

情報数学

中山クラス 第10週

<今日の内容>

- ◇演習問題(前回)の解説
- ◇第1章 ベイズ統計の準備
 - 3. 有名な確率分布
 - 4. 尤度関数と最尤推定法
- ◇第2章 ベイズの定理とその応用
 - 1. ベイズの定理とは
- ◇演習問題

演習問題(前回)の解説

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。

解答例

(1) 乗客の中で男性客は60%であるから,

$$P(A) = 0.6$$

(2) 乗客の中で日本人は50%で, 日本人女性は20%であるから, 日本人男性は

$$50\% - 20\% = 30\%$$

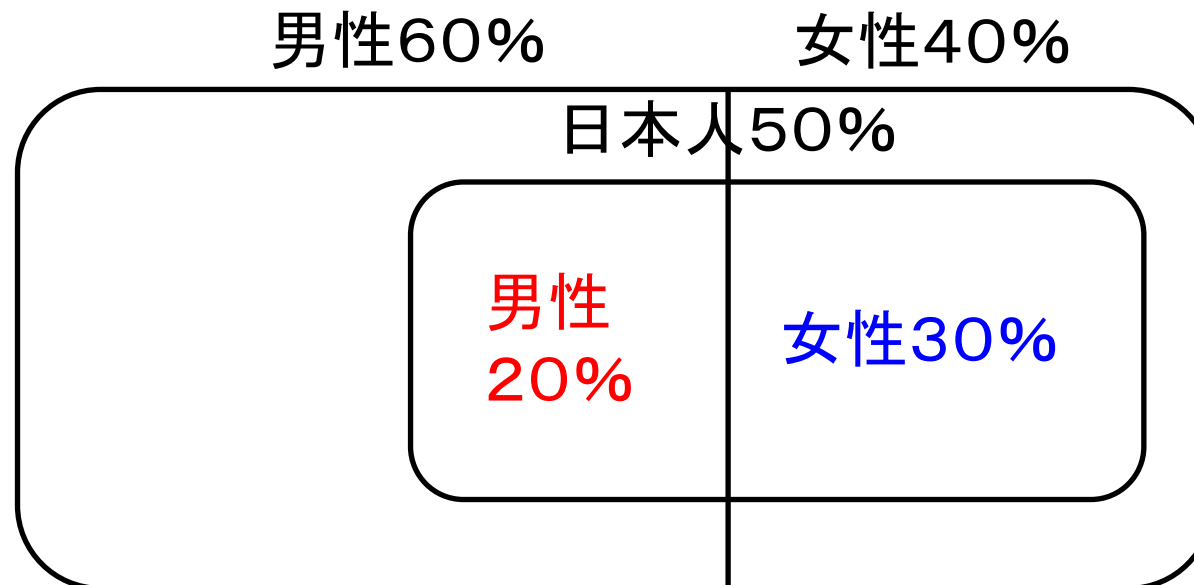
である.

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

(3)

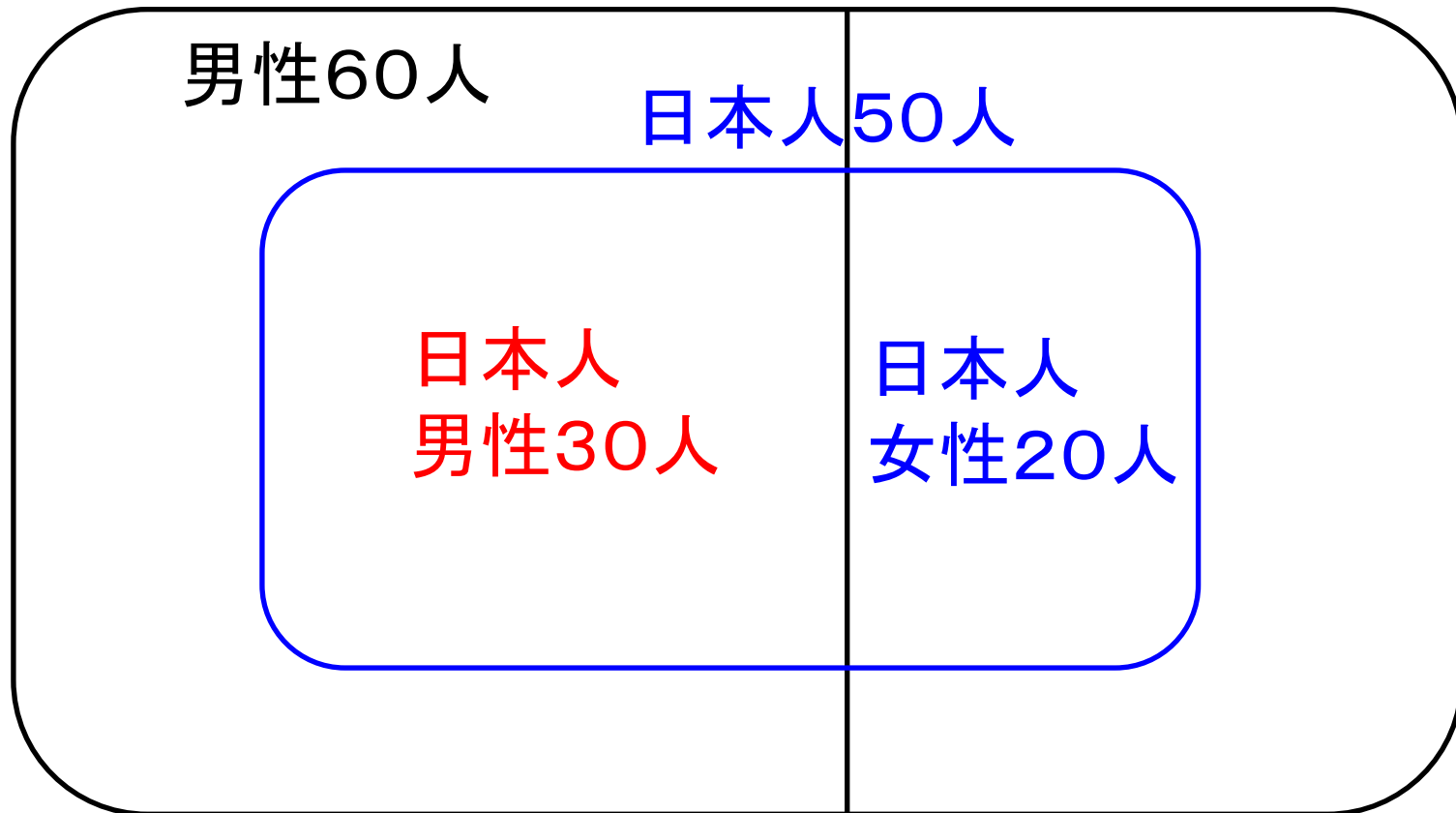
$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.3}{0.6} = 0.5$$

(参考1) $P(A \cap B) \neq P(A) \times P(B)$



(参考2)

