン 1 で表が出る」 and 「コイン 2 で表が出る」}] \\ &= 0.25 = 0.50 \times 0.50 \\ &= Pr[\text{「コイン 1 で表が出る」}] \times Pr[\text{「コイン 2 で表が出る」}] \end{aligned} \tag{16}$$

の関係が

$$(\text{コイン } 1, \text{ コイン } 2) \in \{(\text{表}, \text{表}), (\text{表}, \text{裏}), (\text{裏}, \text{表}), (\text{裏}, \text{裏})\} \tag{17}$$

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	コイン 2 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

$$\begin{aligned} Pr[X_1 = 1] &= Pr[X_1 = 1 \text{ and } X_2 = 0] + Pr[X_1 = 1 \text{ and } X_2 = 1] \\ &= 0.2 + 0.3 = 0.5 \end{aligned} \tag{18}$$

$$\begin{aligned} Pr[X_1 = 0] &= Pr[X_1 = 0 \text{ and } X_2 = 0] + Pr[X_1 = 0 \text{ and } X_2 = 1] \\ &= 0.3 + 0.2 = 0.5 \end{aligned} \tag{19}$$

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、

$$\begin{aligned} Pr[X_2 = 1] &= Pr[X_1 = 0 \text{ and } X_2 = 1] + Pr[X_1 = 1 \text{ and } X_2 = 1] \\ &= 0.2 + 0.3 = 0.5 \end{aligned} \tag{20}$$

$$\begin{aligned} Pr[X_2 = 0] &= Pr[X_1 = 0 \text{ and } X_2 = 0] + Pr[X_1 = 1 \text{ and } X_2 = 0] \\ &= 0.3 + 0.2 = 0.5 \end{aligned} \tag{21}$$

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。

$$\begin{aligned} & Pr[\text{「コイン 1 で表が出る」 and 「コイン 2 で表が出る」}] \\ &= 0.3 \neq 0.50 \times 0.50 \\ &\neq Pr[\text{「コイン 1 で表が出る」}] \times Pr[\text{「コイン 2 で表が出る」}] \end{aligned} \tag{22}$$

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、従属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 (15) の対偶

$$X \not\perp\!\!\!\perp Y \Leftrightarrow Pr[X = x \text{ and } Y = y] \neq Pr[X = x] \times Pr[Y = y] \tag{23}$$

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。

2.3 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

$$Pr[X_2 = x_2 | X_1 = x_1] = \frac{Pr[X_1 = x_1 \text{ and } X_2 = x_2]}{Pr[X_1 = x_1]} \tag{24}$$

を用いれば、表 7 のような条件付き確率表を得ることができる。

表 7 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 7 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、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

$$Pr[X_1 = x_1 \text{ and } X_2 = x_2 \text{ and } \cdots X_{100} = x_{100}] = p(x_1, x_2, \ldots, x_{100}) \tag{25}$$

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、

$$Pr\left[X_1 = x_1 \text{ and } X_2 = x_2 \text{ and } \cdots X_{100} = x_{100}\right] = p\left(x_1, x_2, \ldots, x_100\right) = p_1\left(x_1\right) \times p_2\left(x_2\right) \times \cdots \times p_{100}\left(x_{100}\right) \tag{26}$$

と、1 次元の確率を 100 個、で定義が完了する。

表 8 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[3]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。

2.4 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。

3 確率分布

離散分布の方が考え易いので、離散分布から入る。

3.1 二項分布再び

3.1.1 任意の確率分布の二項化

確率分布の二項化、という操作がある。任意の一次元の確率変数 X と定数の関数として、$X \leq a$ ならば 0、$X > a$ ならば 1 という二値をとる関数 $y(X)$ を考える。

$$y(x) = \begin{cases} 1 & \text{if} x \geq a \\ 0 & \text{if} x < a \end{cases} \tag{27}$$

この変換を施した確率変数 $Y = y(X)$ は、ベルヌーイ試行と同じ確率関数を持つ。

$$p(y) \quad = \quad q^y (1-q)^{1-y}, \;\; q = Pr[X > a] \tag{28}$$

[3]じょう、と読む。10^28 の単位のこと。一、万、億、兆、京 (けい)、垓 (がい)、予 (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。

パラメータ q が、確率変数 X の値が $X > a$ を満たす確率、という意味を持つが、これは a を超えるか否かという試行の成功確率に他ならない。同じ試行を独立に繰り返し観測すれば、成功回数は二項分布、負の二項分布など、ベルヌーイ試行に基づく確率分布に従う。

