情報数学

中山クラス
第９週

＜今日の内容＞
「道具としてのベイズ統計」
◇第１章　ベイズ統計の準備
１．条件付き確率と乗法定理
２．確率変数と確率分布
３．有名な確率分布
◇演習問題

第1章　ベイズ統計の準備　p.22

１　条件付き確率と乗法定理

■確率の意味とその記号

試行　　サイコロを投げる操作，箱を選ぶ操作

事象　　試行によって得られる結果

A:　偶数の目の出る事象

事象Aの起こる確率

$$p(A)=\frac{\text{偶数の目の出る場合の数}}{\text{起こりえる目の全ての場合の数}}=\frac{3}{6}=\frac{1}{2}$$

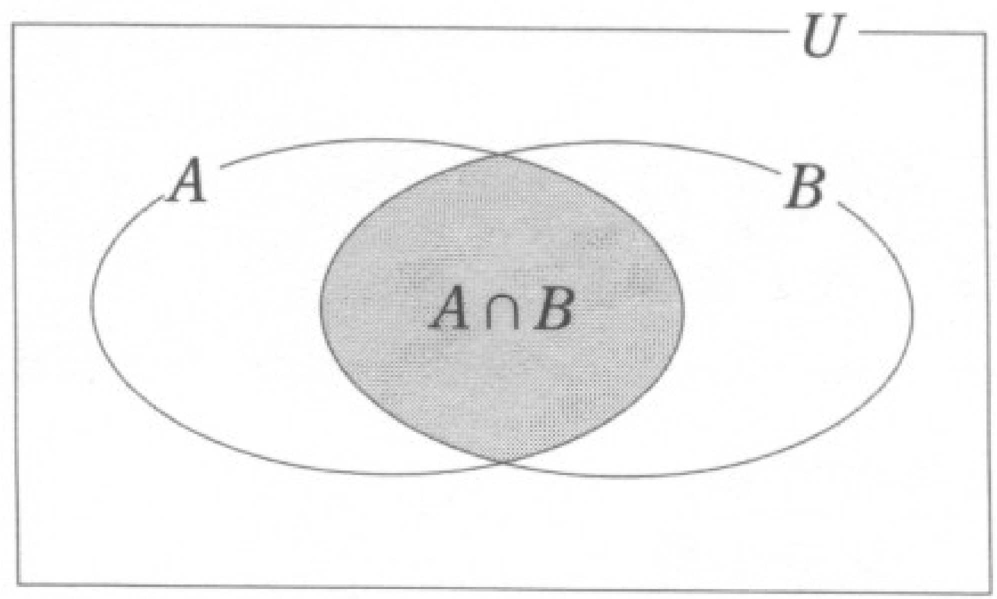
一般化

$$p(A)=\frac{\text{事象}A\text{の起こる場合の数}}{\text{起こり得る全ての場合の数}}$$

同時確率　p.23

事象A:　偶数の目が出る(2, 4, 6)
事象B:　3の倍数の目が出る(3, 6)
事象A&B:偶数&3の倍数(6)
事象A, Bが同時に起こる確率$= P(A \cap B), P(A, B)$

$$P(A \cap B) = \frac{1}{6}$$



U:確率論では標本空間

周辺確率　p.24

同時確率$P(A \cap B)$に対して$P(A), P(B)$を周辺確率

＜男女のビールの好き嫌い＞
事象A＝男性，事象B＝女性
事象C＝好き，事象D＝普通，事象E＝嫌い

$P(A) = 0.3 + 0.2 + 0.1 = 0.6$
$P(B) = 0.2 + 0.1 + 0.1 = 0.4$
$P(A) + P(B) = 0.6 + 0.4 = 1$
$P(C) = 0.3 + 0.2 = 0.5$
$P(D) = 0.2 + 0.1 = 0.3$
$P(E) = 0.1 + 0.1 = 0.2$
$P(C) + P(D) + P(E)$
$= 0.5 + 0.3 + 0.2 = 1$

