医学統計学

確率分布１：
一様分布・二項分布・ポアソン分布・ガウス分布

芳賀　昭弘

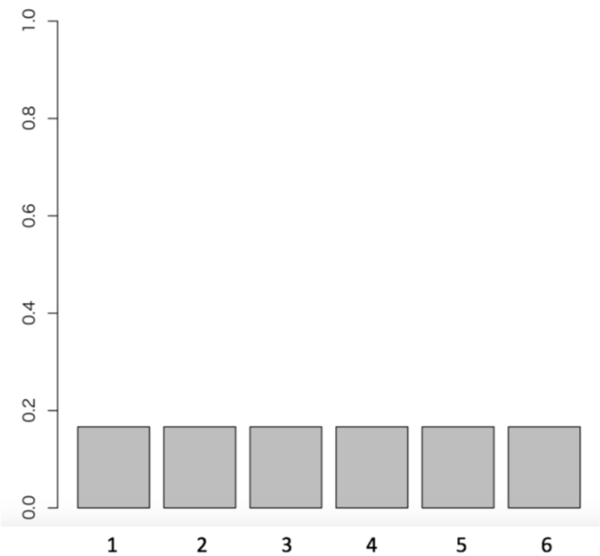
:（前期）月曜日6講時　15:30 - 16:30

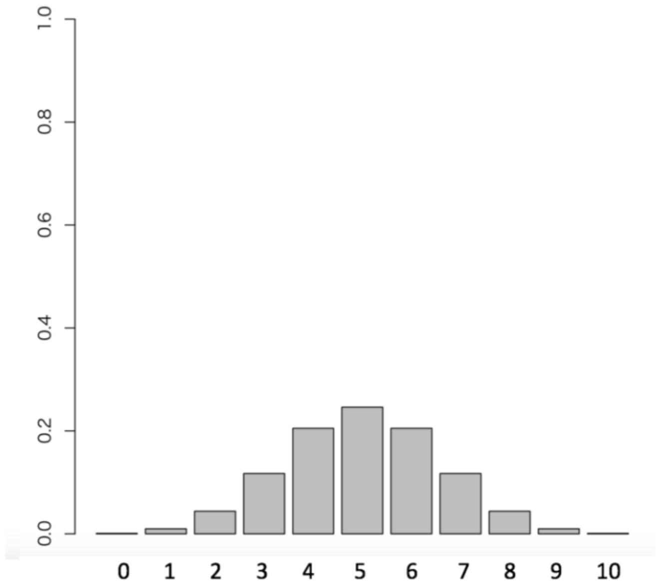
確率分布

- 事象が発生する確率を分布で表したもの

例1） サイコロのそれぞれの目が出る確率は？

例2） 10回コインを投げた時に、表が出る回数が0~10の確率は？

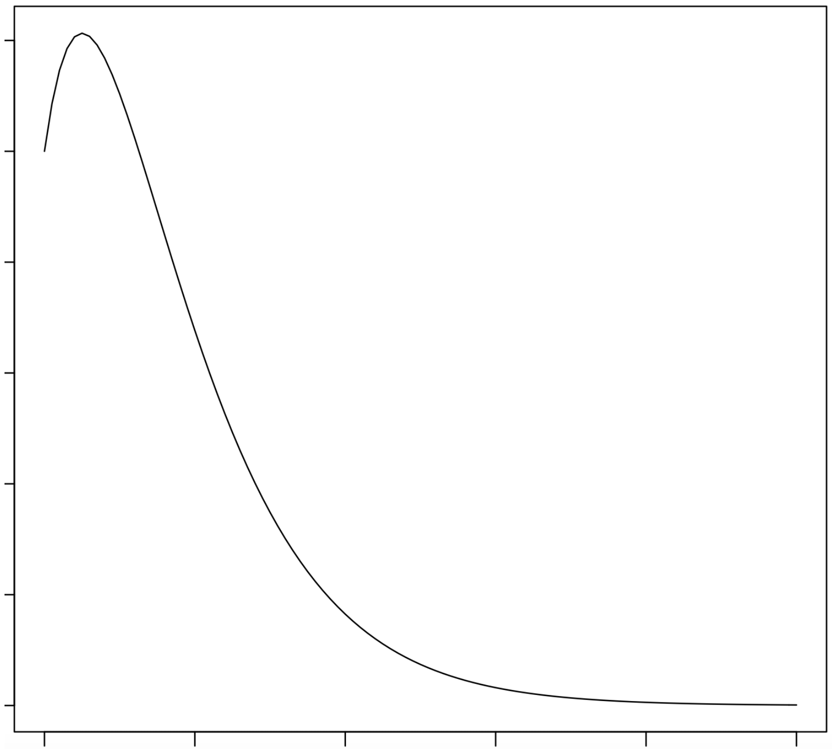




確率分布

- 事象が発生する確率を分布で表したもの

例3）　ある高さからコインを落とした時に、コインが転がって拡がる半径は？



連続

確率分布

- 離散的な場合もあれば、連続的な場合もある

確率分布の基本的性質

- 全ての事象の生起確率を足す（積分）すると、1になる
- 生起確率は負にならない

離散変数

$$\sum_{i=1}^{n} \rho(x_i) = 1$$

$$\rho(x_i) \geq 0, \quad (n = 1, 2, \cdots n)$$

or

連続変数

$$\int_a^b \rho(x) dx = 1$$

$$\rho(x) \geq 0, \quad (a \leq x \leq b)$$

確率分布の性質

- 加法定理

$$p(X) = \sum_{Y} p(X, Y)$$

同時分布

- 乗法定理

$$p(X, Y) = p(Y|X)p(X)$$

条件付き分布

→ ベイズの定理

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

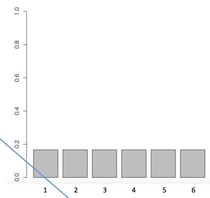
一様分布（Uniform distribution）

- 全ての確率変数（階級値）の生起確率が同じ

離散変数の時

