統計解析応用研究
確率

丸山 祐造 Yuzo Maruyama

神戸大学　大学院経営学研究科

標本空間と事象 I

- 試行：実験や観測
- 標本空間 Ω（オメガと読む）：試行の結果として起こりうるもの全体
  - コイン投げ $\Omega = \{裏, 表\}$
  - サイコロ投げ $\Omega = \{1, 2, 3, 4, 5, 6\}$
- 標本空間 Ω は「数学的には」集合なので，部分集合などの概念や集合に対する演算が重要
- 事象：標本空間 Ω の部分集合

標本空間と事象 II

- 「事象Aが起こる」の意味：試行の結果としてAに含まれる要素の一つが起こる
- 2回のコイン投げ

$$\Omega = \{(裏, 裏), (裏, 表), (表, 裏), (表, 表)\},$$
$$A = \{(裏, 表), (表, 表)\}$$

- 全事象：標本空間Ω自身もΩの部分集合
- 空事象$\emptyset$：要素が何もない

標本空間と事象 III

事象の演算 ベン図

- A の余事象 A^c：A に含まれない要素からなる事象
  $A^c = \{x : x \notin A\}$
- 和事象 $A \cup B = \{x : x \in A \text{ or } x \in B\}$
  $\Uparrow$ 和事象と余事象の定義から $\Omega = A \cup A^c$
- 積事象 $A \cap B = \{x : x \in A \text{ and } x \in B\}$
- A と B は排反：$A \cap B = \emptyset$

標本空間と事象 IV

３つ以上の事象

- 和事象 $A_1 \cup A_2 \cup \cdots \cup A_n$
- 積事象 $A_1 \cap A_2 \cap \cdots \cap A_n$
- 例：Ω ２つのサイコロの出た目の和．$A_1 = \{$奇数$\}$，$A_2 = \{$偶数$\}$，$A_3 = \{3$の倍数$\}$，$A_4 = \{4$の倍数$\}$

$$A_1 \cup A_3 \cup A_4 = \{3, 4, 5, 6, 7, 8, 9, 11, 12\}$$
$$A_2 \cap A_3 \cap A_4 = \{12\}$$

標本空間と事象 V

3つの事象の分配法則，結合法則

$$\text{分配}\ (A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

$$\text{結合}\ (A \cup B) \cup C = A \cup (B \cup C)$$

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

ド・モルガンの法則

$$(A \cup B)^c = A^c \cap B^c,\ (A \cap B)^c = A^c \cup B^c$$

確率 I

確率

- 事象 A の起こりやすさを $[0,1]$ の実数で表す
  $\Uparrow$ 0 に近いほど起こりにくく，1 に近いほど起こりやすい

確率の定義

- 標本空間 Ω の任意の部分集合（事象）A に対して，実数 $P(A)$ が定まっていて，以下の３条件を満たすとき $P(A)$ を事象 A の確率という

確率 II

P1 任意の事象 A に対して $0 \leq P(A) \leq 1$

P2 全事象 Ω に対して，$P(\Omega) = 1$

P3 $A_1 \cap A_2 = \emptyset$ のとき

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

同時に起こりえない事象の和事象の確率
$=$ 各事象の確率の和

確率 III

(P3) について

- $A_1 \cap A_2 = \emptyset$ で「２つ」の排反な集合であることは本質的ではなく,「有限」であることが本質
- (P3) から「n 個の事象 $A_1, A_2, \ldots, A_n$ が互いに排反ならば,

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) \\ = P(A_1) + P(A_2) + \cdots + P(A_n)$$

が成り立つ」が導ける

確率 IV

- 例えば $n=3$ の場合
  - $A_1 \cap A_2 = \emptyset$, $A_2 \cap A_3 = \emptyset$, $A_3 \cap A_1 = \emptyset$
  - $A_1 \cup A_2 \cup A_3 = (A_1 \cup A_2) \cup A_3$ であることに注意.
  - 明らかに $(A_1 \cup A_2) \cap A_3 = \emptyset$
  - $P((A_1 \cup A_2) \cup A_3) = P(A_1 \cup A_2) + P(A_3)$
  - 第一項の $P(A_1 \cup A_2)$ に対して，(P3) を適用

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3)$$

確率の基本的性質 I

(P1)–(P3) から導出される性質・公式

1. $P(A^c) = 1 - P(A)$
   - $\Omega = A \cup A^c$, $A \cap A^c = \emptyset$
   - (P3) より $P(\Omega) = P(A) + P(A^c)$
   - (P2) より $P(\Omega) = 1$
2. $A \subset B \Rightarrow P(A) \leq P(B)$
   - $B = A \cup (B \cap A^c)$ でありまた $A \cap (B \cap A^c) = \emptyset$
   - (P3) より $P(B) = P(A) + P(B \cap A^c)$
   - (P1) より $P(B \cap A^c) \geq 0$

確率の基本的性質 II

3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ ベン図

- $A \cup B = (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$
- 右辺の和事象は互いに排反

(P3) より $P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(A^c \cap B)$

- $A = (A \cap B^c) \cup (A \cap B)$ で右辺の和事象は互いに排反

(P3) より $P(A) = P(A \cap B^c) + P(A \cap B)$

- $B = (A \cap B) \cup (A^c \cap B)$ で右辺の和事象は互いに排反

(P3) より $P(B) = P(A \cap B) + P(A^c \cap B)$

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

確率の基本的性質 III

4. 標本空間が有限集合の場合の公式
   - 標本空間が有限集合 $\Omega = \{x_1, \ldots, x_N\}$
   - $P(\{x_1\}) = P(\{x_2\}) = \cdots = P(\{x_N\})$
     $\Downarrow$
   - 任意の事象 A の確率は

$$P(A) = \frac{A\text{に含まれる要素数}}{N}$$

$\Uparrow$ 当たり前のように感じられるが，「確率の定義」(P1)–(P3) から導出される「性質」（定理）である

条件付き確率 I

条件付き確率の定義

- 事象 B が起きたことが分かっている条件のもとで事象 A の起こる確率「事象 B が与えられたときの事象 A の条件付き確率」

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad \text{ただし } P(B) > 0$$

- 事象 B の確率に対する事象 $A \cap B$ の確率の比
- 事象 B が起きたことが分かっているという条件下では全事象は B
- その場合 A が起きることと $A \cap B$ が起きることは同じ

条件付き確率 II
くじ引きの例

- 10 本中 3 本当たり．B 君，A 君の順に引く

$$P(B) = \frac{3}{10}, \quad P(B^c) = \frac{7}{10},$$ B: B 君が当たり

- このとき，

$$P(A \mid B) = \frac{2}{9}, \quad P(A \mid B^c) = \frac{3}{9},$$ A: A 君が当たり

乗法公式 $\Leftarrow$ 条件付き確率の定義から明らか

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

条件付き確率 III

くじ引きの例（再び）：くじ引きの順番とくじの当たる確率は無関係

$$P(A \cap B) = P(A \mid B)P(B) = \frac{2}{9} \times \frac{3}{10} = \frac{1}{15}$$
$$P(A \cap B^c) = P(A \mid B^c)P(B^c) = \frac{3}{9} \times \frac{7}{10} = \frac{7}{30}$$
$$A = (A \cap B) \cup (A \cap B^c) \quad \text{ただし}, (A \cap B) \cap (A \cap B^c) = \emptyset$$
$$P(A) = P(A \cap B) + P(A \cap B^c) = \frac{1}{15} + \frac{7}{30} = \frac{3}{10} = P(B)$$

全確率公式

全確率公式 $$P(A) = \sum_{i=1}^{n} P(A \mid H_i)P(H_i)$$

ただし，$\begin{cases} \text{標本空間の分割 } \Omega = H_1 \cup \cdots \cup H_n \\ H_i \cap H_j = \emptyset \ \ \text{for } i \neq j \end{cases}$

証明

- $A = A \cap \Omega$ であり，また分配法則より

$$A = A \cap (H_1 \cup \cdots \cup H_n) = (A \cap H_1) \cup \cdots \cup (A \cap H_n)$$

- $(A \cap H_1) \ldots (A \cap H_n)$ は互いに排反なので

$$P(A) = \sum_{i=1}^{n} P(A \cap H_i) = \sum_{i=1}^{n} P(A \mid H_i)P(H_i)$$

最初の等号：(P3)，二番目の等号：乗法公式

3 囚人問題 I

3 囚人問題 $\Leftarrow$ 条件付き確率の有名な例

- 3 人の死刑囚 A,B,C
- 恩赦が出て 3 人のうちランダムに選ばれた 1 人だけ助かることになった
- 誰が恩赦になるかは明かされておらず，それぞれの囚人が「私は助かるのか?」と聞いても看守は答えない
- 囚人 A が恩赦になる確率はこの時点では $1/3$

3 囚人問題 II

- A$\rightarrow$ 看守 :「私以外の 2 人のうち少なくとも 1 人は死刑になるはず．その者の名前は？ 私のことじゃないんだから教えてくれてもよいだろう」
- 看守 $\rightarrow$A :「B は死刑になる」
- それを聞いた囚人 A は喜んだ
  $\Uparrow$ B が死刑になる事は確定した以上，恩赦になるのは A か C のいずれか一方．A が恩赦になる確率は $1/2$
- 囚人 A が喜んだのは正しいか?