乗客全員100人



男性60人中, 日本人男性が30人 $\rightarrow 30/60=1/2$

中央極限定理 p.33

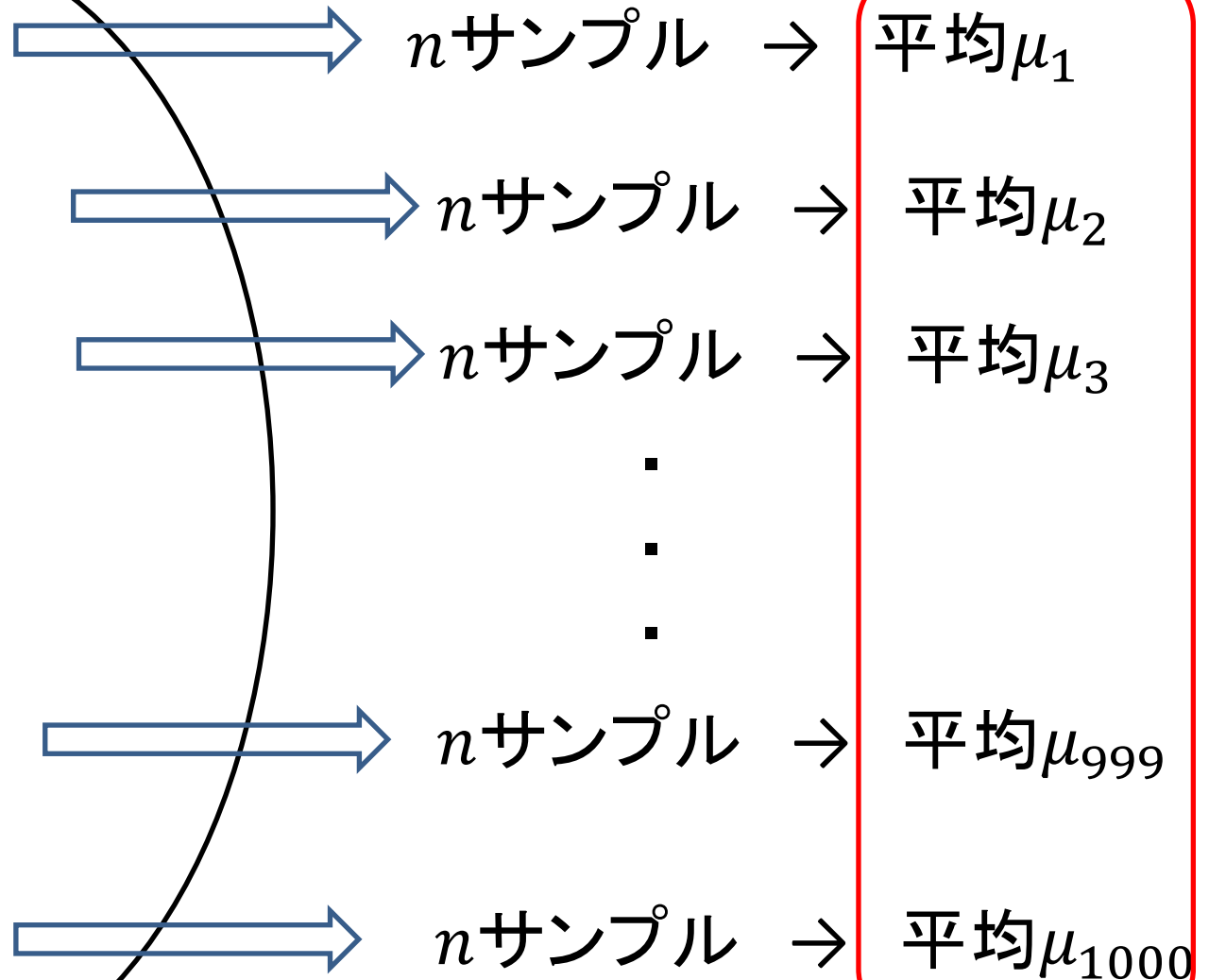
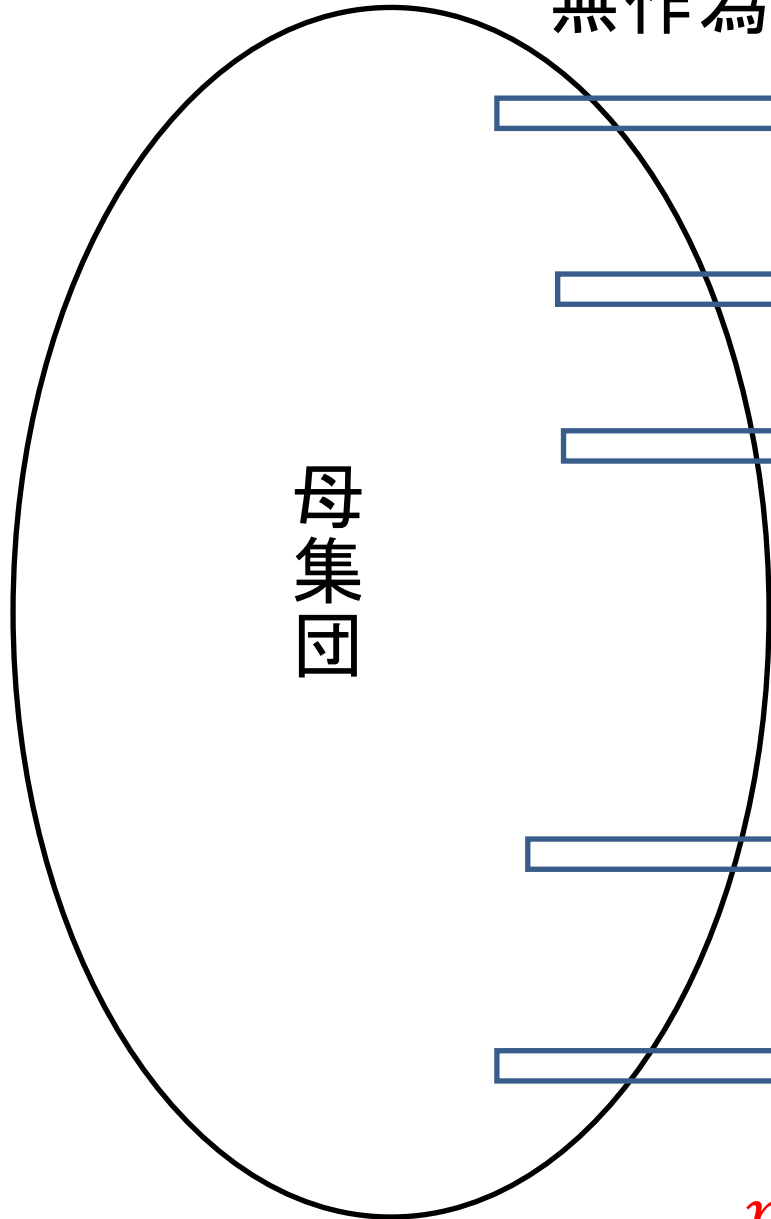
もとの母集団の分布が何であっても、標本の数が多くなるに従って標本平均の分布は正規分布に近づく.

- (1) 元の母集団の分布: 一様分布と正規分布が混在
- (2) 5サンプルの平均を1000回とった時の分布
- (3) 30サンプルの平均を1000回とった時の分布
- (4) 200サンプルの平均を1000回とった時の分布

サンプルサイズが大きくなるほど正規分布に近づく.
同時に標本平均の標準偏差(標準誤差)が小さくなりヒストグラムの幅が狭くなっている.