	ビールの好き嫌い			
	好き	普通	嫌い	計
男	0.3	0.2	0.1	0.6
女	0.2	0.1	0.1	0.4
計	0.5	0.3	0.2	1

男性でビールが好き:$A \cap C$

$$P(A \cap C) = 0.3$$

女性でビール嫌い:$B \cap E$

$$P(B \cap E) = 0.1$$

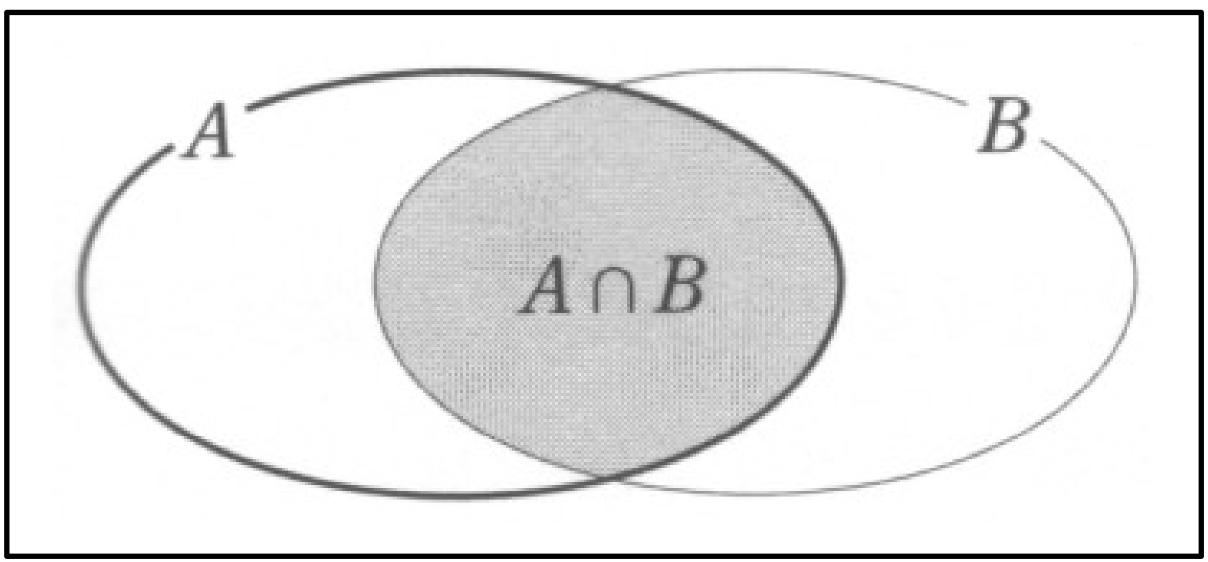
条件付き確率　p.24

「ある事象Aが起こったという条件のもとで別の事象Bが起こる確率」→「AのもとでBが起こる条件付き確率」

→　$P(B|A),\ P_A(B)$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \qquad P(A) \neq 0$$