3 囚人問題 III

- $\tilde{A}$，$\tilde{B}$，$\tilde{C}$：A,B,C が恩赦
- $P(\tilde{A}) = P(\tilde{B}) = P(\tilde{C}) = 1/3$
- $\mathcal{W}$：「看守が『B が死刑』と言う」事象
- 考慮すべき条件付き確率

$$P(\tilde{A} \mid \mathcal{W}) = \frac{P(\tilde{A} \cap \mathcal{W})}{P(\mathcal{W})} = \frac{P(\tilde{A})P(\mathcal{W} \mid \tilde{A})}{P(\mathcal{W})} = \frac{P(\mathcal{W} \mid \tilde{A})}{3P(\mathcal{W})}$$

3 囚人問題 IV

- 全確率公式より

$$
\begin{aligned}
P(\mathcal{W}) &= P(\tilde{A})P(\mathcal{W} \mid \tilde{A}) + P(\tilde{B})P(\mathcal{W} \mid \tilde{B}) + P(\tilde{C})P(\mathcal{W} \mid \tilde{C}) \\
&= \frac{P(\mathcal{W} \mid \tilde{A}) + 1}{3}
\end{aligned}
$$

$\Uparrow$ $P(\mathcal{W} \mid \tilde{B}) = 0$ & $P(\mathcal{W} \mid \tilde{C}) = 1$「看守は嘘をつかない」という前提

- よって，$P(\tilde{A} \mid \mathcal{W}) = \dfrac{P(\mathcal{W} \mid \tilde{A})}{P(\mathcal{W} \mid \tilde{A}) + 1}$

3 囚人問題 V

- 「A が恩赦」つまり $\tilde{A}$ のもとで,「B が死刑」と「C が死刑」のどちらを看守が選ぶかは「他に何も情報がない以上」等しいと考えるのが妥当

$$P(\mathcal{W} \mid \tilde{A}) + P(\text{「看守が『C が死刑』と言う」事象} \mid \tilde{A}) = 1$$
$$\Rightarrow \ P(\mathcal{W} \mid \tilde{A}) = 1/2$$

- つまり $P(\tilde{A} \mid \mathcal{W}) = \dfrac{1/2}{1/2 + 1} = \dfrac{1}{3}$

3囚人問題 VI

- 一方，$\mathcal{W}$ と $\tilde{B}^c$ を混同したとする
  - $\mathcal{W}$:「看守が『B が死刑』と言う」事象
  - $\tilde{B}^c$: B が恩赦ではない

$$P(\tilde{A} \mid \tilde{B}^c) = \frac{P(\tilde{A} \cap \tilde{B}^c)}{P(\tilde{B}^c)} = \frac{P(\tilde{A} \cap (\tilde{A} \cup \tilde{C}))}{P(\tilde{A} \cup \tilde{C})}$$

$$= \frac{P(\tilde{A})}{P(\tilde{A}) + P(\tilde{C})} = \frac{1}{2}$$

- 条件付き確率においては「条件となる事象」をきちんと考えないといけないという例

ベイズの定理 I

ベイズの定理

$$P(H_k \mid E) = \frac{P(E \mid H_k)P(H_k)}{\sum_{i=1}^n P(E \mid H_i)P(H_i)} \quad \text{for } k = 1, \ldots, n$$

$\Downarrow$ 証明 $\qquad \Uparrow \begin{cases} \Omega = H_1 \cup \cdots \cup H_n \\ H_i \cap H_j = \emptyset \quad \text{for } i \neq j \end{cases}$

- $P(H_k \mid E) = \dfrac{P(H_k \cap E)}{P(E)}$
- 分子：乗法公式より $P(E \mid H_k)P(H_k)$ に等しい
- 分母：全確率公式より $\sum_{i=1}^n P(E \mid H_i)P(H_i)$ に等しい

ベイズの定理 II

用語：$P(H_k)$ 事前確率，$P(H_k \mid E)$ 事後確率

ベイズの定理の適用例

- 壺 1：〇〇●，壺 2:〇●●
- 等確率で壺 1,2 のいずれかを選び（$P(H_1) = P(H_2) = 1/2$），玉を抽出したところ「〇」
- $P(〇 \mid H_1) = 2/3$，$P(〇 \mid H_2) = 1/3$
- ベイズの定理より

$$P(H_1 \mid 〇) = \frac{(2/3) \times (1/2)}{(1/2) \times (2/3) + (1/2) \times (1/3)} = \frac{2}{3}$$

- 同様に $P(H_2 \mid 〇) = 1/3$

事象の独立性 I

2 つの事象 A と B が独立

$$
\begin{aligned}
&\Leftrightarrow \; P(A|B) = P(A) \\
&\Leftrightarrow \; P(A \cap B) = P(A)P(B) \\
&\Leftrightarrow \; P(B|A) = P(B)
\end{aligned}
$$

このうち $P(A \cap B) = P(A)P(B)$ は 2 つ以上の事象への拡張に有効

事象の独立性 II

n 個の事象 $E_1, \dots, E_n$ の独立性

$2 \leq m \leq n$ なる整数 m を任意に選び，m 個の事象を $E_1, \dots, E_n$ の中から任意に選んだ事象を $F_1, \dots, F_m$ とする．

$$P(F_1 \cap F_2 \cap \cdots \cap F_m) = P(F_1)P(F_2)\dots P(F_m)$$

が成り立つとき，$E_1, \dots, E_n$ は独立．

$\Uparrow$ 3事象の独立性を示す場合に「いくつの」「どのような」等式を示す必要があるか？

事象の独立性 III

$$
\begin{aligned}
P(A \cap B \cap C) &= P(A)P(B)P(C) \\
P(A \cap B) &= P(A)P(B) \\
P(B \cap C) &= P(B)P(C) \\
P(C \cap A) &= P(C)P(A)
\end{aligned}
$$

離散確率変数 I

確率変数
いろいろな値（実数）をいろいろな確率でとるような変数

$\Uparrow$ 先ほどの節の「確率」との関係が不明瞭．その繋がりは（可測）関数．wikipedia へのリンク

例

- 歪みのないサイコロは $1, \ldots, 6$ の値をそれぞれ確率 $1/6$ でとる
- コインを投げて表が出れば 1，裏ならば 0

離散確率変数 II

離散確率変数
サイコロの目やコインの様に取り得る値が有限個
（あるいは加算無限個）の確率変数

記法のルール
大文字 X：確率変数，$P()$：確率

サイコロ　$P(X = 1) = \cdots = P(X = 6) = 1/6$

離散確率変数 III

サイコロの目の確率関数

$$p(x) = P(X = x) = 1/6, \quad x = 1, \ldots, 6$$

- 小文字 x でその実現値を表す
- 確率関数 $p(x)$：x での確率を x の関数とみたもの

標本空間：実現値の集合

- サイコロ：$\{1, 2, \ldots, 6\}$
- 将棋の駒（7 面体）：まわり将棋
- 宝くじ：$\{0, 200, \ldots, 40000000\} \Leftarrow$ 後で紹介

離散確率変数 IV

確率関数が満たすべき条件

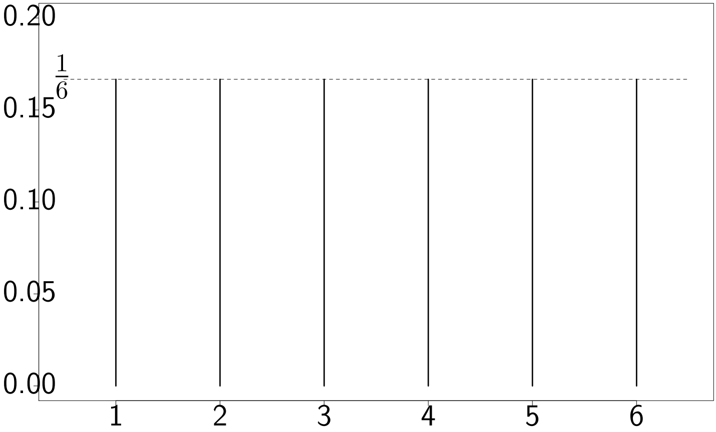
- $p(x) \geq 0$
- $\sum_{x\in\Omega} p(x) = 1$, Ω：標本空間

データの度数分布と確率関数の関係

- $p(x) \Leftrightarrow x$ に落ちた観測値の相対度数
- $p(x)$ のグラフ $\Leftrightarrow$ ヒストグラム

離散確率変数 V

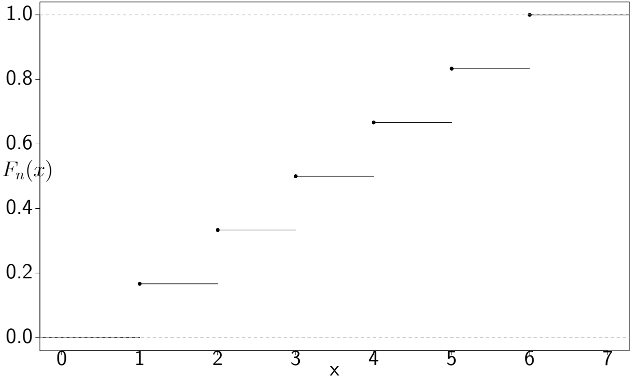
サイコロにおける確率関数 $P(X = x)$



離散確率変数 VI

確率変数 X の累積分布関数 $F(x) = P(X \leq x)$

サイコロの場合



離散確率変数 VII

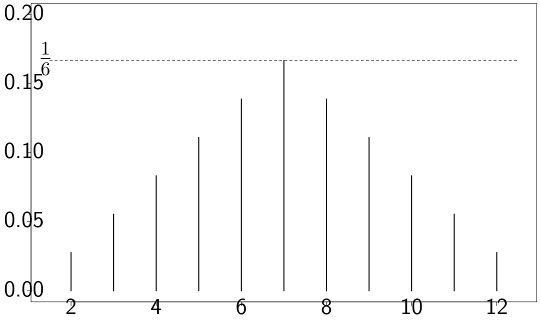
独立なサイコロ２つの目の和の確率関数 $P(X = x)$

- 2: $(1, 1)$
- 3: $(1, 2)$, $(2, 1)$
- 4: $(1, 3)$, $(3, 1)$, $(2, 2)$
- 5: $(1, 4)$, $(4, 1)$, $(2, 3)$, $(3, 2)$
- 6: $(1, 5)$, $(5, 1)$, $(2, 4)$, $(4, 2)$, $(3, 3)$
- 7: $(1, 6)$, $(6, 1)$, $(2, 5)$, $(5, 2)$, $(3, 4)$, $(4, 3)$
- 8: $(2, 6)$, $(6, 2)$, $(3, 5)$, $(5, 3)$, $(4, 4)$
- 9: $(3, 6)$, $(6, 3)$, $(4, 5)$, $(5, 4)$
- 10: $(4, 6)$, $(6, 4)$, $(5, 5)$
- 11: $(5, 6)$, $(6, 5)$
- 12: $(6, 6)$