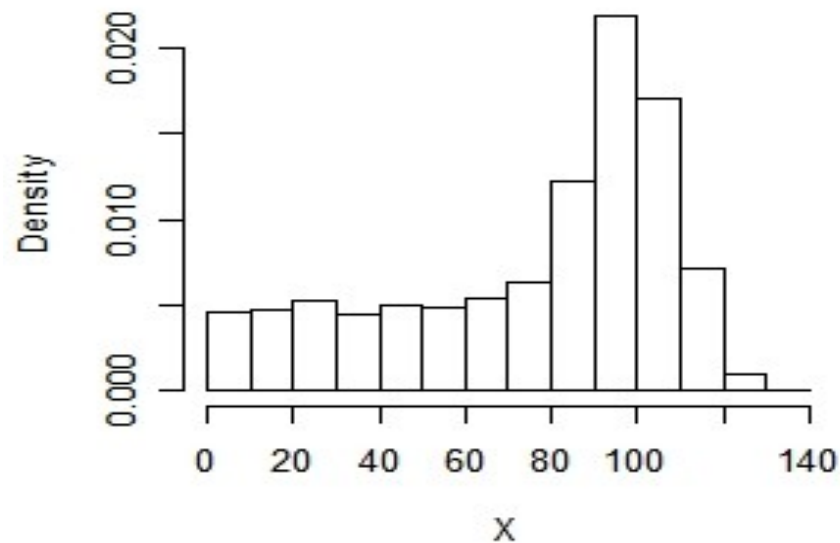
無作為抽出

分布図

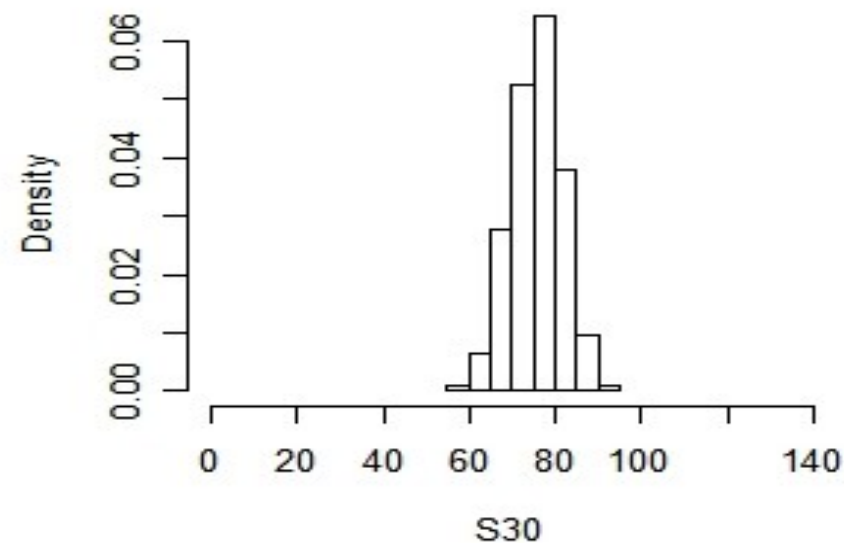


$n = 5, 30, 200$

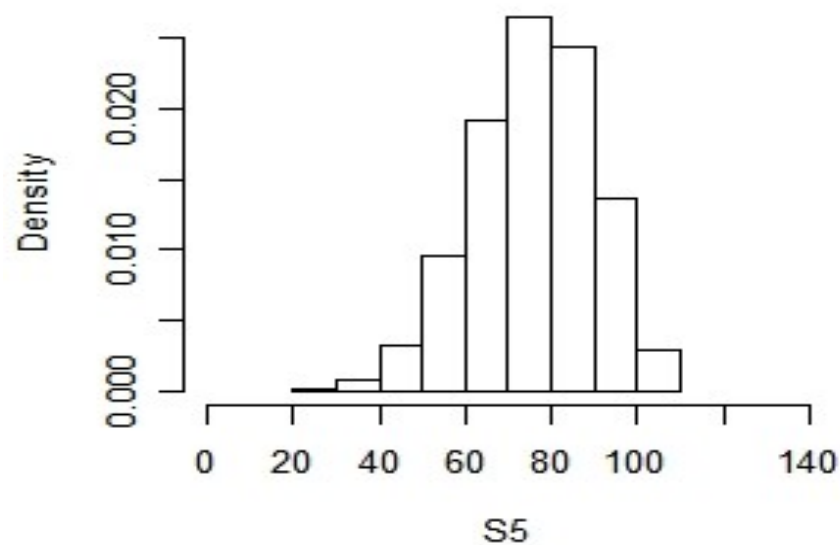
(1) Histogram of X



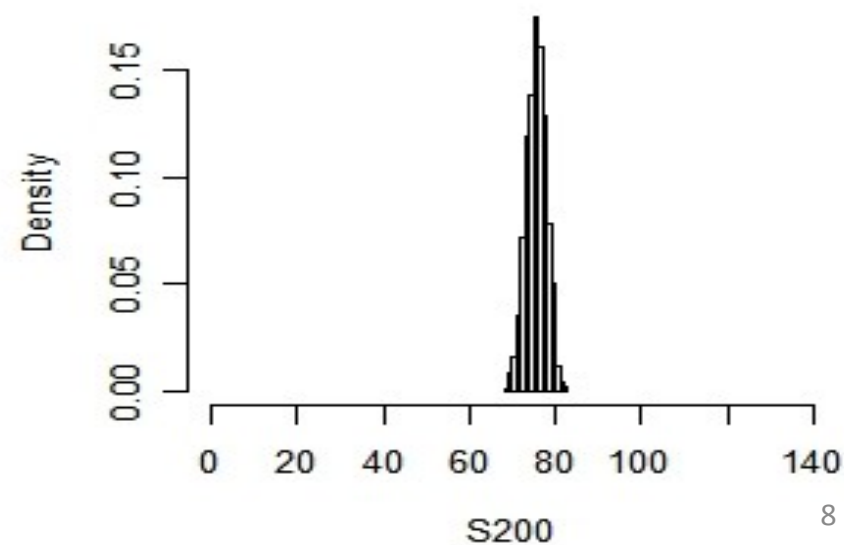
(3) Histogram of S30



(2) Histogram of S5



(4) Histogram of S200



一様分布 p.33

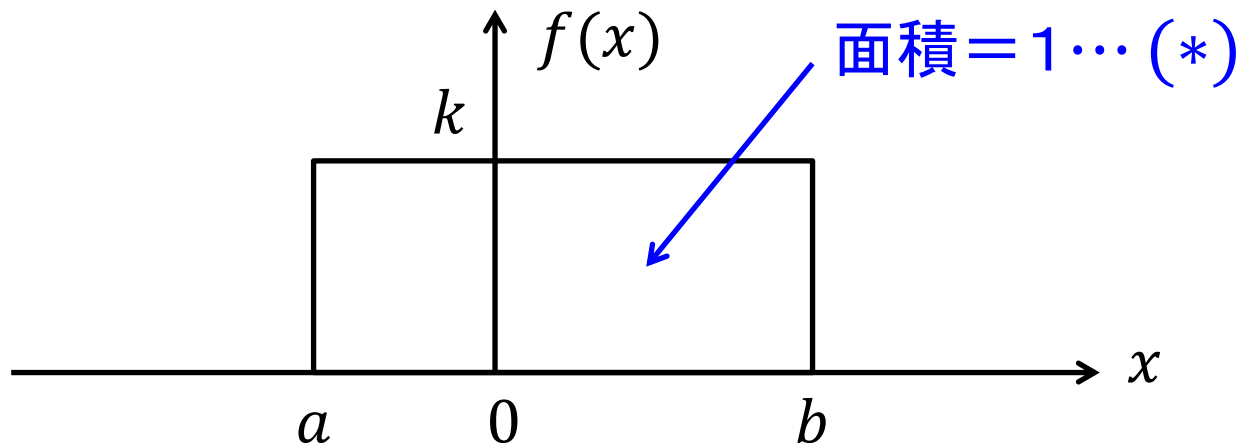
一様分布 $U(a, b)$ の確率密度関数

$$f(x) = \begin{cases} k & \text{一定 } (a \leq x \leq b) \\ 0 & (x < a, b < x) \end{cases}$$

