第7回：正規分布（1回目）

- **正規分布** (Normal Distribution)、**ガウス分布**(Gaussian distribution)、**誤差関数**とも呼ばれる。
- **確率変数x は実数値をとる、連続分布**
- **2つのパラメータ、平均μ、分散σ^2 で特徴付けられる。N(μ, σ^2)**

正規分布の確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in R$$

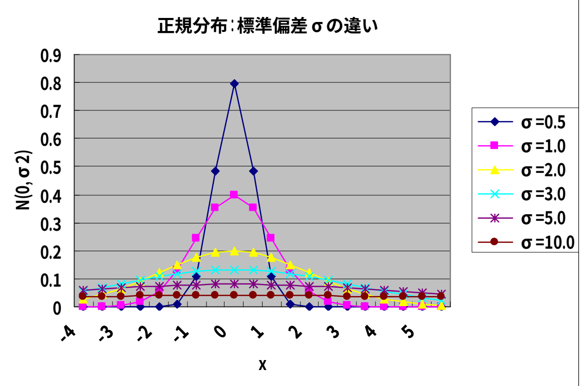
ただし、μ, σは実数。

正規分布は、2つのパラメータμ, σで表される。

N(μ, σ^2)。式より、直ちにわかること；

- x=μに関して、左右対称（偶関数）。
- また、この平均値x=μで最大値をとる。
- x=μから左右にσだけ外れたところでは、最大値のexp(-1/2)=0.606530…倍

正規分布のグラフ



標準正規分布 $N(0,1)$への標準化

一般に、正規分布N(μ, σ^2)に従う確率変数x を、以下のようにuに変数変換すると、確率変数u は μ=0.0, σ=1.0の標準正規分布$N(0,1)$に従う。

$$x \to u, \quad u = \frac{x-\mu}{\sigma}$$

$$E(u) = E\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma}\left(E(x)-\mu\right) = 0,$$

$$E(u^2) = E\left(\left(\frac{x-\mu}{\sigma}\right)^2\right) = \frac{1}{\sigma^2}\cdot E\left((x-\mu)^2\right) = \frac{\sigma^2}{\sigma^2} = 1$$

この変数変換により$N(0,1)$になおすことを、「正規分布の標準化、正規化」という。

μだけ左に平行移動し（それにより、$x=0$が分布の中央に）、σ分の1に縮小する（それにより、分散で表される拡がりが1に）

標準正規分布に変換して計算

$$\mathrm{Prob}\{a \le x \le b\} = \mathrm{Prob}\left\{\frac{a-\mu}{\sigma} \le \frac{x-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right\}$$

$$= \mathrm{Prob}\left\{\frac{a-\mu}{\sigma} \le u \le \frac{b-\mu}{\sigma}\right\} = \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} N(0,1)(u)du$$

u は、標準正規分布$N(0,1)$に従うので、$N(0,1)$について上記の確率を求めることができれば、任意の正規分布N(μ, σ^2)に従うxについて、確率を求めることができる。

そこで、特に$N(0,1)$についてのみ、確率値を与える確率密度関数$N(0,1)$の積分値を、**正規分布表**として用意しておき、数値計算等の手間を省く。

正規分布表

$$\mathrm{Prob}\{u \ge K_p\} = \int_{K_p}^{\infty} N(0,1)(u)du = \int_{K_p}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = P$$

を与える、$K_p \Leftrightarrow P$ の対応表：**正規分布表**

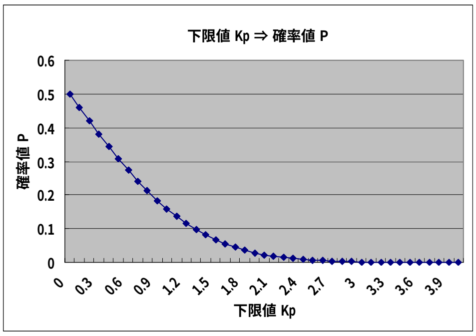
積分値Pが、確率値を与える。

$K_p > 0$ の場合について求めておけば、十分。

⇒ $K_p < 0$については、どうすれば良いか？

Answer: 分布の対称性を利用せよ。

K_pとPの対応：グラフでみると…



正規分布表を用いた確率計算の例題 ：製品の品質管理

- テキスト61ページの例題5．1：「品質管理」

化粧品のボトル。工場でのボトル詰め工程により、実際にボトル詰めされた製品の容量にはバラツキが生じてしまう。いま製品を多数抜き取って調べたところ、平均125ml, 標準偏差2.0mlの正規分布に従っていることがわかった。ボトルは、「容量120ml」と表示して出荷しているので、120ml以下で規格外である製品は、極力出してはいけないが、現状では何%生じているか？