$\Uparrow$ 各事象の確率は $(1/6) \times (1/6) = 1/36$

離散確率変数 VIII
独立なサイコロ２つの目の和の確率関数

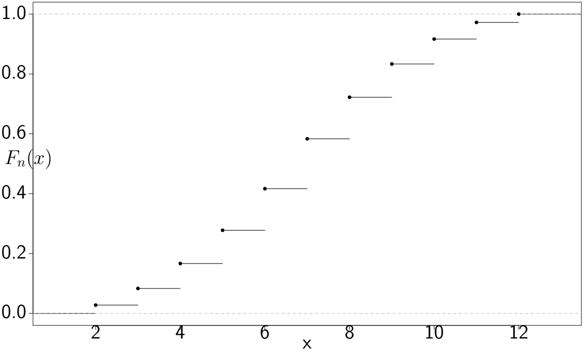
$$P(X=i)=\frac{\min(i-1,13-i)}{36},\quad i=2,3,4,\dots,12$$



離散確率変数 IX

確率変数 X の累積分布関数 $F(x) = P(X \leq x)$

サイコロ 2 個の場合



離散確率変数の期待値 I

確率変数の期待値

- データの平均値と対応させて定義
- サンプルサイズ n のデータを $y_1, \ldots, y_n$ とし，これらの観測値に重複がある場合を考える
- 異なる値 $x_1, \ldots, x_k$， x_i の度数を f_i
- $\bar{y} = \frac{1}{n}(y_1 + \cdots + y_n) = \frac{f_1}{n}x_1 + \cdots + \frac{f_k}{n}x_k$
- 離散確率変数の期待値：相対頻度 f_i/n を $p(x_i)$ で置換

$$\mathrm{E}(X) = \mu = \textstyle\sum_{x \in \Omega} x p(x)$$

$\Uparrow$ 期待値（母平均）は（標本平均と同じく）確率変数が従う分布の位置の指標（分布の中心）

離散確率変数の期待値 II
確率変数の分散

$$\mathrm{Var}(X) = \sigma^2 = \sum_{x \in \Omega} (x-\mu)^2 p(x)$$

- データの標本分散 $\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}$ との比較

$$= \underbrace{\frac{f_1}{n}}_{\text{置換} P(X=x_1)} (x_1 - \underbrace{\bar{y}}_{\text{置換}\mu})^2 + \cdots + \underbrace{\frac{f_k}{n}}_{\text{置換} P(X=x_k)} (x_k - \underbrace{\bar{y}}_{\text{置換}\mu})^2$$

⇑ 分散：（標本分散と同じく）確率変数が従う分布が，『分布の中心である母平均』からどれだけばらつくのかを表す

離散確率変数の期待値【例１】I

宝くじの例 (1990 年ごろの東京都宝くじ)

- 10 万通を一組として，01 組から 130 組までの 1300 万通 26 億円を売り出す（一通 200 円）
- 抽選前のある一通の当選金を X とする

$$P(X = 40000000) = \frac{7}{130000000}, \dots, P(X = 200) = \frac{1}{10}$$

  など．9 割近くが空くじ
- 期待値は一枚当たりの予想当選金

$$\mathrm{E}(X) = 40000000 \times \frac{7}{130000000} + \cdots + 200 \times \frac{1}{10}$$

離散確率変数の期待値【例 1】II

等級	当選金	本数
1 等	4 千万円	7
1 等前後	1 千万円	14
1 等組違い	20 万円	903
2 等	1 千万円	5
2 等組違い	10 万円	645
3 等	100 万円	130
4 等	14 万円	130
5 等	1 万円	1300
6 等	1 千円	26000
7 等	200 円	1300000

離散確率変数 $g(X)$ の期待値

確率変数 $g(X)$

- X が確率変数（いろいろな値をいろいろな確率で取る変数）ならば，$g(X)$ も確率変数
- $g(X)$ の期待値の定義

$$\mathrm{E}(g(X)) = \sum_x g(x)p(x)$$

- 分散は $g(x) = (x - \mu)^2$ の場合

$$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \mathrm{E}\left[(X - \mathrm{E}(X))^2\right]$$

連続確率変数 I

連続確率変数 $\Rightarrow$ 取り得る値が連続な確率変数

例：ルーレットの止まる位置

- 固定された起点と針が止まった位置までの角度 θ
- $X = \dfrac{\theta}{2\pi}$ ：0 から 1 の間の確率変数
- 確率 $P(0 \le X < x)$ は，ちょうど角度の割合 x

$$P(0 \le X < x) = x,\ 0 < x \le 1$$

連続確率変数 II

連続確率変数においては，1 点の確率は 0

- 任意の x と $\epsilon > 0$ に対して，

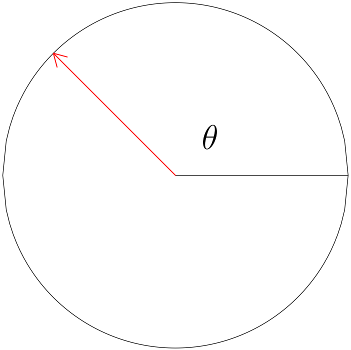
$$P(x \leq X < x+\epsilon)$$
$$= P(0 \leq X < x+\epsilon) - P(0 \leq X < x) = (x+\epsilon) - x = \epsilon$$

- $\epsilon \to 0$ とすると，$P(X = x) = 0$

復習：サイコロ（離散確率変数）

$$p(x) = P(X = 1) = \frac{1}{6} > 0$$

連続確率変数 III



連続確率変数 IV

確率密度関数

- 連続確率変数では確率関数を考えることは出来ない
- かわりに x のまわりに分布している確率の「濃さ」を考える
- 点 x での確率密度関数 $f(x)$ の定義

$$f(x) = \lim_{\epsilon \to 0} \frac{P(x < X \leq x + \epsilon)}{\epsilon} \tag{1}$$

連続確率変数 V

- Δx を微小量とすると，

$$P(x < X < x + \Delta x) \doteq f(x)\Delta x$$

と書けるので，密度関数は x のまわりでの区間の長さと比較した確率の濃さを表す

累積分布関数

$$F(x) = P(X \leq x) = P(-\infty < X \leq x)$$

- 離散，連続とも累積分布関数は上で定義される
- ただし，連続の場合次の頁の関係が重要

連続確率変数 VI

- 連続の場合の f（密度）と F（累積）の関係

$$f(x) = F'(x)$$

- (1) の分子 $P(x < X \leq x + \epsilon)$

$$\begin{aligned} &= P(-\infty < X \leq x + \epsilon) - P(-\infty < X \leq x) \\ &= F(x + \epsilon) - F(x) \end{aligned}$$

- よって $f(x) = \lim_{\epsilon \to 0} \dfrac{F(x+\epsilon) - F(x)}{\epsilon} = F'(x)$

連続確率変数 VII

区間の確率 $P(a < X \leq b)$

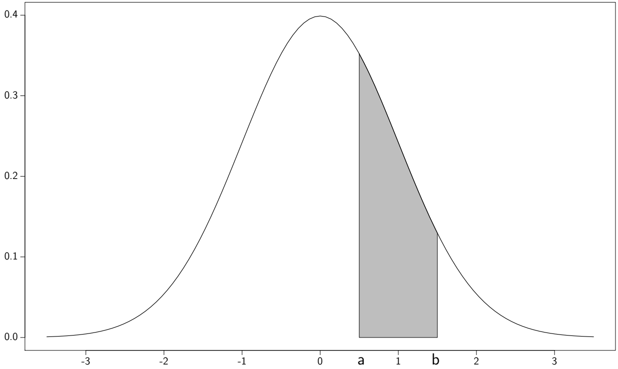
$x = a$, $x = b$, $y = 0$, $y = f(x)$ で囲まれる領域の面積と一致 $\Rightarrow$ 次ページ

$$P(a < X \leq b) = \int_a^b f(x)\mathrm{d}x$$

なお１点の確率はゼロなので，以上及び以後で端の $<, \leq$ は区別しない

$$\begin{aligned} P(a < X < b) &= P(a \leq X \leq b) \\ &= P(a < X \leq b) = P(a \leq X < b) \end{aligned}$$

連続確率変数 VIII



連続確率変数 IX

ルーレットの例

$$F(x) = x, f(x) = \begin{cases} 1, & \text{if } 0 \leq x < 1 \\ 0, & \text{otherwise} \end{cases}$$

- 0 と 1 の間で密度関数が平 $\Rightarrow$ 0 と 1 の間のどの点も「同様に確からしい」ことを表す
- この分布を 0 と 1 の間の一様分布とよぶ
- このような「具体的な」分布は後でまとめて紹介