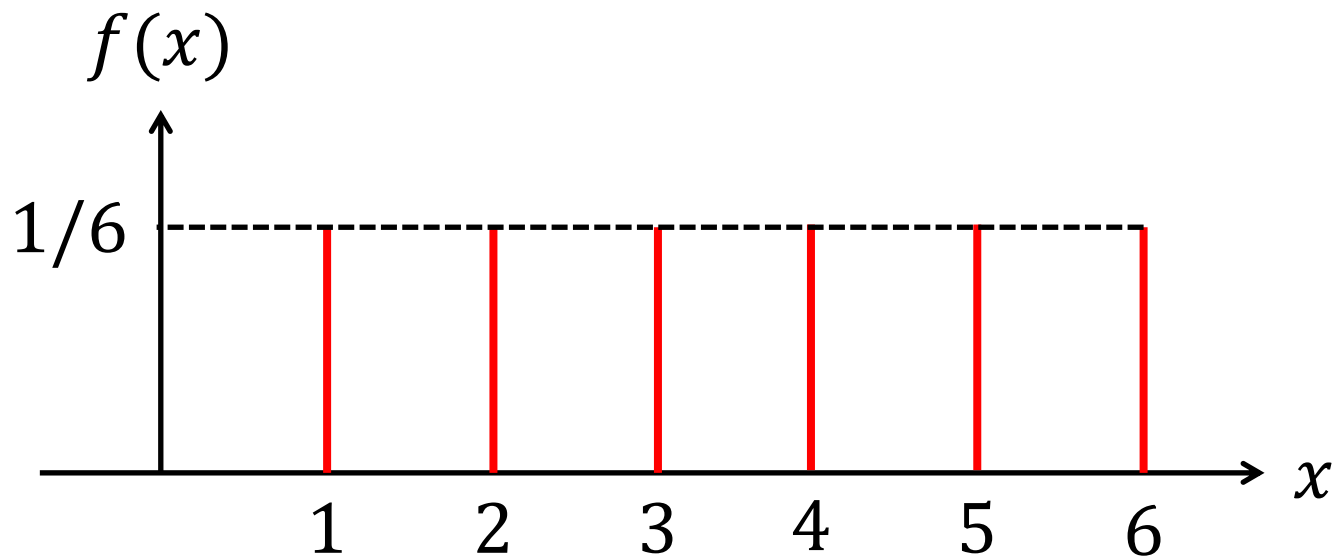
$$k(b - a) = 1 \cdots (*)$$

平均値, 分散

$$\mu = \frac{a + b}{2}, \quad \sigma^2 = \frac{(b - a)^2}{12}$$



一様分布の例：サイコロの目の確率分布



ベータ分布 p.34

ベータ分布 $Be(p, q)$ の確率密度関数

$$f(x) = kx^{p-1}(1-x)^{q-1}$$

k : 定数, $0 < x < 1, 0 < p, 0 < q$

k は確率の総和 = 1 より決まる.

平均値, 分散

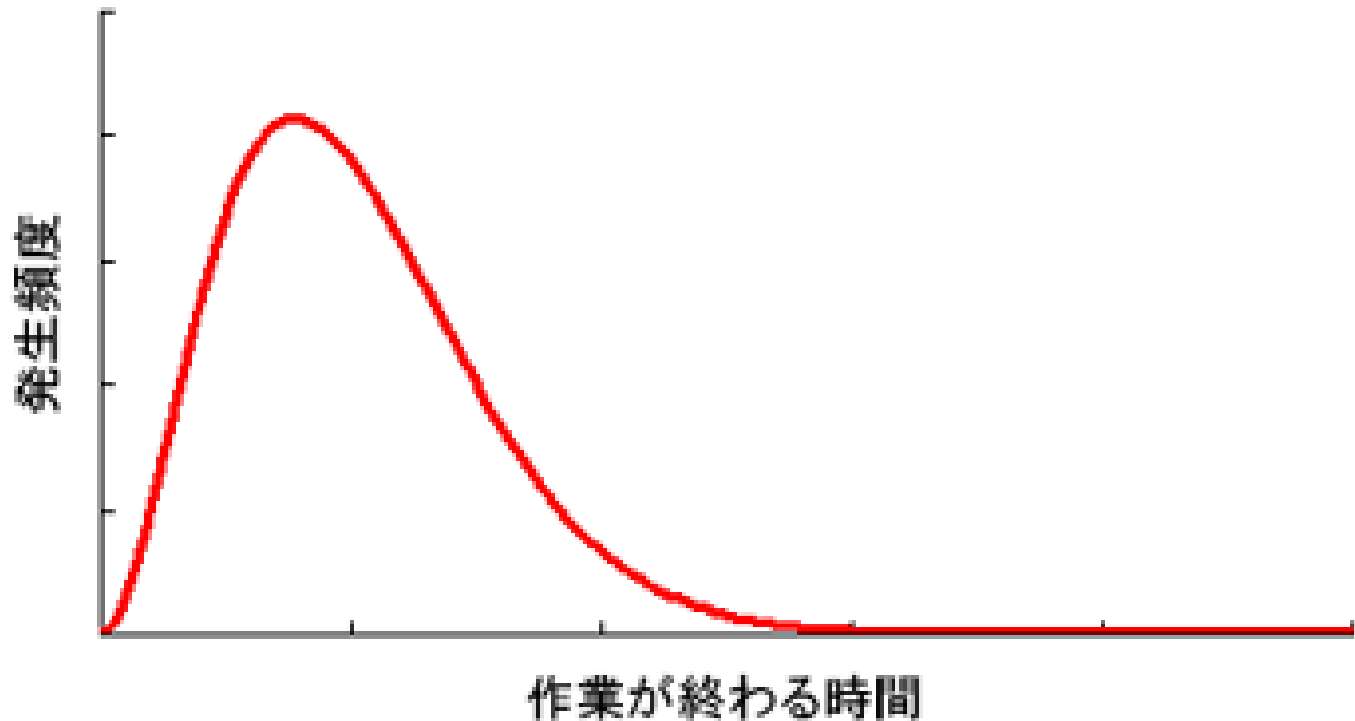
$$\mu = \frac{p}{p+q}, \quad \sigma^2 = \frac{pq}{(p+q)^2(p+q+1)}$$

一様分布 = $Be(1, 1)$

◆ベータ分布の例

プロジェクトの中のひとつひとつのステップ(作業)にかかる期間について

1つのステップが終了するまでの期間は、状況によってバラツキがある。予定よりも早く終わることもあれば、予想していなかった事態が起こり、大幅に遅れることもある。そのバラツキ具合は、「ベータ分布」に従う