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



＜同時確率＞

事象A：　偶数の目が出る（2，4，6）

事象B：　3の倍数の目が出る（3，6）

事象A＆B：偶数＆3の倍数（6）

事象A，Bが同時に起こる確率＝ $P(A \cap B), P(A, B)$

$$P(A \cap B) = 1/6$$

＜同時確率と条件付き確率の関係＞

事象Aの起こる確率：$P(A) = 3/6 = 1/2$

事象Aの下で事象Bが起こる確率：$P(B|A) = 1/3$

事象A，Bが同時に起こる確率：

$$P(A \cap B) = P(B|A)P(A) = (1/3) \times (1/2) = 1/6$$

（例１）

ある飛行機の乗客のうち，60%が日本人，42%が日本人男性である．日本人のなかから一人を選び出したとき，それが男性である確率を求めよ．

＜解答例＞

事象A:一人を選ぶとき，それが日本人である．$P(A) = 0.6$

事象B:一人を選ぶとき，それが男性である．$P(B)$

事象C:一人を選ぶとき，それが日本人の男性である．

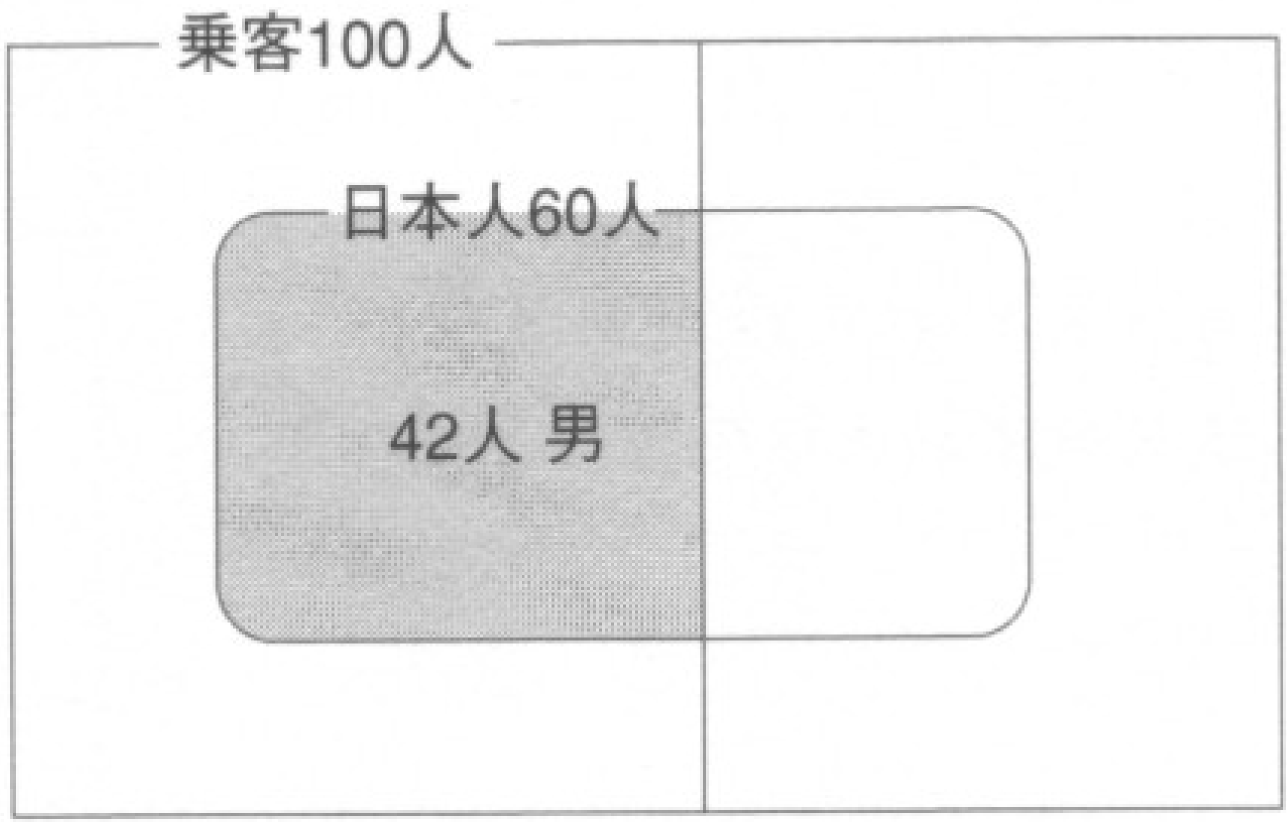
$$P(C) = P(A \cap B) = 0.42$$

事象D:日本人のなかから（条件付き）一人を選ぶとき，それが男性である．$P(D) = P(B|A)$

$$P(D) = P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.42}{0.60} = 0.7$$

