決定理論　と　ベイズ法

不確か　と　未知　の違いを理解しよう

渡辺澄夫

Ｑ ＆ Ａ

学生のみなさまから

　「決定理論とは何ですか」

　「ベイズ統計は決定理論に基づいていますか」

というご質問をいただきました。

このファイルはその質問への回答です。

答えのみを知りたい人は「６まとめ」をご覧ください。

はじめに

このファイルでは、決定理論とベイズ統計との関係についてできるだけ短く簡明に説明します。決定理論は「不確かな状況において合理的な意思決定を行うためのもの」 と説明されることがあります。ここで、「不確かな状況」と「未知の状況」の言葉が表す意味の相違を理解し、正しく利用しましょう。

１　決定理論とは

決定理論とは。

決定理論では

確率的な状況　ただし

確率分布はわかっている

最小化したい損失も定まっている

を考察します。

まず例で説明します。

「不確かな状況」

（1） q は $[0,1]$ 上の一様分布に従う。

（2） X は確率 q で1, $(1-q)$で 0 となる確率変数。

「意思決定」

（3） 損失 $(q-f(X))^2$ の(q,X) に関する平均値を

最小にする関数 $f(X)$ を「意思決定」とする。

（注）損失は、もちろん、定まっていれば二乗誤差でなくても、なんでもよい。

決定理論のフレームワーク

前ページのことを一般的に書きます。

(1) w は既知の P(w) に従う確率変数である。

(2) X は既知の P(x|w) に従う確率変数である。

(3) 行動は決定関数 f(X) で定義される。

(3) 決定 f(X) の損失 L(f(X),w) が定まっている。

(4) リスクを $R(f,w) = \int L(f(x),w)\, P(x|w)\, dx$ と定める。

(5) 平均リスクを $E[R(f,w)] = \int R(f,w)\, P(w) dw$ と定める。

決定理論の問題

決定理論では

P(w), P(x|w), L(f(x),w) をすべて既知とする。

平均リスクを最小にする関数 f(x) は何か。

を問う。

決定理論には解答がある

決定理論では $P(w)$, $P(x|w)$ のどちらも既知であるから $P(x) = \int P(w)P(x|w)\, dw$ も既知であり、また

$$P(w|x) = P(w)P(x|w) / P(x)$$

も既知である（ベイズの定理）。平均リスクは

$$E[R(f,w)] = \int \int L(f(x),w)\, P(w|x)\, dw\, P(x)\, dx$$

であるから $\int L(f(x),w)\, P(w|x)\, dw$ を最小にする $f(x)$ が答えである。

例1（簡単に解ける場合）

(1) w は [0,1] 上の既知の一様分布に従うとする。

(2) X は確率 w で1, (1−w)で0とする。$P(X|w)= w^x(1-w)^{1-x}$

(3) 関数 f(x) の損失を $L(f(X),w) = (w-f(X))^2$ と定める。

(4) 関数 f(X) のリスクを計算すると次のようになる。

$$R(f,w) = \sum_{x=0,1} (w-f(x))^2 \ w^x(1-w)^{1-x}$$

$$= (2f(0)-2f(1)+1)w^2 +(-f(0)^2 -2f(0)+f(1)^2)w +f(0)^2$$

(5) 平均リスクを計算すると次のようになる。

$$E[R(f,w)]=f(0)^2 /2-f(0)/3+f(1)^2 /2-2f(1)/3+1/3$$

平均リスクを最小にする関数は f(0)=1/3, f(1)=2/3 である.

例2（簡単に解ける場合）

(1) w は 平均0分散1の正規分布 $P(w)=N(w|0,1)$ に従う。

(2) X は平均 w 分散1の正規分布 $P(x|w)=N(x|w,1)$ に従う。

(3) 関数 f(x) の損失を $L(f(X),w) = (w-f(X))^2$ と定める。

(4) $P(x) = \int N(w|0,1)N(x|w,1)\,dw = N(x|0,2)$ であり

$$P(w|x) = N(w|0,1)N(x|w,1) / N(x|0,2) = N(w|x/2,2)$$

であるから

$$E[L(f,w)] = \int \int \{w^2 -2wf(x)+f(x)^2\} P(w|x)\,dw\,P(x)\,dx$$

$$= \int \int \{w^2 -xf(x)+f(x)^2\} P(w|x)\,dw\,P(x)\,dx.$$

平均リスクを最小にする関数は $f(x) = x/2$ である.