ポアソン分布 p.35

確率密度関数

$$f(x) = \frac{e^{-\theta} \theta^x}{x!}, x = 0, 1, 2, \dots, 0 < \theta$$

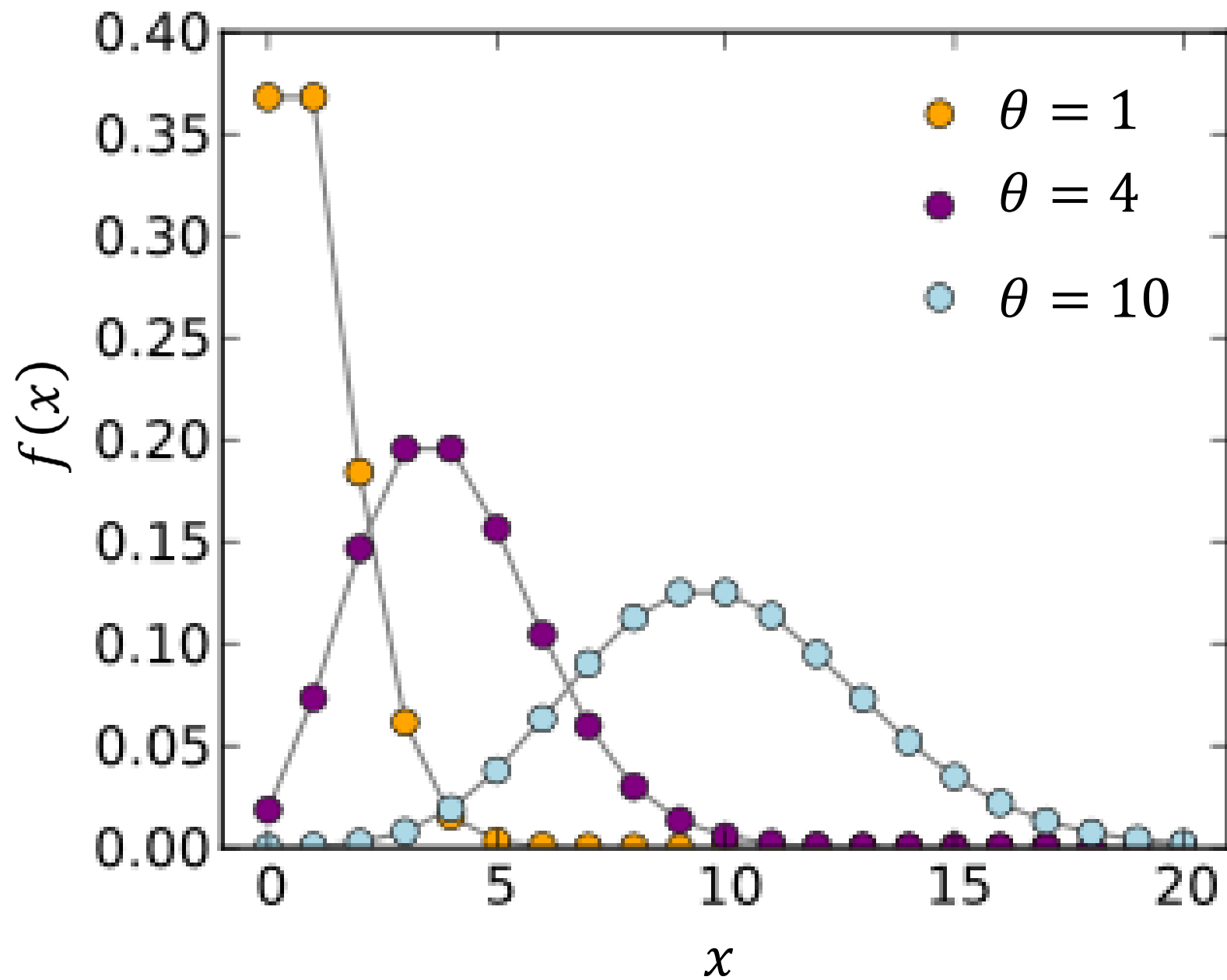
平均値, 分散

$$\mu = \theta, \quad \sigma^2 = \theta$$

θ はある区間内で発生する事象の期待回数

$f(x)$ は「単位時間中に平均で θ 回発生する事象が x 回発生する確率」に相当する.

(例) 事象が1分間で平均1回発生する場合, 6分間で事象が発生する回数に対する確率分布は $\theta = 6$ のポアソン分布に従う.



ポアソン分布は交通事故のような希な現象を説明するための確率分布として利用される.

分布図は, 横軸を回数 x , 縦軸を x 回起きる確率 $f(x)$ としたグラフで表す.

確率変数が離散値なので $f(x)$ は確率そのものを表す.

起こる確率が小さいので, 二項分布の左右に偏ったパターンになる.

起こる確率が高くなると, 正規分布に近づく.

(例)

ある都市の1日の交通事故死亡者数が3日間で1, 2, 3人だとする. このような事象が起こる確率

$$\frac{e^{-\theta} \theta^1}{1!}, \quad \frac{e^{-\theta} \theta^2}{2!}, \quad \frac{e^{-\theta} \theta^3}{3!}$$

事故回数は整数なので, 横軸は離散値になり, グラフは離散値を結んだ折れ線になる.

1日の交通事故死亡者数の期待値が1人($\theta = 1$)である場合→グラフ参照

ガンマ分布 p.35

ガンマ分布 $Ga(\alpha, \lambda)$ の確率密度関数

$$f(x) = kx^{\alpha-1}e^{-\lambda x}, 0 < x, 0 < \lambda, k: \text{定数}$$

k は規格条件 (確率の総和 = 1) より決まる.

