公開講座 2000年7月3日

統計数理研究所 下平英寿

午前： 統計学のおさらい（1–2章）

午後： AIC入門（3–4章）

参考書：情報量統計学，統計学辞典

「情報量統計学」の１章から４章まで（第I編 理論編）の範囲

統計学のおさらい

１章 確率と確率変数

２章 確率分布と統計モデル

AIC入門

３章 推定

４章 AIC

1 確率と確率変数

この章のトピック —— 確率変数

- 事象と確率
- 条件付確率と独立性
- 確率変数と分布関数
- 期待値
- 多次元の確率分布

1.1 事象と確率

「確率」の定義 —— あることがらが起こりうる割合．確からしさの度合い．

おなじ条件でくりかえし実験をおこなう（頻度論）

現代では「測度論」という数学を用いて，抽象的に定義される．

現象が必ずしも「確率的」でなくても，情報不足による不確実さを表現する手段として，現象を確率的と見なすと便利なことがある．

[例] 画びょうをなげる（上，下）

1	2	3	4	5	6	7	8	9	10
上	下	上	下	下	上	下	下	下	上
11	12	13	14	15	16	17	18	19	20
下	下	下	上	上	上	下	下	下	上
91	92	93	94	95	96	97	98	99	100
下	上	上	上	下	上	下	下	上	下

上＝43回，下＝57回

$$\frac{43}{100} = 0.43$$

画びょうが上になる割合

	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	1-100
上	4	4	4	3	6	3	3	5	6	5	43
下	6	6	6	7	4	7	7	5	4	5	57
上の割合	0.4	0.4	0.4	0.3	0.6	0.3	0.3	0.5	0.6	0.5	0.43

画びょうが上になる確率 ＝ 実験回数を増やしたときの極限

- **母集団 ＝ 実験回数が非常に多い(無限)のとき**
- **確率 ＝ 母集団における割合**

実際には有限回しか実験は行えない

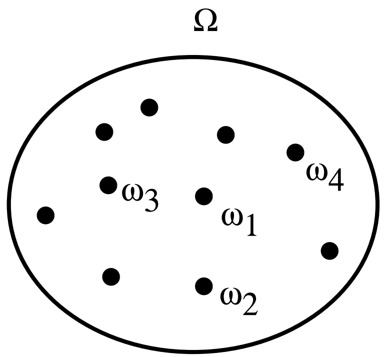
同じ条件で実験を多数繰り返すのは大変難しい

$\Rightarrow$ **確率というのは抽象的な概念**

実験の数学的モデル

測度論（集合，面積をあつかう数学理論）

確率空間 $(\Omega, \mathcal{F}, P)$ $\Omega =$ **標本空間，** $\mathcal{F} =$ **事象の全体，** $P =$ **確率**

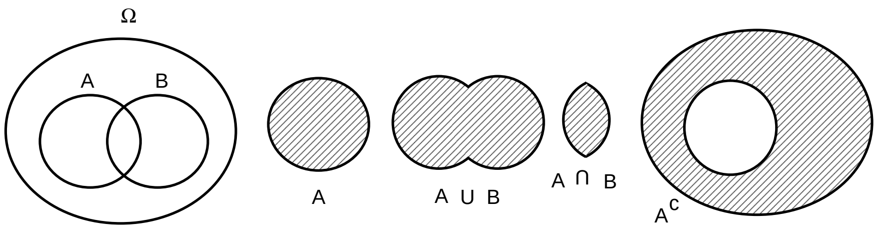


標本空間 $\Omega = \{\omega \mid \omega$ **は可能な標本点**$\}$

[標本空間の例]

- $\Omega = \{\omega_1, \omega_2\}$; $\omega_1 =$ **画びょうが上，** $\omega_2 =$ **画びょうが下**
- $\Omega = \{\omega_1, \ldots, \omega_6\}$; $\omega_1 =$ **サイコロの目が1,. . . ,** $\omega_6 =$ **サイコロの目が6**
- $\Omega = \{\omega_1, \ldots, \omega_{52}\}$; $\omega_1 =$ **ハートの1，. . . ，** $\omega_{52} =$ **スペードのキング**
- $\Omega = \{$**「表表」,「表裏」,「裏表」,「裏裏」**$\}$; **２枚のコイン**
- $\Omega = \{$**たろう, はなこ**$, \ldots$ **(クラスの人)**$\}$
- $\Omega = \{0, 1, 2, \ldots\}$; **0以上の整数**
- $\Omega = \{x \mid x \in$ **実数全体**$\}$

事象（Ωの部分集合）



- **事象** $A \subset \Omega$，**事象** $B \subset \Omega$
- A**と**B**の和事象** $A \cup B$ **（**A**と**B**のすくなくともどちらかが起こる）**
- A**と**B**の積事象** $A \cap B$ **（**A**と**B**の両方が起こる）**
- A**の余事象** A^c **（**A**が起こらない）**

[事象の例]

- $A = \{\omega\}$; **根元事象**
- $A =$ **画びょうが上** $= \{\omega_1\}$
- $A =$ **サイコロの目が偶数** $= \{\omega_2, \omega_4, \omega_6\}$
- $A =$**マークがスペード，**$B =$**数字が 1**
- A**と**B**の和事象: 少なくとも一方に属する標本の集合**
  $A \cup B =$ **スペードの 1 から 13 までと，ハート，ダイヤ，クラブの 1（16 枚）**
- A**と**B**の積事象: いずれにも属する標本の集合**
  $A \cap B =$ **スペードの 1**
- $A \cap B = \phi$**なら，**A**と**B**は排反（同時にはおこらない）**
- **余事象** A^c**は**A**に属さない標本の集合**
- **部分事象** A**が**B**に含まれるとき** $A \subset B$

