統計数理

石川顕一

http://ishiken.free.fr/lecture.html

11/15は演習のみ

参考書

[1] 理工系の数学入門コース7・薩摩順吉著
「確率・統計」岩波書店　2001年

[2] コルモゴロフ、ジュルベンコ、プロホロフ　共著
丸山　哲郎、馬場　良和　共訳
「コルモゴロフの確率論入門」森北出版 2003年

[3] 岩波基礎物理シリーズ8・北原和夫著
「非平衡系の統計力学」岩波書店　1997年

統計数理

石川顕一

11/1　　代表的な確率分布

- ２項分布
- ポアソン分布
- 正規分布
- 中心極限定理

3−1　2項分布

→ 参考書[1] p.68

- 2項分布の定義

[例] サイコロを5回振る。このとき、1の目が出る回数を確率変数Xとする。Xはどのような確率分布に従うであろうか。

X=2となる場合の数 ⇨ ${}_5C_2 = 10$ 通り

1つ1つの場合の起こる確率 ⇨ $\underbrace{\frac{1}{6}\times\frac{1}{6}}_{1の目2回}\times\underbrace{\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}}_{1以外の目3回}=\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3$

確率密度　$f(2)=10\times\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3=\frac{625}{3888}=0.161$ ← 1の目が2回出る確率

$f(0)=\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^5=\frac{3125}{7776}=0.402$　　$f(1)=\underbrace{5}_{{}_5C_1}\times\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^4=\frac{3125}{7776}=0.402$

3ー1　2項分布

- ２項分布の定義

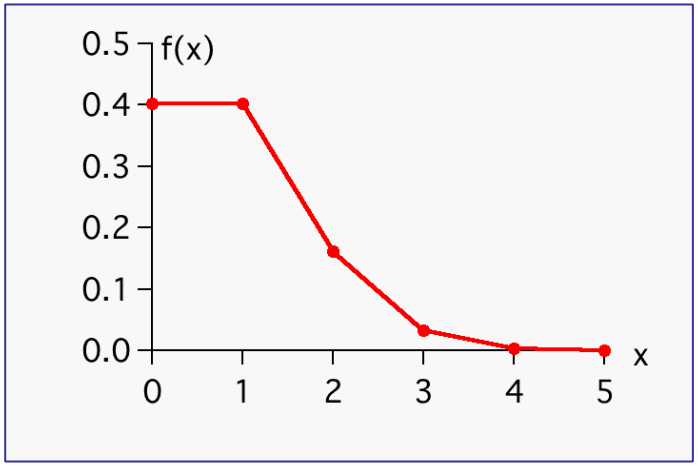
[例] サイコロを5回振る。このとき、1の目が出る回数を確率変数Xとする。Xはどのような確率分布に従うであろうか。

$$f(2)=10\times\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3=\frac{625}{3888}=0.161$$

$$f(0)=\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^5=\frac{3125}{7776}=0.402$$

$$f(1)=5\times\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^4=\frac{3125}{7776}=0.402$$

$$f(3)=\frac{125}{3888}\qquad f(4)=\frac{25}{7776}\qquad f(5)=\frac{1}{7776}$$



２項分布（ベルヌーイ分布）

ある事象Aの起こる確率$P(A)=p$が与えられているとき、n回独立試行を行ってAがx 回起こる確率は、

$$Bin(n,p) \Rightarrow f(x)={}_nC_x p^x(1-p)^{n-x}\qquad (x=0,1,2,\mathcal{L},n)$$