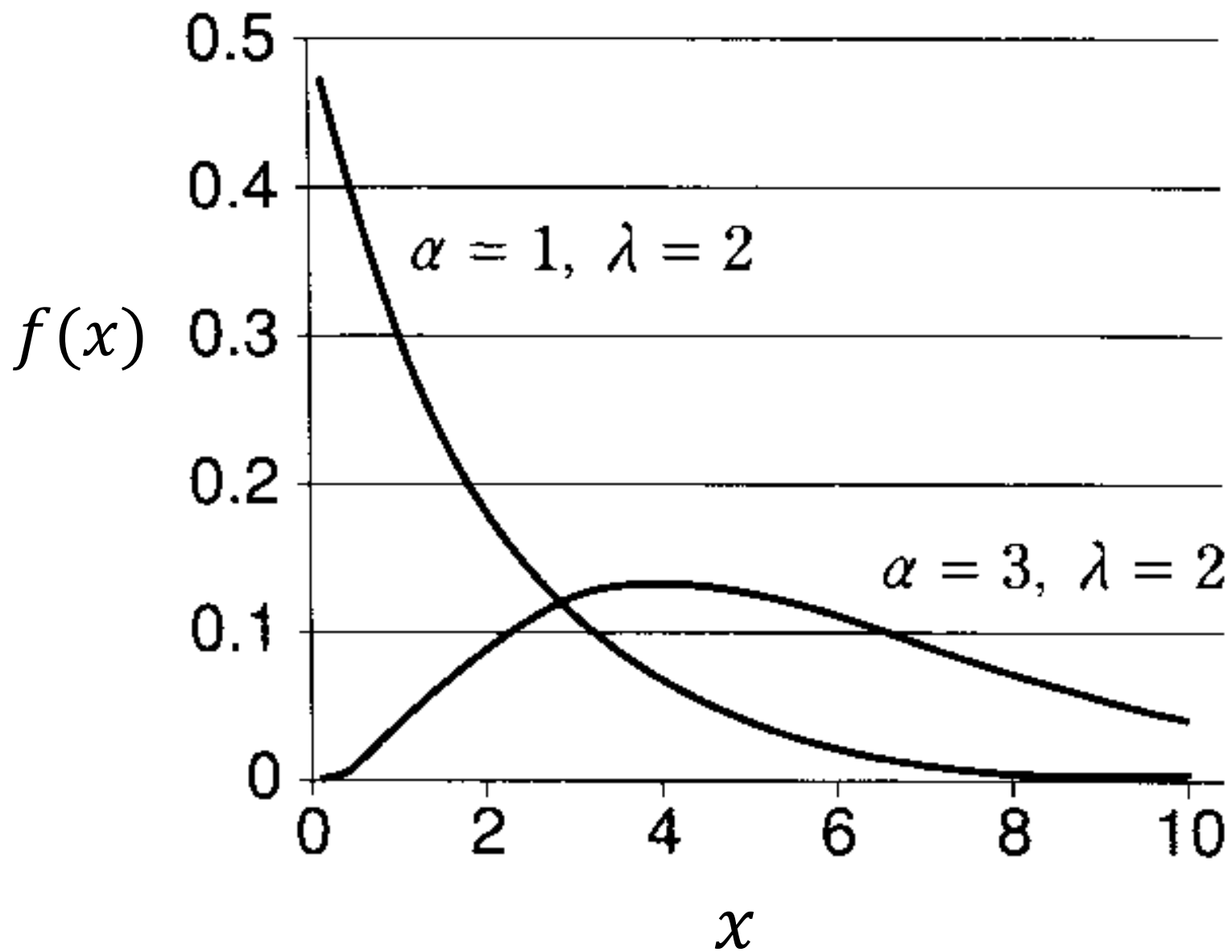
α : 形状母数, λ : 尺度母数 (母数: パラメータ)

平均値, 分散

$$\mu = \frac{\alpha}{\lambda}, \quad \sigma^2 = \frac{\alpha}{\lambda^2}$$

(応用分野)

信頼性工学における電子部品の寿命分布や通信工学におけるトラフィックの待ち時間分布



◆ガンマ分布の例

単位時間に λ 人の訪問者があるWebの場合、 α 人が訪問するまでの時間 x はガンマ分布に従う.

λ : 発生率, α : 事象の生起回数

x : 事象が発生するまでに要する時間

逆ガンマ分布 p.36

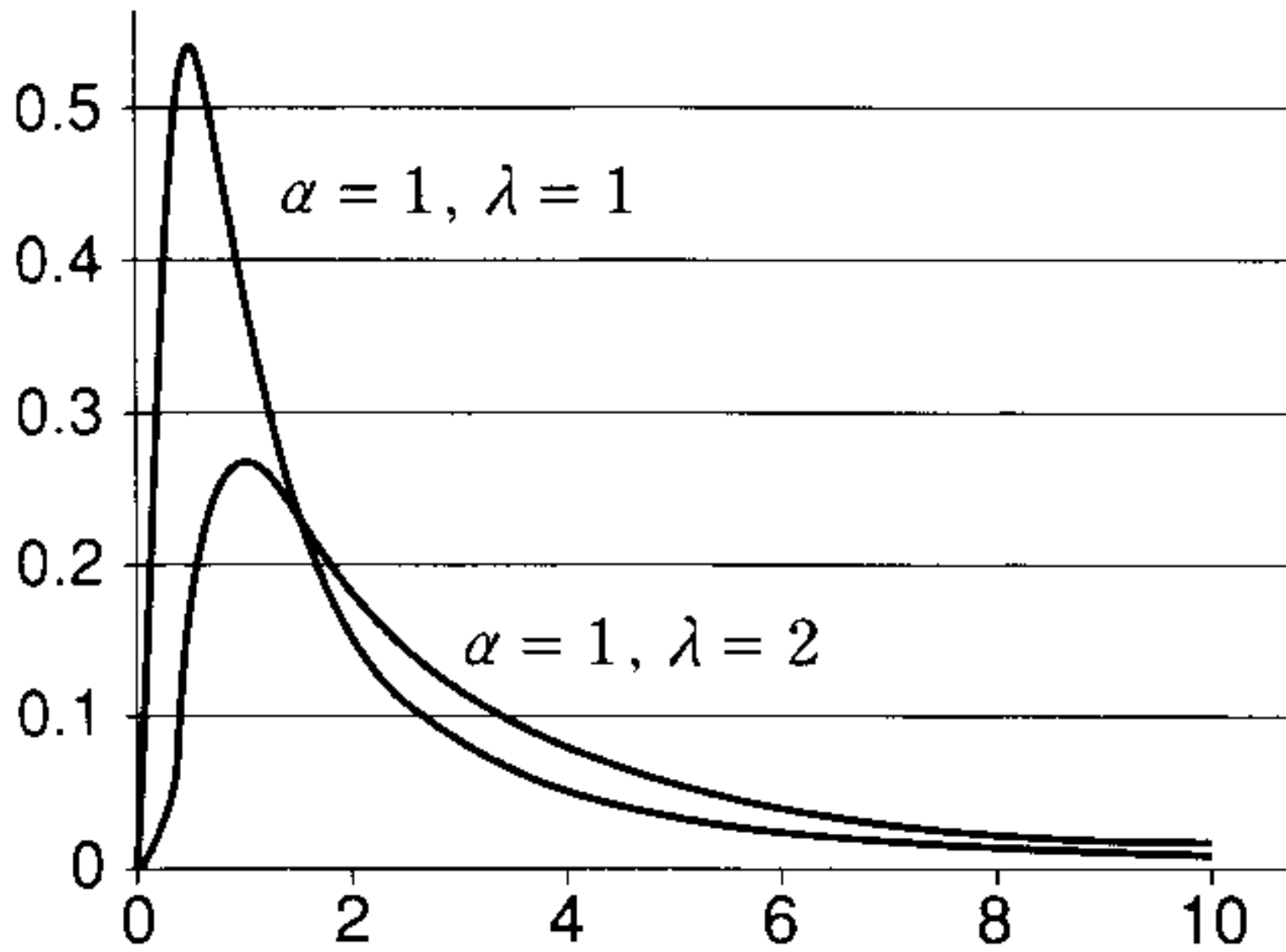
逆ガンマ分布 $IG(\alpha, \lambda)$ の確率密度関数

$$f(x) = kx^{-\alpha-1}e^{-\frac{\lambda}{x}}, \quad 0 < x, 0 < \lambda$$
$$f(x) = kx^{\alpha-1}e^{-\lambda x}, \quad 0 < x, 0 < \lambda, k: \text{定数}$$

k は規格化条件で決まる.

平均値, 分散

$$\mu = \frac{\lambda}{\alpha - 1}, \alpha > 1, \quad \sigma^2 = \frac{\lambda^2}{(\alpha - 1)^2(\alpha - 2)}$$



4 尤度関数と最尤推定法 p.38

統計資料の分析・・・統計モデルを作って分析

統計モデルには母数(パラメータ)が付随

母数の例: 正規分布の平均と分散(標準偏差)

統計的な分析→統計モデルの選択+母数の決定(推定)

母数の決定 → 最尤推定法

尤度: もっともらしさ

最尤推定法: もっともな値の推定法

最尤推定法の例題 p.38

コインの表の出る確率 p を最尤推定法で推定する.

