確率・統計の初歩(1)

別所俊一郎

確率・統計がなぜ必要か。

- 実証分析にはさまざまな「でたらめさ」が付随
  - 標本抽出（サンプリング）
  - 考慮できない／捨象された要因たち
  - （決定論的立場には立たない）
- 確率論
  - 「でたらめさ」を数量化して扱う数学的手法
  - 回帰分析や計量経済学に必要な部分だけ
- 母集団と標本の区別

確率論の基礎用語

- 根源事象outcome
  - 相互に排他的な、潜在的に起こりうる結果。どれもが同じように起きやすいということではない
- 標本空間
  - 起きうる結果すべてから成る集合
- 事象event
  - 標本空間の部分集合。根源事象の組み合わせで表現できる。
- 確率
  - 事象が起きる頻度の割合
- 確率変数
  - 事象の数学的表現。標本空間から実数への関数として定義される。（確率変数を大文字で、実現値を小文字で表す）

離散確率変数のばあい

- 確率分布
  - 確率変数が取りうるすべての値（根源事象）と、それに対応する確率のリスト。確率をすべて足すと1。
- 各事象の確率
  - 確率分布から求められる
- 累積確率分布 c.d.f.
  - 確率変数がある値より小さい値をとる確率。累積分布関数とも。
- 例：Table 2.1
- ベルヌーイ分布
  - 2値変数のとる分布
  - ベルヌーイ家は天才を輩出した家系として知られる。

連続確率変数のばあい

- 累積確率分布
  - 離散のばあいと同じ。確率変数がある値より小さい値をとる確率。
- 確率密度関数 p.d.f.
  - 「下」の面積が確率に等しくなるような関数
  - 累積分布を微分したもの。全区間を積分すると1。
  - 離散変数と同じようには定義できないことに注意。
- 例：Figure 2.2.
  - 連続変数を扱うケースのほうが多い。

期待値・モーメント

- 期待値 $E[Y]$, μ_Y
  - 何度も試行を繰り返したときの長期的な平均値
  - 離散変数のばあいは確率を重みとする加重平均
  - 連続変数のばあいも“加重平均”みたいなもの

  - 成功確率pのベルヌーイ分布のばあい
- モーメント（積率）
  - 確率変数の累乗の期待値
  - r次のモーメント $E[Y^r]$

分散・標準偏差

- 分散
  - 2次のモーメントのこと。var(Y) =
  - 確率分布の「広がり」を表す
- 標準偏差
  - 分散の平方根 σ_y =
- ベルヌーイ分布のばあい

- 3次モーメントは歪度、4次モーメントは尖度
  - 数学的な仮定として用いられる

期待値の線形性

- 確率変数の線形関数の期待値と分散
  - 確率変数X、定数a, bに対して、$Y = a + bX$ とする
  - このとき、$E[Y] = a + b\,E[X]$ 、var(Y) = b^2var(X)。

［証明］

同時分布、周辺分布

- 同時分布 Pr(X=x, Y=y)
  - 確率変数Xがxという値をとり、かつ、確率変数Yがyという値をとる確率
  - 全ての組合せについて確率を足すと1
  - 例：Table 2.2.
- 周辺分布 Pr(X)
  - ひとつの確率変数だけに注目したときの分布
  - 同時分布の和として表現できる。
  - Pr(X) =

条件付き分布

- 条件付き分布 Pr(Y=y | X=x)
  - 確率変数Xの実現値を所与にしたときの確率変数Yの分布
  - 確率変数Xの実現値の関数となる
  - Pr(Y=y | X=x) =
  - 例：Table 2.3.
- 条件付き期待値 E[Y | X=x] =

期待値のくりかえしの法則

- 確率変数*Y*の期待値は、確率変数*X*の実現値*x*で条件付けられた*Y*の条件付き期待値を*X*の分布で加重平均したものに等しい。
- $E[Y] = E[E[Y|X]]$

［証明］

条件付き分散

- 確率変数*X*の実現値を所与にしたときの確率変数*Y*の分散
- $\mathrm{var}(Y \mid X{=}x) =$

独立

- 確率変数XとYが互いに独立に分布しているとは、
  - Xの実現値についての情報がYの条件付き分布についての情報とならないこと。
  - 実現値x, yの全てに対して以下が成り立つ
  - それゆえ、条件付き分布の定義式より、確率変数XとYの同時分布について以下が成り立つ

共分散、相関

- 共分散
  - 2つの変数が似たような動きをするかどうかを示す指標。同じ方向に動くときに正の値。
  - 定義：σ_{xy} = cov(X, Y) =
- 相関
  - 共分散の単位をそろえたもの。ゼロであるとき、2つの確率変数は「無相関である」という（独立⇒無相関）
  - 定義：corr(X, Y) =

共分散・相関係数の性質

- 相関係数は1を越えない
  - 証明は appendix 2.1.
- 条件付き平均が無条件平均に等しいとき、共分散はゼロ。逆は必ずしも成り立たない。

［証明］

確率変数の線形結合

- 2つの確率変数X, Yの和の期待値、和の分散は、

$$E[X+Y] = E[X] + E[Y]$$

$$\mathrm{var}(X+Y) = \mathrm{var}[X] + \mathrm{var}[Y] + 2\,\mathrm{cov}(X, Y)$$

［証明］