確率と統計学

- 確率の問題：ある条件の下で、どのような結果が期待されるか？
- 統計学の問題：ある結果を導くために、どのような条件が必要か？
- 統計的推定とは「逆確率」を求めること。

頻度主義とベイズ主義における確率

- 頻度主義：統計的推定の道具を組み立てるために確率を使う。
- ベイズ主義：統計的推定の道具として確率を使う。

ベイズ統計学と確率

- ベイズ統計学は**すべての**未知数の不確実性を表現するモデルとして確率を用いる。
- データを確率過程（サンプリング誤差と測定誤差に由来するノイズを含む確率論的な生態学的プロセス）の実現値と考える。

確率的表現の難しさ

- πの 100 万桁目の数を X とする。でも、X は確率変数ではない。
- そのため、「$X = 5$ である確率は 10%」というのは不自然な感じがする。
- しかし、我々の中にある不確実性の表現としては問題なさそう。

第 2 章の内容

- 確率論は不確実性を表現するモデルを提供する。
- でも、統計学のエンドユーザーがその詳細に立ち入る必要はない。
- 2.1 では、不確実性を表現するモデルとしての確率の概念を整理する。
- 2.2 では、ベイズ推定に必要な確率論の概念と手法を整理する。

2.1 確率とは何か？

「明日雨が降る確率は 70%」？

- 「明日雨が降るかどうかについて自分の中の不確実性を表現するモデル」が必要。

「コイン投げで表が出る確率は 50%」の意味すること

- 2 つの結果が等しくなりそう。だから、「表」になる確率は 0.50。
- 「なりそう」って何？
- コイン投げではそうなるという長年の経験があるって言ってしまいそう。
- 「過去のコイン投げではなく、今やろうとしているコイン投げ」が問題。
- 考えているプロセスの不確実性を表現するモデルが必要になる。

確率論は不確実な事象の数学的抽象化

- 「不確実な事象」とは、さまざまな形で起こり、必然的な結果を伴わない事象。
- 事象と結果を区別する。明日の天気は事象、雨や雪は結果。
- 確率論では、**結果**は全体集合Ωの要素ω、**事象** E は結果の集合として定義される。

確率の定義

- 確率は、Ωの部分集合 $\boldsymbol{F}$ の集まりの関数 P として定義される。
- 部分集合 $\boldsymbol{F}$ の集まりは**σ-代数**（補集合をとると非空で閉じており、加算な和集合）でなければばならない。
- 関数 P は確率測度であり、直感的に確率を記述する 3 つの基本的特徴をもつ。
- $E \in \boldsymbol{F}$ のとき、

$$0 \le P(E) \le 1 \tag{2.1}$$

$$P(E^C) = 1 - P(E) \tag{2.2}$$

$$P(E_*) = \textstyle\sum_i P(E_i) \tag{2.3}$$

$E^C = \Omega - E$ は E の補集合、$E_1, E_2, E_3, \ldots$が素集合のとき、E_*は和集合になる。

- 式(2.1)の意味：確率は 0 と 1 の間にある。
- 式(2.2)の意味：事象が起こら**ない**確率は、1 から事象が起こる確率を引いたもの。
- 式(2.3)の意味：相互に排他的な事象の確率を足すことができる。
- 全体集合Ω、σ-代数 $\boldsymbol{F}$、確率測度 P は確率空間を構成する。
- 確率変数 X はΩの関数であり、すべての t について、$\{\omega : X(\omega) \le t\}$が集合 $\boldsymbol{F}$ の中にあるという特性をもつ。

例：確率空間と確率変数

確率関数と確率変数の定義

- Ω = [0, 1]の部分集合からなる集まりを考える。
- $\boldsymbol{F} = \{E_1 = [0, 0.50], E_2 = (0.50, 1], E_3 = [0, 1], E_4 = \emptyset\}$, このとき$\emptyset$は空集合。
- $P(E_1) = 0.5, P(E_2) = 0.5, P(\Omega) = 1, P(\emptyset) = 0$ を満たすよう $\boldsymbol{F}$ の確率関数 P を定義できる。
- このとき、$X(\omega) = \begin{cases} 0, \omega \le 0.5 \\ 1, \omega > 0.5 \end{cases}$ と定義する。

確率空間

- 集合 $\boldsymbol{F}$ の集まりはσ-代数、P は(2.1)-(2.3)を満たす確率測度：$\{\Omega, \boldsymbol{F}, P\}$が確率空間。
- 関数 X は値 0 と 1（各確率 1/2）をとる確率変数。

コイン投げ再考

- 確率空間$\{\Omega, \boldsymbol{F}, P\}$のアナロジーによって結果の不確実性を記述してみる。
- E_1 は「表」、E_2 は「裏」、E_3 は「表か裏のどちらか」、E_4 は「表でも裏でもない」。
- 「表」と「裏」を確率変数 X の 2 つの値とする。