連続確率変数 X

f が確率密度関数であるための必要十分条件

$$f(x) \geq 0,\ \int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$$

- 連続確率変数 X が与えられたとき，

$$f(x) = \lim_{\epsilon \to 0} \frac{P(x < X \leq x + \epsilon)}{\epsilon}$$

  と定義すると，$f(x) \geq 0,\ \int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$ を満たす
- $f(x) \geq 0,\ \int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$ を満たす f が与えられたとき，f を確率密度関数とするような確率変数 X が存在する

補足：広義積分

広義積分：無限区間での積分 $\Leftarrow$ 定積分の極限

$$\int_{-\infty}^{b} f(x)\mathrm{d}x = \lim_{a\to-\infty}\int_{a}^{b} f(x)\mathrm{d}x,$$

$$\int_{a}^{\infty} f(x)\mathrm{d}x = \lim_{b\to\infty}\int_{a}^{b} f(x)\mathrm{d}x$$

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \lim_{\substack{a\to-\infty\\ b\to\infty}}\int_{a}^{b} f(x)\mathrm{d}x$$

例：$f(x) = 1/x^2$ の $(1,\infty)$ での広義積分．原始関数 $F(x) = -1/x$

$$\int_{1}^{b} f(x)\mathrm{d}x = [-1/x]_1^b = -1/b + 1 \to 1 \text{ as } b \to \infty$$

連続確率変数の期待値 I

連続確率変数の期待値

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$$

離散確率変数に近似する解釈

- 実数軸上に Δx の間隔で点をとる $x_i = i\Delta x,\ i = 0, \pm 1, \pm 2, \ldots$ 添字は正の整数が普通であるが，ここではやや一般化 $\Uparrow$
- 実数軸 $= \cup (x_i, x_{i+1}]$ と考える

連続確率変数の期待値 II

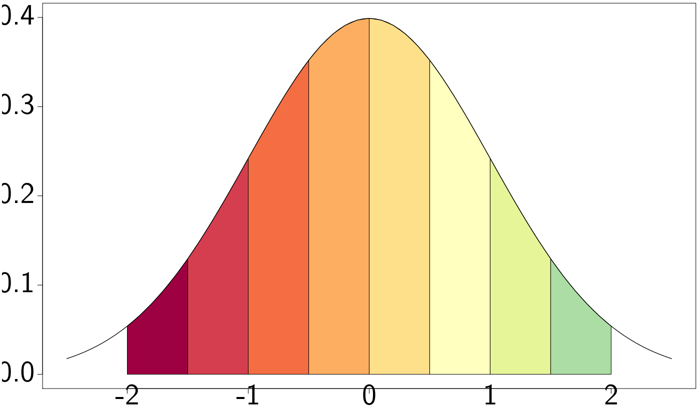
- $(x_i, x_{i+1}]$ の区間の左端の値を取るルールにより離散確率変数を作る

$$\tilde{X} = x_i \text{ if } x_i < X \leq x_{i+1} \tag{2}$$

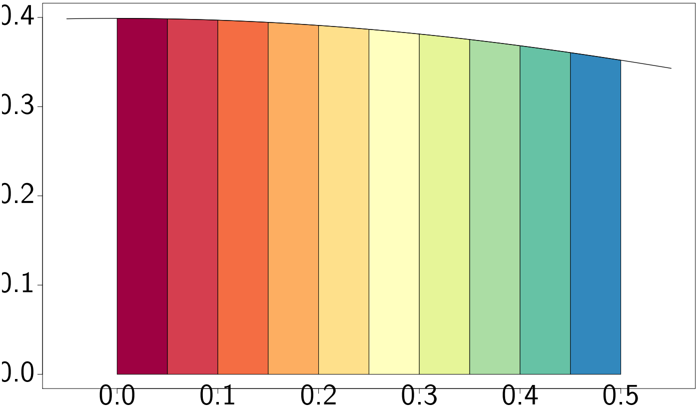
- このとき $P(\tilde{X} = x_i) = P(x_i < X \leq x_{i+1})$
- $p_i = P(\tilde{X} = x_i)$ とすると，離散確率変数 $\tilde{X}$ の期待値

$$\mathrm{E}(\tilde{X}) = \sum_{i=-\infty}^{\infty} x_i p_i \tag{3}$$

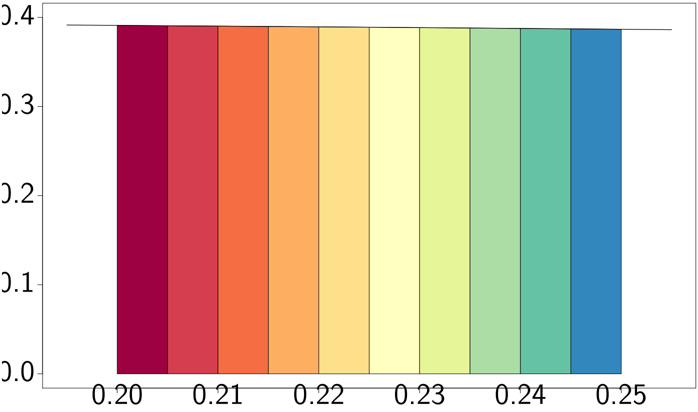
連続確率変数の期待値 III



連続確率変数の期待値 IV



連続確率変数の期待値 V



連続確率変数の期待値 VI

- 3 枚目のように 0.005 刻みくらいになると，密度関数はその区間においてほぼ水平なので以下の近似が有効

$$
\begin{aligned}
p_i = P(x_i < X \le x_{i+1}) &= \int_{x_i}^{x_{i+1}} f(x)\mathrm{d}x \\
&\approx f(x_i)\int_{x_i}^{x_{i+1}} \mathrm{d}x = f(x_i)\{x_{i+1} - x_i\} = f(x_i)\Delta x
\end{aligned} \tag{4}
$$

- $\tilde{X}$ の期待値の近似値

$$
\mathrm{E}(\tilde{X}) = \sum_{i=-\infty}^{\infty} x_i p_i \approx \sum_{i=-\infty}^{\infty} x_i f(x_i)\Delta x \tag{5}
$$

連続確率変数の期待値 VII

- $\Delta x \to 0$ のとき,
  - $\tilde{X} \to X$
  - $\mathrm{E}(\tilde{X}) \approx \sum_{i=-\infty}^{\infty} x_i f(x_i)\Delta x \to S$（ある値）

  ⇑ 数学的にはもう少しきちんと説明する必要あり
- S を $\int_{-\infty}^{\infty} x f(x) \mathrm{d}x$ と書く

$$\lim_{\Delta x \to 0} \mathrm{E}(\tilde{X}) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x \tag{6}$$

この $\int_{-\infty}^{\infty} x f(x) \mathrm{d}x$ が連続確率変数の期待値の定義

連続確率変数の期待値 VIII

分散や一般の期待値 $\mathrm{E}[g(X)]$

$$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\mathrm{d}x$$

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

$\Uparrow$ 離散の場合と同様

確率変数 $a + bX$ の期待値とその線形性

X が離散，連続のいずれにおいても，，，，
$a + bX$ の期待値と分散

$$\mathrm{E}(a + bX) = a + b\mathrm{E}(X)$$
$$\mathrm{Var}(a + bX) = b^2\mathrm{Var}(X)$$

- 本質的には数列の和，積分に関する線形性
- 特に離散の場合の証明 math.pdf 2.3節

$a+bX$ の確率密度関数 I

- $a+bX$ の確率密度関数（X を連続確率変数とする）
- $Y=a+bX$ として，$P(Y \leq y)$ の y に関する微分が Y の確率密度関数

$$
\begin{aligned}
P(Y \leq y) &= P(a+bX \leq y) \\
&= P\left(X \leq \frac{y-a}{b}\right) = F\left(\frac{y-a}{b}\right)
\end{aligned} \tag{7}
$$

$\Uparrow$ $F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t$ は X の累積分布関数

$a+bX$ の確率密度関数 II

- 「累積分布関数の微分 = 確率密度関数」の関係
- $F\left(\dfrac{y-a}{b}\right)$ の y に関する微分

合成関数の微分より $\dfrac{\mathrm{d}}{\mathrm{d}y}P(Y \leq y) = f\left(\dfrac{y-a}{b}\right)\dfrac{1}{b}$

$Y = a + bX$ の確率密度関数 $\Uparrow$

$a + bX$ の確率密度関数 III

- $\mathrm{E}[Y] = \mathrm{E}[a + bX]$ のはず

$$\begin{aligned}\mathrm{E}[Y] &= \int_{-\infty}^{\infty} y \left\{ f\left(\frac{y-a}{b}\right)\frac{1}{b}\right\} \mathrm{d}y = \int_{-\infty}^{\infty}(a+bz)f(z)\mathrm{d}z \\ &= a\int_{-\infty}^{\infty} f(z)\mathrm{d}z + b\int_{-\infty}^{\infty} zf(z)\mathrm{d}z = a + b\mathrm{E}[X]\end{aligned}$$

$\Uparrow$ 二番目の等号は置換積分による $z = \dfrac{y-a}{b}$，$\dfrac{\mathrm{d}y}{\mathrm{d}z} = b$

次に何を学ぶか I

- 離散確率変数　確率関数 $p(x)$ によって特徴づけられる（必要十分条件）

$$p(x) \geq 0, \quad \sum_{x \in \mathbb{R}} p(x) = 1$$

  - 離散確率変数の確率関数 $p(x)$ は上記を満たす
  - $p(x) \geq 0,\ \sum_{x \in \mathbb{R}} p(x) = 1$ を満たす関数を確率関数とする離散確率変数 X が存在する

次に何を学ぶか II

- 連続確率変数　確率密度関数 $f(x)$ によって特徴づけられる（必要十分条件）

$$f(x) \geq 0,\ \int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$$

  - 連続確率変数の確率密度関数 $f(x)$ は上記を満たす
  - $f(x) \geq 0,\ \int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$ を満たす関数を確率密度関数とする連続確率変数 X が存在する

