



ベイズのはなし

(ネタバレ防止案)

ゆっきん

(すうがく徒のつどい 第5回)

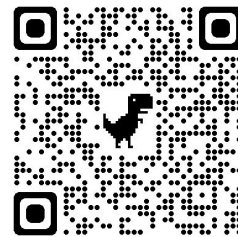
前置き

- ゆっきんについて

- 理学部数学科卒→高校数学教師→SE→高校プログラミング講師(NOW)
- 数学と情報の学習コミュニティとボードゲームコミュニティを運営
- QRコードからXアカウントをフォローしていただけると嬉しいです
(仲良くしていただけると、もっと嬉しいです)

- 本講座について

- 本講座はベイズ統計学の入門講演です
- 入門講演として、初学者の方でも楽しめるように心がけます
ぜひ楽しい時間にしましょう！！



目次

1. 条件付き確率とベイズの定理

条件付き確率の復習と、本講座の根幹であるベイズの定理について説明します。

2. 確率分布

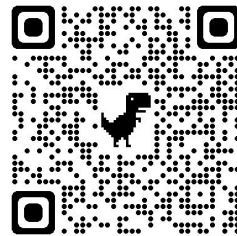
ベイズ推論のための準備として、確率分布について説明します。

3. ベイズ推論

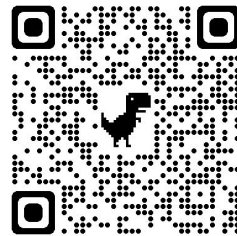
コイン投げの例を通して、ベイズ推論の基本的な考え方を説明します。

4. ベイズ線形回帰

ベイズ推論の機械学習への応用として、ベイズ線形回帰を紹介します。



1. 条件付き確率とベイズの定理



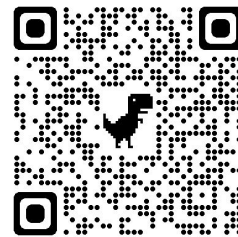
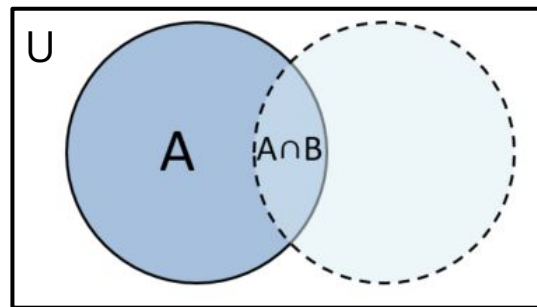
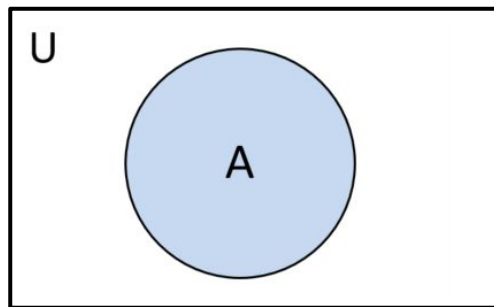
根元事象がすべて同様に確からしい試行において、

事象 A の起こる確率は

$$P(A) = \frac{n(A)}{n(U)}$$

事象 A が起こった時に事象 B が起こる条件付き確率は

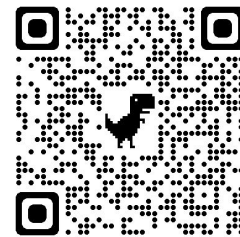
$$P(B|A) = \frac{n(A \cap B)}{n(A)} = \frac{P(A \cap B)}{P(A)}$$



条件付き確率の例

さいころを1回振ります。

出た目が4以上のとき、その目が偶数である確率を求めなさい。



問題「2人の子ども①」

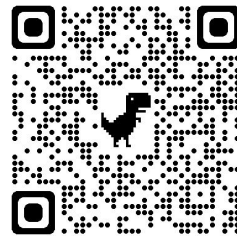
スミスさんには子どもが2人います。2人のうち、**年上の子は女の子**です。
では、2人とも女の子である確率を求めなさい。
ただし、男の子と女の子は等確率で生まれるものと仮定します。

答えはどれでしょうか？

① $\frac{1}{2}$

② $\frac{1}{3}$

③ $\frac{1}{4}$



問題「2人の子ども②」

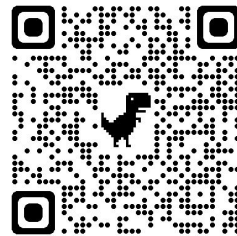
スミスさんには子どもが2人います。2人のうち、**少なくとも1人は女の子**です。
では、2人とも女の子である確率を求めなさい。
ただし、男の子と女の子は等確率で生まれるものと仮定します。