＜別の解法＞

乗客が100人いると仮定する．条件より，うち60人が日本人で，42人が日本人男性である．従って，日本人のなかで男性である割合は42／60＝0.7である．



乗法定理　p.26

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

より

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

→ベイズ統計の基本公式

（例２）

１００本のなかに１０本の当たりがあるくじをＡ君，Ｂ君の順に引く．このとき，Ａ君が当たりくじを引き，続いてＢ君も当たりくじを引く確率を求めよ．ただし，引いたくじは戻さない物とする．

＜解答例＞

事象Ａ：　Ａ君が当たる　$P(A)$

事象Ｂ：　Ｂ君が当たる　$P(B)$

事象Ｃ：　Ａ君とＢ君が同時に当たる　$P(C) = P(A \cap B)$

Ａ君が当たる確率＝ $P(A) = 10/100$

引き続いてＢ君が当たる確率

$$= P(B|A) = (10-1)/(100-1) = 9/99$$

これらが同時に起こる確率$P(B|A)P(A)$

$$P(A \cap B) = P(B|A)P(A) = \frac{9}{99} \times \frac{10}{100} = \frac{1}{110}$$

２　確率変数と確率分布　p.27

■確率変数：　確率的に値が決まる変数

（例）サイコロの目

サイコロを１００回振った.

１０回目に出る目は？・・・分からない！

目が１，２，３と出た，次に出る目は４？・・・分からない！

１００回のうち，１００／６→１６，１７回は１の目がでるはず

・・・もっとも！

サイコロの目の出方は「確定的」にはわからない．「確率的」にのみ分かる→確率変数.

■確率分布とその平均値，分散

確率分布：　　確率変数の値に対する確率
確率分布表：　確率変数とその値の対応表

（例）サイコロの目の確率分布表

サイコロの目	１	２	３	４	５	６
確率	1/6	1/6	1/6	1/6	1/6	1/6

平均（期待値）

$$\mu = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = \frac{21}{6} = 3.5$$

分散

$$\sigma^2 = (1-3.5)^2 \times \frac{1}{6} + \cdots + (6-3.5)^2 \times \frac{1}{6} = \frac{35}{12} \cong 2.9$$

標準偏差　　$\sigma = \sqrt{35/12} \cong 1.7$

一般式による表現

確率変数	x_1	x_2	x_3	$\cdots$	x_n
確率	p_1	p_2	p_3	$\cdots$	p_n

平均（期待値）

$$\mu = x_1 p_1 + x_2 p_2 + \cdots + x_n p_n$$

分散

$$\sigma^2 = (x_1 - \mu)^2 p_1 + (x_2 - \mu)^2 p_2 + \cdots + (x_n - \mu)^2 p_n$$

