

第2回: 確率分布と確率変数(2)

身長や体重の分布はどう表されるか(連続分布)

確率変数 x がとる値が連続量(実数値など)
 x のとりうる値が、連続量の一定幅のある区間のとき、
 x がある値 $x=a$ (区間中の1点)をとる確率は、0 である。
 x がある区間 $a \leq x \leq b$ にある確率は、有限なので、
その確率 $\text{Prob}\{a \leq x \leq b\}$ を考える。

$\text{Prob}\{x=a\}$, $\text{Prob}\{x=b\}$ なので、
 $\text{Prob}\{a < x \leq b\}$, $\text{Prob}\{a \leq x < b\}$, $\text{Prob}\{a < x < b\}$
は、いずれも同じ。

累積分布関数(cumulative distribution function)

確率変数 x がある数 b 以下をとる確率 $F(b) = \text{Prob}\{x \leq b\}$

先の確率 $\text{Prob}\{a < x \leq b\}$ は、この
累積分布関数 $F(b)$ を用いて

$$\begin{aligned}\text{Prob}\{a < x \leq b\} &= \text{Prob}\{x \leq b\} - \text{Prob}\{x \leq a\} \\ &= F(b) - F(a) \\ &= \int_a^b f(x) dx\end{aligned}$$

ここで、 $f(x) = \frac{d}{dx} F(x)$ **確率密度関数**
(probability density function)

累積分布関数 $F(x)$ が満たすべき性質

$$0 \leq F(x) \leq 1, \quad F(\infty) = 1$$

確率密度関数 $f(x)$ が満たすべき性質

$$f(x) \geq 0, \quad \int_{-\infty}^{+\infty} f(x) dx = 1$$

参考:(離散)確率分布が満たすべき条件

$$p(x) \geq 0, \quad \sum_x p(x) = 1$$

銀行窓口へのお客の到着時間間隔 (連続分布の例としての指数分布)

- お客さんどうしは、互いに独立にばらばらに到着する→このとき、**到着時間の間隔 t** は、**指数分布**(exponential distribution) となる。

つまり、 $\text{Prob}\{t_1 \leq t \leq t_2\} = \int_{t_1}^{t_2} f(t) dt$ は、

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & (t \geq 0) \\ 0 & (t < 0) \end{cases} \quad F(t) = 1 - e^{-\lambda t}$$

λ は、ただ1つのパラメータ。平均値が $1/\lambda$ (後出)。

平均と分散(mean/average and variance)

- 確率変数が従う確率的規則 \Leftrightarrow
 - 確率関数 $p(x)$ (離散分布に対して)
 - 確率密度関数 $f(x)$ (連続分布に対して)
- これらの関数を特徴付ける量: **特性量**
- 最も基本的な特性量: **平均と分散**
標準偏差(分散の平方根)

(1) 平均 μ (mean)

$$\mu = \begin{cases} \sum_x x \cdot p(x) & (\text{離散分布のとき}) \\ \int_{-\infty}^{+\infty} x \cdot f(x) dx & (\text{連続分布のとき}) \end{cases}$$

x の値として期待される平均の値
平均値、期待値

確率変数 x がおおよそどんな値をとるか、
を特徴付ける

(2) 分散 σ^2 (variance)

$$\sigma^2 = \begin{cases} \sum_x (x-\mu)^2 \cdot p(x) & (\text{離散分布のとき}) \\ \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x) dx & (\text{連続分布のとき}) \end{cases}$$

$(x-\mu)^2$ の値として期待される値
 →平均 μ からのずれ(上下ともに)の2乗

標準偏差(必ず非負): σ^2 の(正の)平方根
 (standard deviation)

確率分布が平均 μ のまわりにどの程度広がっているか、
 を特徴付ける

例題2.4:[1]「コイン投げ」での平均と分散

2値を等確率でとる離散分布 $p(x) = \begin{cases} 1/2 & (x=1) \\ 1/2 & (x=0) \end{cases}$ に対して、

$\mu = \sum_x x \cdot p(x)$ と $\sigma^2 = \sum_x (x-\mu)^2 \cdot p(x)$ を計算する

$$\Rightarrow \mu = \frac{1}{2}, \sigma = \frac{1}{2}$$

例題2.4:[2]「複数回コイン投げ」での平均と分散

x 回目に表が出る確率 $p(x)$.
 p : 1回のコイン投げで表が出る確率。

幾何分布 $p(x) = (1-p)^{x-1} \cdot p$, $x = 1, 2, 3, \dots$ に対して、

$\mu = \sum_x x \cdot p(x)$ と $\sigma^2 = \sum_x (x-\mu)^2 \cdot p(x)$ を計算する

$$\Rightarrow \mu = \frac{1}{p}, \sigma^2 = \frac{1-p}{p^2}$$

$$p = \frac{1}{2} \text{ のとき、} \mu = 2, \sigma^2 = 2 \quad (\sigma = \sqrt{2})$$

例題2.4:[3]指数分布での平均と分散

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 $\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x) dx$ を計算する

$$\Rightarrow \mu = \frac{1}{\lambda}, \sigma^2 = \frac{1}{\lambda^2}$$

時間間隔 t の平均は $1/\lambda$

よく使う公式(分散の計算に便利)

$$\begin{aligned} \sigma^2 &= \sum_x (x-\mu)^2 \cdot p(x) \\ &= \sum_x (x^2 - 2\mu x + \mu^2) \cdot p(x) \\ &= \sum_x x^2 \cdot p(x) - \mu^2 \end{aligned}$$

分散は、 x^2 の平均 - 平均の2乗

連続分布の場合にも、同様の式を立てて確認せよ。

この授業で呈示した資料は、

以下のURLで、pdfファイルとして公開しています。

授業HP <http://sys.ci.ritsumei.ac.jp/probability/>

あるいは

1. 知能情報学科HP <http://www.ci.ritsumei.ac.jp> の右上にある「▲在学生へのサポート」をクリック、
2. 一番上にある「2回生科目:前期」の欄から「■確率統計D1」をクリックする。

および

コースツールmanaba+Rの本科目「コンテンツ」