次に何を学ぶか III

- 離散確率変数も連続確率変数も無数に考えられるが，数学的に扱いやすく綺麗な性質を持ち，意味のある解釈が出来る確率変数はそれほど多くない
- まさにそのような「数学的に扱いやすく綺麗な性質を持ち，意味のある解釈が出来る確率変数」を学ぶ

$\Uparrow$ 主要な確率分布

次に何を学ぶか IV

- また，以上は（1 次元）確率変数についての説明
- 2 次元確率変数（ベクトル）の一般論
  2 つの確率変数の共分散，相関係数
- 「数学的に扱いやすく綺麗な性質を持ち，意味のある解釈が出来る確率変数」ベクトル

多次元分布，多次元確率変数

- 複数の確率変数を同時に考えたときの分布
  $\Uparrow$2 次元が基本
- X と Y の二つの確率変数を同時に考える
- (X, Y) の組
  $\Uparrow$ 2 次元確率ベクトルまたは 2 次元確率変数
- 離散，連続の順に考える

多次元離散分布 I

確率的九九

- U,V：$1 \sim 9$ の整数を等確率 $1/9$ でとる確率変数
- $10X + Y = UV$，$0 \le X \le 8$，$0 \le Y \le 9$
- $X = 3, Y = 6 \Leftrightarrow UV = 36$

$$36 = 4 \times 9 = 9 \times 4 = 6 \times 6 \quad \text{3通り}$$

$$P(X = 3, Y = 6) = P(UV = 36) = \frac{3}{81} = \frac{1}{27}$$

$\Uparrow P(X = x, Y = y)$ の「,」は and の意味

多次元離散分布 II

$P(X=x, Y=y)$ の 81 倍

$x\backslash y$	0	1	2	3	4	5	6	7	8	9	
0		1	2	2	3	2	4	2	4	3	23
1	2		4		2	2	3		4		17
2	2	2			4	1		2	2		13
3	2		2			2	3				9
4	2		2			2			2	1	9
5					2		2				4
6				2	1						3
7			2								2
8		1									1
	8	4	12	4	12	9	12	4	12	4	81

多次元離散分布 III

同時確率：複数の確率変数を同時に考えたときの確率

$$p(x,y) = P(X = x, Y = y) \quad \text{同時確率関数}$$

もちろん次の２条件を満たす

- $p(x,y) \geq 0$ for any $(x,y) \in \Omega$，Ω は標本空間
- $\sum_{(x,y)\in\Omega} p(x,y) = 1$

多次元離散分布 IV

$$\text{周辺確率} \begin{cases} \text{最下行は}Y\text{だけの確率} \\ \text{最右列は}X\text{だけの確率} \end{cases}$$

X の周辺確率関数：同時確率関数 $p(x, y)$ に対し,

$$P(X = x) = p(x) = \sum_y p(x, y)$$

例

$$P(X = 3) = \sum_{y=0}^{9} p(3, y) = \sum_{y=0,2,5,6} p(3, y) = \frac{9}{81}$$

多次元離散分布 V
条件付確率

- $X = x$ が与えられたときの Y の確率関数

$$p(y|x) = P(Y = y|X = x) = \frac{p(x, y)}{p(x)} = \frac{p(x, y)}{\sum_z p(x, z)}$$

同時確率の表で $X = x$ の行内の相対的な確率

分母は $\sum_y p(x, y)$ でも良いが，y の関数ではないことを分かりやすくするため，$\sum_z p(x, z)$ と書く．$\sum_w p(x, w)$ でも良い

例：$X = 3$ が与えられたとき

$$p(0|3) = p(2|3) = p(5|3) = \frac{2}{9}, \quad p(6|3) = \frac{1}{3}$$

多次元離散分布 VI

X と Y が互いに独立

$$\Leftrightarrow \begin{cases} p(y|x) = p(y) \\ p(x,y) = p(x)p(y) \end{cases} \quad \forall (x,y) \in \Omega \tag{8}$$

解釈：$X = x$ という値を観測しても，Y の分布についても何も情報が得られない

X と Y は独立でない例：次ページ $p(y)$ と $p(y|X=3)$ が相異なる．$X = 3$ を観測することによって，Y について情報が得られている

多次元離散分布 VII

Y の周辺確率関数

$$p(y) = \begin{cases} 8/81 & y=0 \\ 4/81 & y=1 \\ 12/81 & y=2 \\ 4/81 & y=3 \\ 12/81 & y=4 \end{cases}, \quad \begin{cases} 9/81 & y=5 \\ 12/81 & y=6 \\ 4/81 & y=7 \\ 12/81 & y=8 \\ 4/81 & y=9 \end{cases} \tag{9}$$

$X=3$ が与えられたときの Y の条件付き確率関数

$$p(y|X=3) = \begin{cases} 2/9 & y=0 \\ 0 & y=1 \\ 2/9 & y=2 \\ 0 & y=3 \\ 0 & y=4 \end{cases}, \quad \begin{cases} 2/9 & y=5 \\ 3/9 & y=6 \\ 0 & y=7 \\ 0 & y=8 \\ 0 & y=9 \end{cases} \tag{10}$$

多次元連続分布 I

多次元連続確率変数

$$\left\{\begin{array}{l}\text{一次元連続確率変数}\\ \text{多次元離散確率変数}\end{array}\right. \quad \text{との対応}$$

(累積) 分布関数

$$\begin{aligned} F(x, y) &= P(X \leq x, Y \leq y) \\ &= P(-\infty < X \leq x, -\infty < Y \leq y) \end{aligned} \tag{11}$$

$\Uparrow$ 離散分布，連続分布で共通

多次元連続分布 II

同時確率密度関数：$f(x,y) = \dfrac{\partial^2}{\partial x \partial y} F(x,y)$

解釈

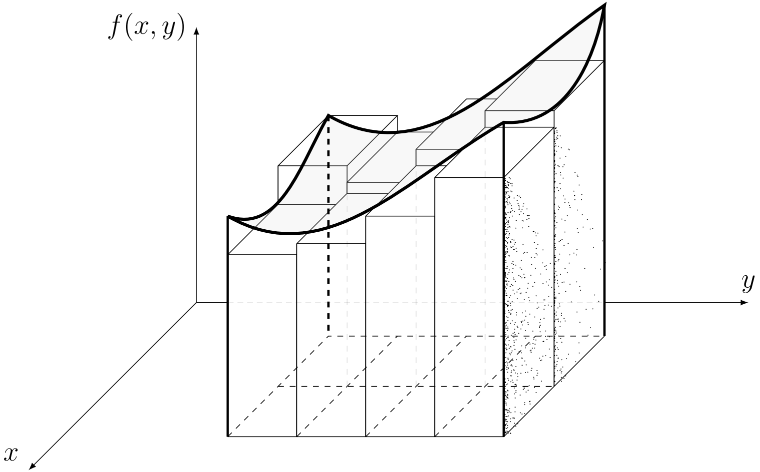
- 1次元連続確率変数：面積

$$P(a < X \le a + \Delta a) \doteq \overbrace{f(a)}^{\text{高さ}} \; \overbrace{\Delta a}^{\text{底辺の長さ}}$$

- 2次元連続確率変数：体積

$$P(a < X \le a + \Delta a, c < Y \le c + \Delta c) \doteq \overbrace{f(a,c)}^{\text{高さ}} \overbrace{\Delta a \Delta c}^{\text{底面積}}$$

多次元連続分布 III



多次元連続分布 IV

- 周辺密度関数：周辺確率関数に相当するもの

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y \tag{12}$$

$f_X(x)$ は一変量確率変数 X の密度関数

- $X = x$ を与えたときの Y の条件付き確率密度関数

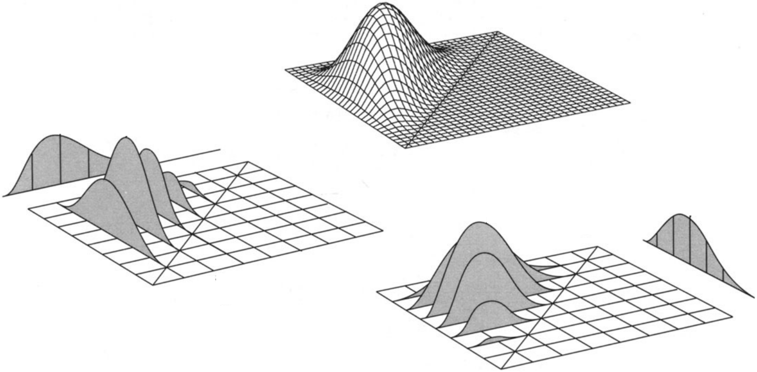
$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,z)\mathrm{d}z} \tag{13}$$

- 独立性　どのような (x,y) に対しても，

$$f(x,y) = f_X(x)f_Y(y), \text{ または } f(y|x) = f_Y(y)$$

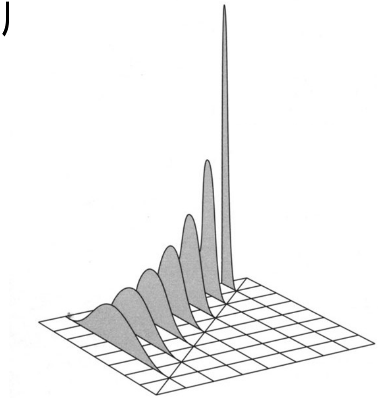
同時確率密度，周辺確率密度，条件付き確率密度 I

(X, Y) の同時確率密度

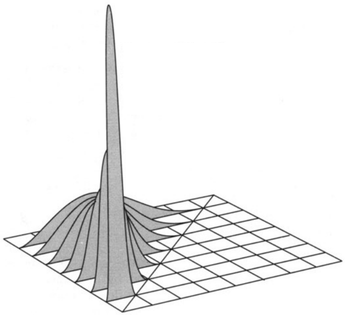


X の周辺確率密度　　　　Y の周辺確率密度

同時確率密度，周辺確率密度，条件付き確率密度



$X = x$ を与えたときの
Y の条件付き確率密度



$Y = y$ を与えたときの
X の条件付き確率密度

多次元分布の期待値 I

- $g(x, y)$ を 2 変数の実数値関数とするとき，その期待値は，離散・連続の場合，それぞれ

$$\mathrm{E}[g(X, Y)] = \begin{cases} \sum_{x,y} g(x, y)p(x, y) \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)f(x, y)\mathrm{d}x\mathrm{d}y \end{cases} \tag{14}$$