標準偏差

$$\sigma = \sqrt{\sigma^2}$$

■連続的な確率変数と確率密度関数　p.29

確率変数xが離散的な値を取る場合（例：サイコロの目）は確率変数の値に対して，確率そのものを表や棒グラフで表現できる．

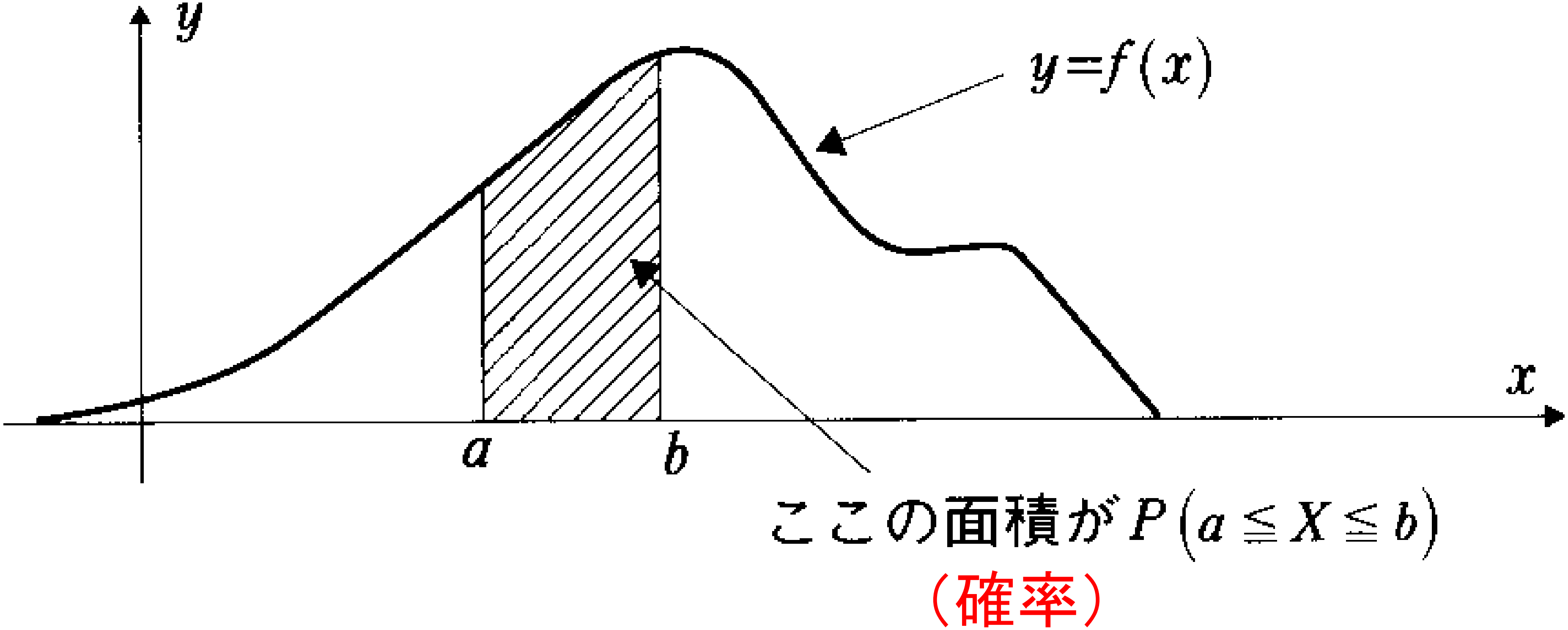
確率変数xが連続的な値を取る場合
（例）長さ，重さ，株価，為替レートなど
確率分布→確率密度関数$f(x)$で表現

確率変数xが$a \leq x \leq b$の値を取る確率

$$P(a \leq x \leq b) = \int_a^b f(x)dx$$

（注意）$x = a$となる確率は零である．

確率密度関数$f(x)$と確率の計算



■連続的な確率変数の平均，分散，標準偏差

平均（期待値）

$$\mu = \int_{-\infty}^{\infty} x f(x) dx$$

分散

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

標準偏差

$$\sigma = \sqrt{\sigma^2}$$

＊積分範囲は確率密度関数が定義されている範囲

３　有名な確率分布

実際に与えられた資料を分析するために
「そのデータがどのような確率分布に従って生まれたか」
を仮定する.

仮定として用いられる確率分布
　ベイズ統計でよく利用されるものを紹介する
　具体的な利用方法は2章以降で述べる.