発生し得る事象xがN通り存在する場合，
事象x_iの生起確率$p(x_i)$は

$$p(x_i) = \frac{1}{N}$$



$$\sum_{i=1}^{n} \rho(x_i) = 1$$

期待値： $E[x] = \sum_{i}^{n} P(x_i) x_i$

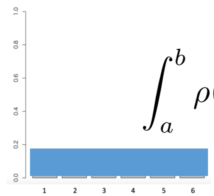
分散： $V[x] = E[x^2] - E[x]^2$

ここで $E[x^2] = \sum_{i}^{n} P(x_i) x_i^2$

連続変数の時

発生し得る事象xの範囲が$[a, b]$の，
事象xの生起確率p(x)は

$$p(x_i) = \frac{1}{b-a}$$



$$\int_a^b \rho(x) dx = 1$$

期待値： $E[x] = \int_a^b x P(x) dx$

分散： $V[x] = E[x^2] - E[x]^2$

ここで $E[x^2] = \int_a^b x^2 P(x) dx$

二項分布（Binomial distribution）

- n回のベルヌーイ試行※における成功回数xの分布

表・裏：　合格・不合格：　有り：無し　・・・などを考える時に現れる

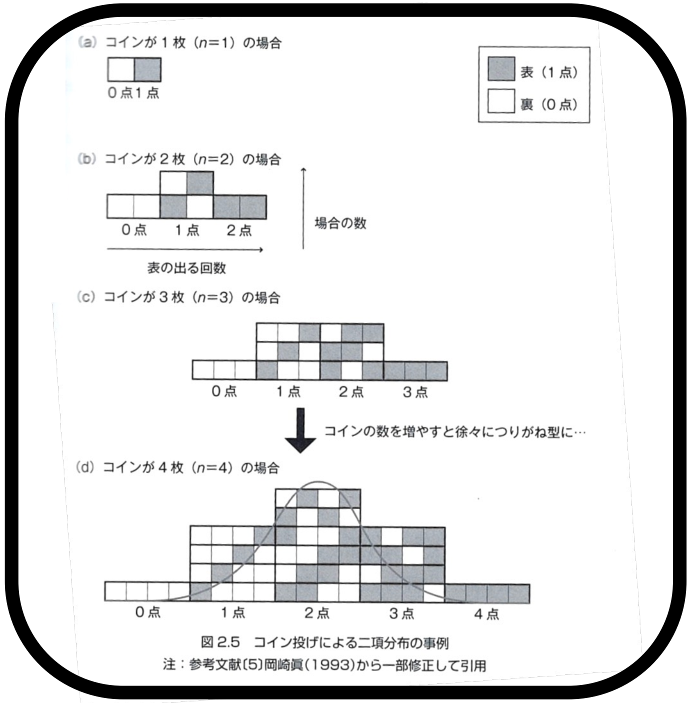


図2.5　コイン投げによる二項分布の事例
注：参考文献[5]岡崎眞（1993）から一部修正して引用

ヤコブ・ベルヌーイ

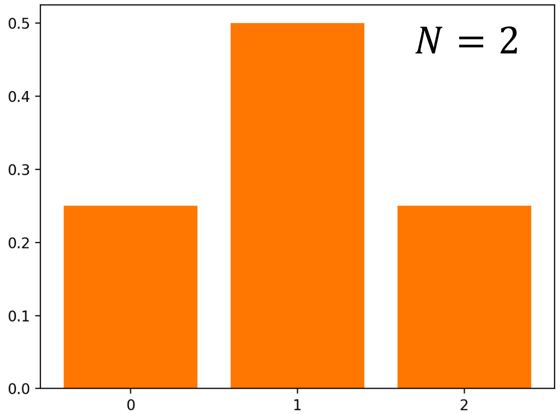
Wikipedia

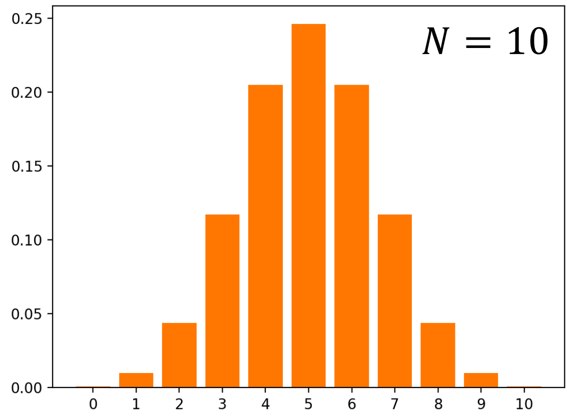
※**ベルヌーイ試行**：コイン投げの表と裏のように2つの結果（二値）しかない試行のこと。コイン投げにおいて、表と裏の出る確率は等しいと考えるかもしれないが、もちろんより一般に異なる確率（つまり出現確率がどちらかに偏っている）でも良い。