事象の全体 $\mathcal{F}$

- $\Omega = \{\omega_1, \omega_2\}$ **のとき** $\mathcal{F} = \{\phi, \{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_2\}\}$
- $\Omega = \{\omega_1, \ldots, \omega_6\}$ **のとき** $\mathcal{F} = \{\phi, \{\omega_1\}, \ldots, \{\omega_1, \ldots, \omega_6\}\}$

完全加法族
- $\phi \in \mathcal{F}$
- $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
- $A_n \in \mathcal{F}, n = 1, 2, \ldots \Rightarrow A_1 \cup A_2 \cup \cdots \in \mathcal{F}$

確率の公理系
- $\mathcal{F}$**の上で定義された実数値関数**P **（これが確率を与える関数）**
- **任意の**$A \in \mathcal{F}$**に対して**$P(A) \geq 0$
- $P(\Omega) = 1$
- $A_i \cap A_j = \phi$ **ならば** $P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$

1.2 条件付確率と独立性

二つの事象A**と**B**について，**

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

ただし $P(B) > 0$ **とする．**
$P(A|B)$**は事象**B**を条件とする事象**A**の「条件付確率」という．**

[例] トランプで全てのカードの確率が等しく$\frac{1}{52}$**とする．**

$A =$**数字が 1 か 2，**$B =$**数字が奇数**

$P(A \cap B) = \frac{1}{13}$，$P(B) = \frac{7}{13}$

従って$P(A|B) = \frac{1}{7}$**であり，**$P(A) = \frac{2}{13}$**より小さい．**

“chain rule”

$$P(A|B) = P(A \cap B)/P(B)$$

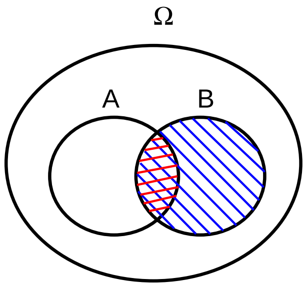
と書き換えると

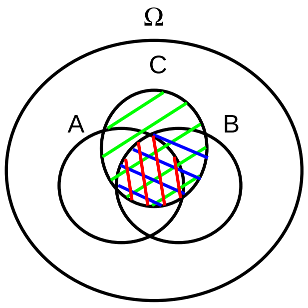
$$P(A \cap B) = P(A|B)P(B)$$

を得る．これを繰り返し用いると，

$$P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C)$$

などが得られる．





ベイズの定理

$H_1, H_2, \ldots$**のどの二つも互いに排反で，かつ** $H_1 \cup H_2 \cup \cdots = \Omega$**であるとき，**

$$P(H_k|A) = \frac{P(A|H_k)P(H_k)}{\sum_i P(A|H_i)P(H_i)}$$

H_k **原因，**A **結果**
$P(H_k)$ **事前確率，**$P(H_k|A)$ **事後確率**

[証明]

$$\sum_i P(A|H_i)P(H_i) = \sum_i P(A \cap H_i) = P(A)$$

これをベイズの定理の式の両辺にかけると

$$P(H_k|A)P(A) = P(A \cap H_k) = P(A|H_k)P(H_k)$$

となる．

適用上の注意点： 事前分布の決めかた

Thomas Bayes (1702-1761) イギリスの牧師

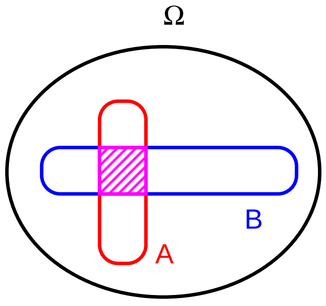
独立性

一般に$P(A \cap B) = P(A|B)P(B)$**であるが，とくに**$P(A|B) = P(A)$**のとき，**
事象A**と**B**は互いに「独立」という．このとき，**$P(A \cap B) = P(A)P(B)$**とかける．**

事象$A_1, A_2, \ldots$**が独立とは，**

$$P(A_1 \cap A_2 \cap \cdots) = P(A_1)P(A_2)\cdots$$

が成り立つとき．



P(A ∩ B) = P(A) P(B)

[独立性の例]

- $A =$**エース，**$B =$**ハートとする．全てのカードの確率は** $\frac{1}{52}$ **とする．**

  $P(A) = \frac{4}{52} = \frac{1}{13}$，$P(B) = \frac{13}{52} = \frac{1}{4}$，$P(A \cap B) = \frac{1}{52}$．

  $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{52}}{\frac{1}{4}} = \frac{1}{13} = P(A)$$

  したがってA**と**B**は独立．**

- **10個のコインを投げてすべてが表である確率は？**

  $$\frac{1}{2} \times \cdots \times \frac{1}{2} = \left(\frac{1}{2}\right)^{10} = \frac{1}{1024}$$