答えはどれでしょうか？

① $\frac{1}{2}$

② $\frac{1}{3}$

③ $\frac{1}{4}$

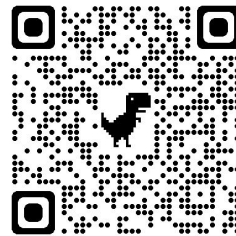


問題「2人の子ども③ 一火曜日に生まれた少女」

スミスさんには子どもが2人います。2人のうち、**少なくとも1人は火曜日に生まれた女の子**です。では、2人とも女の子である確率を求めなさい。
ただし、男の子と女の子は等確率で生まれるものと仮定します。

火曜日に生まれたという一見関係なさそうな条件が、確率に影響を与えるのでしょうか？ ぜひ考えてみてください！

答えは、「火曜日に生まれた少女」と検索すると出てきます。



$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$ より、

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$$

この式を**ベイズの定理**と呼びます。

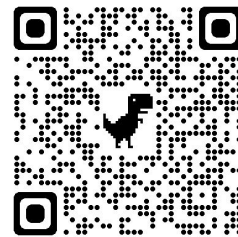
B が何らかの観測、A をその原因としたとき、

$P(A)$: 原因Aの発生確率

$P(B)$: 観測結果Bの発生確率

$P(B|A)$: Aが発生した際に観測結果Bが発生する確率(時間順行)

$P(A|B)$: Bが発生した際に原因Aが起こっていた確率(時間逆行)



問題「病気に罹患している確率」

ある病気の罹患率は1%です。この病気に罹患しているか検査する方法があり、罹患している人は99%の確率で陽性と診断され、健康な人は97%の確率で陰性と診断されます。この検査で陽性と診断されたとき、実際に罹患している確率を求めなさい。

答えはどれでしょうか？

①

25%

②

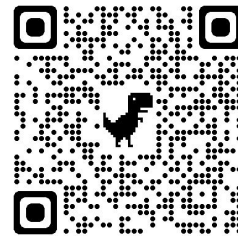
50%

③

70%

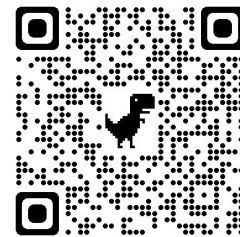
④

99%

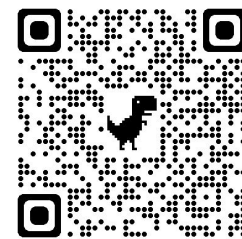


ベイズの定理の活用例「迷惑メールフィルター」

届いたメールはどのようにして、迷惑メールかどうか判断されるのでしょうか？



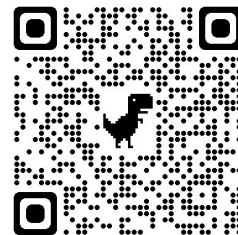
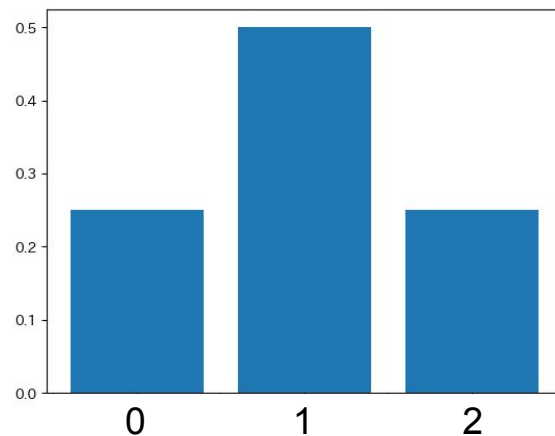
2. 確率分布



どの値を取るかが確率的に決まる変数のことを**確率変数**と呼びます。
また、確率を表す関数のことを**確率分布**と呼びます。

(例)コインを2回投げたときの表が出た回数 X

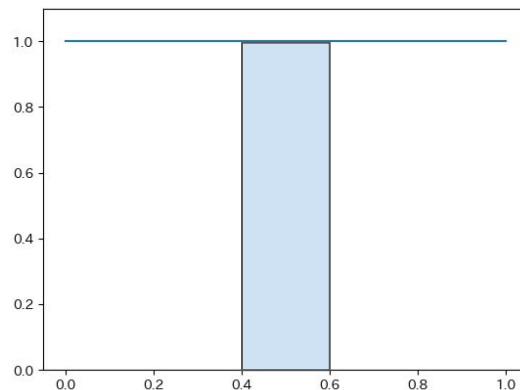
$$P(X) = \begin{cases} \frac{1}{4} & (X = 0) \\ \frac{1}{2} & (X = 1) \\ \frac{1}{4} & (X = 2) \end{cases}$$