表・裏の出る確率が等しくないときでも良い

$$P(x) = \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}$$

表裏の出る確率が等しい($p = 0.5$)の時の二項分布





表が出る確率：p

裏が出る確率：$q\ (= 1 - p)$

$(p+q)$ 表 or 裏 × $(p+q)$ 表 or 裏 × ⋯ (N)

$q = 1 - p$ なので$(p+q)^N = 1^N = 1$

確率変数xで二項分布の和を取ると確かに1となっていることが確認できる。

$$= (p+q)^N = \sum_{x=0}^{N} \frac{N!}{(N-x)!\,x!} p^x q^{N-x} = \sum_{x=0}^{N} \mathrm{Bin}(x|N,p)$$

このうち p^x 項がx回表が出る確率 $= \mathrm{Bin}(x|N,p)$

二項分布（Binomial distribution）

例題１：　サイコロを6回投げて1の目が2回以上出る確率はいくらか？

~0.26

例題２：以下の二つのうちどちらの賭けをしないといけない時に、あなたはどちらに賭けますか？
・細工のないコインを5回投げて2回表が出ると100円もらえる
・細工のないコインを5回投げて3回以上表が出ると200円もらえる

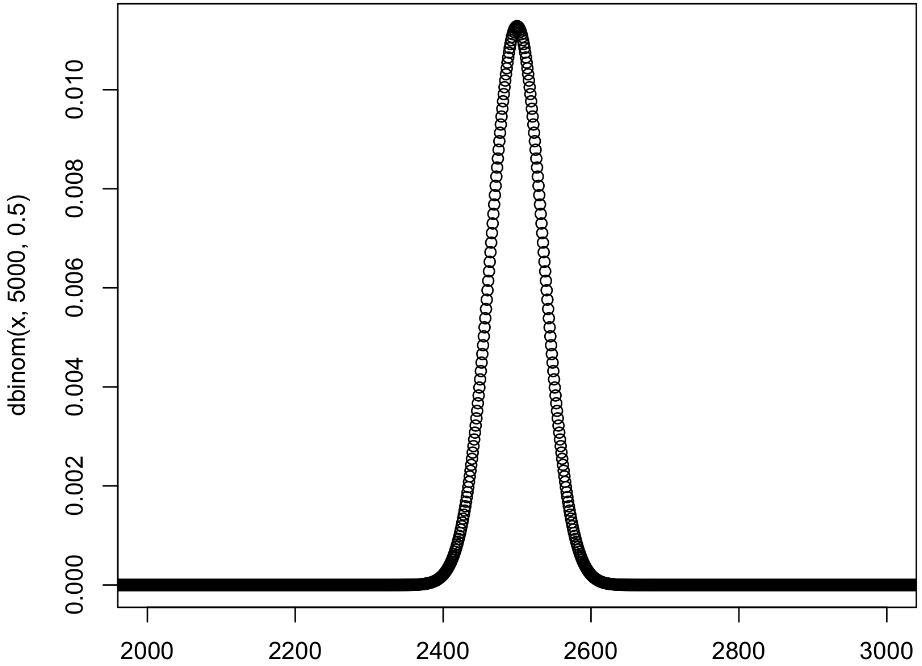
0.3125 vs 0.5

例題３：日本人のコロナに感染している確率をpとしよう。その時、感染していない日本人の確率qは1-pとなる。

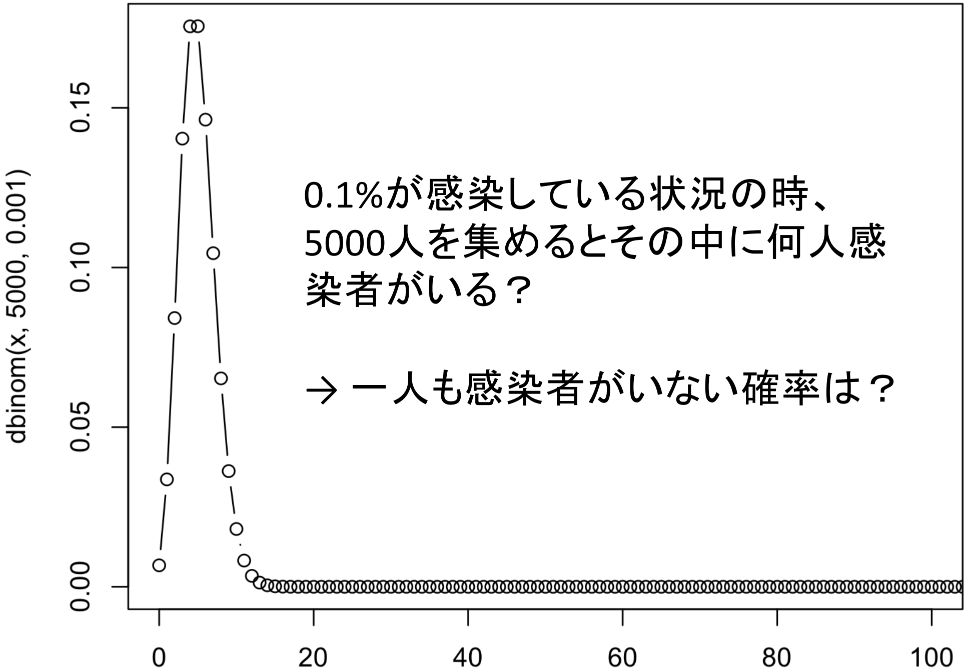
では、n人の日本人を適当に集めた時、x人がコロナに感染しているという確率は？pとnで表せ。