二項分布

１回の試行で，ある事象Aが起こる確率がpである．この試行をn回繰り返したとき，事象Aがk回起こる確率

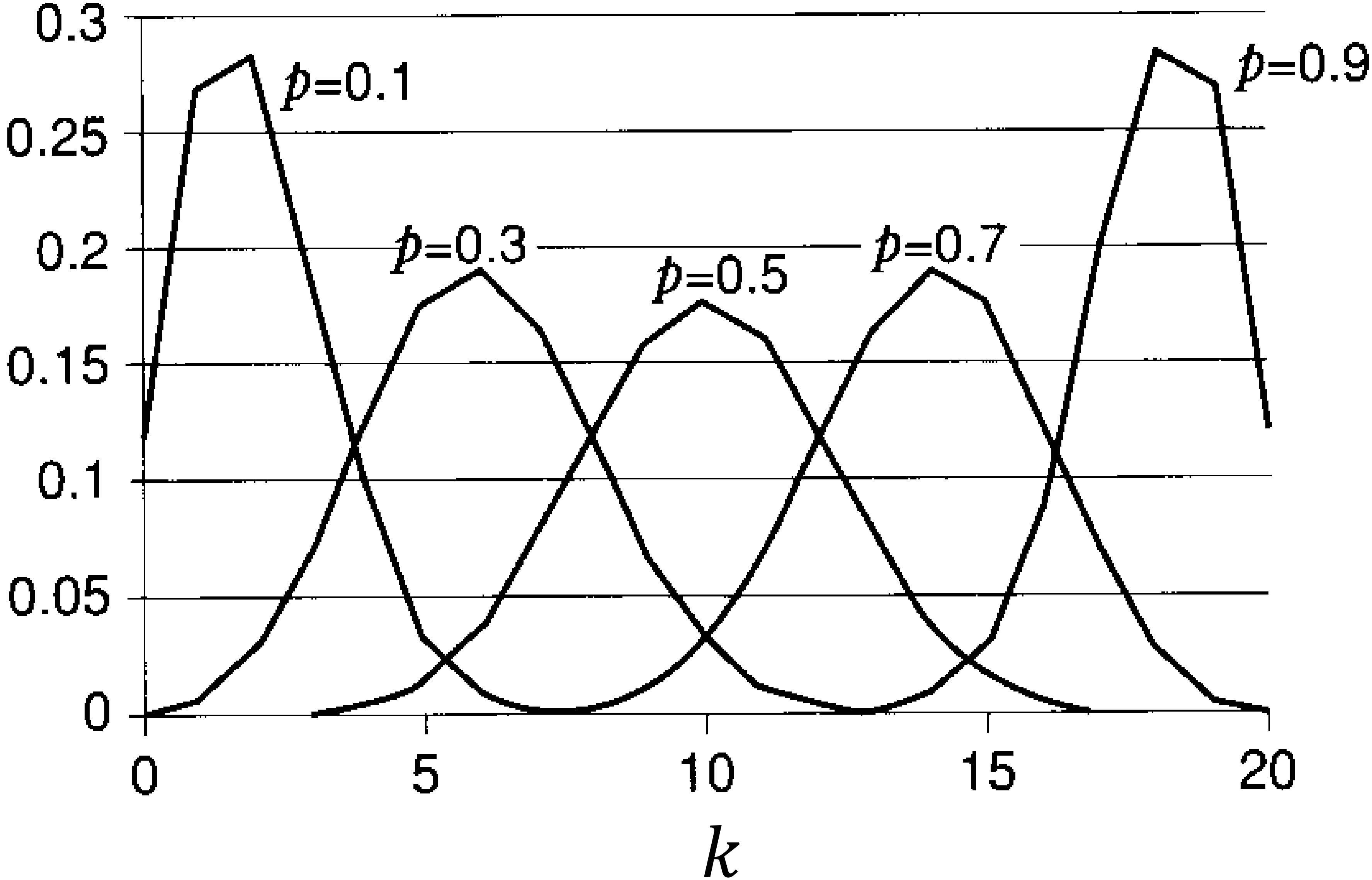
$$ {}_nC_k p^k (1-p)^{n-k} $$

確率変数Xが$X = k$という値をとる確率がこの式で与えられるとき，この確率分布を二項分布という.

平均値　$\mu = np$
分散　　$\sigma^2 = np(1-p)$

＊正規分布で近似される.

二項分布$B(n, p), n = 20$



二項分布の例

（1）サイコロ

サイコロを20回投げ，そのうち1の目が7回出る確率

$$p_7 = {}_{20}C_7 \left(\frac{1}{6}\right)^7 \left(1-\frac{1}{6}\right)^{20-7}$$

1の目が7回出る確率，1以外の目が20-7=13回出る確率

$$\left(\frac{1}{6}\right)^7, \qquad \left(1-\frac{1}{6}\right)^{20-7}$$

これらは同時に起こる

$$\left(\frac{1}{6}\right)^7 \left(1-\frac{1}{6}\right)^{20-7}$$

1が何回目に出るか→ ${}_{20}C_7$通りある.