3.1.2　ポアソン分布からの導出

ところで二項分布には、独立な 2 つのポアソン確率変数の和を条件づけたときの、片方の確率変数が従う条件付き分布、という導出もある。まず、2 つの確率変数 X, Y が互いに独立にポアソン分布 $Po(\lambda_x), Po\left(\lambda_y\right)$ に従うとする。このとき $X+Y$ もポアソン分布に従う。

$$X \sim Po(\lambda_x),\ \ Y \sim Po\left(\lambda_y\right),\ \ X \perp\!\!\!\perp Y \Rightarrow X+Y \sim Po\left(\lambda_x+\lambda_y\right) \tag{29}$$

すると $X+Y=c$ との条件を与えた時の X の条件付き分布は

$$\begin{aligned}
Pr[X=k|X+Y=c] &= p(x|X+Y=c) \\
&= Pr[X=k \ \text{and} \ \ Y=(c-k)] \\
&= \frac{Pr[X=k \ \text{and} \ \ Y=(c-k)]}{Pr[X+Y=c]} \\
&= \frac{Pr[X=k]\,Pr[Y=(c-k)]}{Pr[X+Y=c]} \\
&= \frac{{\lambda_x}^k}{k!}e^{-\lambda_x}\frac{{\lambda_y}^{c-k}}{(c-k)!}e^{-\lambda_y}\frac{c!}{\lambda_x+{\lambda_y}^c}e^{\lambda_x+\lambda_y} \\
&= \frac{c!}{k!\,(c-k)!}\frac{{\lambda_x}^k{\lambda_y}^{c-k}}{\left(\lambda_x+\lambda_y\right)^k} \\
&= \frac{c!}{k!\,(c-k)!}\left(\frac{\lambda_x}{\lambda_x+\lambda_y}\right)^k\left(\frac{\lambda_y}{\lambda_x+\lambda_y}\right)^{c-k}
\end{aligned} \tag{30}$$

となる。最後の式で $n=c, p=\lambda_x/\left(\lambda_x+\lambda_y\right)$ と置けば

$$1-p = 1-\frac{\lambda_x}{\lambda_x+\lambda_y} = \frac{\lambda_y}{\lambda_x+\lambda_y} \tag{31}$$

であり、二項分布の確率関数

$$p(k) = \frac{n!}{k!\,(n-k)!}p^k(1-p)^{n-k} \tag{32}$$

を得る。この導出により、ランダムな 2 つのポアソン確率変数の和だけが把握できている状況での、それぞれの内訳の分布、がやはり二項分布に従うことが分かる。

3.2　多項分布

3.2.1　多項分布の紹介

表 3 から表 4 に書き直す代わりに、別のことを考える。(X_1, X_2) の組み合わせの事象を、次の表のように番号付けする。

表 2 つの試行の組み合わせ事象

事象番号	事象
1	$(0,0)$
2	$(0,1)$
3	$(1,0)$
4	$(1,1)$

そしてこれら 4 つの事象に関する確率分布を導出したい。まず思いつくのは、事象番号をとる確率変数 Z を定義することだろうか。確かに Z と X_1, X_2 の間には

$$Z = 2X_1 + X_2 + 1 \tag{33}$$

という関数関係があるが、この関係を用いても、うまく Z の確率関数を定義することはできない。

ここでは新たに $\{1,2,3,4\}$ の何れかの値をとったかを、事象ごとに数える確率ベクトル $\boldsymbol{Y}$ を考える。この $\boldsymbol{Y}$ は

$$\boldsymbol{Y} \quad = \quad (Y_1, Y_2, Y_3, Y_4) \tag{34}$$

と確率ベクトルであり、事象 k が起きたときに Y_k のみが 1 で残りが 0 となる。

表 9 2つの試行の組み合わせ事象

事象番号	事象	確率	$\boldsymbol{Y}$
1	(0, 0)	p_1	(1, 0, 0, 0)
2	(0, 1)	p_2	(0, 1, 0, 0)
3	(1, 0)	p_3	(0, 0, 1, 0)
4	(1, 1)	p_4	(0, 0, 0, 1)