- **一台あたりの故障の確率が0.001の装置が100台すべて故障の無い確率は？**

  $$(1 - 0.001)^{100} = 0.905$$

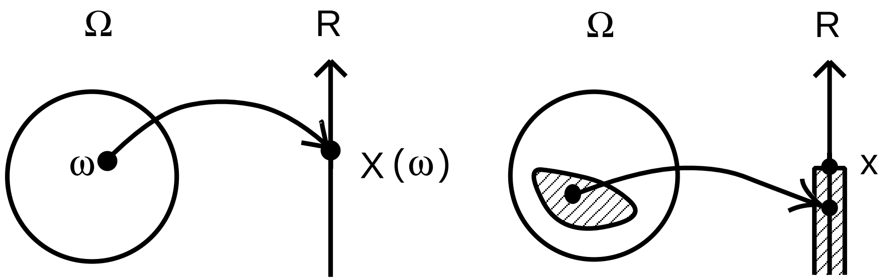
1.3　確率変数と分布関数

標本空間 Ω 上で定義された実数値関数 $X(\omega)$ があって任意の実数 x に対して $X(\omega)$ が x 以下となる様な標本点の集合

$$\{\omega | X(\omega) \leq x\}$$

に対してその確率が常に定義できるとき，$X(\omega)$ を「確率変数」と呼ぶ．

実験の結果 $X(\omega) = x$ が観測されたとき，x を確率変数 X の「実現値」と呼ぶ．



[確率変数の例]

- 画びょうが上のとき1，下のとき0となる確率変数は，

$$X(\omega) = \begin{cases} 1 & \omega = \omega_1 \text{のとき} \\ 0 & \omega = \omega_2 \text{のとき} \end{cases}$$

- 2枚のコインなげで，表の枚数

$$X(\omega) = \begin{cases} 2 & \omega = \text{「表表」のとき} \\ 1 & \omega = \text{「表裏」又は「裏表」のとき} \\ 0 & \omega = \text{「裏裏」のとき} \end{cases}$$

- $\Omega = \{$クラスの人$\}$

$$X(\omega) = \text{身長}, \quad \text{または} \quad X(\omega) = \text{体重}$$

- $\Omega =$ 実数全体のときに，$X(\omega) = \omega$

- 「離散型確率変数」　$X(\omega)$ のとる値が有限個または可算無限個
- 「連続型確率変数」　$X(\omega)$ のとる値が連続的

確率変数 X の「分布関数」

$$F(x) = P(\{\omega | X(\omega) \leq x\})$$

ただし，簡単のために

$$F(x) = P(X \leq x)$$

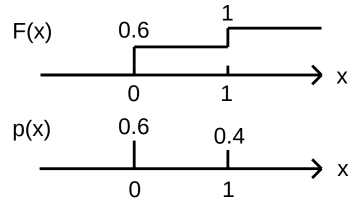
のように ω を明記しないことが多い．

実際の応用では，確率空間，分布関数を通して確率変数を定義することはあまりない．

後述の確率関数，密度関数を通して確率変数を導入するのが一般的．

[例] 画びょうなげ（上の確率が0.4とする）

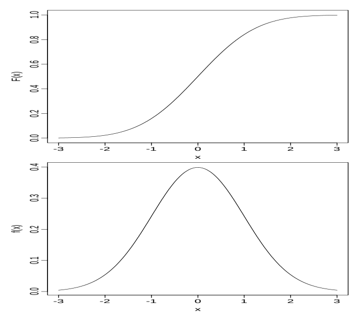
$$F(x) = \begin{cases} 0 & -\infty < x < 0 \\ 0.6 & 0 \leq x < 1 \\ 1 & 1 \leq x < \infty \end{cases}$$



[例] 正規分布（平均0，分散1）

$$F(x) = \int_{-\infty}^{x} \phi(y)\, dy$$

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$



確率関数

Xが離散型確率変数のときは,「確率関数」

$$p(x) = P(X = x)$$

を用いるのが便利．確率分布関数は

$$F(x) = \sum_{x_i \leq x} p(x_i)$$

と表現される．ただしx_i，$i = 1, 2, \ldots$は，xの取り得る値．

$$\sum_{x_i} p(x_i) = 1; \quad p(x_i) \geq 0$$

[例] 2枚のコインの表の枚数の確率関数は，表裏の確率が等しいとして，

$$p(0) = 1/4, \quad p(1) = 1/2, \quad p(2) = 1/4$$

(確率) 密度関数

Xが連続型確率変数のときは，(確率) 密度関数$f(x)$が便利．確率分布関数は

$$F(x) = \int_{-\infty}^{x} f(y)\, dy$$

と表現される．

$$\int f(x)\, dx = 1; \quad f(x) \geq 0$$

直観的には，ヒストグラムの極限 (データ数無限，分点の間隔ゼロ)

$$f(x) = \lim_{h \to +0} \frac{P(x < X \leq x + h)}{h} = \lim_{h \to +0} \frac{F(x+h) - F(x)}{h} = \frac{dF}{dx}$$

1.4 期待値

離散型確率変数Xが，$x_1, x_2, \ldots$という値を確率$p(x_1), p(x_2), \ldots$でとるものとする．

$$\mu = E[X] = \sum_i x_i p(x_i)$$

を確率変数の「期待値」あるいは平均と呼ぶ．

[例] サイコロの目がすべて同じ確率$\frac{1}{6}$ならば，
$x_1 = 1, x_2 = 2, \ldots, x_6 = 6$; $p(x_i) = \frac{1}{6}$とおいて，その期待値は

$$E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3.5$$

Xが連続型確率変数の場合には密度関数$f(x)$を用いて，

$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x)\, dx$$

確率変数Xのある関数を$g(X)$とするとその期待値は

$$E[g(X)] = \sum_i g(x_i) p(x_i) \quad \text{又は} \quad E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

とくに$g(X) = (X - \mu)^2$とおくと，Xの「分散」が得られる．

$$\sigma^2 = V[X] = E[(X - \mu)^2] = \sum_i (x_i - \mu)^2 p(x_i) \quad \text{又は} \quad \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx$$

$(X - \mu)^2$は，Xのμからのへだたりをあらわしている．

サイコロの目の分散は

$$V[X] = (1 - 3.5)^2 \times \frac{1}{6} + (2 - 3.5)^2 \times \frac{1}{6} + \cdots + (6 - 3.5)^2 \times \frac{1}{6} = \frac{35}{12} \approx 2.92$$

