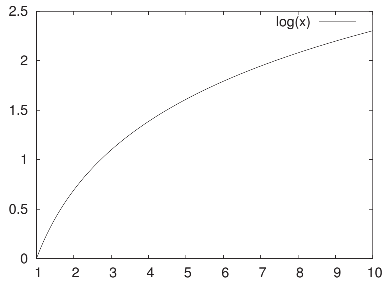
第 7 章 正規分布

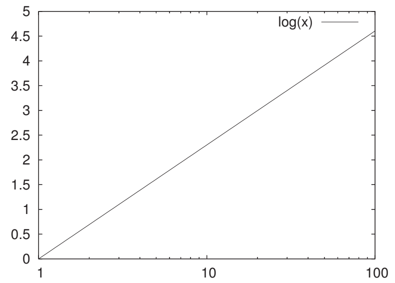
浅川 伸一

2006 年 05 月 24 日

6.5 若干の数学概念

- 置換積分: $x = g(t)$ とおくと
$$\int f(x)dx = \int f(g(t))\frac{dx}{dt}dt = \int f(g(t))g'(t)dt \tag{1}$$
- 部分積分:
$$\int f(x)g'(x)dx = f(x)g(x) - \int f'(x)g(t)dt \tag{2}$$
- 指数関数、対数関数: 成長曲線、自信の規模を表すマグニチュード。フェヒナーの法則： 感覚量は刺激強度の対数に比例する
$$E = k \log I \tag{3}$$
ここで E は心理量 (感覚量)、k は定数、I は刺激強度である。

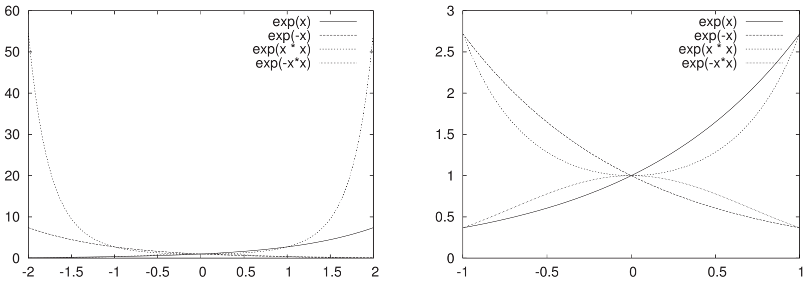




- 二重積分 (多重積分): 一つ一つの積分をするだけ
- 極座標: 位置の別表現
- 関数行列式 (ヤコビアン): 杏仁豆腐の体積、多重積分と置換積分の組合せ
- 偏微分: 自分だけを考慮して、他を顧みないわがままな微分
- 行列式二本のベクトルで定義される平行四辺形の面積。

7　正規分布 normal distribution

以下の図は $\exp(x)$, $\exp(-x)$, $\exp(x^2)$, $\exp(-x^2)$ を描いたものである。



図からわかることは e^{-x^2} が釣鐘形の曲線になっていることである。この曲線を変形することによって正規分布が導出される。

平均 μ 分散 σ^2 に従う正規分布の確率密度関数 $N(\mu,\sigma^2)$ は次のとおり

$$f(x) = N(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{4}$$

に書くことができる。あるいは e の肩の上が小さい文字で見にくいとき、

$$f(x) = N(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] \tag{5}$$

などと表現されることもある。

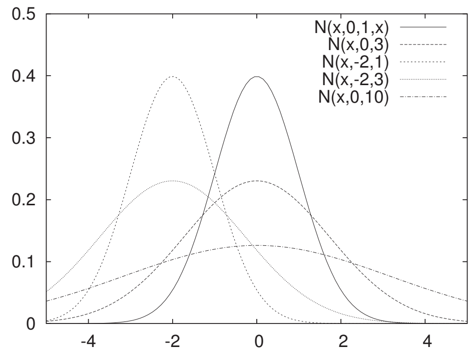
正規分布は統計学において中心的な役割を果たす重要な分布である。正規分布のことをガウス分布、あるいは誤差分布と呼ぶこともある。これはある変数（心理学的なデータであると考えても良い）の観測値が、真の値に誤差 $\mu + \epsilon$ を加えたものと見なして統計的な検討がなされることが多いからである。この場合 ϵ は平均 0, 分散 σ^2 に従うものとして扱われる。データが離散型の確率でさえ誤差が正規分布に従うことが仮定される場合も多い。このようなデータの扱い方は、心理学のデータ解析でしばしば登場する。

正規分布に「従う」とは、正規分布から導き出される、あるいは、確率変数がが正規分布を仮定したとき、などの意味合いである。

μ, σ の値を変化さてたときのグラフを示す。

図より、平均の違いは釣鐘形カーブの頂上の位置として表され、分散 (あるいは標準偏差) の値は釣鐘形カーブの裾野の広さとして表される。どのカーブも確率変数であるから定義域 $(-\infty,\infty)$ 全域で積分すると 1 になることは変わらない。