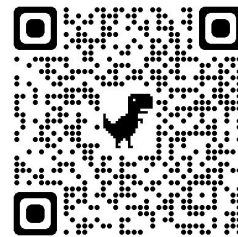
連続型の確率変数、確率分布を考えることもできます。

(例)0 から 1 までのランダムな実数 X

- $P(X = 0.5) = 0$ (一点の確率は0)
- $P(0 \leq X \leq 1) = 1$ (全体の確率は1)
- $P(0.4 \leq X \leq 0.6) = 0.2$ (区間の確率は面積)



$f(x) = 1$ ($0 \leq x \leq 1$) を X の **確率密度関数** と呼びます。

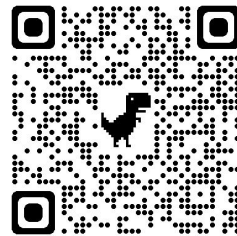
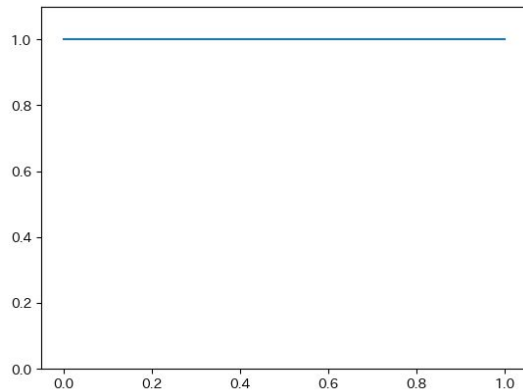


確率分布の例① 連続一様分布

確率変数がどのような値でも、確率密度関数が一定の値をとる分布

$$f(x) = \frac{1}{b-a} \quad (a \leq x < b)$$

(例)0 から 1 までの実数をランダムで決定するときの値が従う確率分布



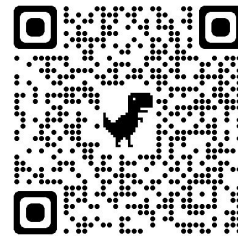
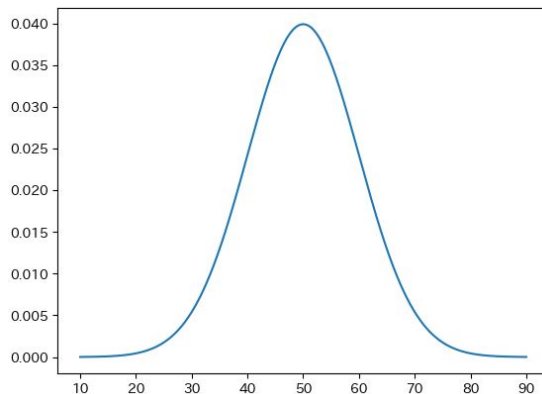
確率分布の例② 正規分布

統計における最重要分布(中心極限定理)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

μ : 平均
 σ : 標準偏差

(例) 平均50、標準偏差10の正規分布(偏差値)



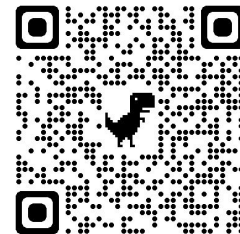
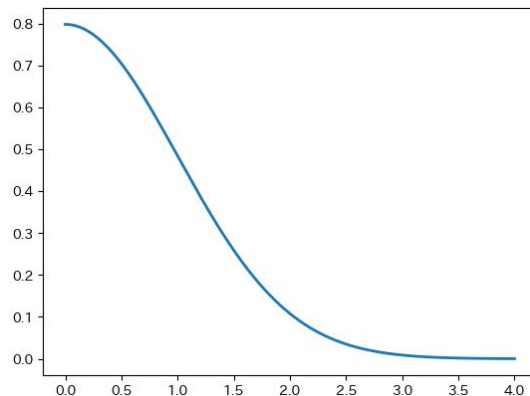
確率分布の例③ 半正規分布

正の値のみをとる正規分布を考えたい時に使用する分布

σ : 標準偏差

$$f(x) = \sqrt{\frac{2}{\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (x > 0)$$