「標準偏差」 (standard deviation) は

$$\sigma = \sqrt{V[X]} = \sqrt{\frac{35}{12}} \approx 1.71$$

(公式)

$$V[X] = E[X^2] - (E[X])^2$$

[例] サイコロの目では $E[X^2] = 91/6$，$E[X]^2 = (7/2)^2$

$$\frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

[証明]

$$E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2\mu E[X] + \mu^2 = E[X^2] - \mu^2$$

[参考]

$$E[X+Y] = E[X] + E[Y]$$

$$E[aX] = aE[X]$$

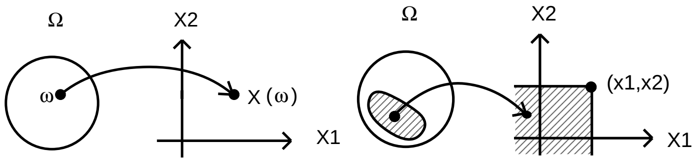
1.5　多次元の確率分布

多次元確率変数（k 次元確率ベクトル）

$$\mathbf{X}(\omega) = (X_1(\omega), \ldots, X_k(\omega))$$

同時分布関数

$$\begin{aligned} F(x_1, \ldots, x_k) &= P(X_1 \le x_1, \ldots, X_k \le x_k) \\ &= P(\{\omega | X_1(\omega) \le x_1, \ldots, X_k(\omega) \le x_k\}) \end{aligned}$$



[例] $\Omega = \{$**クラスの人**$\}$

$$X_1(\omega) = \omega\text{さんの身長}, \quad X_2(\omega) = \omega\text{さんの体重}$$

- X **が離散型の場合は，「同時確率関数」**

$$p(x_1, \ldots, x_k) = P(X_1 = x_1, \ldots, X_k = x_k)$$

をもちいて分布関数は

$$F(x_1, \ldots, x_k) = \sum_{y_1 \le x_1} \cdots \sum_{y_k \le x_k} p(x_1, \ldots, x_k)$$

- X **が連続型の場合は，「同時密度関数」**

$$f(x_1, \ldots, x_k)$$

をもちいて分布関数は

$$F(x_1, \ldots, x_k) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f(y_1, \ldots, y_k)\, dy_1 \cdots dy_k$$

同時確率関数（離散型）; 同時確率密度（連続型）

$$p(x,y) = P(X = x, Y = y); \qquad f(x,y) = \frac{d^2 F}{dxdy}$$

(X, Y) **の同時分布**

$$F(x,y) = P(X \le x, Y \le y) = \sum_{x' \le x} \sum_{y' \le y} p(x', y'); \qquad \int_{-\infty}^{x} \int_{-\infty}^{y} f(x', y')\, dx' dy'$$

X **の「周辺分布」の確率関数; 密度関数**

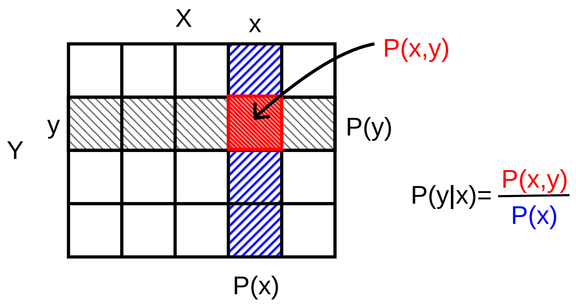
$$p(x) = \sum_y p(x,y); \qquad f(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$$

$X = x$ **の時の** Y **の「条件付」確率関数; 密度関数**

$$p(y|x) = \frac{p(x,y)}{p(x)}; \qquad f(y|x) = \frac{f(x,y)}{f(x)}$$

[注意] 本来なら $p_X(x)$，$p_Y(y)$，$p_{XY}(x,y)$，$p_{Y|X}(y|x)$ **などと添字をつけて区別すべきだが，特に混乱の恐れがない限り，簡単のため省略する．**

周辺分布，条件つき分布



$$A=\{\omega|Y(\omega)=y\},\qquad B=\{\omega|X(\omega)=x\}$$

$$P(B)=\sum_y p(x,y)=p(x),\qquad P(A\cap B)=p(x,y)$$

$$p(y|x)=P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{p(x,y)}{p(y)}$$

確率関数の "chain rule"

- $(X_1,\ldots,X_k)=(Y,Z)$

$$Y=(X_1,\ldots,X_m),\qquad Z=(X_{m+1},\ldots,X_k)$$

  と分割すれば

$$p(x_1,\ldots,x_m)=p(y)=\sum_z p(y,z)=\sum_{x_{m+1},\ldots,x_k} p(x_1,\ldots,x_k)$$

$$p(x_{m+1},\ldots,x_k|x_1,\ldots,x_m)=p(z|y)=\frac{p(y,z)}{p(y)}=\frac{p(x_1,\ldots,x_k)}{p(x_1,\ldots,x_m)}$$

- 繰り返し適用すると，

$$\begin{aligned}p(x_1,\ldots,x_k) &= p(x_k|x_1,\ldots,x_{k-1})p(x_{k-1}|x_1,\ldots,x_{k-2})\cdots\\ &\qquad\cdots p(x_3|x_1,x_2)p(x_2|x_1)p(x_1)\end{aligned}$$