注意

(1) 例１例２は、f(x) が簡単に求まる場合である。

答えが簡単に求まらないとき数理的な研究が必要である。

(2) 重要な問題で f(x) を見つけたり、そのための定理を作ると

新しい研究になる。

(3) 問題の前提として、P(w) と P(x|w) が与えられていること

が必要である。その前提で、問題を考える。

(4) P(w) あるいは P(x|w) が不明な「未知の状況」では

決定理論はそのままでは適用できない。

「不確かな状況」≠「未知の状況」

決定理論における「不確かな状況」という言葉の意味は、確率はすべて既知であるということであり、「不確か」ではあっても「未知」ではない。

→ 「未知の状況」、すなわち P(w), P(x|w) が不明な場合には決定理論を適用することはできない。

「不確か」と「未知」の違いの例

「不確かな状況」の例

(1) w は平均0分散1の正規分布 $N(w|0,1)$ に従う。

(2) X は平均 w 分散1の正規分布 $N(x|w,1)$ に従う。

(3) $(w-f(X))^2$の平均値を最小にしたい。

「未知な状況」の例

(1) W の確率分布は不明である。

(2) X の確率分布は不明である。

(3) $(W-f(X))^2$の平均値を最小にしたい。

２　MIN MAX 法

$P(x|w)$ が既知で $P(w)$ が不明なとき

決定理論では　$P(w)$, $P(x|w)$ のどちらも既知であることが必要である。

　$P(x|w)$ が既知であり、w が集合 W に含まれているということだけが既知であるときには $P(w)$ の複数の候補を比較することや、w について損失が最大になる場合などが研究されている。

さらに例で説明します。

「不確かな状況」

（1） 真の q は区間 [0,1] のどこかにある。

（2） X は確率 q で1, (1−q)で 0 とする。

「MinMax意思決定」

(3) 最大リスク $\mathrm{Max}_{q}\, E_{x}(q-f(X))^2$ を最小にする

関数 f(X) を X に対する意思決定とする。

MinMAX 理論の設定

前ページのことを一般的に書きます。

(1) 未知の w は集合 W の中のどこかにある。

(2) X は既知の P(x|w) に従う確率変数である。

(3) 行動は決定関数 f(X) で定義される。

(3) 決定 f(X) の損失 L(f(X),w) が与えられている。

(4) リスクを $R(f,w) = \int L(f(x),w)\, P(x|w)\, dx$ と定める。

(5) 最大リスクを $\mathrm{Max}_w R(f,w)$ と定める。

例３（簡単に解ける場合）

(1) 未知の w は $[0,1]$ のどこかにある。

(2) (3) (4) 関数 $f(X)$ のリスクは例１と同じで

$$R(f,w) = (2f(0)-2f(1)+1)w^2 +(-f(0)^2 -2f(0)+f(1)^2)w +f(0)^2$$

(5)最適な $f(x)$は $0 \leqq f(0), f(1) \leqq 1$を満たす。$R(f,w)$ の w についての最大値は $\max_w R(f,w) \geqq R(f,1/2) \geqq 1/4$ である。

特に $f(0)=1/4$, $f(1)=3/4$ で等号が成立し、それ以外では2番目の等号は成立しない。したがってMAXリスクを最小にするのは $f(0)=1/4$, $f(1)=3/4$ である.

例4

(1) 未知のW は実数の集合のどこかにある。

(2) X は平均 w 分散1の既知の正規分布 $N(x|w,1)$ に従う。

(3) 関数 $f(x)$ の損失を $L(f(X),w) = (w-f(X))^2$ と定める。

(4) リスク $R(f,w) = \int (w-f(x))^2 \, N(x|w,1)\, dx$

$f(x)=x$ ならば w によらず $R(f,w)=1$ である。

$\mathrm{Max}_w R(f,w)$ を最小にする関数は $f(x)=x$ だろうか？

Min MAX 法の定理

<u>**定理**</u>。ある確率分布 Q(w) をひとつ固定し、その分布を用いて定義された平均リスク $\int\int L(f(x),w)\, P(x|w)\, dx\, Q(w)dw$ を最小にする決定関数を g(x) とする。
もしも リスク　$\int L(g(x),w)\, P(x|w)\, dx$　が w の定数関数であれば、その g(x) は min max 最適な関数である。