二項分布（Binomial distribution）

P=0.5の時の分布



P=0.001の時の分布



例題4

- ５人の患者のうち４人が介護が必要という病気がある。実際に５人の患者を集めた時に4人介護が必要であった確率はどれか？最も近いものを選べ。

1. 0.9997
2. 0.6723

〇3. 0.4096

4. 0.3277
5. 0.2048

医学統計学

確率分布２：
一様分布・二項分布・ポアソン分布・ガウス分布

芳賀　昭弘

:（前期）月曜日6講時　15:30 - 16:30

離散型分布の一つ

ポアソン分布（Poisson distribution）

シメオン・ドニ・ポアソン
Wikipedia

- 二項分布において, nが非常に大きくかつpが非常に小さいという極端な時を考えてみよう（これは前回示した感染症の例が当てはまる）。

$$P(x) = \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}$$

- この時、$np = \lambda$とおくと上記の式は以下のようになる※

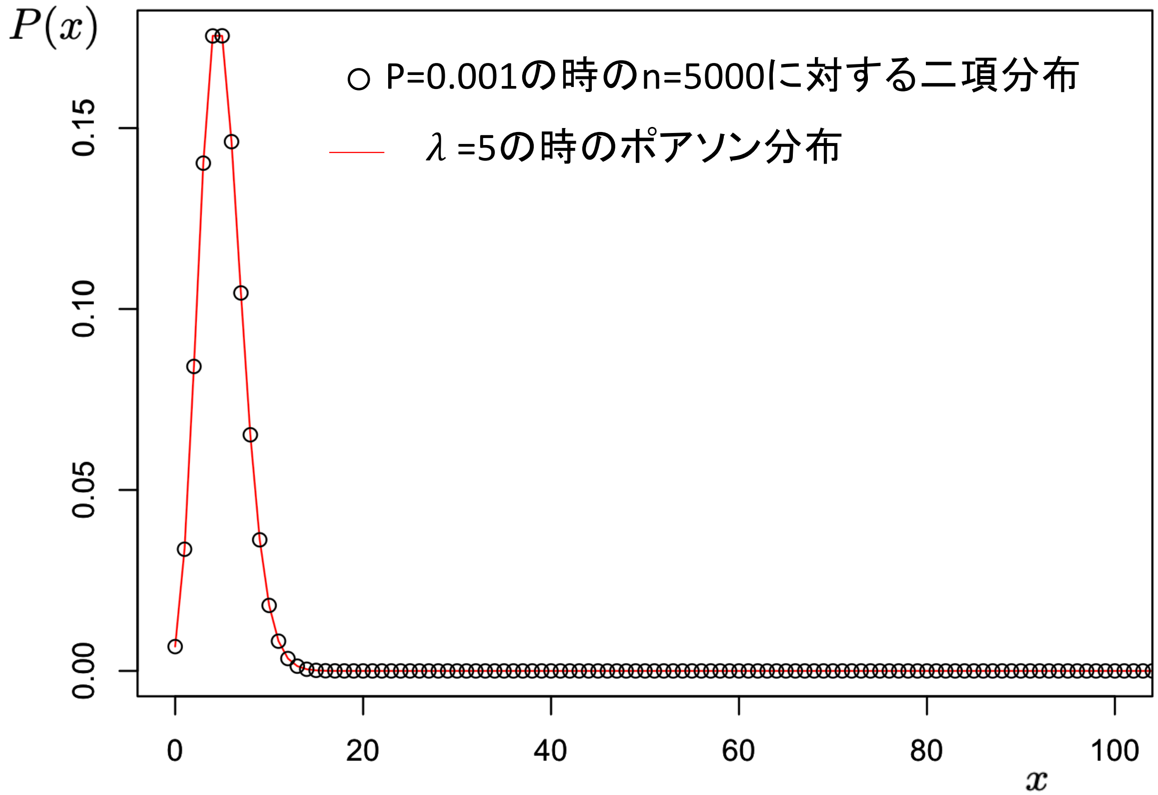
$$P(x) = \frac{\lambda^x}{x!}\mathrm{e}^{-\lambda}$$

これをポアソン分布という

※ eが次のように書けることを利用することで証明できる（テイラー展開も参照）

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-\frac{n}{\lambda}} = \mathrm{e}$$

ポアソン分布（Poisson distribution）

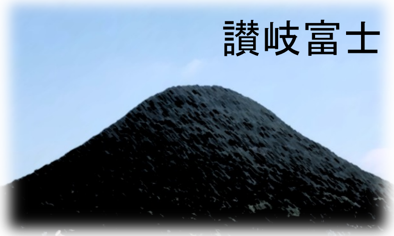


ポアソン分布の例；

- 感染率が小さいときにある大規模集団に感染者が何人いるかを見積もる
- 透過力が強い（＝反応確率が低い）X線を被写体に大量に射出して透過後のX線を検出器で検出する量を見積もる
- 水試料中に含まれる非常に少ないある特定の細菌を1ミリリットル中に見出す数