時系列（マルコフモデル）

$$x_1,x_2,\ldots,x_k$$

が時系列のとき，しばしば１時刻前の影響だけを考えれば良いことがある．

$$p(x_m|x_1,\ldots,x_{m-1})=p(x_m|x_{m-1})$$

という仮定が用いられる．このとき "chain rule" より，

$$p(x_1,\ldots,x_k)=p(x_k|x_{k-1})p(x_{k-1}|x_{k-2})\cdots p(x_3|x_2)p(x_2|x_1)p(x_1)$$

もし２時刻まえの影響までを考慮するなら，

$$\begin{aligned}p(x_1,\ldots,x_k) &= p(x_k|x_{k-1},x_{k-2})p(x_{k-1}|x_{k-2},x_{k-3})\cdots\\ &\qquad\cdots p(x_4|x_3,x_2)p(x_3|x_2,x_1)p(x_2|x_1)p(x_1)\end{aligned}$$

確率変数の独立性

確率変数 X と Y が独立とは，事象 $\{\omega|X(\omega)\le x\}$ と $\{\omega|Y(\omega)\le y\}$ が独立のとき．

$$P(X\le x,Y\le y)=P(X\le x)P(Y\le y)$$

これより

$$p(x,y)=p(x)p(y);\qquad f(x,y)=f(x)f(y)$$

$X_1,X_2,\ldots,X_k$ が「独立」

$$p(x_1,\ldots,x_k)=p(x_1)\cdots p(x_k);\qquad f(x_1,\ldots,x_k)=f(x_1)\cdots f(x_k)$$

$x_1,\ldots,x_k$ を「時系列」とみなすと，これは前の時刻の影響が全くないモデルに相当する．

独立性と無相関

もし X と Y が独立なら，

$$E[g(X)h(Y)] = \int f(x)g(x)\,dx \int f(y)h(y)\,dy = E[g(X)]E[h(Y)]$$

したがって「共分散」(covariance)

$$C(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E[X-\mu_X]E[Y-\mu_Y] = 0$$

もしくは「相関」(correlation)

$$\rho(X,Y) = \frac{C(X,Y)}{\sqrt{V(X)V(Y)}} = 0$$

[注意] ただし $C(X,Y)=0$ だからといって X と Y が独立とは言えない．独立性の指標には「相互情報量」（後述）が便利．

(X,Y) が多変量正規分布の場合には，独立性と無相関はおなじ．

条件つき独立とマルコフ性

- 確率変数 X,Y,Z の同時分布が

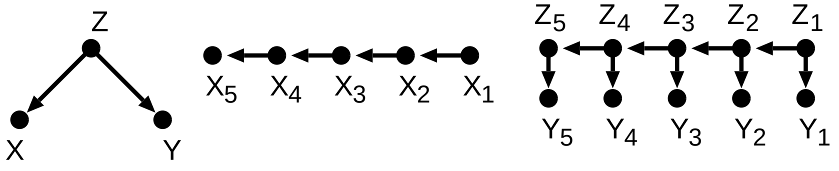
$$p(x,y,z) = p(x|z)p(y|z)p(z)$$

  を満たすとき，X と Y は Z の下で「条件つき独立」であるという．

- 同時分布のグラフ表現 (グラフィカルモデル)

$$p(x_m|x_1,\ldots,x_{m-1}) = p(x_m|x_i, i \in C_m)$$

  のとき X_m に向かう矢印を X_i，$i \in C_m$ から引く



2　確率分布と統計モデル

この章のトピック — パラメトリック・モデル

- 離散型確率分布
- 連続型確率分布
- 統計的モデル

2.1　離散型確率分布

1A ２項分布

1B ポアソン分布

1C 多項分布

2.2　連続型確率分布

2A 一様分布

2B 正規分布

2C 多次元正規分布

2D カイ２乗分布

1A．2項分布

n回コイン (画びょう) を投げたとき，表の出る回数kの分布．

$$p(k|\theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}; \quad k = 0, 1, \ldots, n$$

θ: １回のコイン投げで表の出る確率

[例] コイン投げ $\theta = 0.5$，$n = 10$として計算すると，

	0	1	2	3	4	5	6	7	8	9	10
２項分布	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001
正規近似*	.002	.010	.042	.113	.207	.252	.207	.113	.042	.010	.002
数値実験**	.001	.010	.045	.120	.199	.254	.204	.112	.047	.008	.001

＊ 後述の正規分布を利用して近似計算

＊ コンピュータによるシミュレーション 10回の画びょう投げ＝1セットとして，10000セットの実験 (3，4，6，6，5，7，2，3，5，6，. . . ，6，7，2，6，7，5，4，7，3，6.)

[導出] $(x_1, \ldots, x_n)$を各コイン投げでの表$x_i = 1$か裏$x_i = 0$を表すベクトルとする．

例えば，全て表なら$(1, \ldots, 1)$，裏なら$(0, \ldots, 0)$となる．

$$k = \sum_{i=1}^{n} x_i$$

$$\begin{aligned} p(x_1, \ldots, x_n|\theta) &= \theta^{x_1}(1-\theta)^{(1-x_1)} \times \cdots \times \theta^{x_n}(1-\theta)^{(1-x_n)} \\ &= \theta^{(\sum_i x_i)} \times (1-\theta)^{\sum_i (1-x_i)} \\ &= \theta^k (1-\theta)^{n-k} \end{aligned}$$

これに組合わせの数をかけて$p(k|\theta)$を得る．

平均，分散

$$\sum_{k=0}^{n} kp(k|\theta) = n\theta; \quad \sum_{k=0}^{n} (k - n\theta)^2 p(k|\theta) = n\theta(1-\theta)$$