多次元分布の期待値 II

- 特に $g(x,y)=x$，$g(x,y)=y$ のとき，

$$\mathrm{E}[X]=\mu_X,\ \mathrm{E}[Y]=\mu_Y$$

$\Uparrow$ $\sum_{x,y} xp(x,y) = \sum_x x\{\sum_y p(x,y)\} = \sum_x xp(x)$ のように計算

- 特に $g(x,y)=(x-\mu_X)^2$，$g(x,y)=(y-\mu_Y)^2$ のとき，

$$\mathrm{Var}[X]=\sigma_X^2,\ \mathrm{Var}[X]=\sigma_Y^2$$

$\Uparrow$ $\sum_{x,y}(x-\mu_X)^2 p(x,y) = \sum_x (x-\mu_X)^2\{\sum_y p(x,y)\} = \sum_x (x-\mu_X)^2 p(x)$ のように計算

多次元分布の期待値 III

- 共分散と相関係数
  - 共分散は $g(x,y) = (x-\mu_X)(y-\mu_Y)$ に対応

$$\mathrm{COV}(X,Y) = \mathrm{E}[(X-\mu_X)(Y-\mu_Y)] \quad (15)$$

  - X と Y の相関係数 ρ_{XY}

$$\rho_{XY} = \frac{\mathrm{COV}(X,Y)}{\sigma_X \sigma_Y} \quad (16)$$

    - $|\rho_{XY}| \leq 1$ を満たす
    - X と Y の線形的な関係の強さの指標

多次元分布の期待値 IV

- $|\rho_{XY}| \leq 1$ の証明
  - 任意の t について $\mathrm{E}[\{t(X-\mu_X)+(Y-\mu_Y)\}^2] \geq 0$
  - $\mathrm{E}[\{t(X-\mu_X)+(Y-\mu_Y)\}^2]$

$$
\begin{aligned}
&= \mathrm{E}[t^2(X-\mu_X)^2 + 2t(X-\mu)(Y-\mu_Y) + (Y-\mu_Y)^2] \\
&= t^2\mathrm{E}[(X-\mu_X)^2] + 2t\mathrm{E}[(X-\mu)(Y-\mu_Y)] + \mathrm{E}[(Y-\mu_Y)^2] \\
&= t^2\sigma_X^2 + 2t\mathrm{COV}(X,Y) + \sigma_Y^2 \\
&= t^2\sigma_X^2 + 2t\sigma_X\sigma_Y\rho_{XY} + \sigma_Y^2 \\
&= \sigma_X^2(t + \{\sigma_Y/\sigma_X\}\rho_{XY})^2 + \sigma_Y^2(1-\rho_{XY}^2)
\end{aligned}
$$

  - 最小値を取る $t = -\{\sigma_Y/\sigma_X\}\rho_{XY}$ の場合でも非負

多次元分布の期待値 V

- X と Y が独立 $\Rightarrow$ 共分散・相関係数は 0

$$
\begin{aligned}
&\mathrm{E}((X-\mu_X)(Y-\mu_Y)) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f_X(x)f_Y(y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}(x-\mu_X)f_X(x)\mathrm{d}x\int_{-\infty}^{\infty}(y-\mu_Y)f_Y(y)\mathrm{d}y = 0
\end{aligned}
$$

$\int_{-\infty}^{\infty}(x-\mu_X)f_X(x)\mathrm{d}x = \int_{-\infty}^{\infty}xf_X(x)\mathrm{d}x - \mu_X\int_{-\infty}^{\infty}f_X(x)\mathrm{d}x$

$= \mu_X - \mu_X \times 1 = 0$

多次元分布の期待値 VI

- 確率変数の和の性質 $Z = X + Y$

$$
\begin{aligned}
\mathrm{E}(Z) &= \mathrm{E}(X) + \mathrm{E}(Y) \\
\mathrm{Var}(Z) &= \mathrm{E}[(Z - \mathrm{E}(Z))^2] \\
&= \mathrm{E}\left[(X + Y - \{\mathrm{E}(X) + \mathrm{E}(Y)\})^2\right] \\
&= \mathrm{E}\left[(\{X - \mathrm{E}(X)\} + \{Y - \mathrm{E}(Y)\})^2\right] \\
&= \mathrm{E}\left[\{(X - \mathrm{E}(X)\}^2 + \{Y - \mathrm{E}(Y)\}^2\right. \\
&\quad \left. +2\{X - \mathrm{E}(X)\}\{Y - \mathrm{E}(Y)\}\right] \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{COV}(X, Y)
\end{aligned}
$$

math.pdf 2.2節

多次元分布の期待値 VII

- $X_1, \ldots X_n$ の n 個の確率変数の和

$$
\begin{aligned}
&\mathrm{E}\left[X_1 + \cdots + X_n\right] = \mathrm{E}\left[X_1\right] + \cdots + \mathrm{E}\left[X_n\right] \\
&\mathrm{Var}(X_1 + \cdots + X_n) \\
&= \mathrm{E}\left[(X_1 + \cdots + X_n - \{\mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]\})^2\right] \\
&= \mathrm{E}\left[(\{X_1 - \mathrm{E}[X_1]\} + \cdots + \{X_n - \mathrm{E}[X_n]\})^2\right] \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{COV}(X_i, X_j)
\end{aligned}
$$

多次元分布の期待値 VIII

- $X_1, \ldots X_n$ が独立ならば，$\mathrm{COV}(X_i, X_j) = 0$ で

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) \tag{17}$$

  - 独立の場合に限り，「和の分散 = 分散の和」
  - 期待値はいつでも，「和の期待値 = 期待値の和」

多次元分布の期待値 IX

- 同じサイコロを n 回投げるときのように各 X_i が同じ分布（しかも独立）であれば，「$X_1, \ldots, X_n$ は独立同一分布に従う」という
- 各 X_i の期待値 μ，分散 σ^2 とすると，

$$\mathrm{E}[X_1 + \cdots + X_n] = n\mu$$

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = n\sigma^2$$

$$\begin{cases} \mathrm{E}[\bar{X}] = \mu & \Leftarrow \mathrm{E}[aY] = a\mathrm{E}[Y]\text{を使う} \\ \mathrm{Var}(\bar{X}) = \sigma^2/n & \Leftarrow \mathrm{Var}[bY] = b^2\mathrm{Var}[Y]\text{を使う} \end{cases}$$

二項分布 I

- ベルヌーイ試行 一回のコイン投げの言い替え
  - X: 1 か 0 の値をとる確率変数

$$P(X=1)=p,\ P(X=0)=1-p$$

  - 1 を便宜的に「成功」と解釈
  - p を成功確率という
- n 回の独立なベルヌーイ試行 $X_1, \ldots, X_n$

$$Y = X_1 + \cdots + X_n \quad \text{成功回数}$$

$\Uparrow$ Y の分布を二項分布 $\mathrm{Bin}(n,p)$ という

二項分布 II

- Y の確率関数 $p(y)$
  - 成功をS, 失敗をFで表記すると, n 回のベルヌーイ試行の結果は例えばFSSSFS$\cdots$FS (y 回のS).
  - このような結果が得られる確率 $p^y(1-p)^{n-y}$

$$\underbrace{p \times \cdots \times p}_{y\text{ 個}}\underbrace{q \times \cdots \times q}_{n-y\text{ 個}} = p^y q^{n-y}$$

$\Uparrow$ 各試行は互いに独立であることに注意

二項分布 III

- これは成功回数が y 回となる組合せの一つ
- 組合せの総数 $\binom{n}{y} = \frac{n!}{y!(n-y)!} = {}_n\mathrm{C}_y$
- この講義では $\binom{n}{y}$ を使う

- 二項分布の確率関数

$$p(y) = \binom{n}{y} p^y (1-p)^{n-y} \quad y = 0, 1, \ldots, n$$

二項分布 IV

確率関数?