（2）コイン

コインを10回投げ，そのうち3回で表が出る確率

$$p_3 = {}_{10}C_3 \left(\frac{1}{2}\right)^3 \left(1 - \frac{1}{2}\right)^{10-3}$$

正規分布

確率密度関数

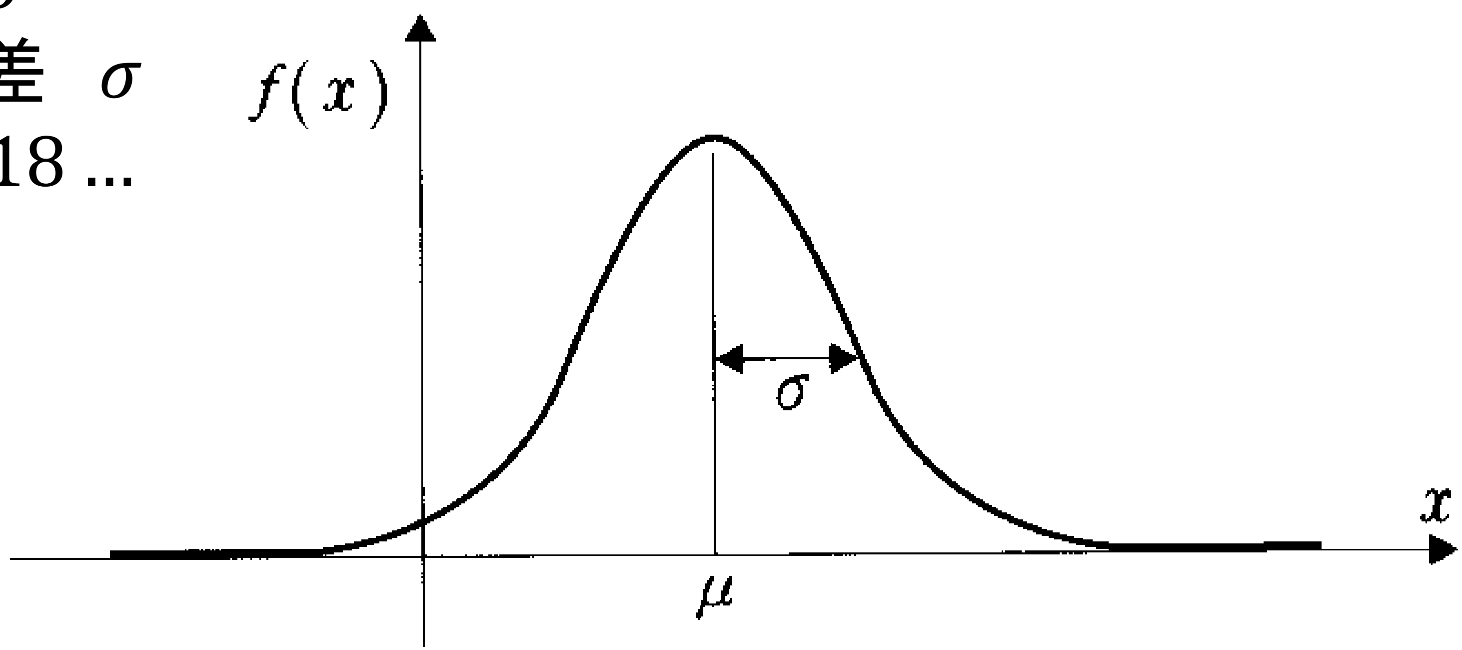
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