また、0.1%以下にするには、標準偏差の値をどこまで小さくすべきか？（平均を上げると損になるので、詰め作業の精度を上げることで標準偏差を小さくしたい）

確率分布が満たすべき2条件の確認

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x\in R$$

- $f(x)\geq 0$
- $\int_{-\infty}^{+\infty} f(x)dx=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}}du \quad \left(x\to u=\frac{x-\mu}{\sigma}, du=\frac{1}{\sigma}\cdot dx, u:-\infty\to+\infty\right)$$

$$\boxed{=1}$$

最後の等号は、もし次式が成り立っているならば(a=1)⇒ $\int_{-\infty}^{+\infty} e^{-\frac{a}{2}u^2}du=\sqrt{\frac{2\pi}{a}}, \quad \text{for } a>0$

$\int_{-\infty}^{+\infty} e^{-\frac{a}{2}u^2}du=\sqrt{\frac{2\pi}{a}}$ の証明

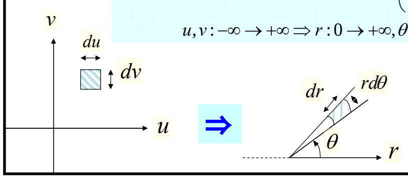
左辺$=I$ とおくと、

$$I^2=\left(\int_{-\infty}^{+\infty} e^{-\frac{a}{2}u^2}du\right)\cdot\left(\int_{-\infty}^{+\infty} e^{-\frac{a}{2}v^2}dv\right)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{a}{2}(u^2+v^2)}dudv$$

ここで、変数変換：直交座標系(u,v)から極座標系(r,θ)へ

$(u,v)\Rightarrow(r,\theta)$: $u=r\cos\theta, v=r\sin\theta$, $\left(r^2=u^2+v^2, \theta=\tan^{-1}\frac{v}{u}\right)$

$u,v:-\infty\to+\infty\Rightarrow r:0\to+\infty, \theta:0\to 2\pi$



$$du\cdot dv=rd\theta\cdot dr$$

$\int_{-\infty}^{+\infty} e^{-\frac{a}{2}u^2}du=\sqrt{\frac{2\pi}{a}}$ の証明（続き）

$$du\cdot dv=rd\theta\cdot dr$$

$$I^2=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{a}{2}(u^2+v^2)}dudv=\int_0^{\infty} re^{-\frac{a}{2}r^2}dr\int_0^{2\pi}d\theta$$

$$u,v:-\infty\to+\Rightarrow r:0\to\infty, \theta:0\to 2\pi$$

$$=2\pi\left[-\frac{1}{a}\cdot e^{-\frac{a}{2}r^2}\right]_0^{\infty}\boxed{=\frac{2\pi}{a}}$$

証明終わり

正規分布の平均

$$平均=\int_{-\infty}^{+\infty} x\cdot f(x)dx$$

パラメータμは平均値

$$=\int_{-\infty}^{+\infty}\{(x-\mu)+\mu\}\cdot\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

$$=\int_{-\infty}^{+\infty} x'\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x'^2}{2\sigma^2}}dx'+\mu\cdot\int_{-\infty}^{+\infty} f(x)dx$$

$$=0+\mu$$

$$\boxed{=\mu}$$

正規分布の分散

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

パラメータ σ² は分散

$$= \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= \sigma^2 \cdot \int_{-\infty}^{+\infty} u^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \quad \left(\text{by } x \Rightarrow u = \frac{x-\mu}{\sigma}, du = \frac{dx}{\sigma}, u : -\infty \to +\infty \right)$$

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

最後の等号は、もし次式が成り立っているならば ⇒ $\int_{-\infty}^{+\infty} u^2 e^{-\frac{u^2}{2}} du = \sqrt{2\pi}$

$\int_{-\infty}^{+\infty} u^2 e^{-\frac{u^2}{2}} du = \sqrt{2\pi}$ の証明（2つの方法）

証明方法1：

先に証明した $\int_{-\infty}^{+\infty} e^{-\frac{a}{2}u^2} du = \sqrt{\frac{2\pi}{a}}$ の両辺を a で微分すると、

$$\int_{-\infty}^{+\infty} \left(-\frac{u^2}{2} \right) \times e^{-\frac{a}{2}u^2} du = \left(-\frac{1}{2} \right) \frac{1}{a} \times \sqrt{\frac{2\pi}{a}}$$

