# 第1章　確率分布

## 1.1　確率密度関数

累積分布関数（単に分布関数）$F(x)$ は $x$ までの累積の確率（確率変数が $X < x$）を表す。連続形確率変数を扱う際には確率密度関数が必要で、

$$F(x) = \int_{-\infty}^{x} f(t)dt \tag{1.1}$$

で、$f(x)$ を確率密度関数という。確率密度関数は

$$f(x) = \frac{d}{dx}F(x) \tag{1.2}$$

のように、$F(x)$ を微分することで得られる。

## 1.2　変数変換

確率変数 $X$ とそれが従う確率密度関数 $f_X(x)$ が分かっているときに、例えば $y = x^2$ で変換される確率変数 $Y$ はどんな確率分布になるか？を考える。

何となく何も考えずに話を進めてしまうと陥った例を挙げておく。$N(0,1)$ に従う確率変数 $X$ があったときに、$y = x^2$ で変換した値が従う確率密度関数は、

$$f(x) = \frac{1}{\sqrt{2\pi}}\exp\{-x^2/2\} \to f(y) = \frac{1}{\sqrt{2\pi}}\exp\{-y/2\} \tag{1.3}$$

と考えてしまって、「ん?? あれ何も考えずとも求まってる？」となった。間違っているのだが、正しくは右辺は $\pm\sqrt{y}$ が従うガウス分布であり、$f(y)$（$Y$ の確率密度関数）ではない。今求めたいのは、変数変換した後の確率変数が従う密度関数であり（$f(y)$ もしくはそのまま $f(\cdot)$ を使うとややこしい場合は $g(y)$）、単純に置換しただけでは求まらないことに留意。

ここで確率変数は $Y = g(X)$ で変換される場合を考える。$Y$ の分布関数は

$$F_Y(y) = P(g(X) \leq y) = P(X \in \{x|g(x) \leq y\}) \tag{1.4}$$

で表される[1]。$X$ が連続型確率変数の場合には両辺を $y$ で微分することで、$Y$ の確率密度関数を計算することができる。

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}P(X \in \{x|g(x) \leq y\}) \tag{1.5}$$

# 第2章　回帰分析

説明変数と目的変数のくみを一つの観測値としてとらえ、$n$ 個の観測値からなるデータ集合を使用して分析を行う。説明変数が単一の場合を単回帰、複数の場合を重回帰と呼ぶ。

回帰式は以下の式で定義される：

$$\hat{y}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{ip}, \quad i = 1, 2, ..., n \tag{2.1}$$

$\hat{y}_i$ は予測値、$\beta$ は回帰係数、$x_{pi}$ は説明変数である。予測精度の良い回帰式とは、$\epsilon$ は誤差を表す。で表される。、誤差の最小化を行う。

最小二乗法は残差（residual, 実測値と予測値の差分）の平方和を最小化する回帰式を選択し、最尤法では残差の確率分布を仮定して尤度が最大となる回帰式を選択する。

## 2.1　重回帰分析

## 2.2　カイ二乗分布

自由度 $n$ のカイ二乗分布

$$f(x) = \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} x^{n/2-1} \exp\{-x/2\} \tag{2.2}$$