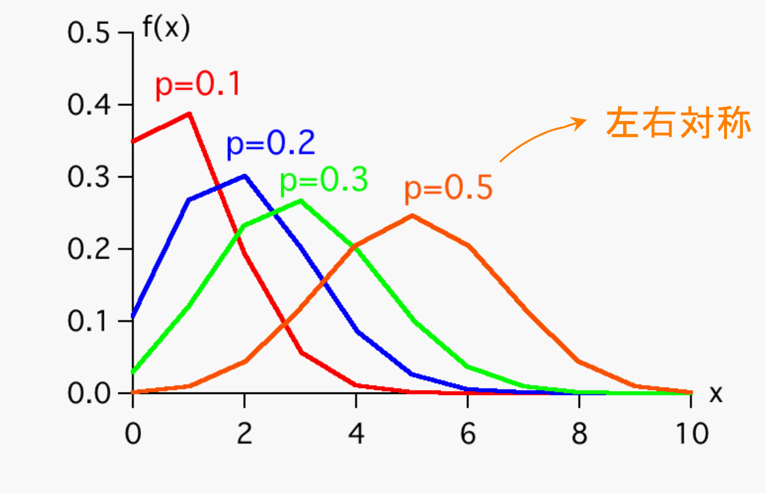
3—1　2項分布

- [例] 5択の問題が10題あり、配点は各問10点である。全くでたらめに答えたとき、80点以上とれる確率は？　→ 参考書[1] p.69

$Bin(10,1/5)$ ⇨

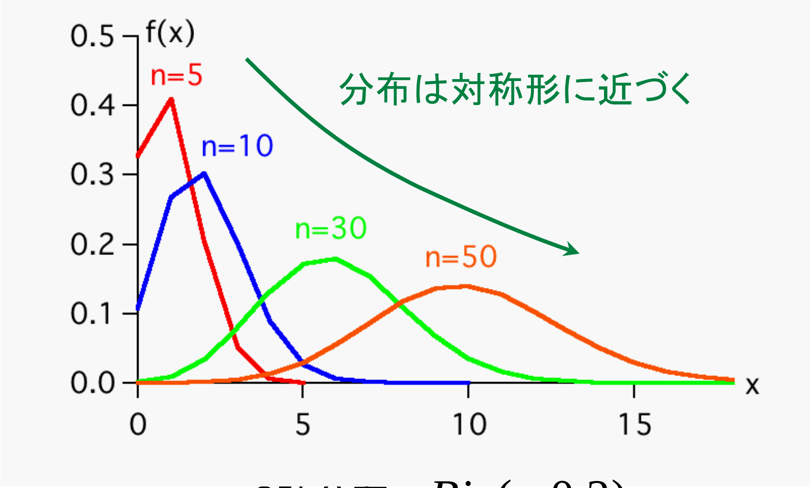
$$f(8)+f(9)+f(10)={}_{10}C_8\left(\frac{1}{5}\right)^8\left(\frac{4}{5}\right)^2+{}_{10}C_9\left(\frac{1}{5}\right)^9\left(\frac{4}{5}\right)^1+{}_{10}C_{10}\left(\frac{1}{5}\right)^{10}\left(\frac{4}{5}\right)^0$$

$$=\frac{761}{9765625}=0.000078$$



2項分布　$Bin(10,p)$

各問の正解率



2項分布　$Bin(n,0.2)$

3—1　2項分布

- 2項分布の性質

2項定理と関係

$$f(x) = {}_nC_x p^x (1-p)^{n-x} \xrightarrow{q=1-p} f(x) = {}_nC_x p^x q^{n-x}$$

$$\Longrightarrow (p+q)^n = \sum_{x=0}^{n} {}_nC_x p^x q^{n-x}$$ の2項展開式の各項

$$\sum_{x=0}^{n} f(x) = 1$$

$$(p+q)^n = \sum_{x=0}^{n} {}_nC_x p^x q^{n-x} \xrightarrow{p\text{で微分}} n(p+q)^{n-1} = \sum_{x=0}^{n} x\, {}_nC_x p^{x-1} q^{n-x} \xrightarrow{p\text{をかける}}$$

期待値 $\mu_x = np$

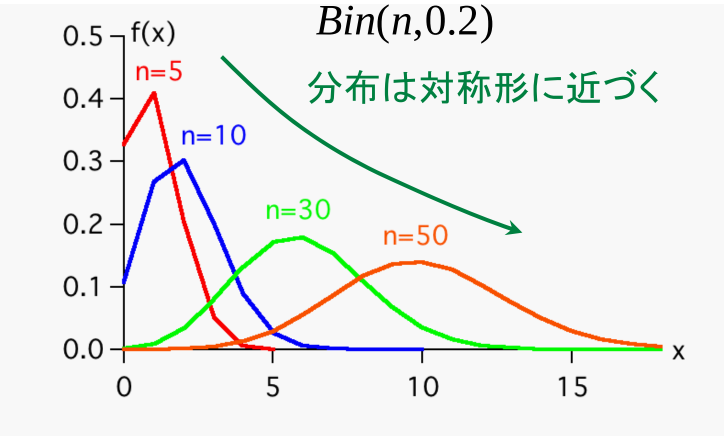
pで微分

$$n(n-1)(p+q)^{n-2} = \sum_{x=0}^{n} x(x-1)\, {}_nC_x p^{x-2} q^{n-x} \xrightarrow{p^2\text{をかける}} n(n-1)p^2 = \sum_{x=0}^{n} (x^2 - x) f(x)$$