- $\binom{n}{y} p^y (1-p)^{n-y}$ の各項は非負．$p(y) \geq 0$
- 二項展開より，

$$\sum_{y=0}^{n} p(y) = \sum_{y=0}^{n} \binom{n}{y} p^y (1-p)^{n-y} = \{p + (1-p)\}^n = 1 \tag{18}$$

二項分布 V

- 二項分布の期待値と分散
  - 定義通り計算可能

$$\mathrm{E}(Y) = \sum_{y=0}^{n} y p(y) = \sum_{y=0}^{n} y \binom{n}{y} p^y (1-p)^{n-y}$$

$$\mathrm{Var}(Y) = \sum_{y=0}^{n} (y - \mathrm{E}(y))^2 p(y)$$

  - ベルヌーイ試行を通じて計算した方が簡単

二項分布 VI

- ベルヌーイ試行 X

$$\mathrm{E}(X) = 1 \times p + 0 \times (1-p) = p$$
$$\begin{aligned}\mathrm{Var}(X) &= (1-p)^2 \times p + (0-p)^2 \times (1-p) \\ &= p(1-p)\end{aligned}$$

- Y : n 個の独立なベルヌーイ試行の和

$$\mathrm{E}(Y) = np, \ \mathrm{Var}(Y) = np(1-p)$$

- 確率関数のアニメーション

n 固定 p 変動　　n 変動 p 固定

幾何分布 I

- X：成功確率（表が出る確率）p のコイン投げで，初めて表が出るまでの試行回数　幾何分布

$$\Omega = \{1, 2, 3, \ldots,\}$$

- 違った定義：初めて表が出るまでの裏の回数

$$Y = X - 1, \quad \Omega = \{0, 1, 2, \ldots,\}$$

幾何分布 II

- 確率関数 $P(X=x|p)=p(1-p)^{x-1}$

$$=\overbrace{(1-p)(1-p)\dots(1-p)}^{x-1}\underbrace{p}_{1},\ x=1,2,\dots,$$

  - $p(1-p)^{x-1}\geq 0$
  - $\sum_{x=1}^{\infty}P(X=x|p)=1$ は以下より従う

$$\sum_{x=0}^{\infty}a^x=1/(1-a)\text{ for }|a|<1$$

幾何分布 III

- $\mathrm{E}[X] = 1/p$

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_{x=1}^{\infty} xp(1-p)^{x-1} \\
&= 1p(1-p)^{1-1} + \{p(1-p)^{2-1} + p(1-p)^{2-1}\} \\
&\quad + \{p(1-p)^{3-1} + p(1-p)^{3-1} + p(1-p)^{3-1}\} + \dots \\
&= \sum_{x=1}^{\infty} p(1-p)^{x-1} + \sum_{x=2}^{\infty} p(1-p)^{x-1} + \sum_{x=3}^{\infty} p(1-p)^{x-1} + \dots \\
&= p\left(1/p + (1-p)/p + (1-p)^2/p + \dots\right) \\
&= 1 + (1-p) + (1-p)^2 + \dots \\
&= 1/p
\end{aligned}
$$

幾何分布 IV

- 同様に $\mathrm{Var}[X] = \dfrac{1-p}{p^2}$
- $Y = X - 1$ の場合

$$\mathrm{E}[Y] = \mathrm{E}[X-1] = \frac{1-p}{p}, \quad \mathrm{Var}[Y] = \frac{(1-p)}{p^2}$$

- Y で定義された幾何分布の確率関数のアニメーション

幾何分布 V

幾何分布の無記憶性

- $P(X>n)=P(\text{最初の}n\text{回が全部裏})=(1-p)^n$

$$P(X>s|X>t)=\frac{P(X>s \text{ and } X>t)}{P(X>t)}$$

$$=\frac{P(X>s)}{P(X>t)}=(1-p)^{s-t}$$

$$=P(X>s-t) \text{ for } s>t>0$$

- t 回の裏の後，さらに $s-t$ 回の裏が続く確率は，最初の $s-t$ 回が裏で裏の確率と同じ

多項分布 I

- 二項分布のコインをサイコロに置き換える
- k 面体のサイコロ投げが多項分布
- 面 i の出る確率 p_i, $p_i \geq 0,\ p_1 + \cdots + p_k = 1$
- n 回投げたとき面 i の出た回数 X_i $(i = 1, \ldots, k)$

$$p(x_1, \ldots, x_k) = \frac{n!}{x_1! \ldots x_k!} p_1^{x_1} \ldots p_k^{x_k}, \text{ただし} \sum_{i=1}^{k} x_i = n$$

$\Uparrow$ 同時確率関数

多項分布 II

- 各成分は二項分布と同じ解釈
  - $p_1^{x_1} \dots p_k^{x_k}$ は特定の結果の列が得られる確率
  - 面 i の回数が x_i となるような場合の総数

$$\text{多項係数} \quad \frac{n!}{x_1! \dots x_k!}$$

- 【例】一等が確率 0.05，二等が確率 0.2，その他がはずれのくじ．10 回引いて一等と二等が一回ずつ当たる確率

$$p(1,1,8) = \frac{10!}{1!1!8!} 0.05 \times 0.2 \times 0.5^8 = 0.0901$$

- 【例】将棋の駒 math.pdf 2.4節

多項分布 III

- X_i の期待値，分散

X_i の周辺分布：面 i が出るか否かの二項分布 Bin(n, p_i)

$$\mathrm{E}(X_i) = np_i,\ \mathrm{Var}(X_i) = np_i(1-p_i)$$

- 共分散 $\mathrm{COV}(X_i, X_j) = -np_ip_j$
  - $\mathrm{Var}(X_i + X_j) = \mathrm{Var}(X_i) + \mathrm{Var}(X_j) + 2\mathrm{COV}(X_i, X_j)$
  - 左辺の確率変数 $X_i + X_j$ は，二項分布 Bin$(n, p_i + p_j)$
  - 従って

$$\begin{aligned} & n(p_i + p_j)(1 - p_i - p_j) \\ & = np_i(1-p_i) + np_j(1-p_j) + 2\mathrm{COV}(X_i, X_j) \end{aligned}$$

多項分布 IV

- ３項分布のアニメーション（$n = 10$．x_1 と x_2 の同時確率関数を見る．x_3 は自動的に決まるため．）
  - 例１:$p_3 = 1/3$ に固定．p_2/p_1 が変動する時（大きくなっていく）の x_1 と x_2 の同時確率関数

    アニメーションのタイトルでは p_1/p_2 となっているがタイポ
  - 例２:$p_1/p_2 = 1$ に固定．p_3 が変動する時の x_1 と x_2 の同時確率関数

ポアソン分布 I

- 二項分布において
  - $n \to \infty$, $p \to 0$
  - np（期待値）は一定．$np = \lambda$とする

$$p(y) = \frac{n(n-1)\dots(n-y+1)}{y!}\left(\frac{\lambda}{n}\right)^y\left(1-\frac{\lambda}{n}\right)^{n-y}$$

$$= \frac{\lambda^y}{y!}\frac{n}{n}\frac{n-1}{n}\dots\frac{n-y+1}{n}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-y}$$

$$\to \frac{\lambda^y}{y!}e^{-\lambda} \quad \left(\text{ただし} \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}\right)$$

ポアソン分布 II

- 極限の確率関数

$$p(y) = \frac{\lambda^y}{y!}e^{-\lambda}, \quad y = 0, 1, 2, \ldots \tag{19}$$

この確率関数をもつ分布を**ポアソン分布**という

- 導出から分かるように期待値は λ
- $n \to \infty$ のとき，$np(1-p) = \lambda(1-\lambda/n) \to \lambda$ となり，分散も λ
- 上記の確率関数に基づいた期待値と分散がどちらも λ であることの確認 math.pdf 2.5節
- 確率関数のアニメーション

ポアソン分布 III

例１：東京都の 1992 年の交通事故死亡者は 509 人

- 一日平均 $509/365 = 1.39$ 人の死亡者
- 交通事故にあって死ぬことを運の悪い宝くじに当たることと考える
- 東京都の人口 n 人
- n 人のそれぞれが非常に小さい確率 p で，一日のうちに事故で死亡する可能性がある
- p が小でも n が大なので，平均死亡者数 λ は 1.39 人程度

$\Uparrow$ 事故数の分布はポアソン分布か?

ポアソン分布 IV

例２：歴史的に有名なデータ　プロシャ陸軍の14個の連隊において馬に蹴られて死んだ兵士の数（1875年から20年間）

死亡者数	0	1	2	3	4
頻度	147	90	31	10	2
相対頻度	0.525	0.3214	0.110	0.036	0.0007

- 総死亡者数は196人
- 「年・連隊」あたりの平均死亡者数 $0.7 = 196/280$

ポアソン分布 V

- $\lambda = 0.7$ のポアソン分布の確率

$$
\begin{aligned}
p(0) &= \exp(-0.7) = 0.49659 \\
p(1) &= \exp(-0.7) \cdot 0.7 = 0.34761 \\
p(2) &= \exp(-0.7)0.7^2/2 = 0.12166 \\
p(3) &= \exp(-0.7)0.7^3/6 = 0.02839 \\
p(\geq 4) &= 0.00575
\end{aligned}
$$

$\Uparrow$ 相対頻度とよく一致している

正規分布 I

- 正規分布は統計学で最も重要な分布
  $\Uparrow$ 数学的に扱いやすく，美しい性質を持つ
- 基準となる標準正規分布 $N(0,1)$ の密度関数

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

  原点を中心とする対称な「つり鐘型」
- 密度関数を全域で積分すると 1（証明略）

$$\int_{-\infty}^{\infty} \phi(x) \mathrm{d}x = 1 \text{ or } \int_{-\infty}^{\infty} e^{-x^2/2} \mathrm{d}x = \sqrt{2\pi}$$

正規分布 II

- 指数関数の記法 $e^2 = \exp(2)$　指数の肩が複雑な場合「右辺」の記法がしばしば用いられる
- 期待値 $\mathrm{E}[X] = 0$
  - $\phi(x) = (2\pi)^{-1/2} e^{-x^2/2}$
  - 被積分関数 $x\phi(x)$ が奇関数

$$\begin{aligned} x\phi(x) &= -1 \times (-x)\phi(-x) \\ \mathrm{E}(X) &= \int_{-\infty}^{\infty} x\phi(x)\mathrm{d}x = 0 \end{aligned} \tag{20}$$

- 分散 $\mathrm{Var}[X] = 1$ math.pdf 2.6節

正規分布 III

標準正規分布の累積分布関数

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \mathrm{d}u = \int_{-\infty}^{x} \phi(u) \mathrm{d}u \tag{21}$$