このとき (Y_1, Y_2, Y_3, Y_4) の従う確率分布は

$$\begin{aligned} Pr\,[Y_1 = y_1 \ \text{ and } \ Y_2 = y_2 \ \text{ and } \ Y_3 = y_3 \ \text{ and } \ Y_4 = y_4] & \\ = \ & p\,(y_1, y_2, y_3, y_4) \\ = \ & \frac{1!}{1!0!0!0!} {p_1}^{y_1} {p_2}^{y_2} {p_3}^{y_3} {p_4}^{y_4} \end{aligned} \tag{35}$$

となる。ここで $y_1, \ldots, y_4$ は、どれかひとつのみが 1 となり、残りは 0 となる。

これを少し一般化すると、1 回の試行で k 種類の事象 $A_1, \ldots, A_k$ のうちどれか 1 つだけが生起する確率事象を考える。

表 10 多項分布の事象

事象	確率	$\boldsymbol{Y}$
A_1	p_1	$(1, 0, \ldots, 0)$
A_2	p_2	$(0, 1, \ldots, 0)$
$\vdots$	$\vdots$	$\vdots$
A_k	p_k	$(0, 0, \ldots, 1)$

各事象の起こる確率は p_j とし、確率の公理から

$$\sum_{j=1}^{k} p_j = 1, \ \ p_j \geq 0, j = 1, \ldots, k \tag{36}$$

が成り立つ。この確率変数ベクトル $\boldsymbol{Y}$ が従う確率分布は

$$\begin{aligned} Pr\,[Y_1 = y_1 \ \text{ and } \ Y_2 = y_2 \ \text{ and } \ \cdots \ \text{ and } \ Y_k = y_k] & \\ = \ & p\,(y_1, y_2, \ldots, y_k) \\ = \ & \frac{1!}{1!0!0!0!} {p_1}^{y_1} {p_2}^{y_2} \cdots {p_k}^{y_k} \end{aligned} \tag{37}$$

という確率関数を持つ。

次に、上の試行の n 回の独立な繰り返しを考える。各回の試行を $\boldsymbol{Y}_i, i = 1, \ldots, n$ とすると

$$\boldsymbol{Y}_i \quad = \quad \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{jk} \end{pmatrix}, \quad i = 1, \ldots, n \tag{38}$$

であり、それぞれは 0 か 1 の値を取る。するとこれらの和のベクトルを $\mathbf{Z}$ とすると

$$\mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_k \end{pmatrix} = \sum_{i=1}^{n} \boldsymbol{Y}_i \quad = \quad \begin{pmatrix} \sum_{i=1}^{n} Y_{i1} \\ \sum_{i=1}^{n} Y_{i2} \\ \vdots \\ \sum_{i=1}^{n} Y_{ik} \end{pmatrix} \tag{39}$$

は、

$$\sum_{j=1}^{k} Y_{ij} = 1, \quad i = 1, \ldots, n \quad , \sum_{i=1}^{n} \sum_{j=1}^{k} Y_{ij} = \sum_{j=1}^{k} Z_k = n \tag{40}$$

より、

$$\begin{aligned} & Pr\left[Z_1 = z_1 \;\text{ and }\; Z_2 = z_2 \;\text{ and }\; \cdots \;\text{ and }\; Z_k = z_k\right] \\ & \quad = \quad p\left(z_1, z_2, \ldots, z_k\right) \\ & \quad = \quad \frac{n!}{z_1!z_2!\cdots!z_k!} {p_1}^{z_1} {p_2}^{z_2} \cdots {p_k}^{z_k} \end{aligned} \tag{41}$$

を得る。この確率関数を持つ確率分布を、多項分布 という。独立なベルヌーイ試行列の和から二項分布を得たように、この独立な試行列から多項分布を得る。

なお総試行回数 n は所与であるため、

$$Z_k = n - \sum_{j=1}^{k-1} Z_j, \tag{42}$$

また総確率も 1 は所与であるため、

$$p_k = 1 - \sum_{j=1}^{k-1} p_j, \tag{43}$$

として、

$$\begin{aligned} & Pr\left[Z_1 = z_1 \;\text{ and }\; Z_2 = z_2 \;\text{ and }\; \cdots \;\text{ and }\; Z_k = z_k\right] \\ & \quad = \quad p\left(z_1, z_2, \ldots, z_k\right) \\ & \quad = \quad \frac{n!}{z_1!z_2!\cdots!z_k!} {p_1}^{z_1} {p_2}^{z_2} \cdots {p_{k-1}}^{z_{k-1}} \left(1 - \sum_{j=1}^{k-1} p_j\right)^{n - \sum_{j=1}^{k-1} z_j} \end{aligned} \tag{44}$$