$\theta = 0.5$，$n = 10$のとき　$E(k) = 5, V(k) = 2.5$

1B．ポアソン分布

毎日の事故の回数の分布．

$$p(k|\lambda) = \frac{\lambda^k}{k!} \exp(-\lambda); \quad k = 0, 1, 2, \ldots$$

平均，分散

$$E(k) = \sum_{k=0}^{\infty} kp(k|\lambda) = \lambda; \quad V(k) = \sum_{k=0}^{\infty} (k-\lambda)^2 p(k|\lambda) = \lambda$$

[参考] 先程の２項分布をp_{Bi}，ポアソン分布をp_{Po}と書くと，

$$\lim_{n\to\infty} p_{Bi}(k|\lambda/n) = p_{Po}(k|\lambda)$$

である．

1C．多項分布

（かたよりのある）サイコロをn回投げたときの出る目の分布．

$$p(k_1, \ldots, k_c|\theta_1, \ldots, \theta_c) = \frac{n!}{k_1! \cdots k_c!} \theta_1^{k_1} \cdots \theta_c^{k_c}$$

cは目の数，$k_1, \ldots, k_c$はそれぞれの目の出た回数，$\theta_1, \ldots, \theta_c$はそれぞれの目の確率．

$$\sum_{i=1}^{c} k_i = n, \quad k_i \geq 0; \quad \sum_{i=1}^{c} \theta_i = 1, \quad \theta_i \geq 0$$

$$\sum_{k_1, \ldots, k_c} p(k_1, \ldots, k_c|\theta_1, \ldots, \theta_c) = 1$$

[導出] 第i回目の試行でとる値をx_iとすると

$$p(x_1, \ldots, x_n|\theta_1, \ldots, \theta_c) = \theta_{x_1} \cdots \theta_{x_n} = \theta_1^{k_1} \cdots \theta_c^{k_c}$$

これに組合わせの数をかける．

[参考] $c=2$のとき，2項分布になる．

$$p(k_1,k_2|\theta_1,\theta_2)=\frac{n!}{k_1!k_2!}\theta_1^{k_1}\theta_2^{k_2}$$

ここで

$$k_1=k,\quad k_2=n-k;\quad \theta_1=\theta,\quad \theta_2=1-\theta$$

[参考] c個の独立なポアソン分布

$$p_{Po}(k_i|\lambda_i),\quad i=1,\dots,c$$

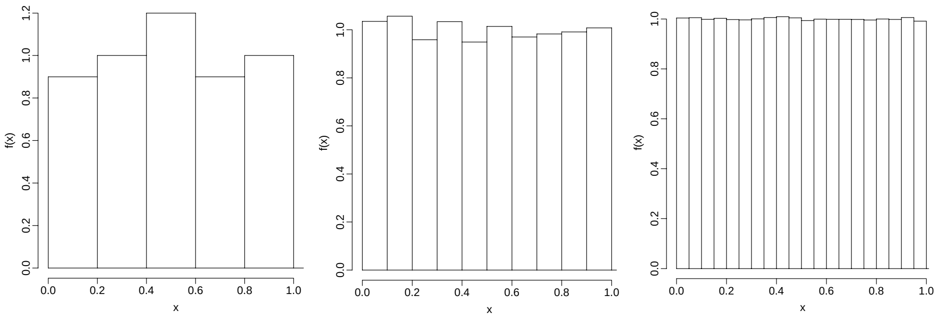
$\sum_i k_i=n$という条件付きでの$(k_1,\dots,k_c)$の分布は，多項分布

$$\frac{p_{Po}(k_1|\lambda_1)\cdots p_{Po}(k_c|\lambda_c)}{\sum_{k_1+\cdots+k_c=n}p_{Po}(k_1|\lambda_1)\cdots p_{Po}(k_c|\lambda_c)}=\frac{n!}{k_1!\cdots k_c!}(\lambda_1/n)^{k_1}\cdots(\lambda_c/n)^{k_c}$$

2A．一様分布

区間$[0,1]$上の一様分布の密度関数

$$f(x)=\begin{cases}1 & 0\le x\le 1\\ 0 & x<0,\ x>1\end{cases}$$



0.17, 0.66, 0.62, 0.48, 0.97, 0.65, 0.09, 0.10, 0.46, 0.28, 0.78, 0.73, 0.65, 0.37, 0.79, 0.39, 0.59, 0.36, 0.41, 0.64, ...

2B．正規分布

平均μ，分散σ^2の正規分布の密度関数

$$f(x|\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

確率変数Xがこの分布に従うとき

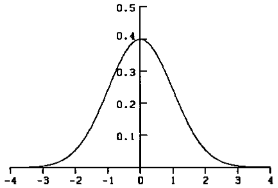
$$X\sim N(\mu,\sigma^2)$$

とかく．

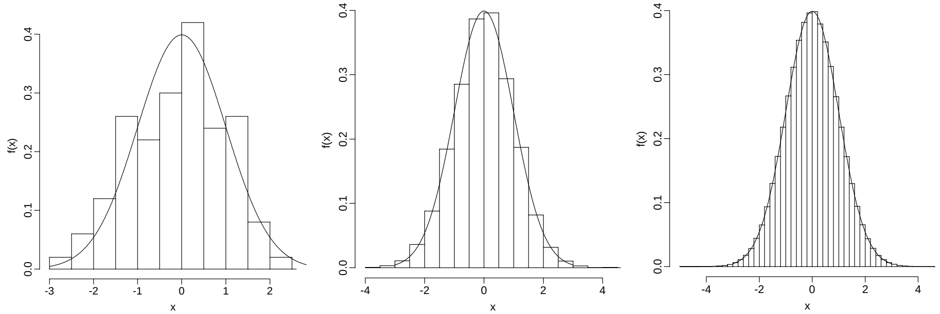
$$E(X)=\int_{-\infty}^{\infty}xf(x|\mu,\sigma^2)\,dx=\mu,\quad V(X)=\int_{-\infty}^{\infty}(x-\mu)^2f(x|\mu,\sigma^2)\,dx=\sigma^2$$