$$n(n-1)p^2 + \mu_x = \sum_{x=0}^{n} x^2 f(x)$$

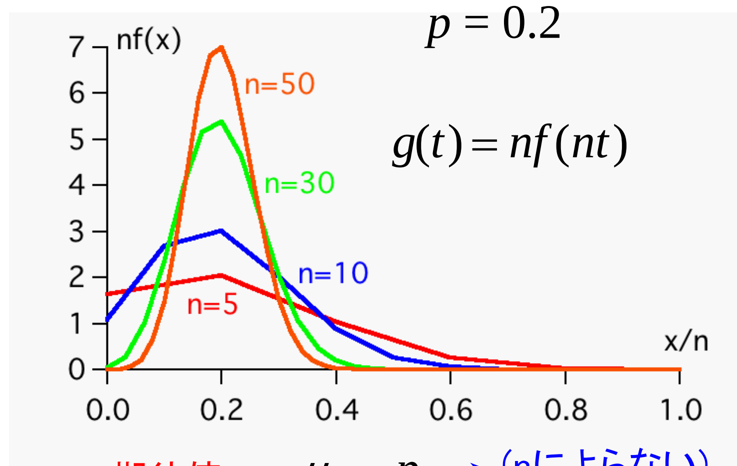
分散 $\sigma_x^2 = n(n-1)p^2 + np - n^2p^2$
$= np(1-p)$

- 大数の法則



$$T = \frac{X}{n}$$

横軸を1/n倍
縦軸をn倍



期待値 $\mu_x = np$

分散 $\sigma_x^2 = np(1-p)$

期待値 $\mu_t = p \longrightarrow$ (nによらない)

分散 $\sigma_t^2 = \dfrac{p(1-p)}{n} \xrightarrow[n\to\infty]{} 0$

- 大数の法則 経験的確率を数学的に扱う大切な根拠！ → 参考書[1] p.74
  - 1回1回の試行で、ある事象*A*が起こるかどうかは確率的にしか分からないが、試行回数を増やせば増やすほど、その事象の起こる割合は一定の値*p*に近づく。

東京大学工学部 システム創成学科Cコース Program for Biological & Information Systems

3−2　ポアソン分布

→ 参考書[1] p.75

２項分布　　$f(x) = {}_nC_x p^x (1-p)^{n-x}$

平均 $\mu = np$ を一定値に保ったまま $n \to \infty, p \to 0$ の極限をとる。

$$f(x) = \frac{n(n-1)\mathcal{L}\,(n-(x-1))}{x!}\left(\frac{\mu}{n}\right)^x \left(1-\frac{\mu}{n}\right)^{n-x}$$

$$= \frac{n^x}{x!}\cdot 1 \cdot \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\mathcal{L}\left(1-\frac{x-1}{n}\right)\left(\frac{\mu}{n}\right)^x \left(1-\frac{\mu}{n}\right)^{n-x}$$

$$= \frac{\mu^x}{x!}\cdot 1 \cdot \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\mathcal{L}\left(1-\frac{x-1}{n}\right)\left(1-\frac{\mu}{n}\right)^{n} \left(1-\frac{\mu}{n}\right)^{-x}$$

$$= \frac{\mu^x}{x!}\cdot 1 \cdot \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\mathcal{L}\left(1-\frac{x-1}{n}\right)\left[\left(1-\frac{\mu}{n}\right)^{-n/\mu}\right]^{-\mu} \left(1-\frac{\mu}{n}\right)^{-x}$$