（注意）例３はこの定理の例であるが、例４はこの定理の例ではない。

Min Max 法　注意(1)

(1) Wの分布がわかっていないとき リスクの最大値を最小にする f を見つけることが研究されるときがある。Min Max 理論という。

$$\mathrm{Min}_f \ \mathrm{Max}_w \ R(f,w)$$

「ゲームの理論」にしばしば登場する設定である。

Min Max　法　注意(2)

(2) Min Max 法は P(w) が不明でも利用できる。

しかし「集合 W に w が含まれている」という仮定と

「モデル P(x|w) が正しい」という仮定は必要である。

Min Max 法の答えが、ある P(w) を用いた決定理論

の答えと一致している場合もある。

Min Max 法　注意(3)

(3) より多くの未知の状況に対応するために

W の集合を大きくして $P(x|w)$ の関数としての集合を大きくすると、$\mathrm{Max}_w\ R(f,w)$ が、それにつれて大きくなる。

３　決定理論とベイズ法

統計的推測は　確率分布を推定する。

確率変数であるサンプルに対して、その確率分布を推測することを統計的推測という。

確率変数は定義より　必ず確率分布を持つ。
最尤法でもベイズ法でも確率分布が推測される。

真の事前分布とモデルが既知のとき決定理論は　ベイズ法の基礎になる

決定理論をベイズ法に使うためには、

(1) 真の事前分布と真のモデルが既知である

(2) サンプルが確率変数である

の両方が必要である。

「真の事前分布」と「真のモデル」がどちらも

わかっているときに限り、決定理論が使える。

統計的推測における決定理論

<u>定理</u>。 真の事前分布と真のモデルが既知であるとする。KL情報量を損失とする決定理論では、最適な決定関数はベイズ予測分布である。

（注）真の事前分布も真のモデルも既知である（主観ではなく真である）と仮定できる場合も<u>ごく稀に</u>ある。

http://watanabe-www.math.dis.titech.ac.jp/users/swatanab/q_and_a1.pdf

ベイズ法を決定理論の枠組みで記述する

（簡単のためサンプルが独立な場合を述べる）

(1) W は事前分布 $\Phi(w)$ に従う確率変数。

(2) $X^n = (X_1, \cdots, X_n)$ は $\prod_i P(x_i|w)$ に従う確率変数。

(3) 損失関数はKL情報量

$$L(f(\cdot|X^n), w) = \int P(x|w) \log (P(X|w) / f(x|X^n)) \, dx$$

(4) リスクは $R(f,w) = \int L(f(\cdot|x^n), w) \prod_i P(x_i|w) \, dx_i$

(5) 平均リスクは $E[R(f,w)] = \int R(f,w) \Phi(w) dw$

決定理論の統計的推測への応用

定理. 前ページの(1)−(5) のもとで平均リスクを最小にするのはベイズ予測分布である。

(a) ベイズ事後分布の定義

$$P(w|X^n)=(1/Z_n)\ \Phi(w)\ \Pi\ P(X_i|w)$$

(b) ベイズ予測分布の定義

$$P(x|X^n)= \int P(x|w)\ P(w|X^n)\ dw.$$

Min Max 法のベイズ推測への応用

真のモデルがわかっていて、事前分布が不明であるとする。
このとき Min Max法の解が、ある事前分布を用いたベイズ法と一致する場合がある。

その場合の事前分布は、サンプルサイズに依存したり、積分が有限でない（プロパーでない）等になることがある。
（通常のベイズ法を拡張したものになっていることがある）。

事前分布かモデルが不明のとき
数学的理論がある

真のモデルあるいは真の事前分布が不明であるときは、候補として用いているモデルと事前分布について汎化損失や周辺尤度の挙動を解明するための数理がある。検定やモデル選択を行うための理論的な基盤がある。