変数は整数となる（すなわちポアソン分布は離散型分布）

ガウス分布（Gaussian distribution）



- ガウス分布は統計学において様々な場面に現れる連続変数の確率分布である。最も重要な分布と言って過言ではない。
- 二項分布においてnが大きいとき、ポアソン分布においてλが大きい時、ガウス分布となる（各自証明してみよう）。
- 正規分布（Normal distribution）とも呼ばれる。

$$A\,e^{bx^2}$$

ガウス分布

$$\frac{1}{(2\pi\sigma^2)^{1/2}}\,e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

正規分布

xの期待値 $\mathrm{E}[x]=\mu$

xの分散 $\mathrm{V}[x]=\sigma^2$

$z=\dfrac{x-\mu}{\sigma}$ とおく

$$\frac{1}{(2\pi)^{1/2}}\,e^{-\frac{1}{2}z^2}$$

標準正規分布

zの期待値 $\mathrm{E}[z]=0$

zの分散 $\mathrm{V}[z]=1$

ガウス分布（Gaussian distribution）

カール・フリードリヒ・ガウス

Wikipedia

- 表記

$$\mathcal{N}(x|\mu,\sigma) = \frac{1}{(2\pi\sigma^2)^{1/2}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

- xの定義される範囲 $[-\infty, +\infty]$

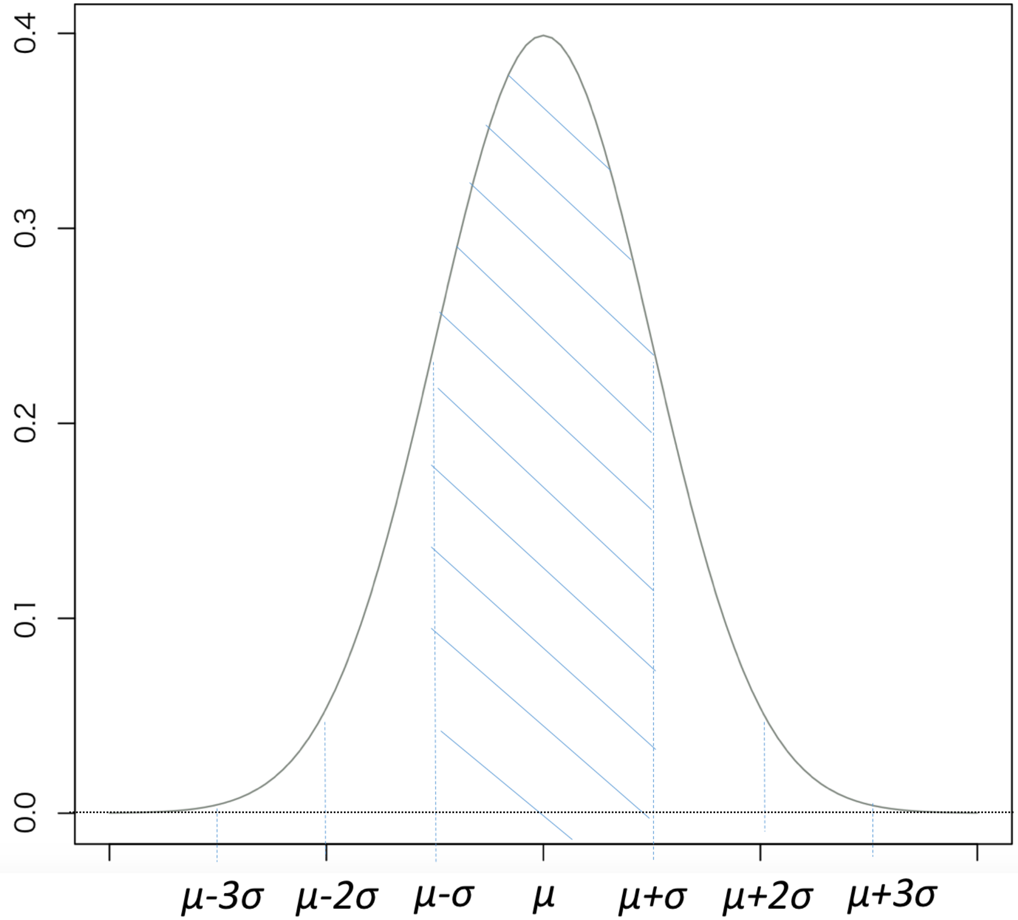
範囲 $[-\infty, +\infty]$で積分すると、

$$\int_{-\infty}^{+\infty} \mathcal{N}(x|\mu,\sigma)\, dx = \int_{-\infty}^{+\infty} \frac{1}{(2\pi\sigma^2)^{1/2}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = 1$$