$$\xrightarrow[n\to\infty]{} \frac{\mu^x}{x!}e^{-\mu}$$

ポアソン分布　　$f(x) = \dfrac{\mu^x}{x!}e^{-\mu}$

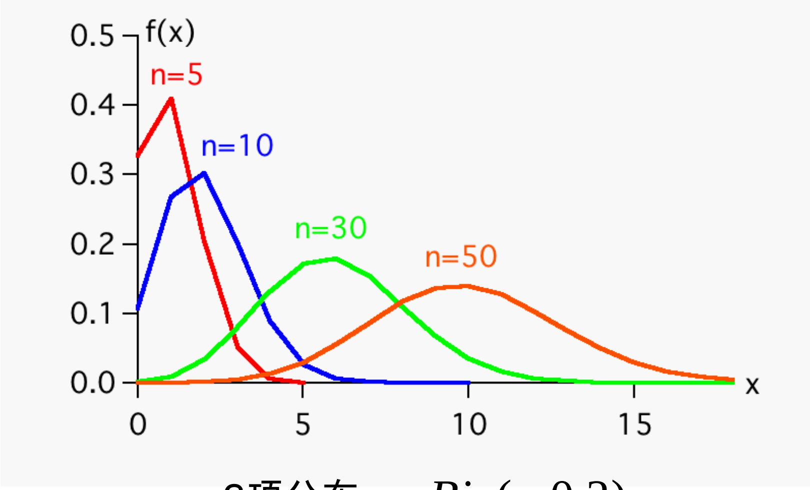
3−2 ポアソン分布

ポアソン分布 $P(\mu)$ $$f(x) = \frac{\mu^x}{x!} e^{-\mu}$$

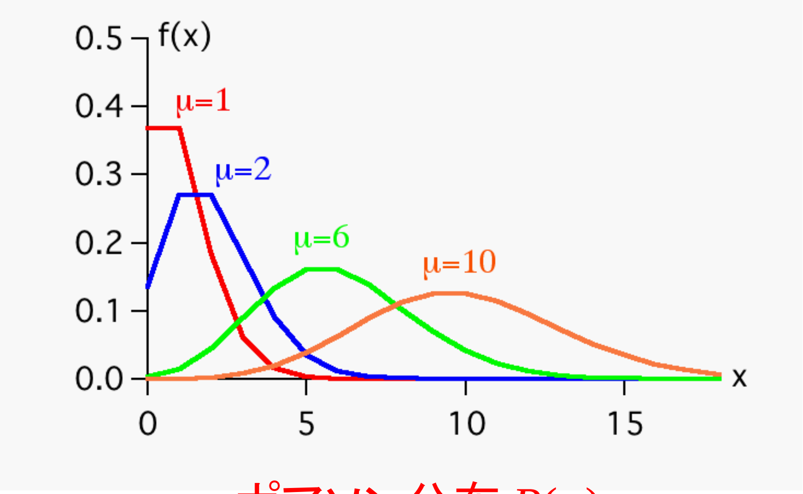
- 起こる確率の小さい事象（p が小さい）
- 多数回独立試行（n が大きい）

分散 $$\sigma^2 = np(1-p) = \mu\left(1 - \frac{\mu}{n}\right) \xrightarrow[n\to\infty]{} \mu$$

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



２項分布 $Bin(n, 0.2)$



ポアソン分布 $P(\mu)$

3−2 ポアソン分布

ポアソン分布 $P(\mu)$ $$f(x)=\frac{\mu^x}{x!}e^{-\mu}$$

[例] プロイセンにおいて、1875年から1894年までの20年間に、馬に蹴られて死亡した兵士の数（出典：参考書[1] p.76）

めったにないこと

死亡者数	0	1	2	3	4	計
部隊数	109	65	22	3	1	200

$$\mu=(0\times 109+1\times 65+2\times 22+3\times 3+4\times 1)/200=0.61$$

ポアソン分布 $P(0.61)$ の場合の理論値を表にすると…

死亡者数	0	1	2	3	4
部隊数	108.7	66.3	20.2	4.1	0.6

3−2 ポアソン分布

ポアソン分布 $P(\mu)$　　　$f(x)=\frac{\mu^x}{x!}e^{-\mu}$

[例] ある2年間のプロ野球セ・パ両リーグの公式戦1560試合について、各試合で何回逆転が起こったか。（出典：参考書[1] p.77）

「めったにないこと」

逆転回数	0	1	2	3	4以上	計
頻度	944	457	128	25	6	1560

平均　$\mu=0.52$

ポアソン分布 $P(0.52)$ の場合の理論値を表にすると…

逆転回数	0	1	2	3	4以上
頻度	927.0	482.5	125.6	21.8	3.1

3−2 ポアソン分布

ポアソン分布 $P(\mu)$ $f(x)=\dfrac{\mu^x}{x!}e^{-\mu}$

- 起こる確率の小さい事象(p が小さい)
- 多数回独立試行(n が大きい)