という表現をとることも少なくない。

多項分布に従う確率ベクトルの期待値は、

$$E\left[Z\right] \quad = \quad \begin{pmatrix} E\left[Z_1\right] \\ E\left[Z_2\right] \\ \vdots \\ E\left[Z_k\right] \end{pmatrix} = \begin{pmatrix} np_1 \\ np_2 \\ \vdots \\ np_k \end{pmatrix} \tag{45}$$

また分散共分散行列は

$$\begin{aligned} V\left[Z\right] \quad & = \quad E\left[\left(\mathbf{Z} - E\left[\mathbf{Z}\right]\right)\left(\mathbf{Z} - E\left[\mathbf{Z}\right]\right)'\right] \\ & = \quad \begin{pmatrix} V\left[Z_1\right] & Cov\left[Z_1, Z_2\right] & \cdots & Cov\left[Z_1, Z_k\right] \\ Cov\left[Z_2, Z_1\right] & V\left[Z_2\right] & \cdots & Cov\left[Z_2, Z_k\right] \\ \vdots & \vdots & \ddots & \vdots \\ Cov\left[Z_k, Z_1\right] & Cov\left[Z_k, Z_2\right] & \cdots & V\left[Z_k\right] \end{pmatrix} \end{aligned}$$

$$
= \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & \cdots & -np_1n_k \\ -np_2p_1 & np_2(1-p_2) & \cdots & -np_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_kp_1 & -np_kp_2 & \cdots & np_k(1-p_k) \end{pmatrix} \tag{46}
$$

となる。

3.2.2 任意の確率分布の多項化

確率分布 F に従う確率変数 X の標本空間 $\mathcal{X}$ を、k 分割することを考える。それぞれの領域を $A_1, A_2, \ldots, A_k$ と置く。また新たに k 個の確率変数 $Y_1, \ldots, Y_k$ を考える。Y_j, $j = 1, \ldots, k$ それぞれは、確率変数 X が領域 A_j に含まれたときだけ 1 の値を得て、残りは 0 のままとなる。

$$
Y_j = y_j(X) = \begin{cases} 1 & X \in A_j \\ 0 & X \notin A_j \end{cases}, \quad j = 1, \ldots, k. \tag{47}
$$

この時 $(Y_1, Y_2, \ldots, Y_k)$ は確率が $p_j = Pr\left[X \in A_j\right]$, $j = 1, \ldots, k$ の多項分布に従う。

データを採取してヒストグラムを描くことは、データを区間を区切って数え上げるので、経験分布の多項分布化と見なせなくもない。

3.2.3 ポアソン分布からの導出

二項分布と同様に多項分布も、互いに独立にポアソン分布に従う確率変数の和を所与としたときの各確率変数の条件付き分布、として導出することができる。

3.3 正規分布

どうしようねぇ… 。

4 レポート略解

#9-1 前々回の講義中に解説したように、マルコフの不等式を適用する問題。この不等式による確率の評価には平均が必要で、それをデータから求めて、確率を評価すればいい。

#9-1 前々回の講義中に解説したように、チェビシェフの不等式を適用する問題。この不等式による確率の評価には平均だけでなく分散も必要で、それをデータから求めて、確率を評価すればいい。

#9-3 参考書の通り。

5 レポート課題

#11-1 次の計算を行え。計算過程も理解して解答すること。

(1) 定積分 $\displaystyle\int_{-\infty}^{\infty} e^{-x^2} dx$ を求めよ。

(2) e^{-x^2} を $x = 0$ の回りで 3 次の項までテイラー展開せよ。

(3) $erf(x) = \dfrac{2}{\sqrt{\pi}} \displaystyle\int_0^x e^{-t^2} dt$ は誤差関数と呼ばれる。(2) で得たテイラー近似を $erf(x)$ の定義に代入して、項別に積分し、この積分の近似式を導出してみよ。

#11-2 正規分布の密度関数を

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{48}$$

と書くとき、この確率分布の期待値 $E[X]$ を求めよ。(答えは μ だから、計算過程を調べて、という課題)

6 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	#11 (2012.06.28 出題)
提出期限	2012 年 7 月 2 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 各自が自力で取り組むことを、切に願う 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 表紙はつけないこと 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

下記は見本である。

```
                                             提出日:2012/07/02
                     確率論レポート #11

                                             学籍番号:0000000
                                             氏　　名:電通　大

課題#11-1
```

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.