平均値　μ

分散　　σ^2

標準偏差　σ

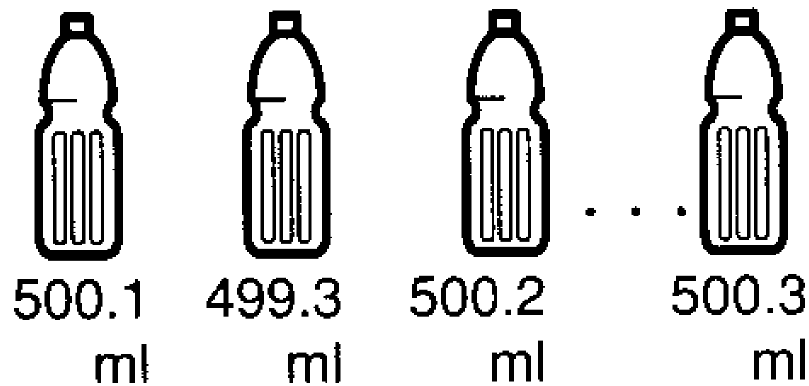
$e = 2.718 \ldots$

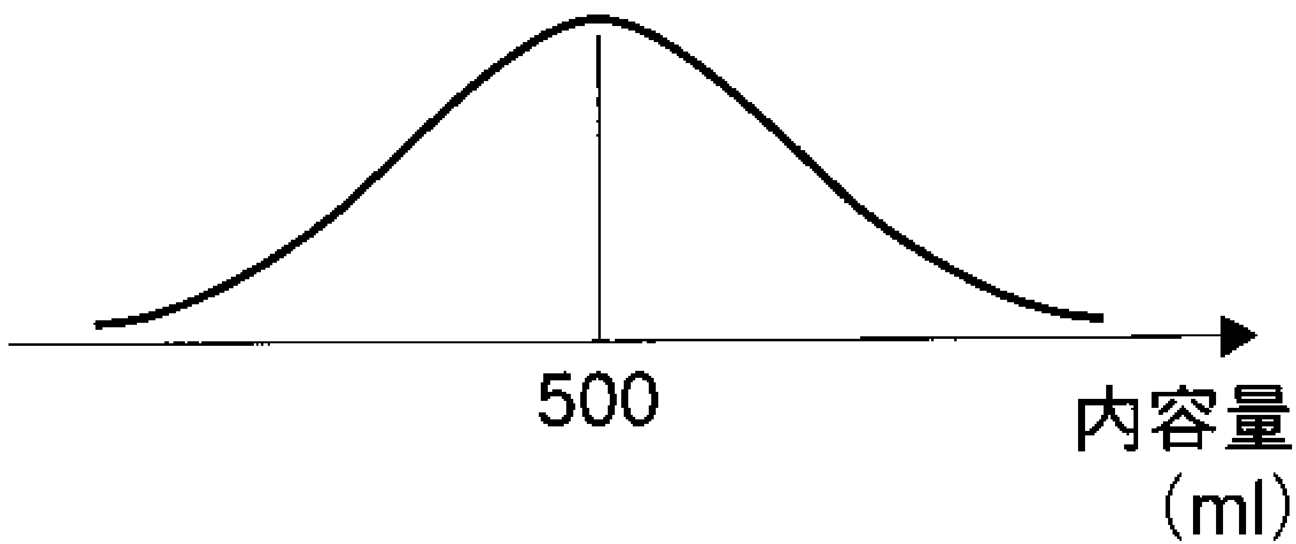


◇正規分布の例：ペットボトルの容量

「内容量500ml」→実際にはばらつきがある.

ばらつきが正規分布に近い.





演習問題

ある客船の乗客のうち，50%が日本人で，60%が男性である．また，日本人女性の乗客は20%である．男性のなかから1人を選び出したとき，それが日本人である確率を求めよ．

事象A：一人を選ぶとき，それが男性である．
事象B：一人を選ぶとき，それが日本人である．
$P(B|A)$：男性から1人を選んだとき，それが日本人である確率
（1）$P(A)$を求めよ．
（2）$P(A \cap B)$を求めよ．
（3）（1），（2）の結果を用いて$P(B|A)$を求めよ．