非常に多数の人や物の中で、あまり起こらない事柄

- 放射性元素の1分間の崩壊数(放射線のカウント)
- 1日の交通事故件数
- 1年の飛行機墜落事故件数
- 1か月の有感地震の回数

[例] あるラーメン屋には10分間に平均4人の割合でお客さんがやってくる。このラーメン屋へ、10分間に6人以上お客さんの来る確率を求めよ。

何人かで連れ立って来る人はいないと仮定

お客さんの数Xはポアソン分布$P(4)$に従うと考える。

$$1-\sum_{x=0}^{5} f(x)=1-\left(\frac{4^0}{0!}+\frac{4^1}{1!}+\frac{4^2}{2!}+\frac{4^3}{3!}+\frac{4^4}{4!}+\frac{4^5}{5!}\right)e^{-4}=1-\frac{643}{15}e^{-4}=0.21$$

東京大学工学部 システム創成学科 C コース Program for Biological & Information Systems

3−2 ポアソン分布

ポアソン分布に従う事象の間隔の分布

[例] ある放射線元素は1分間に平均1回の割合で崩壊する。（1分間に平均1回の割合で放射線がカウントされる）このとき、2つの連続する崩壊（カウント）の間隔の分布g(t)はどうなるか。

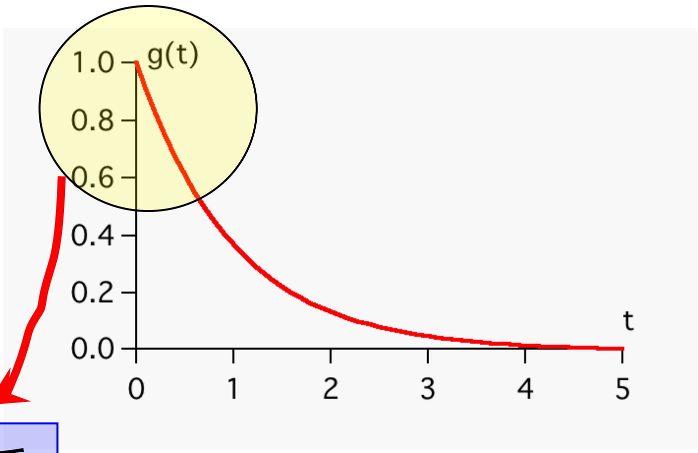
カウントがあってから、t 分間カウントのない確率$p(t)$は、

$$p(t)=e^{-t}$$

一方

$$\int_t^{\infty} g(t')dt' = p(t)$$

$$\boxed{g(t)=-p'(t)=e^{-t}}$$

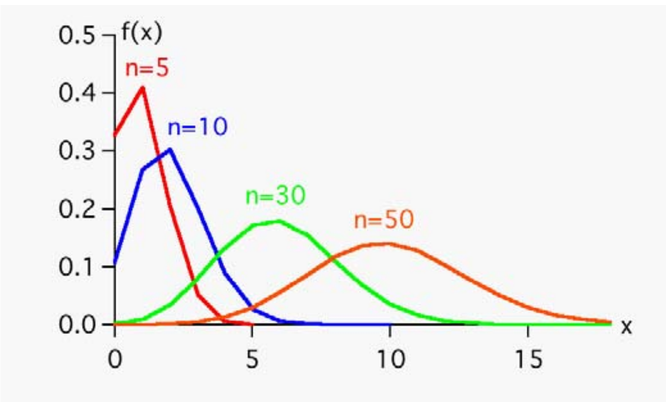


相互に独立に起こる事象は、（直観に反して）立て続けに起こりやすい。

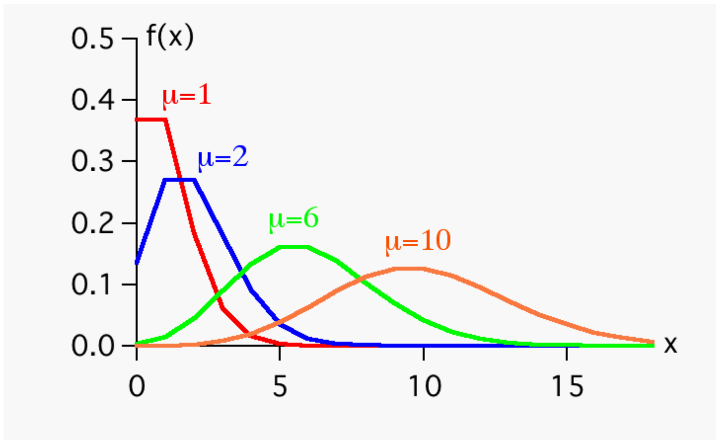
3−3　正規分布

→ 参考書[1] p.82

