2.正規分布（normal distribution）

（学習内容）

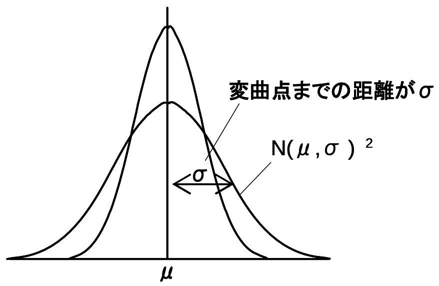
・正規分布とは何か

・正規分布の性質を学ぶ

2.1 確率変数と正規分布

代表的な確率密度関数の一つ

（天体の位置計測の誤差からガウスにより導出：ガウス分布）



(1) 正規分布のグラフ

正規分布の確率密度関数は次式で表される

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

μ、σ^2のパラメータで完全に特徴付けられる。N(μ, σ^2)で表す

計量値として得られるデータの母集団分布の多くは正規分布とみなせる

３シグマのルール：μから3σ以上離れたデータが得られる確率はほとんど0（0.3%以下）

確率変数xがN(μ, σ^2)に従うとき

E(x)=μ

V(x)=σ^2

N(μ, σ^2)は平均μ、分散σ^2の確率分布をとる

(2) 標準正規分布

μ=0、σ=1の正規分布N(0, 1^2)を標準正規分布（standard normal distribution）という

確率変数xがN(μ, σ^2)に従うとき

u=(x−μ)/σ　はN(0, 1^2)に従う　→　標準化

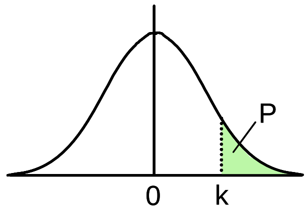
（w=x−μ：中心がμから0へ、u=w/σ：変曲点がσから1へ）

(3) 正規分布表

正規分布表：標準正規分布においてk（uの値）からP（右側の確率）を求めるのに利用

$$P = \Pr\{u \geq k\} = \int_k^{\infty}\frac{1}{\sqrt{2\pi}}\exp\{-\frac{u^2}{2}\}du$$

例）Pr{u≧1.96}　=0.0250



一般の正規分布 N(μ, σ^2)についての確率を求めるためには標準化の変形を行い、u=(x−μ)/σ

が N(0,1^2) に従うことより求める。

Pr{x≧a}=Pr{(x−μ)/σ≧(a−μ)/σ}=Pr{u≧(a−μ)/σ}

例） x が N(2,4^2) に従うとき

Pr{x≧3.8}=Pr{(x−2.0)/4.0≧(3.8−2.0)/4.0}=Pr{u≧0.45}=0.3264

(4) 正規分布の性質

正規分布に従う確率変数について、次の幾つかの性質が成り立つ

a) 確率変数 x が正規分布 N(μ, σ^2) に従い、a、b が定数のとき ax+b は N(aμ+b, $a^2\sigma^2$) に従う

証明）

$$E(ax+b) = \int_{-\infty}^{\infty}(ax+b)f(x)dx = a\int_{-\infty}^{\infty}xf(x)dx + b\int_{-\infty}^{\infty}f(x)dx$$
$$= aE(x) + b = a\mu + b$$

$$V(ax+b) = E[((ax+b)-E(ax+b))^2] = E[(ax+b-aE(x)-b)^2] = E[(a^2(x-E(x))^2]$$
$$= a^2E[(x-E(x))^2] = a^2V(x) = a^2\sigma^2$$

b) ２つの確率変数 x と y がそれぞれ独立に正規分布 N(μ_1, σ_1^2)、N(μ_2, σ_2^2) に従うとき

x+y は N($\mu_1+\mu_2$, $\sigma_1^2+\sigma_2^2$) に従う

証明）

$$E(x+y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x)f(y)dxdy = \int_{-\infty}^{\infty}(f(y)\int_{-\infty}^{\infty}xf(x)dx + yf(y)\int_{-\infty}^{\infty}f(x)dx)dy$$
$$= \int_{-\infty}^{\infty}(f(y)E(x) + yf(y))dy = E(x) + E(y) = \mu_1 + \mu_2$$

$$V(x+y) = E[((x+y)-E(x+y))^2] = E[(x-E(x)+y-E(y))^2]$$
$$= E[(x-E(x))^2 + (y-E(y))^2 + 2(x-E(x))(y-E(y))] = V(x) + V(y) + 2E[(x-E(x))(y-E(y))]$$
$$= {\sigma_1}^2 + {\sigma_2}^2 + 2E[xy - xE(y) - yE(x) + E(x)E(y)]$$
$$= {\sigma_1}^2 + {\sigma_2}^2 + 2(E(xy) - E(y)E(x) - E(x)E(y) + E(x)E(y))$$
$$= {\sigma_1}^2 + {\sigma_2}^2$$

2.2 統計量の分布

データは確率変数の実現値であるため、データから計算される統計量も何らかの確率分布に従う

(1) $\bar{x}$ の分布

確率変数 $x_1, x_2, \cdots, x_n$ が互いに独立に N(μ, σ^2) に従うとき、$\bar{x}$ は N(μ, σ^2/n) に従う

証明）

$$E(\bar{x}) = E(\frac{x_1}{n} + \frac{x_2}{n} + \cdots \frac{x_n}{n}) = E(\frac{x_1}{n}) + E(\frac{x_2}{n}) + \cdots + E(\frac{x_n}{n}) = \frac{\mu}{n} + \frac{\mu}{n} + \cdots + \frac{\mu}{n} = \mu$$

$$V(\bar{x}) = V(\frac{x_1}{n} + \frac{x_2}{n} + \cdots \frac{x_n}{n}) = V(\frac{x_1}{n}) + V(\frac{x_2}{n}) + \cdots + V(\frac{x_n}{n}) = \frac{\sigma^2}{n^2} + \frac{\sigma^2}{n^2} + \cdots + \frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

(2) $\bar{x}$ の分布の標準化

$$u = \frac{\bar{x} - \mu}{\sqrt{\sigma^2 / n}}$$ は N(0, 1^2) に従う

○ 偏差値

$$z = 10 \times \frac{x - \mu}{\sigma} + 50$$ の変換を行った値

テストの点数は正規分布 N(μ, σ^2) に従う

($x-\mu$)/σ は標準正規分布 N(0, 1^2) に従う

10($x-\mu$)/σ+50 は N(50, 10^2) に従う→平均 50、標準偏差 10 となるように変換