[参考] $X\sim N(0,1)$なら，$\sigma X+\mu\sim N(\mu,\sigma^2)$

標準正規分布 $N(0,1)$　　密度関数$\phi(x)$，分布関数$\Phi(x)=\int_{-\infty}^{x}\phi(y)\,dy$



$$\phi(x)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}$$



0.37, -0.57, 0.65, 1.89, 1.32, 1.52, 0.37, 0.36, -0.63, 0.90, 3.34, -0.71, -2.06, -0.09, -0.29, -0.60, 2.20, -1.43, -0.49, 0.02, ...

2C．多次元正規分布

平均ベクトル μ，共分散行列 Σ の k 次元正規分布の密度関数

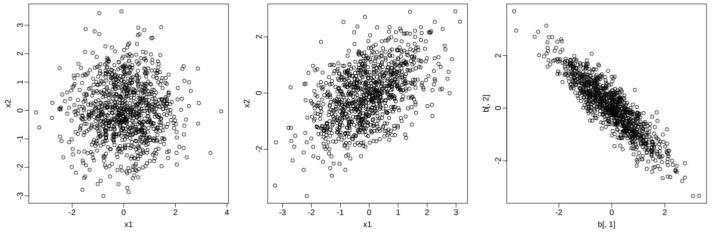
$$f(\boldsymbol{x}|\boldsymbol{\mu},\Sigma) = (2\pi)^{-k/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

確率変数 $\boldsymbol{X} = (X_1,\ldots,X_k)'$ がこの分布に従うとき，

$$\boldsymbol{X} \sim N(\boldsymbol{\mu},\Sigma)$$

とかく．ただし $\boldsymbol{x} = (x_1,\ldots,x_k)'$，$\boldsymbol{\mu} = (\mu_1,\ldots,\mu_k)'$，$\Sigma = (\sigma_{ij}; i,j = 1,\ldots,k)$．

$$E(\boldsymbol{X}) = \boldsymbol{\mu}, \quad V(\boldsymbol{X}) = E((\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})') = \boldsymbol{\Sigma}$$



$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}$$

- とくに $\mu = 0$，$\Sigma = \boldsymbol{I} = \mathrm{diag}(1,\ldots,1)$ ならば

$$\boldsymbol{X} \sim N(0,\boldsymbol{I})$$

$$f(\boldsymbol{x}|0,\boldsymbol{I}) = \prod_{i=1}^{k}\phi(x_i)$$

$X_1,\ldots,X_k$ は独立に $N(0,1)$ に従う．

- 任意の Σ（正定値行列）は

$$\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}'$$

と分解できるので，

$$\boldsymbol{L}\boldsymbol{X} + \boldsymbol{\mu} \sim N(\boldsymbol{\mu},\Sigma); \quad \boldsymbol{X} \sim N(0,\boldsymbol{I})$$

とかける．

2D．カイ２乗分布

$\boldsymbol{X} \sim N(0,\boldsymbol{I})$ のとき，$\|\boldsymbol{X}\|^2 = \sum_{i=1}^{k} X_i^2$ は，自由度 k のカイ２乗分布に従う．

$$\|\boldsymbol{X}\|^2 \sim \chi_k^2$$

とかく．

$$E(\|\boldsymbol{X}\|^2) = k; \quad V(\|\boldsymbol{X}\|^2) = 2k$$

[平均の導出]

$$E(\|\boldsymbol{X}\|^2) = E(X_1^2 + \cdots + X_k^2) = E(X_1^2) + \cdots + E(X_k^2) = k$$

$\boldsymbol{X} \sim N(\boldsymbol{\mu},\Sigma)$，$\Sigma = \boldsymbol{L}\boldsymbol{L}'$ ならば，

$$\boldsymbol{L}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) \sim N(0,\boldsymbol{I})$$

従って

$$\|\boldsymbol{L}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\|^2 = (\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) \sim \chi_k^2$$

極限定理

$\{X_1,X_2,\ldots\}$ が独立に同一の分布に従い，その平均と分散を μ，σ^2 とし，$n\to\infty$ の極限を考える．

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

[大数の法則]

$$\frac{S_n}{n} \to \mu$$

[中心極限定理]

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \to N(0,1)$$

2.3　統計的モデル

- パラメトリック・モデル
- モデルの合成
- モデルの制約
- リパラメトリゼイション
- 条件つき分布モデル

パラメトリック・モデル

$$X \sim g(\cdot); \quad g(x) \text{ は真の密度関数（or 確率関数）}$$

目的： データ x から $g(\cdot)$ を推定すること

「モデル」を用いる

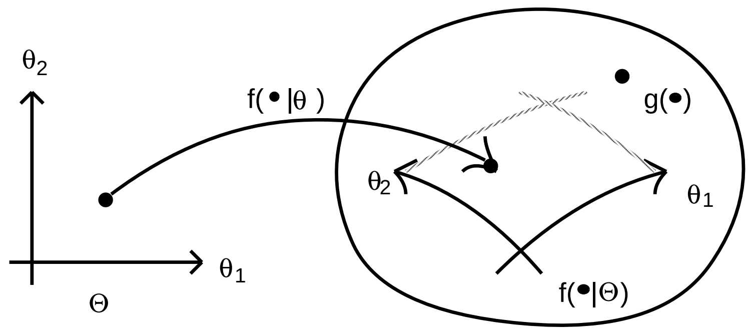
$$f(x|\boldsymbol{\theta})$$

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K); \quad \boldsymbol{\theta} \in \Theta = \{\boldsymbol{\theta} | \boldsymbol{\theta} \text{ に関する制約}\}$$