ガウス積分公式: $\int_{-\infty}^{\infty} \mathrm{e}^{-ax^2} dx = \sqrt{\pi/a}$

から、上記が成り立つことを示してください

ガウス分布（Gaussian distribution）



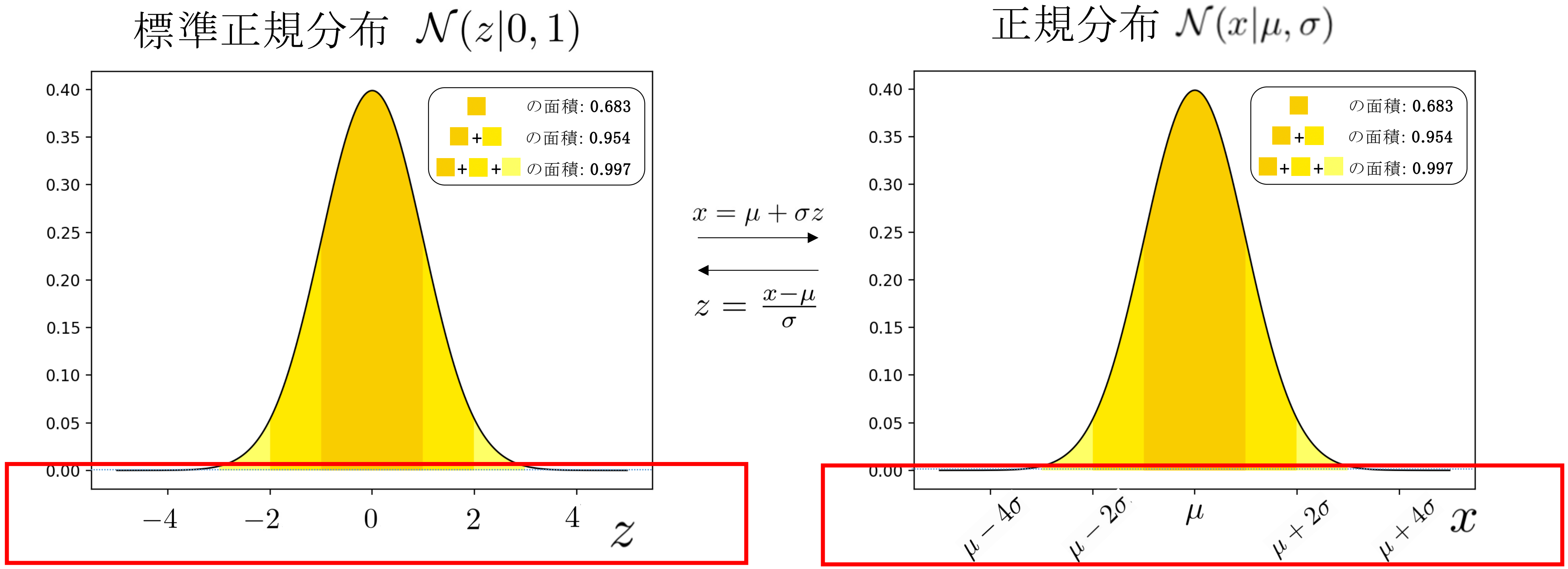
95%を与える範囲は？

$[-1.96\,\sigma, +1.96\,\sigma]$

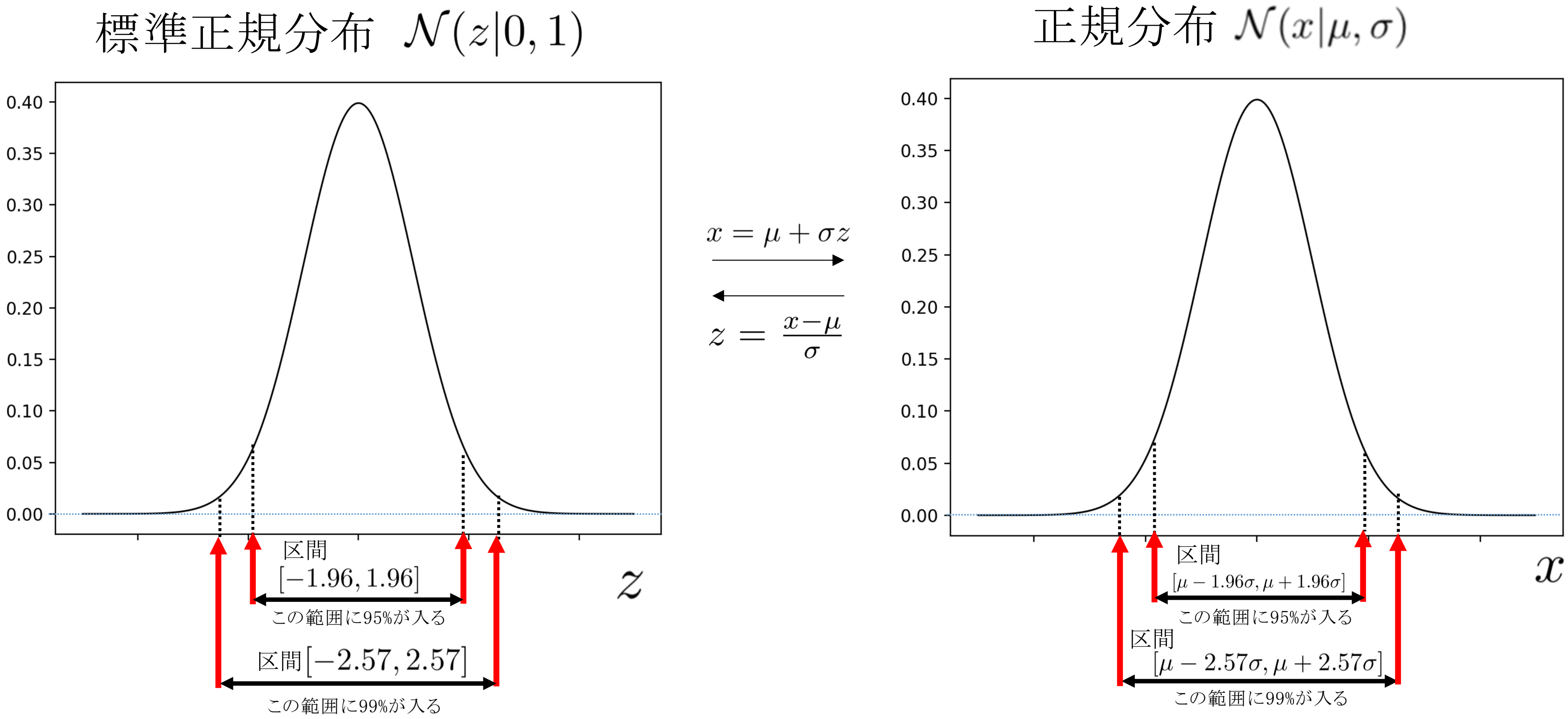