(例) 標準偏差1の半正規分布

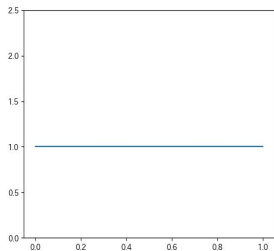


確率分布の例④ ベータ分布

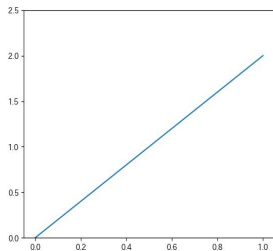
α と β の2つのパラメータによって特徴づけられる分布

$$f(x) = C \cdot x^{\alpha-1}(1-x)^{\beta-1} \quad C = \frac{(\alpha + \beta - 1)!}{(\alpha - 1)!(\beta - 1)!}$$

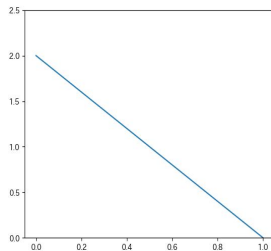
面積を1にする
ための調整係数



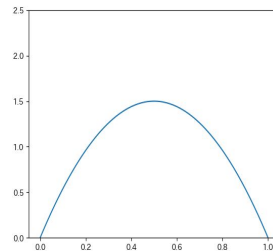
$\alpha=1, \beta=1$



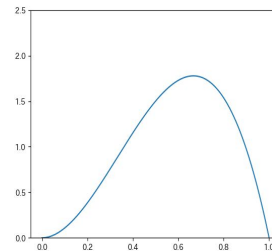
$\alpha=2, \beta=1$



$\alpha=1, \beta=2$

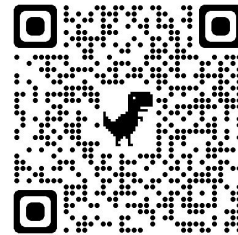


$\alpha=2, \beta=2$

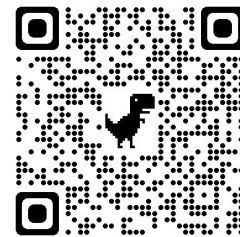


$\alpha=3, \beta=2$

連続一様分布



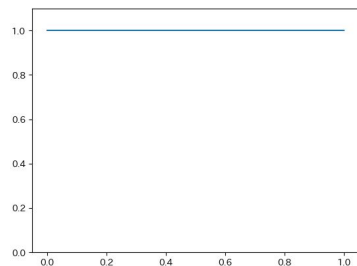
3. ベイズ推論



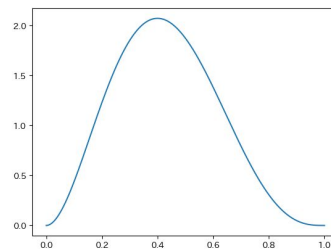
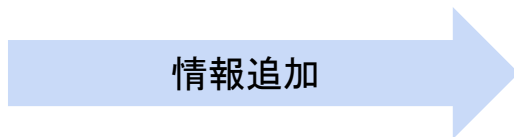
ベイズ推論とは、パラメータ p の確率分布を推論することです。

ベイズ推論の流れは以下となります。

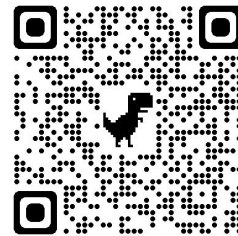
1. パラメータ p の事前の確率分布を設定する(事前分布)
2. 情報が得られる
3. 情報によって、パラメータ p の確率分布が更新される(事後分布)



事前分布



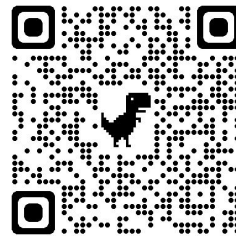
事後分布



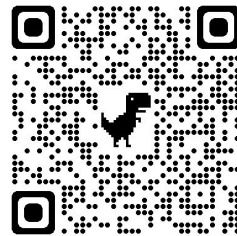
問題「くじ引きで当たりを引く確率」

当たりの確率が一定のくじを5回引いたとき、結果は「当たり・外れ・外れ・当たり・外れ」でした。このくじが当たる確率 p はどのくらいでしょうか？

この問題を最尤推定という方法と、ベイズ推論という方法の2通りで考えましょう。



4. ベイズ線形回帰



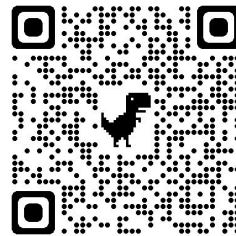
AI(人工知能)は、**ルールベース**の手法と**機械学習**の手法に分かれます。