パラメタを動かして得られるモデル全体を

$$f(\cdot|\Theta) = \{f(x|\boldsymbol{\theta}) \,|\, \boldsymbol{\theta} \in \Theta\}$$

と書く．後述の「最尤法(さいゆうほう)」によって θ を推定する．



モデルは必ずしも正しいとは限らない

むしろ，現実の応用ではモデルが「正しい」ことはまれ

misspecification

自由なパラメタの数（自由パラメタ数）k

「パラメタのうち k 個の値を決めてしまうと残りのパラメタの値が一意に定まる」

[例] 正規分布 $N(\theta_1, \theta_2)$

$$f(x|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(-\frac{(x-\theta_1)^2}{2\theta_2}\right)$$

$$\boldsymbol{\theta} = (\theta_1, \theta_2) = (\mu, \sigma^2); \quad \Theta = \{\boldsymbol{\theta} \mid \theta_2 > 0\}$$

$$\text{自由パラメタ数} = 2$$

[例] 多項分布

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_c); \quad \Theta = \{\boldsymbol{\theta} \mid \sum_{i=1}^{c} \theta_i = 1, \theta_1 \geq 0, \ldots, \theta_c \geq 0\}$$

$$\text{自由パラメタ数} = c - 1$$

モデルの合成

- **「混合モデル」** $f(x|\boldsymbol{\theta}_f)$ **の** $g(x|\boldsymbol{\theta}_g)$ **混合は**

$$h(x|\boldsymbol{\theta}_h) = \alpha f(x|\boldsymbol{\theta}_f) + (1-\alpha) g(x|\boldsymbol{\theta}_g)$$

$$\boldsymbol{\theta}_h = (\alpha, \boldsymbol{\theta}_f, \boldsymbol{\theta}_g)$$

- **「指数型混合モデル」** $f(x|\boldsymbol{\theta}_f)$，$g(x|\boldsymbol{\theta}_g)$ **の混合は**

$$h(x|\boldsymbol{\theta}_h) = \exp\left\{\alpha \log f(x|\boldsymbol{\theta}_f) + (1-\alpha)\log g(x|\boldsymbol{\theta}_g)\right\} / C(\alpha, \boldsymbol{\theta}_f, \boldsymbol{\theta}_g)$$

- **「独立モデル」** $X \sim f(x|\boldsymbol{\theta}_f)$，$Y \sim g(y|\boldsymbol{\theta}_g)$ **の合成は**

$$h(x, y|\boldsymbol{\theta}_f, \boldsymbol{\theta}_g) = f(x|\boldsymbol{\theta}_f) g(y|\boldsymbol{\theta}_g)$$

モデルの制約

$$f(\cdot|\Theta) = \{f(x|\boldsymbol{\theta}) \,|\, \boldsymbol{\theta} \in \Theta\}$$

$$f(\cdot|\Theta') = \{f(x|\boldsymbol{\theta}) \,|\, \boldsymbol{\theta} \in \Theta'\}; \quad \Theta' \subset \Theta$$

リパラメトリゼイション

$$f(\cdot|\Lambda) = \{f(x|\boldsymbol{\lambda}) \,|\, \boldsymbol{\lambda} \in \Lambda\}$$

$$f(\cdot|\Theta) = \{f(x|\boldsymbol{r}(\boldsymbol{\theta})) \,|\, \boldsymbol{\theta} \in \Theta\}; \quad \boldsymbol{r}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta$$

[例] $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{I})$，$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$ **がパラメタ．** $\boldsymbol{\mu} = r(\theta)$ **の例は**

- $\mu_1 = \theta, \ldots, \mu_k = \theta$
- $\mu_i = \theta_1 + (i-1)\theta_2$，$i = 1, \ldots, k$
- $\mu_1 = \theta_1 \cos\theta_2 \sin\theta_3$, $\mu_2 = \theta_1 \sin\theta_2 \sin\theta_3$, $\mu_3 = \theta_1 \cos\theta_3$;
  $\theta_1 \geq 0$, $-\pi < \theta_2 \leq \pi$, $-\pi/2 \leq \theta_3 \leq \pi/2$

条件つき分布モデル

X の実現値 x が与えられたときの Y の条件つき分布 —— 回帰分析

X : **説明変数**, $\quad Y$: **目的変数**

モデル

$$f(y|\boldsymbol{\lambda}); \quad \boldsymbol{\lambda} = \boldsymbol{r}(x; \boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta$$

なお，説明変数のモデル $f(x|\boldsymbol{\theta}_X)$ **とあわせると，** (X, Y) **の同時分布のモデル**

$$f(x|\boldsymbol{\theta}_X) f(y|\boldsymbol{r}(x; \boldsymbol{\theta}))$$

[例] 正規回帰モデル： $Y \sim N(\mu(x), \sigma^2(x))$, $\quad \sigma^2(x) = \theta_1$

線形回帰： $\mu(x) = \theta_2 + \theta_3 x$
3次回帰モデル: $\mu(x) = \theta_2 + \theta_3 x + \theta_4 x^2 + \theta_5 x^3$
フーリエ級数： $\mu(x) = \theta_2 + \sum_{m=1}^{M} (\theta_{2m+1} \sin 2m\pi x + \theta_{2m+2} \cos 2m\pi x)$