- n が大きい極限で2項分布はどんな分布になるか？
- μ が大きい極限でポアソン分布はどんな分布になるか？



2項分布　$Bin(n,0.2)$



ポアソン分布 $P(\mu)$

正規分布

$Z = \dfrac{X-\mu}{\sigma}$ とおくと、Zは

平均0、分散1

$$\boxed{g(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}}$$

に従う。

標準正規分布
$N(0,1)$

3−3　正規分布

標準正規分布　$N(0,1)$

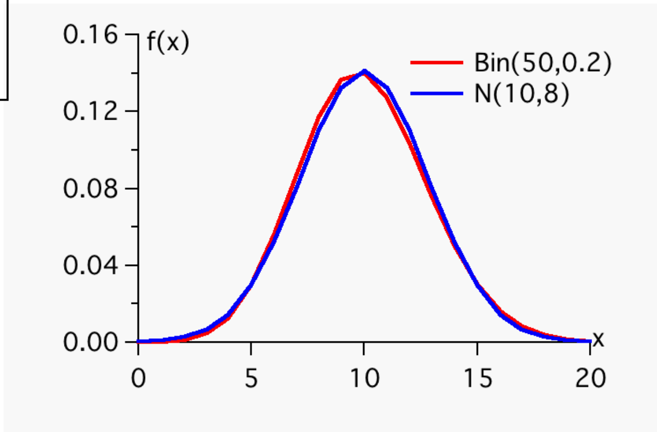
$$g(z)=\frac{1}{\sqrt{2\pi}}e^{-z^2/2}$$

変数変換
（標準化変換）

正規分布　$N(\mu,\sigma^2)$

$$h(y)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$$

ガウス分布とも呼ぶ



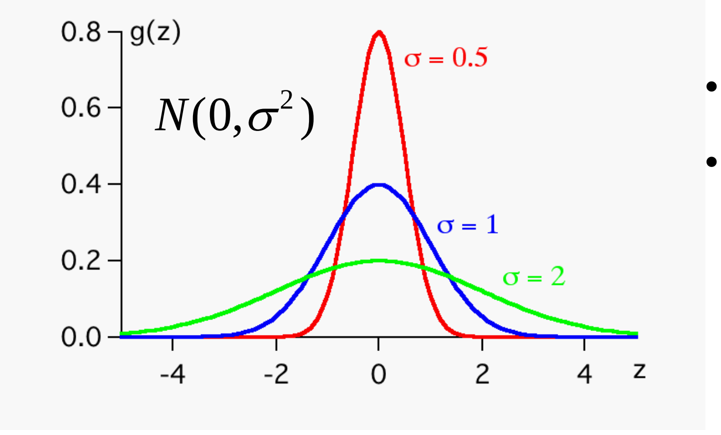
2項分布から正規分布への移行

→　中心極限定理の一例

- いろいろな分布が近似的に正規分布に従っている。
  - 身長、体重、試験の点数
  - 実験の誤差など、理工学の広い分野で現れる。
- 2項分布　→　いろいろな確率分布の出発点
  正規分布　→　実用上もっとも重要

3−3　正規分布

正規分布の性質



- 左右対称
- 標準偏差σが大きいほど、なだらか

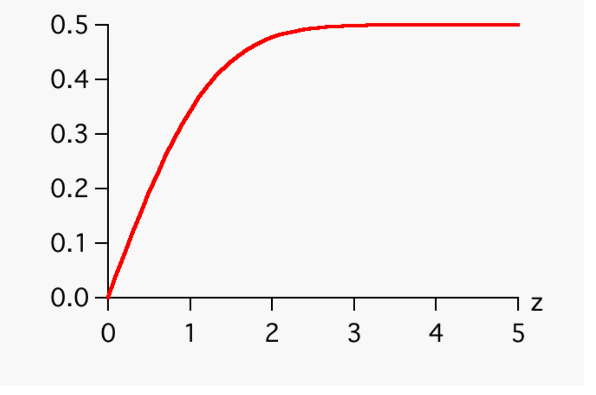
標準正規分布$N(0,1)$に従う確率変数Zが$z_1 < Z < z_2$の間にある確率

$$P(z_1 < Z < z_2) = \int_{z_1}^{z_2} g(z)dz = \int_{z_1}^{z_2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