上式が確率密度関数であるためには $\displaystyle\int f(x)dx = 1$ が成り立っていなければならない。まず $\mu = 0$, $\sigma^2 = 1$ のような単純な形で解を求めることになる。



この計算を確認するために、同じものをかけて、極座標に変換し、置換積分の知識からヤコビアンを求め、最終的に

$$\int_{\infty}^{\infty} e^{-e^2} dx = \sqrt{\pi} \tag{6}$$

を得る。この積分が 1 となるためには、あらかじめ $\sqrt{\pi}$ で割っておけば良い。

7.1　標準正規分布

平均 μ, 分散 σ^2 なる正規分布に従う確率変数 x に対して $z = \dfrac{x-\mu}{\sigma}$ という変換を施すと標準正規分布 $N(0, 1^2)$ となる。教科書に載っている正規分布表は標準正規分布に対する面積 (すなわち確率) を示している。

例題：$N(10, 5^2)$ という正規分布から取り出された、ある特定の値が 15 以下である確率を求めよ。

$$\frac{15-10}{5} = 1 \tag{7}$$

であるから、標準正規分布表を調べることにより

$$\int_{-\infty}^{1} f(x)dx = 0.68 \tag{8}$$

を得る。normdist(x, 平均, 標準偏差, 変数形式)

7.2　z スコア

$N(\mu, \sigma^2)$ から得られたある得点 y を $y = \dfrac{x-\mu}{\sigma}$ と標準正規分布に変換したときの値を z スコアまたは標準化得点と言う。z スコアを 10 倍して 50 を加えた値が偏差値と呼ばれる量になる。norminv(z スコア, 平均, 標準偏差)

例題：日本人の成人男子の身長は平均 168 cm, 分散が 100 であるという。このとき 180 cm 以上の身長のものは何%いるか

180 cm を標準化得点に変換すると $\dfrac{180-168}{10} = 1.2$ であるから、これを標準正規分布の表から調べると（あるいは normdist(180,168,10,1)=0.88493033 より）1- 0.88493033 = 0.11506967 より約 11 % いる。

例題：ある入学試験の成績は平均 100 点、分散 25 であるという。このテストにおいて、得点が 90 点以上 110 点以下であるものは何%いるか。

90 点を標準化得点に変換すると $(90-100)/5 = -2.0$, であり同様に 105 点は $(110-100)/5 = 2.0$ であるから、この値を正規分布の表で調べると（あるいは normdist(90,100,5,1)=0.02275 より）0.5 - 0.02275 = 0.4774 となるので $0.4774 \times 2 = 0.954$ であるから約 96 %を得る。

7.3 中心極限定理

確率変数 x_i が独立に任意の分布に従うとき、確率変数の和 $\sum_{i=1}^{n} x_i$ は n が大きくなれば、平均 μ, 分散 $\dfrac{\sigma^2}{\sqrt{n}}$ 正規分布に近づく。これは元の確率変数がどのような分布に従うかを問わない。このことを中心極限定理という。

7.4 母集団と統計量

心理学のデータは無限母集団からのサンプル（無限母集団から無作為に抽出されたデータ）である、と仮定される場合が多い。推測統計学の主たる目標の一つは、心理実験や調査などから得られたデータを元に、無限母集団の母数を推定することである。

母数とは母集団の分布を記述するために用いられる数のことである。一方、得られたデータの分布を記述する数は統計量と呼ばれる。心理統計学においては母数はギリシャアルファベットが用いられ、統計量についてはアルファベットが使われることが多い。例えば、母集団における平均である母平均を μ, 母分散を σ^2 と表す。一方統計量である標本平均は $\bar{x} = 1/n \sum x_i$、標本分散を $s^2 = 1/n \sum (x_i - \mu)^2$ などと表記する。標本平均、標本分散とも母数を含まない量である。

心理学においては有限母集団を仮定することは稀である。従って統計学のテキストの中で、しばしば詳細に記述されている無限母集団の例を覚える必要はない。しかし、有限母集団と無限母集団との区別を知っていることは重要である。繰り返すが、推測統計学の目標の一つは、得られたデータから、無限母集団の母数を推測することである。

有限母集団としては、例えば早稲田大学の学生の平均年齢などが挙げられる。一方、無限母集団の例としては、ある刺激に対する被験者の反応などが挙げられる。

心理学におけるデータ解析の大前提は、1) データが適切な母集団を想定できること、および 2) データが無作為抽出（ランダムサンプリング）によって

得られたものであること、の 2 点である。これらを認めることができれば統計学の各種技法を用いることが許される。逆に言えば、満たされない場合は結果が保証されない。