もちろん $\Phi(-\infty) = 0$, $\Phi(\infty) = 1$，$\Phi(x)$ は増加関数，$\Phi'(x) = \phi(x)$

- この積分は陽に評価できない
- 様々な x に対して積分値が数表として与えられる
- R では $\begin{cases} \text{pnorm}(x) = \Phi(x) \\ \text{qnorm}(u) = \Phi^{-1}(u) \end{cases}$

正規分布 IV

- $X \sim N(0,1)$ の線形和である

$$Y = \mu + \sigma X \quad \text{for } \sigma > 0$$

  が従う分布 $\Rightarrow$ 正規分布 $N(\mu, \sigma^2)$
  - $\mathrm{E}[Y] = \mathrm{E}[\mu + \sigma X] = \mu + \sigma \mathrm{E}[X] = \mu$
  - $\mathrm{Var}[Y] = \sigma^2 \mathrm{Var}[X] = \sigma^2$

正規分布 V

- 確率密度関数 $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$
  - 累積分布関数 $P(Y \leq y)$

$$P(Y \leq y) = P(\mu + \sigma X \leq y) = P(X \leq \frac{y-\mu}{\sigma}) = \Phi(\frac{y-\mu}{\sigma})$$

  - 合成関数の微分より

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} P(Y \leq y) = \phi(\frac{y-\mu}{\sigma})\frac{1}{\sigma}$$

- 確率密度関数のアニメーション
  μ 固定 σ^2 変動　　μ 変動 σ^2 固定

正規分布 VI

- $Y \sim N(\mu, \sigma^2)$ の線形和

$$Z = a + bY \quad \text{for } b \neq 0$$

  - $X \sim N(0, 1)$ に戻って

$$Z = a + bY = a + b(\mu + \sigma X) = a + b\mu + b\sigma X$$

  - 従って，$Z \sim N(a + b\mu, b^2\sigma^2)$
  - 特に $a = 0$，$b = -1$ の場合，

$$Y \sim N(\mu, \sigma^2) \text{ ならば } -Y \sim N(-\mu, \sigma^2)$$

正規分布 VII

- 正規分布の再生性 math.pdf 2.7節
  - $X_1 \sim N(\mu_1, \sigma_1^2)$，$X_2 \sim N(\mu_2, \sigma_2^2)$ のとき

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

  - $X_1 - X_2 = X_1 + (-1)X_2$ は，

$$X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 - \sigma_2^2)$$

    でなく

$$X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$

一様分布 I

- $[a,b]$ の区間の一様分布 $U(a,b)$ の確率密度関数

$$f(x|a,b) = \begin{cases} \dfrac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

- 連続確率変数なので一点の確率はゼロ.
- 区間 $[a,b]$ で $f(x|a,b)$ が一定ということは, 区間 $[a,b]$ でどの値も同様に出やすい
- $\displaystyle\int_{-\infty}^{\infty} f(x|a,b)\mathrm{d}x = \int_a^b f(x|a,b)\mathrm{d}x = 1$ は明らか

一様分布 II

- 期待値と分散 $\mathrm{E}[X] = \dfrac{a+b}{2},\ \mathrm{Var}[X] = \dfrac{(b-a)^2}{12}$

$$\mathrm{E}[X] = \int_a^b \frac{x}{b-a}\mathrm{d}x = \frac{1}{2}\frac{b^2-a^2}{b-a} = \frac{1}{2}(b+a),$$
$$\mathrm{E}[X^2] = \int_a^b \frac{x^2}{b-a}\mathrm{d}x = \frac{1}{3}\frac{b^3-a^3}{b-a} = \frac{1}{3}(b^2+ab+a^2)$$
$$\mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \frac{b^2+ab+a^2}{3} - \left\{\frac{b+a}{2}\right\}^2 = \frac{(b-a)^2}{12}$$

- 確率密度関数のアニメーション $a=-b$

ガンマ分布 I

- 区間 $[0, \infty)$ のフレキシブルな分布のクラス
- 確率密度関数

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, 0 < x < \infty,$$

  - ガンマ関数 $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$
  - 便利な隣接等式（部分積分により簡単に証明）

$$\Gamma(1) = 1, \ \Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$

ガンマ分布 II

- 階乗 $n!$ の一般化 n は自然数

$$n! = n \cdot (n-1) \cdot \cdots \cdot 1 = n(n-1)!$$

- $f(t) = \dfrac{t^{\alpha-1}e^{-t}}{\Gamma(\alpha)}, t \geq 0$

  $\Uparrow$ 確率密度関数とみなせる（非負，積分して1）
- $X = \beta T$ の確率密度関数は？

$$\frac{(x/\beta)^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)} \times \frac{1}{\beta}$$

ガンマ分布 III

- ２つの母数（パラメータ）
  - $\alpha > 0$：形状母数
  - $\beta > 0$：尺度母数
- $\mathrm{E}[X]$，$\mathrm{E}[X^2]$，$\mathrm{Var}[X]$

$$\begin{aligned}\mathrm{E}[X^n] &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha+n-1}e^{-x/\beta}dx \\ &= \frac{\Gamma(\alpha+n)\beta^{\alpha+n}}{\Gamma(\alpha)\beta^\alpha} = \beta^n\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\end{aligned}$$

$$\mathrm{E}[X] = \alpha\beta,\ \mathrm{Var}[X] = \beta^2\alpha(\alpha+1) - \{\alpha\beta\}^2 = \alpha\beta^2$$

ガンマ分布 IV

- （特別）$\alpha = p/2$, $\beta = 2$ 自由度 p のカイ二乗分布

$$X_1^2 + \cdots + X_p^2 \text{ where } X_i \sim N(0,1)$$

確率密度関数のアニメーション

- （特別）$\alpha = 1$ 指数分布
  確率密度関数 $\beta^{-1} e^{-x/\beta}$ for $x > 0$
- 尺度変換
  - $X \sim \mathrm{Gamma}(\alpha, \beta) \Rightarrow \nu X \sim \mathrm{Gamma}(\alpha, \beta\nu)$
  - $\dfrac{\text{自由度}p\text{のカイ二乗分布}}{p} \sim \mathrm{Gamma}(p/2, 2/p)$

たたみこみと再生性 I

- たたみこみの公式 math.pdf 2.7節
  - X と Y は独立な連続確率変数
  - 確率密度はそれぞれ $f(x)$ と $g(y)$
    $\Rightarrow X+Y$ の確率密度 $\int_{-\infty}^{\infty} f(x)g(z-x)\mathrm{d}x$
- たたみこみ公式を使った例 math.pdf 2.7節
  - $N(\mu_1,\sigma_1^2)+N(\mu_2,\sigma_2^2) \overset{\mathrm{d}}{=} N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$
  - $\mathrm{Ga}(a,\tau)+\mathrm{Ga}(b,\tau) \overset{\mathrm{d}}{=} \mathrm{Ga}(a+b,\tau)$
  - $\sum_{i=1}^p \mathrm{Ga}(a_i,\tau) \overset{\mathrm{d}}{=} \mathrm{Ga}(\sum_{i=1}^p a_i,\tau)$

  $\Uparrow$ $\overset{\mathrm{d}}{=}$ の意味：$X_1 \sim N(\mu_1,\sigma_1^2)$ と $X_2 \sim N(\mu_2,\sigma_2^2)$ が独立であるとき，$X_1+X_2 \sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$ である

たたみこみと再生性 II

- 離散確率変数
  - $\mathrm{Bi}(m,p) + \mathrm{Bi}(n,p) \overset{\mathrm{d}}{=} \mathrm{Bi}(m+n,p)$
  - $\mathrm{Po}(\lambda_1) + \mathrm{Po}(\lambda_2) \overset{\mathrm{d}}{=} \mathrm{Po}(\lambda_1+\lambda_2)$
- カイ二乗分布とたたみこみ公式
  - 定義 平方和 $X_1^2 + \cdots + X_p^2$ の分布

$$\Uparrow \quad X_1, \ldots, X_p \text{は独立で}, X_i \sim N(0,1)$$

  - $X \sim N(0,1)$ のときの X^2 の分布が $\mathrm{Ga}(1/2, 2)$

    math.pdf 2.8節

    $\Rightarrow$ たたみこみ公式より $X_1^2 + \cdots + X_p^2$ の分布は $\mathrm{Ga}(p/2, 2)$ となる

t 分布

- 定義：student の t 分布 $T = \dfrac{X}{\sqrt{V/m}}$
  （ここで $X \sim N(0,1)$ 標準正規分布，$V \sim \chi_m^2$ 自由度 m のカイ二乗分布，X と V は独立）の従う分布
  - 密度関数 math.pdf 2.9節
  - 再生性はない
  - $m = 1$ の場合コーシー分布
  - $m \to \infty$ で $N(0,1)$ に収束
  - 確率密度関数のアニメーション

飽きてきた？ 次は？ I

- 多くの具体的な確率分布を学んだ
- 確率論は，与えられた確率分布の性質を調べる
- 統計では，得られた有限個のデータ $x_1, \ldots, x_n$ の背後に確率的な構造を想定
- 複数の確率分布の候補を用意
- それらの候補の中から，データにあった確率分布を選ぶ $\Leftarrow$ 統計的推測

飽きてきた？ 次は？ II

- 複数の確率分布？
  - 非負で和が 1，積分して 1 になる関数（確率関数，確率密度関数）では集合として多すぎる
  - そこで役に立つのが，「有限個のパラメータで制御できる」「具体的な」確率分布
  - 「データにあった確率分布を選ぶ」というのは，「パラメータを選ぶ」ことに帰着
- 統計的推測の例
  - 1 点で言い当てる「推定」
  - パラメータが入る空間を 2 つに分けてどちらに入るか当てる「検定」