99%を与える範囲は？

$[-2.57\,\sigma, +2.57\,\sigma]$

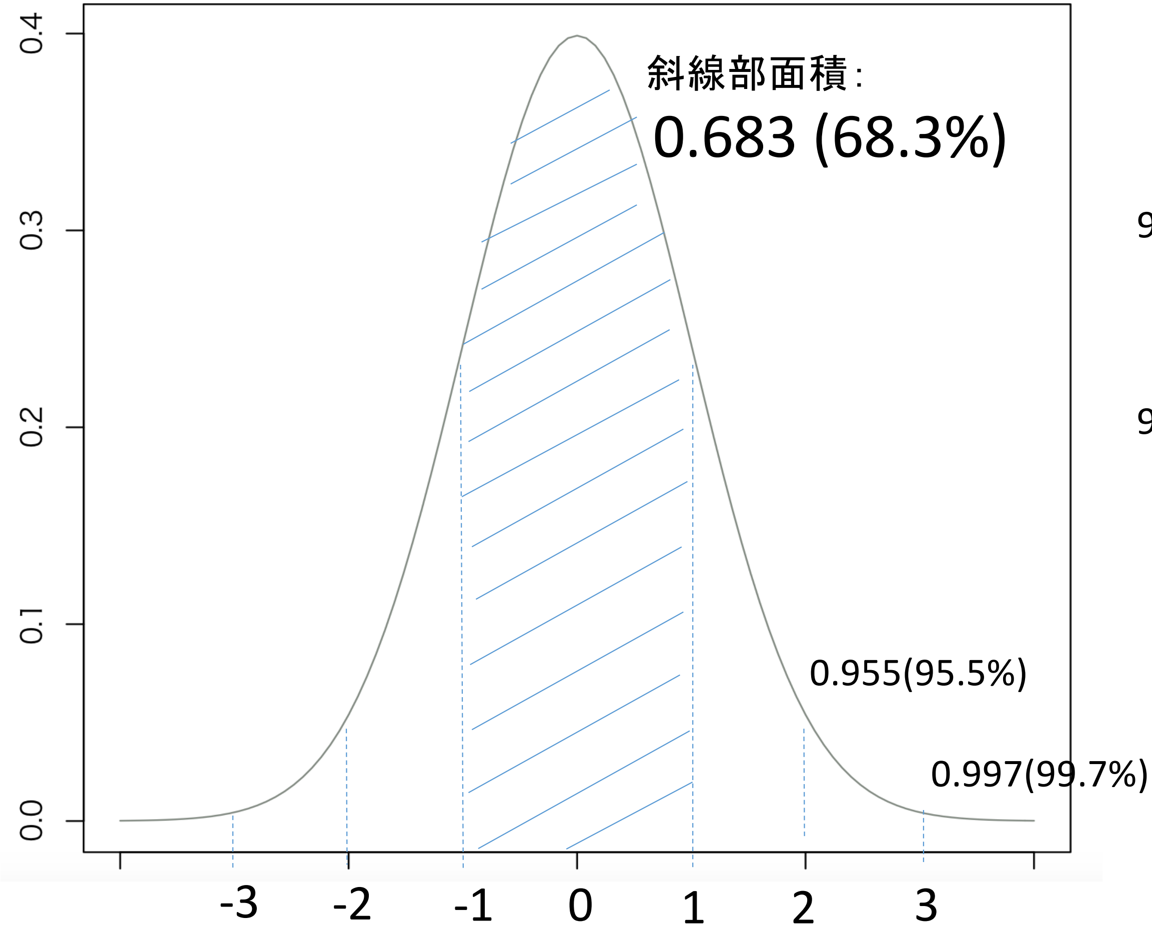
積分区間	面積	
[μ-σ, μ+σ]	0.683 (68.3%)	1 σ（いちシグマ）
[μ-2σ, μ+2σ]	0.955 (95.5%)	2 σ（にシグマ）
[μ-3σ, μ+3σ]	0.997 (99.7%)	3 σ（さんシグマ）
$[-\infty, +\infty]$	1.0 (100%)	