コイン投げにおける確率

- 確率空間$\{\Omega, \boldsymbol{F}, P\}$は不必要な抽象化のように見える。
- 頻度主義者のように、つい長期的な頻度と結びつけて確率を考えがち。
- 頻度主義者は、コイン投げを多数回繰り返すとき、「表」の相対頻度が 0.5 に収束するという意味で「Pr(表) = 1/2」と説明する。

2 つの解説

- 抽象的な確率空間によって「Pr(表) = 1/2」と定義すると、長期的な相対頻度が 0.5 に収束するのは数学的な帰結。なので、確率の**定義**として示されるのは確率の**性質**。
- 長期的な頻度による定義では、「確率」の適用が限定される。「明日雨が降る確率が 70%」という言明を定義するために、何回も明日を繰り返すことはできない。

結局どう考えたらよいのか？

- 必ずしも確率空間やσ-代数、集合関数を考える必要はない。
- でも、これらの抽象化によって生じる確率の幅広い定義を採用するのが便利。
- 確率は不確実性を表現する数理モデルであり、3 つの単純な規則（2.1-2.3）に従う。
- 長期的な頻度による確率の定義を超えて、あらゆる不確実性を記述するために確率を使うという考え方が、ベイズ推定の基礎になる。

2.2 ベイズ推定のための基礎的な確率概念

ベイズ推定に必要な確率概念

- ベイズ推定とは、データと未知数の同時分布の特定である。
- ベイズ推定は、データが与えられたときの、未知数の条件付き分布に基づく。

2.2.1 同時分布・周辺分布・条件付き分布

確率分布関数

- $\boldsymbol{X}$ をベクトル値の確率変数とする。
- **確率分布関数** $f(x)$は、積分もしくは総和すると、領域 R に対して、$\Pr(\boldsymbol{X} \in R)$となる。
- $\boldsymbol{X}$ は $f(x) > 0$ となる値 x の集合をもつ。
- 例：ベクトル $\boldsymbol{X} = (Y, Z)'$の分布関数

$$f_{Y,Z}(y, z) = \frac{y}{5z}, \begin{cases} 0 < y < z \\ z = 1, 2, 3, 4 \end{cases} \tag{2.4}$$

- $\boldsymbol{X}$ が、$\{Y \le 1.5, Z \le 2\}$と定義される領域 R に存在する確率

$$\sum_{Z=1}^{2} \int_0^{\min\{1.5, z\}} f_{Y,Z}(y, z)dy = \int_0^1 f_{Y,Z}(y, 1)dy + \int_0^{1.5} f_{Y,Z}(y, 2)dy = \frac{17}{80}$$

- ベクトル値の確率変数の分布関数はその要素の同時分布になることもある。
- ベクトルの個々の要素の分布を周辺分布という。

周辺分布関数

- 同時分布(2.4)について、$Z = z$ となる確率が周辺分布になる。
- $z = 1, 2, 3, 4$ のとき、

$$f_Z(z) = \Pr(Z = z) = \Pr(0 < Y < z, Z = z) = \int_0^z f_{Y,Z}(y,z)dy = \int_0^z \frac{y}{5z}dy = \frac{z}{10'} \quad (2.5)$$

- 周辺分布を得る：「同時分布の y を integrate out する」or「y を周辺化する」

条件付き分布関数

- Z を与えたときの Y の条件付き分布は、同時分布を Z の周辺分布で割ったもの。
- $0 < y < z$ のとき、

$$f_{Y|Z}(y|z) = \frac{f_{Y,Z}(y,z)}{f_Z(z)} = \frac{2y}{z^2} \quad (2.6)$$

- 条件付き分布 $f_{Y|Z}(y|z)$は z を固定したときの y の関数。
- 同時分布と条件付き分布はどちらも分子に y をもち、比例関係にある。
- 同時分布を与えると、周辺分布$f_{Y|Z}(y|z) \propto g(y) = y$であるとすぐにわかる。
- 条件付き分布を特定するには､積分すると 1 になるように $g(y)$をスケーリングする。
- $0 < y < z$ のとき、$g(y)$の積分は $z^2/2$ なので、条件付き分布は $y/(z^2/2) = 2y/z^2$ となる。

括弧表記

- 同時分布：$[Y, Z]$、Z を与えたときの Y の条件付き分布：$[Y|Z]$、Z の周辺分布：$[Z]$
- 式(2.6)の括弧表記

$$[Y|Z] = \frac{[Y,Z]}{[Z]}$$

- $\boldsymbol{X}$ がデータ、$\boldsymbol{\theta}$がパターメーターのとき、ベイズ推定は$\boldsymbol{\theta}$の条件付き分布に基づく。

$$[\boldsymbol{\theta}|\boldsymbol{X}] \propto [\boldsymbol{X}, \boldsymbol{\theta}] = [\boldsymbol{X}|\boldsymbol{\theta}][\boldsymbol{\theta}] \quad (2.7)$$