誤差関数(error function)　erf(z)

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt = 2\int_0^{\sqrt{2}z} g(t)dt$$

$$\Rightarrow \quad \int_0^z g(t)dt = \frac{\mathrm{erf}(z/\sqrt{2})}{2}$$

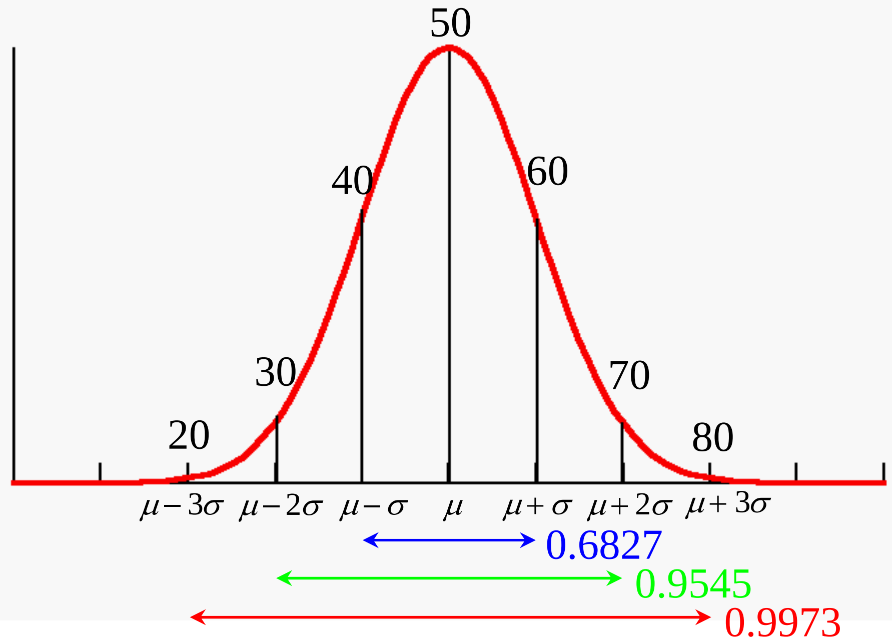


3−3 正規分布

正規分布$N(\mu,\sigma^2)$について実用上よく使われる確率

確率変数Yが正規分布$N(\mu,\sigma^2)$に従う $\Longrightarrow$ $Z=\dfrac{Y-\mu}{\sigma}$ は正規分布$N(1,0)$に従う

$$P(\mu-\sigma<Y<\mu+\sigma)=P(-1<Z<1)=\mathrm{erf}(1/\sqrt{2})=0.6827$$
$$P(\mu-2\sigma<Y<\mu+2\sigma)=P(-2<Z<2)=\mathrm{erf}(2/\sqrt{2})=0.9545$$
$$P(\mu-3\sigma<Y<\mu+3\sigma)=P(-3<Z<3)=\mathrm{erf}(3/\sqrt{2})=0.9973$$



[例] ある試験の平均点は60点、標準偏差は10点であった。この試験の点が正規分布に従っていると仮定すると、

- 80点以上の人は、2.3%
- 50点以下の人は、16%
- 40点から80点の人は、95%

3−4 中心極限定理

→ 参考書[1] p.86

確率変数X_1, X_2, …, X_nがたがいに独立で、平均μ, 分散σ^2をもつ同一の分布に従っているとする。X_1, X_2, …, X_nの単純平均

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

に対して、

$$Z_n = \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu)$$

とすると、nを大きくしたとき、Z_nの分布は標準正規分布に$N(0,1)$に近づく。

→ 参考書[1] p.86

$$\bar{X} \approx N(\mu, \sigma/\sqrt{n})$$

3−4　中心極限定理

[例] コインを1000回投げたとき、表の出る回数が485回以上515回以下である確率は？

表の出る回数Xは、２項分布$Bin(1000,1/2)$に従う。

$$f(x) = {}_{1000}C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{1000-x} = \frac{1000!}{x!(1000-x)!}\left(\frac{1}{2}\right)^{1000}$$

$$P(485 \le X \le 515) = \sum_{x=485}^{515} f(x) = 0.673$$

表の出る回数Xを、正規分布で近似

$$\mu = 1000 \times \frac{1}{2} = 500,\ \sigma = \sqrt{1000 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{250} = 15.8$$

$$P(485 \le X \le 515) \approx P(\mu - \sigma < X < \mu + \sigma) = 0.683$$