４　主観ベイズ法

真の事前分布を主観で決める。
モデルは正しい。
サンプルは確率変数である。
⇒　決定理論が適用できる。

主観ベイズ法　とは

事前分布を主観で決めて、モデルを正しいとし、

サンプルを確率変数とする方法を **主観ベイズ法** という。

このとき　決定理論から　ベイズ予測分布は

主観のもとで KL 情報量を損失として最適である。

主観ベイズ法　の定義

主観ベイズ法では、個人または集団が主観で事前分布を真とし、モデルは正しいとする。主観ベイズ法では

（1）真の分布は主観分布Φ(w) に従う w を用いた

$\Pi\ P(X_i|w)$である。

(2) サンプルは上記の分布に従う独立な確率変数

あるいはその実現値である。

「主観ベイズ法は真の分布がなくても使える」は適切な説明ではない。

主観ベイズ法は、主観を真の事前分布とし、モデルは正しいとし、サンプルはそれに従う確率変数であると決めているので、未知の分布の想定を容認しない。

主観ベイズ法　での禁止事項

主観ベイズ法では、事前分布が主観により真であり
モデルは正しいと決められているので
データを得たのちに事前分布やモデルの変更を
行うことは禁止されている。

（注意）主観ベイズ法では、モデルも事前分布も、データを見る前に
決めておくことが必要である。そうしないと決定理論が適用できない。

主観ベイズでの　具体的禁止事項

主観ベイズ法では、以下の操作を許可しない。

○統計的検定。

○ハイパーパラメータの最適化。

○モデルのチェックや選択。

○汎化誤差を調べて小さくなるようにする。

（注）これらの機構を見かけ上は含むようにモデルを作ることもできるが
それはモデルと事前分布の集合を変えただけで実質は変わらない。

（例）無限混合モデルでハイパーパラメータ調節を行うと主観ベイズでない。

主観ベイズ法(1)　主観の抽出法

主観ベイズ法では、個人あるいは集団の主観を確率分布として取り出す手続きが用意されている。(ある人が賭けに応じるかどうかのバランス点を主観確率と定義する(※))。もしも主観が取り出せたとしたら、それは主観である。

この定義(※)が　主観確率の現実の設計に用いられることはない(？)

「この定義では主観を正しく表現できない(非加法的測度が必要)」という意見もある。賭けが嫌いな人は A にも A^C にも賭けたくないので非加法的 $P(A)+P(A^c)<1=P(A \cup A^c)$であるが主観ベイズでは非合理的であるとして許容しない。

主観ベイズ法(2)
真の事前分布、真のモデル、サンプルは確率変数

「主観ベイズ法」は、個人または集団が主観で決めた事前分布を真であると決め、モデルは正しいとし、サンプルをモデルに従う確率変数であると仮定している。

以上のいずれかを否定すると「主観ベイズ法」は最適性の論拠を失い「主観に基づく合理的な意思決定」ではなくなる。

（※）最尤法でもベイズ法でもサンプルは確率変数である。サンプルがいったん実現値として確定したあとでは、最尤法でもベイズ法でも、どんな方法でも、サンプルは定数である。「主観ベイズではサンプルは定数である」は適切な説明ではない。

５ ベイズ法

未知の状況では
決定理論が適用できない。

そのときの数理がある。

未知で不確かな状況

一般の　統計的推測では、データを発生している分布について何もわかりません（未知の状況）。

事前分布もモデルもわからないので、決定理論を適用することはできません。その場合の数理があります。

一般のベイズ法の理論

未知の状況で　ベイズ法を用いるときには

真の分布は未知であり、サンプルは未知の分布に
従う確率変数であるという仮定の上で成立する
理論が存在しています。

未知で不確かな状況

未知で不確かな状況下では次のようになります。

(1) 未知の分布を $Q(x)$ と表記する。

(2) $X^n = (X_1, \ldots, X_n)$ は $\Pi_i\, Q(x_i)$ に従う確率変数。

(3) 損失関数はKL情報量

$$L(f(\cdot|X^n)) = \int Q(x) \log (Q(x) / f(x|X^n)) \, dx$$

(4) リスクは $R(f) = \int L(f(\cdot|x^n)) \, \Pi_i\, Q(x_i)\, dx_i$

☆ $Q(x)$が未知でも $R(f)$ を推測できる数理がある。

未知で不確かな状況での推測

未知で不確かな状況下でのベイズ法

(1) 候補事前分布 $\varphi(w)$ と候補モデル $p(x|w)$ を準備。

(2) 候補からベイズ予測分布 $f(x|X^n)$ を定義する。

(3) $R(f)$ は、交差検証、情報量規準で推測できる。

周辺尤度などの評価法もある。検定もできる。

汎化誤差の計算を行える。チェックと確認をする。

(4) 候補の事前分布と候補のモデルを様々な観点から

比較して未知の分布に対して適切なものを選択。

未知の状況でのベイズ法

未知の状況でのベイズ法では、事前分布も
モデルも　検証されるべき候補である。
予測分布が未知の分布に対して適切であるか
どうかを調べるための数理的な基盤が作られ
ている（決定理論ではない）。