コインを5回投げたとき, 次のような結果になったとする.

表, 表, 裏, 表, 裏

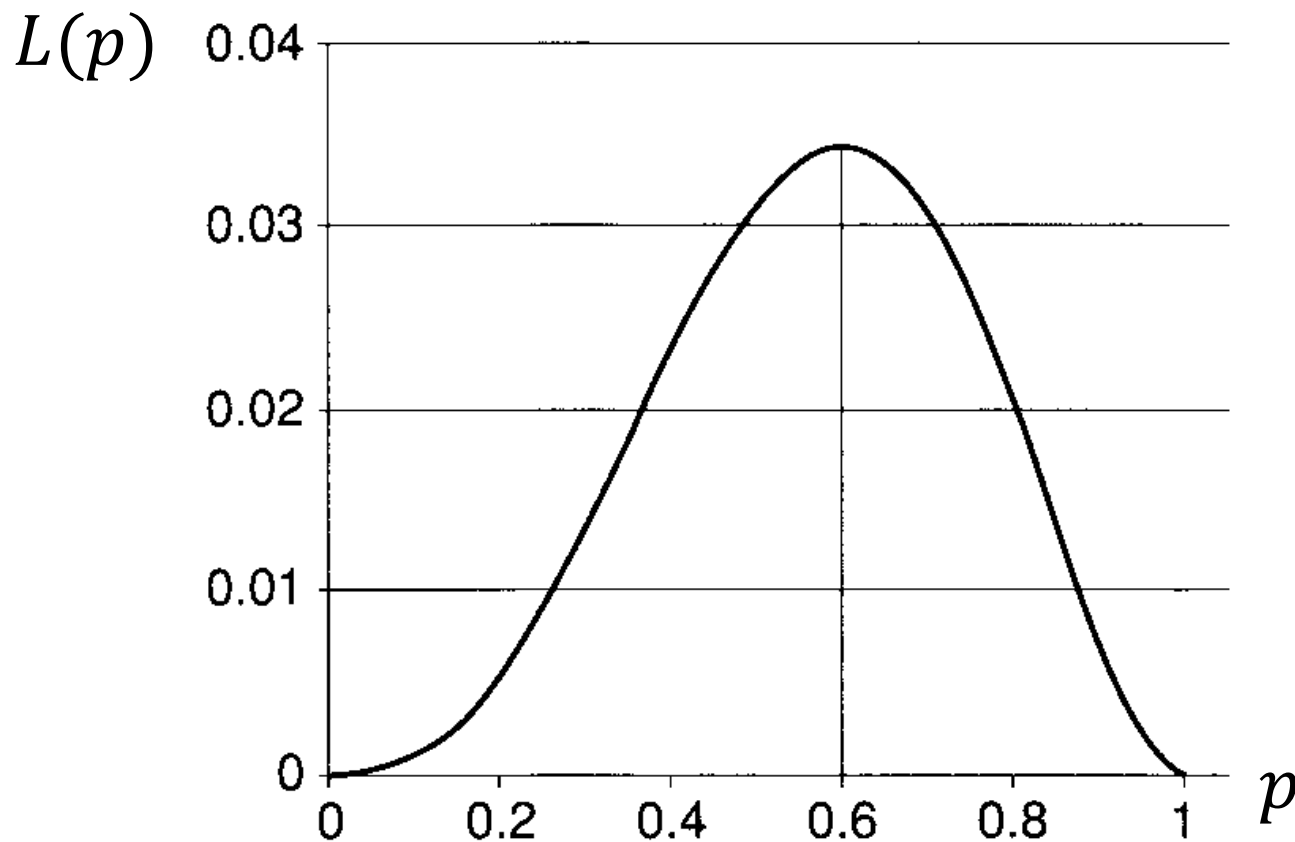
この結果をもたらす確率 p を求める.

この現象(コインを5回投げたとき, 表が3回, 裏が2回出る)が起こる確率は次のようになる.

$$L(p) = p \times p \times (1 - p) \times p \times (1 - p) = p^3(1 - p)^2$$

これを尤度関数と呼ぶ.

最尤推定法では尤度関数 $L(p)$ を最大にする p を求める.
すなわち, この現象が最も起こりやすい確率 p を求める.



$$\frac{dL(p)}{dp} = 3p^2 - 8p^3 + 5p^4 = p^2(5p^2 - 8p + 3) = 0$$

$$p = 0, 0.6, 1 \rightarrow 0.6$$

現象(表3回, 裏2回)→表の出る確率 p がやや高くなる.

対数尤度 p.39

尤度 $L(p)$ に対する対数尤度

$$\log_e L(p) = \log L(p)$$

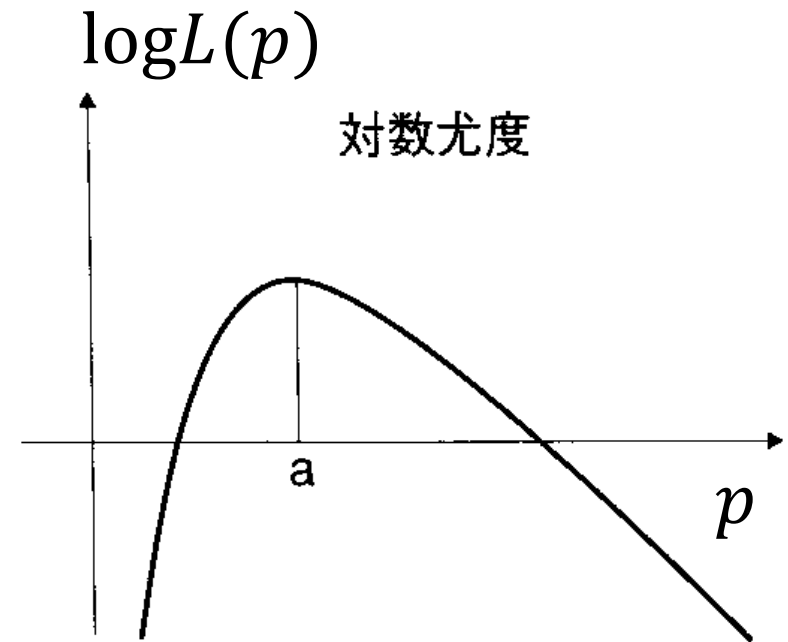
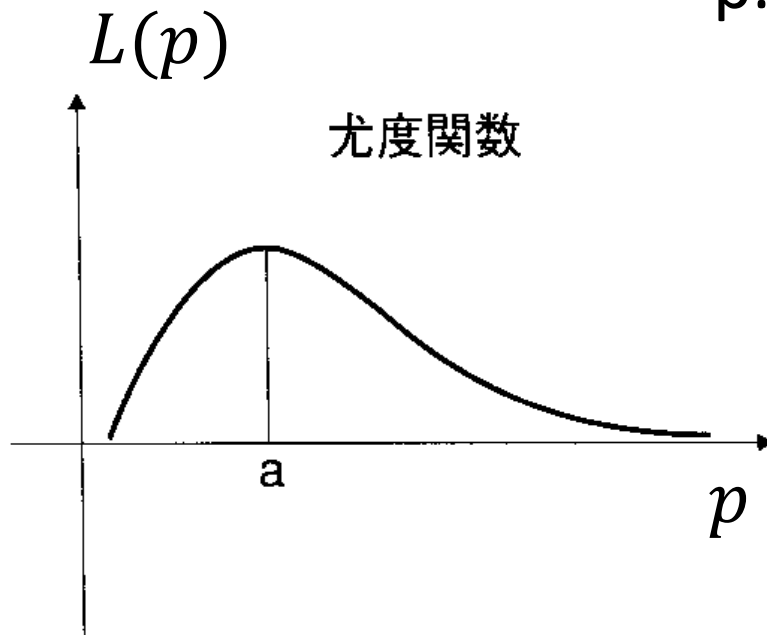
$$\log L(p) = \log p^3 (1 - p)^2 = 3 \log p + 2 \log(1 - p)$$

統計分析で利用される関数の多くは指数関数や積の形をしている.