- **ルールベースの手法**:人がルールを定める
- **機械学習の手法**:機械(コンピュータ)がデータをもとにルールを学習する

機械学習はさらに教師あり学習、教師なし学習、強化学習に分かれます。

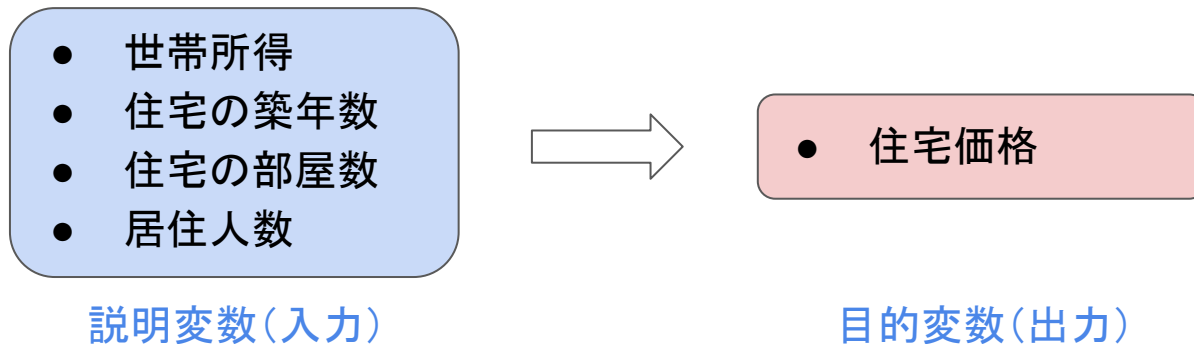
このうち**教師あり学習**とは、正解付きのデータをもとにルールを学習し、未知のデータの正解を予測することを言います。

学習とは、**与えられた入力に対して適切な出力を生成する良い関数を見つける**ことです。



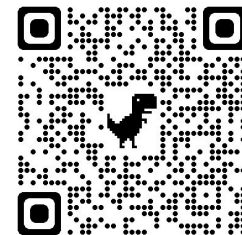
予測の中でも数値を予測することを、**回帰**と呼びます。

回帰の例 カリフォルニアの住宅価格



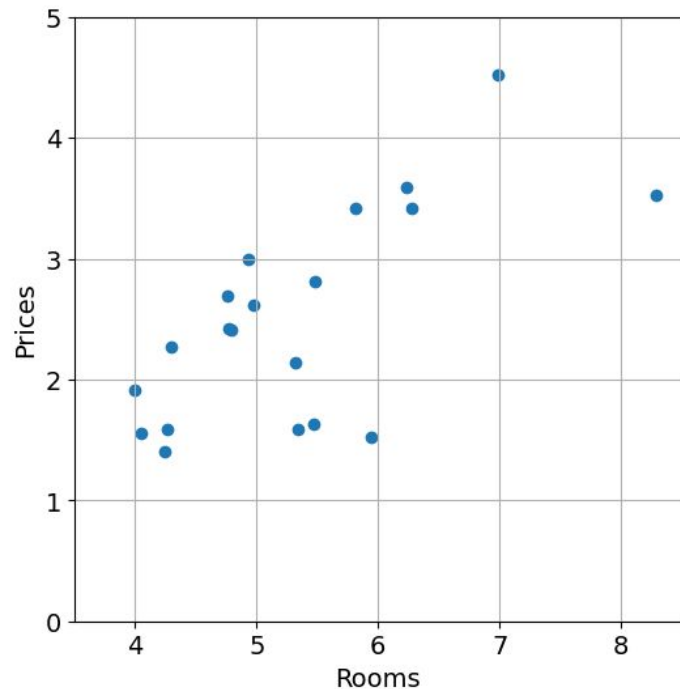
学習時のデータは「説明変数」+「目的変数」です。

予測時のデータは「説明変数」のみで、そこから「目的変数」を求めます。



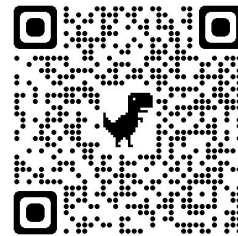
回帰の例「カリフォルニアの住宅価格」

住宅の部屋数と住宅価格の関係性を調べてみましょう。



1 次式 $y = \alpha x + \beta$ で 2 変数の関係を近似しましょう。これを**線形回帰**といいます。

最小 2 乗法を用いた一般的な線形回帰と、
ベイズ線形回帰の 2 通りの方法で考えます。



まとめ

- 確率は情報を得ることで更新されます。(ベイズの定理)
- ベイズ推論は、ベイズの定理を土台とした推論の方法です。パラメータの事前分布と得られたデータをもとに、パラメータの事後分布を推論します。
- ベイズ推論には次のようなメリットがあります
 - 推論の結果が確率分布であることから、不確実性が表現されている
 - データが不十分な場合にも使うことができる
 - 事前知識や経験を事後推定に組み込むことができる