標準正規分布 $\mathcal{N}(z|0,1)$
正規分布 $\mathcal{N}(x|\mu,\sigma)$
の面積: 0.683
+ の面積: 0.954
+ + の面積: 0.997
$x = \mu + \sigma z$
$z = \frac{x-\mu}{\sigma}$
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-4
-2
0
2
4
z
$\mu-4\sigma$
$\mu-2\sigma$
μ
$\mu+2\sigma$
$\mu+4\sigma$
x


標準正規分布 $\mathcal{N}(z|0,1)$
正規分布 $\mathcal{N}(x|\mu,\sigma)$
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
$x = \mu + \sigma z$
$z = \frac{x-\mu}{\sigma}$
区間 $[-1.96, 1.96]$
この範囲に95%が入る
区間 $[-2.57, 2.57]$
この範囲に99%が入る
z
区間 $[\mu - 1.96\sigma, \mu + 1.96\sigma]$
この範囲に95%が入る
区間 $[\mu - 2.57\sigma, \mu + 2.57\sigma]$
この範囲に99%が入る
x

標準正規分布の場合

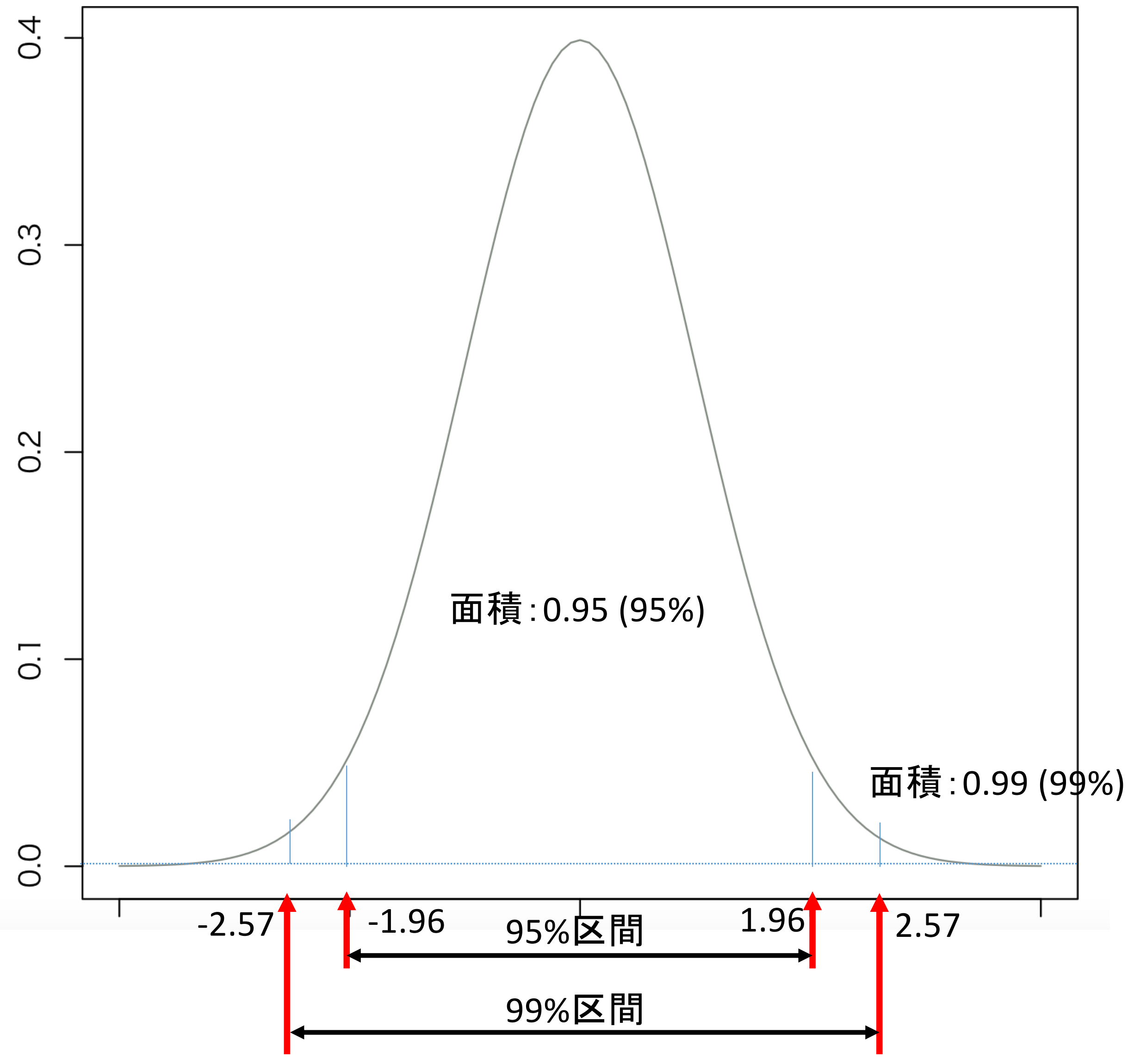


95%を与える範囲は？

$[-1.96, +1.96]$

99%を与える範囲は？

$[-2.57, +2.57]$


0.4
0.3
0.2
0.1
0.0
面積：0.95 (95%)
面積：0.99 (99%)
-2.57
-1.96
1.96
2.57
95%区間
99%区間

例題

(1) 平均が50点で標準偏差が10点の正規分布において、上位0.5%に入るために必要な点数はいくらか。

(2) (1)のテストで偏差値60をとるには上位何%に入らなければならないか。