→対数では倍数や和に変換され簡単な式で表現できる.

対数は単調増加関数であり, $L(p)$ と $\log L(p)$ に対する最尤推定値は一致する. ある p で $L(p)$ が最大となるとき, 同じ p に対して $\log L(p)$ も最大となる.

p.40



尤度関数の最大値を与える最尤推定値と、対数尤度の最尤推定値は一致する。

第2章 ベイズの定理とその応用

1 ベイズ定理とは p.42

■条件付き確率と乗法定理 <省略>

■シンプルなベイズの定理 p.43

乗法定理より

$$P(A \cap B) = P(B|A)P(A) \cdots \textcircled{1}$$

$$P(A \cap B) = P(A|B)P(B) \cdots \textcircled{2}$$

事象A, Bの同時確率であるからAとBを入れ替えることが出来る. ①の右辺=②の右辺より,

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

これを $P(A|B)$ について解く.

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

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

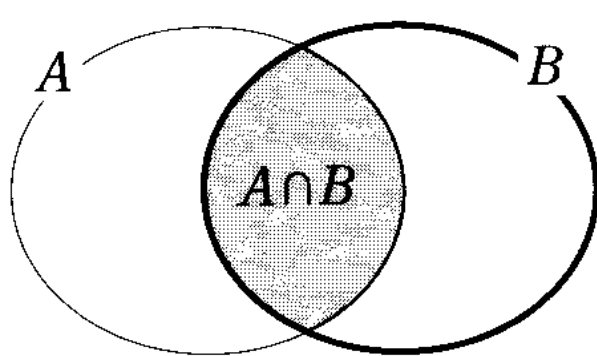
原因：事象A

結果：事象B

$P(A|B)$ とは「結果が事象Bであるとき、その原因が事象Aである確率」→逆確率，原因の確率

$P(A)$ ： 結果Bが起こる前の確率 →事前確率

$P(A|B)$ ：結果Bが起こった後の確率→事後確率



$P(A|B)$

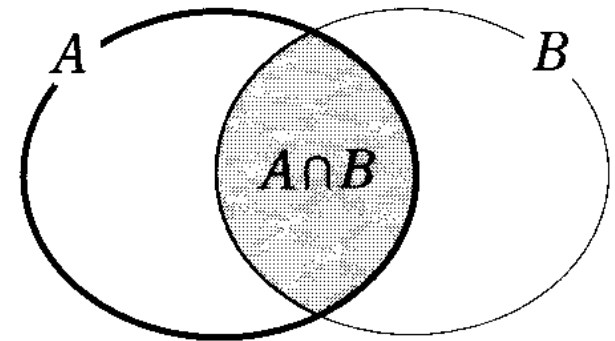
(Bが起こったときのAの確率)

ベイズの定理

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



変換



$P(B|A)$

(Aが起こったときのBの確率)

p.44

ベイズ定理の確認 p.45

(例) 3枚のカード「e」, 「f」, 「g」が箱に入っている.

カード「e」: 両面が白

カード「f」: 片面が白, 片面が黒

カード「g」: 両面が黒

(問題) 1枚のカードを箱から無作為に取り出して, 机上に置く. 取り出したカードの上面が白のとき, そのカードが「f」である確率はいくらか.

＜解1＞確率の定義を用いる

事象 F : 取り出したカードが「f」である.

事象 W : 取り出したカードの上面が白である.

求めるもの: 取り出したカードの上面が白であるとき, そのカードが「f」である確率 = $P(F|W)$

取り出されたカードの上面が白である場合は以下の3通りである.

①「e」表一白, ②「e」裏一白, ③「f」一白

これらは同様に確からしいので, ③が起こる確率は $1/3$.

$$P(F|W) = 1/3$$

＜解2＞ベイズの定理を用いる p.46

事象A→事象F(取り出したカードが「f」である)

事象B→事象W(取り出したカードの上面が白色である)

求めるもの: 取り出したカードの上面が白であるとき, そのカードが「f」である確率

ベイズの定理より,

$$P(F|W) = \frac{P(W|F)P(F)}{P(W)}$$

$P(F)$: 3枚のカードから1枚のカード「f」を取り出す確率
= 1/3

$P(W|F)$: カード「f」が取り出されたとき, その上面が白である確率 = 1/2

$P(W)$: 取り出したカードの上面が白である確率

以下のように、カードを取り出す全ての場合(①～⑥)を考えると、事象 W は①, ②, ③に該当する.

① e表一白, ② e裏一白, ③ f一白, ④ f一黒,
⑤ g表一黒, ⑥ g裏一黒

$$P(W) = 3/6 = 1/2$$

以上より,

$$P(F|W) = \frac{(1/2) \times (1/3)}{1/2} = 1/3$$

原因と結果の関係

カードを選択する → 色の原因

色 → カード選択による結果

結果の色から, 原因のカード選択の確率を求めている.

・・・事後確率(原因の確率)

確率の計算に必要なもの

- ・原因の確率 $P(F)$ ・・・事前確率
- ・原因→結果の確率 $P(W|F)$ ・・・条件付き確率
- ・結果の確率 $P(W)$

$$\text{事後確率} = \frac{\text{条件付き確率} \times \text{事前確率}}{\text{結果の確率}}$$

演習問題

パン屋が3軒あり，売っている種類は以下の通りである．

A店 あんパン，メロンパン，クロワッサン

B店 サンドウィッチ，フランスパン，あんパン

C店 メロンパン，あんパン，クリームパン

<ベイズの定理を用いて計算すること>

1. ある人があんパンを買ったとき，それをA店で買った確率を求めよ．
2. ある人がメロンパンを買ったとき，それをC店で買った確率を求めよ．
3. ある人がフランスパンを買ったとき，それをB店で買った確率を求めよ．

<1. の問題について>

事象 F (カード f である)・・・ A店で買う

事象 W (白色である)・・・ あんパンを買う

$$P(F|W) = \frac{P(W|F)P(F)}{P(W)}$$

$P(F)$: A店で買う確率→3店から1店を選ぶ→ $1/3$

$P(W|F)$ =A店の中であんパンを買う確率→3種類から1種類を選ぶ→ $1/3$

$P(W)$: あんパンを買う確率→全ての組合せ9通り[①～⑨]からあんパンを含む組合せ[①, ⑥, ⑧]を選ぶ
→ $3/9 = 1/3$

全ての組合せ=9通り

- ①Aーあんパン, ②Aーメロンパン, ③Aークロワッサン,
④Bーサンドウィッチ, ⑤Bーフランスパン, ⑥Bーあんパン,
⑦Cーメロンパン, ⑧Cーあんパン, ⑨Cークリームパン