これに $a=1$ を代入して得られる。

証明方法2：

そのまま、部分積分

$$\int_{-\infty}^{+\infty} \underbrace{u}_{\downarrow} \overbrace{\,^2 e^{-\frac{u^2}{2}}}^{\uparrow} du = \left[u \cdot (-1) \cdot e^{-\frac{u^2}{2}} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du = 0 + \sqrt{2\pi} = \sqrt{2\pi}$$

中心極限定理

とはどんな定理か…

● **中心極限定理** (central limit theorem)

n個の確率変数$x_1, x_2, \ldots, x_n$ が、
互いに独立に同じ確率分布に従うとする。
その分布の平均をμ、分散をσ²とする；

$$E(x_i) = \mu, \quad E\left((x_i - \mu)^2\right) = \sigma^2 \quad (i = 1, 2, \cdots, n)$$

このn個の変数の平均 $\boxed{\bar{x} = \frac{1}{n} \cdot (x_1 + x_2 + \cdots + x_n)}$
（これも確率変数）は、nが大きいときには、
正規分布N(μ，σ²/n)に従う。

中心極限定理

- **n 個の確率変数の平均（'中心'）は、**
- **n が十分大きな極限においては、**
- **（もとの各x_iが従う確率分布がどうであれ）**

正規分布N(μ，σ²/n)に従う

…という定理。

平均はμのまま（分布の中心は同じところ）、
分散は$1/n$で小さくなってゆく（バラツキは小さくなる）。

同じ対象を、（念のために）何度も繰り返し測定して、その平均をとることにする。すると、より多くの測定値を平均するほうが、値のバラツキが小さくなり、精度が上がる…

定理の証明は、行わない
（教科書では特性関数を用いて証明）

- **正規分布が、多くの現象に見られる理由**
- **「誤差」（例えば、測定時に生じる測定誤差）：**

一般に、各回で実際に得られる値は、複雑な原因が複数（測定値を左右するn個の要因）積み重なった結果として与えられる。

⇒ 中心極限定理が適用できる

正規分布は、ガウス分布、誤差分布とも呼ばれる

第8回後半：今までの復習課題

問1．平均と分散（答えは既約分数で）

(a) **以下の確率分布 $p(x)$の平均と分散を求めよ。**

$$p(x) = \begin{cases} \frac{1}{5} & \text{for } x = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1 \\ 0 & \text{otherwise} \end{cases}$$

(b) **以下の確率分布$f(x)$の平均と分散を求めよ。**

$$f(x) = \begin{cases} 2x & \text{for } 0 < x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

問2．指数分布の平均と分散

指数分布 $f(t)=\begin{cases}\lambda e^{-\lambda t} & (t\geq 0)\\ 0 & (t<0)\end{cases}$ に対して、

$\mu=\int_{-\infty}^{+\infty}x\cdot f(x)dx$ と $\sigma^2=\int_{-\infty}^{+\infty}(x-\mu)^2\cdot f(x)dx$ を求めなさい。

問3．シュート成功率

シュート成功率が2/3のサッカー選手がいる。この選手について、以下を求めなさい。
（答えは、既約分数で）

(a) 4回シュートして、3回以上成功する確率。

(b) 9回シュートするときの、平均と分散。

p　5割の二項分布の場合は、回数N が大きいときには、正規分布で近似し、計算を簡略化できる。

問4．ランダムな寿命

惑星Xでは、ランダムに降ってくる隕石が頭にあたることで寿命が尽き、X星人の寿命は、平均８０歳の指数分布に従っている。あるX星人が、６０歳以上生きる確率は何%だろう。

必要に応じて以下の式や値を用いなさい。答えは、少数以下1桁まで。

$\exp(-\frac{1}{2})=0.607,\quad \exp(-\frac{3}{4})=0.472,\quad \exp(-1)=0.368$

問5．クッキーのチョコチップ

チョコチップがたくさん入っていることで人気のクッキーを買った。クッキーの製造工程により、クッキー1枚あたりに入っているチョコチップ数は、平均１０粒、標準偏差２粒の正規分布に従っていると店のホームページに書いてある。

(a)　今食べたクッキーには、１３粒もあったが、これ（13粒以上）は何%の幸運だろう。

(b) 逆に、5粒以下しか入っていないクッキーにあたる確率は何%か。

テキストp.149の正規分布表を使い、答えは少数以下1桁まで。

次々回6月１２日は中間試験

- 教室：コラーニング2階　Ｃ２０２
- 授業開始の定刻に教室に着席
- 試験時間は７０分
（途中退室なし。遅刻は試験開始後２０分まで）
- 着席は、学籍番号順に指定
- 学生証、筆記用具を持参
- 持ち込み許可物件はなし
（計算に必要な数値等は、配布用紙に記載）