- 式(2.7)の左辺は$\boldsymbol{\theta}$の事後分布、右辺は尤度と事前分布になる。

2.2.2 パーセンタイル・モーメント・独立性

パーセンタイル

- p を 0 と 1 の間の数とする。
- X が連続確率変数のとき、$[X]$の 100p 番目のパーセンタイルは、$\Pr(X \leq x_p) = p$ かつ $\Pr(X \geq x_p) = 1 - p$ である数 x_p になる。
- 50 番目のパーセンタイルが**中央値**、25 番目と 75 番目のパーセンタイルが**四分位数**。
- 離散確率変数の場合、式(2.5)の分布関数の確率変数 Z は、$z = 1, 2, 3, 4$ のとき、$\Pr(Z \leq z) = 0.10, 0.30, 0.60, 1.00$ となってしまう。
- より柔軟なパーセンタイルの定義：

  $[X]$の 100p 番目のパーセンタイルは、

  $\Pr(X < x) \leq p$ かつ $\Pr(X > x) \geq 1 - p$ である数 x
- 式(2.5)の$[Z]$の 50 番目のパーセンタイルは 2 と 3 の間にある数となる。

モーメント

- X がベクトル値の確率変数で、密度関数 $f_X(x)$と関数 $g(x)$をもつとき、$g(X)$の期待値の定義は、

$$\mathrm{E}\big(g(X)\big) = \int \mathrm{g(x)} f_X(x) dx \tag{2.8}$$

- 本来、式(2.8)では、X の連続要素に対する多重積分と離散要素に対する総和が必要。
- 1 変量の確率変数 X のとき、$g(X) = X^k$ の期待値は X の k 番目のモーメントになる。
- 1 変量の確率変数について、

  平均：$\mathrm{E}(X)$

  分散：$\mathrm{Var}(X) = \sigma^2(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$

  標準偏差：$\mathrm{SD}(X) = \sigma(X) = \sqrt{\mathrm{Var}(X)}$
- 確率変数のペア(X, Y)について、

  共分散：$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$

$$\text{相関：Corr}(X, Y) = \rho(X, Y) = \frac{\text{Cov}\,(X,Y)}{\sigma(\text{X})\sigma(\text{Y})}$$

独立性

- 同時分布$[X, Y] = [X][Y]$のとき、確率変数のペア(X, Y)は**独立**である。
- $[X_1, \ldots, X_n] = [X_1] \cdots [X_n]$のとき、$n$ 個の確率変数の集合は**相互に独立**である。
- 相互に独立な場合はペアも独立である。
- 相互に独立でなくてもペアが独立な場合はある。
- $[X, Y|Z] = [X|Z][Y|Z]$のとき、X と Y は Z の**条件付き独立**である。
- 未知パラメーターθをもつ分布からの確率標本は、未知パラメーターの条件付きで相互に独立である：$[X_1, \ldots, X_n|\theta] = [X_1|\theta] \cdots [X_n|\theta]$
- このような確率変数を「独立同分布（iid）」と呼び、独立性は未知パラメーターの条件付きである。

2.2.3　確率分布

- 生態学モデルで使われる確率分布は限られている（Table 2.1 参照）。
- この本では、多項分布（カテゴリカルデータ）、ポアソン分布（カウントデータ）、指数分布（生存時間データ）が特に重要。
- ベータ分布とガンマ分布は未知パラメーターの事前分布としてよく用いられる。

2.2.4　変数変換

- 統計学における共通の問題 : 確率分布関数 $f_X(x)$をもつ確率変数 X を与えたとき、逆関数 g について $Y = g(X)$となる Yの分布を見つける必要がある。このとき、$X = g^{-1}(Y)$。
- X が離散確率変数のとき、Y も離散的である。Y の確率分布関数は、

$$\Pr(Y = y) = \Pr(X = g^{-1}(y))$$

- y の標本空間（S_y）は、$x \in S_x$（x の標本空間）のときに計算される値$\{g(x_1), g(x_2), \ldots\}$

の集合に等しい。

- X が連続確率変数であり、$g(X)$が微分できるとき、Y の確率分布関数は、

$$f_Y(y) = f_X(g^{-1}(y)) \times \left|\frac{dg^{-1}(y)}{dy}\right|$$

- これを**変数変換の定理**と呼ぶ。このとき、$dg^{-1}(y)/dy$ は逆変換 $X = g^{-1}(Y)$の導関数。
- 例として、$X \sim N(\mu,\sigma^2)$のとき、確率変数 $Y = e^X$ から対数正規分布を導出する。
- $y > 0$ のとき、逆変換は $g^{-1}(y) = \ln(y)$であり、導関数 $1/y$ をもつため、

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(\ln(y)-\mu)^2} \frac{1}{y}$$

- 変数変換の定理は多変量に拡張できる。
- 連続確率変数 $\boldsymbol{X} = (X_1, .., X_p)$を $\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X})$に変換したいとき、$g$ が逆関数で微分可能であれば、

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{g}^{-1}(\boldsymbol{y}))\left|J_{\boldsymbol{g}^{-1}}(\boldsymbol{y})\right|$$

- このとき、$\left|J_{\boldsymbol{g}^{-1}}(\boldsymbol{y})\right|$は逆変換のヤコビ行列式（偏導関数の行列式）。