未知の状況に「ベイズ更新」は到達しない。

未知の分布に対して　モデルが適切でない場合

どんなにデータが増えても推測結果は未知の分布に

近づかない。「ベイズ更新」では未知に到達できない。

未知の状況において　モデルと事前分布の適切さを

調べ改善するための数理が作られている。

⇒　多くの統計モデリングをサポート。

統計モデルとは

候補としてあらゆるモデルと事前分布を準備することは難しい。実際には、データが該当する領域（人文社会科学、自然科学など）の知識・経験を活用して候補を用意する。それらも一種の「主観」である（最尤法でも必要になる）が、主観ベイズ法とは異なり、それらは仮説あるいは候補であり正しいと信じるものではなく、選択・調節・検定の対象である。機械学習では、ほぼ必ず汎化誤差が計算され比較される。適切なモデルはサンプルサイズにも依存する。多くの統計学者は次のように考えていると思われる。
「すべてのモデルは正しくないが、中には役立つものもある」。

未知の状況のベイズ法は　広く使われている。

未知の状況でのベイズ法は、広く利用されています。

(1) Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson,
Aki Vehtari, Donald B. Rubin. Bayesian Data Analysis, CRC Press, 2013.
(2) N. Thompson Hobbs, Mevin B. Hooten. Bayesian Models:
A Statistical Primer for Ecologists,Princeton University Press, 2015.
(3) Franzi Korner-Nievergelt, Tobias Roth, Stefanie von Felten, Jerome Guelat,
Bettina Almasi, Pius Korner-Nievergelt. Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and Stan, Academic Press, 2015.
(4) Richard McElreath. Statistical Rethinking: A Bayesian Course with Examples in R and Stan,CRC Press, 2016.
(5) Osvaldo Martin. Bayesian Analysis with Python, Packt Publishing Ltd, 2016.
(6) Ben Lambert. A Student’s Guide to Bayesian Statistics, SAGE, 2018.
(7) Xiaofeng Wang, Yu Yue Ryan, Julian J. Faraway. Bayesian Regression Modeling with INLA, CRC Press, 2018.
(8) Peter D. Congdon. Bayesian Hierarchical Models, CRC Press, 2019.
(9) M. Antonia Amaral Turkman, Carlos Daniel Paulino, Peter Muller.
Computational Bayesian Statistics, Cambridge University Press, 2019.
(10) Brian J. Reich, Sujit K.Ghosh. Bayesian Statistical Methods, CRC Press, 2019.

未知の状況で　ベイズ法が使われている。

未知の状況でのベイズ法は、新型コロナウイルスの
研究にも使われています。

http://watanabe-www.math.dis.titech.ac.jp/users/swatanab/index-j3.html

６　まとめ

回答1

Q.　決定理論とは何ですか。

A.　確率分布が既知で、行動に対する損失が与えられているとき、平均損失を最小にする行動を選ぶための理論です。

回答2

Q.　主観ベイズ法は決定理論に基づいていますか。

A. はい。主観ベイズ法は、真の事前分布を主観で決めて、モデルを正しいとし、サンプルを確率変数と仮定し、決定理論に基づいて行動を定めます。
<u>主観で事前分布を決め、モデルは正しく、サンプルは確率変数であるとした人の意思決定をサポートします。</u>

回答3

Q.　一般のベイズ法は決定理論に基づいていますか。

A. いいえ。サンプルの分布が未知である場合のベイズ法は決定理論ではない数理に基づいています。事前分布も統計モデルも　候補であり、検定、選択、確認、調節を行うことができます。

未知の分布の統計的推測をサポートします。

まとめ　決定理論を正しく理解しよう。

決定理論は「不確かな状況」での意思決定を与えるものですが、「未知の状況」を対象としていません。未知の状況では決定理論ではない数理が基盤になります。

主観ベイズ法＝主観が真である＋決定理論

一般のベイズ法＝未知の分布を推測する

もっといろいろと調べたい人に

◆ 「未知の環境」での行動決定に関心があるかたは「強化学習」についても調べてみましょう。

◆ 人間の主観そのものの表現のために確率測度の一般化を考えたい人は「非加法的測度」を調べてみましょう（主観ベイズ法ではありません）。