

因果推論

2016 年夏期集中講義資料（作成者：大塚 淳）

1 確率の復習

確率変数の条件付確率と独立・従属

確率変数 X と Y のすべての値 x, y について以下がなりたつとき、 X と Y を独立という。

$$P(x|y) = P(x)$$

ただし $P(x|y) := P(x, y)/P(y)$ は $Y = y$ のもとでの $X = x$ の条件付確率である。独立でないとき、両者は従属しているという。従属であるとは、両変数の値が伴って変化しているということであり、したがって片方からもう一方を予測することが可能になる。

条件付独立

三つ以上の変数の条件付確率も考えることができる。例えば、 $Y = y$ かつ $Z = z$ のもとで $X = x$ である確率は、以下のようになる

$$P(x|y, z) = \frac{P(x, y, z)}{P(y, z)}$$

例えば X, Y, Z をそれぞれ身長、年齢、体重とすると、 $P(X|y = 20, z = 70)$ は 20 歳で 70kg ある人の身長の分布である。同様にその期待値 $E(X|y = 20, z = 70)$ は同齢同体重の人の平均身長となる。

X, Y, Z のすべての値 x, y, z について以下がなりたつとき、 X と Z は Y のもとで条件付独立 (conditionally independent) になるという。

$$P(x|y, z) = p(x|y)$$

またこれを $X \perp_P Z|Y$ と書く。二変数の独立性のときと同様、これは $P(x, z|y) = P(x|y)P(z|y)$ と同値である（確かめてみよ）。

条件付独立は、ある情報を得ることによって、二つの変数間の関係性が消失することを表している。例えば、もし身長と体重が年齢のもとで条件付独立になるのだとしたら、ある人の身長を予測したいときに、その人の年齢さえ知ってしまえば、もはやその人の体重について知ることは身長予測にとっては必要ない、ということになる。

周辺化

任意の確率変数 X, Y, \dots について

$$\sum_y P(X, y, \dots) = P(X, \dots)$$

が成り立つ。つまり変数の全ての値の確率を足し合わせることによって、その変数を消去することができる。これを周辺化といい、周辺化によって得られる確率を周辺分布 (marginal distribution) という。これは例えば、 X を身長、 Y を年齢としたとき、身長の分布が全年齢におけるその分布を足し合わせて得られることから直感的に理解できるであろう。

2 処置効果

We may define a cause to be an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second. Or in other words where, if the first object had not been, the second never had existed.

— Hume, *An Enquiry Concerning Human Understanding*

X を処置、 Y を関心のある結果とする。このとき、次のような潜在変数を導入する

- Y : 処置を受けた ($X = 1$) のときの結果
- Y' : 処置を受けなかった ($X = 0$) のときの結果

これが「潜在」変数と呼ばれるのは、各被験者につきこのどちらか一方しか観測できないからである。もし i さんが処置を受けたならば、 y_i は観測できるが y'_i はできない。逆に受けなかったならば、 y'_i は得られるが y_i は不可知である。よって「もし i さんが処置を受けたとしたら・・・」という反事実仮想は、 y'_i という観測された事実をもとに、実現しなかった y_i について思いを巡らせているわけである。

我々が関心あるのは、 X の効果、つまり平均して処置 X の有無は Y にどれだけの違いをもたらすのか、ということである。これは平均処置効果 (average treatment effect)

$$E(Y) - E(Y')$$

で定義される。しかしこれも直接推定することはできない。平均処置効果の最も素直な推定量は、被験者全体における Y と Y' の差の平均

$$\bar{Y} - \bar{Y}' = \frac{1}{n} \sum_i^n (y_i - y'_i)$$

だろう。しかし上述のように同時に処置を受けかつ受けないことはできないので、それぞれの被験者 i について y_i, y'_i のうち片方は必ず欠測しており、よって上の推定量は計算できないのである。この制約に対処するアプローチとして、無作為化と傾向スコアがあるが、授業では前者を解説する。

3 因果モデル

V を確率変数の集合とする。 V の要素をノードとして、それらを矢印で結んだグラフを因果グラフ G^{*1} という。ここで矢印は変数間の因果関係を表す。例えば $X \rightarrow Y$ は X が Y の直接の原因 (ないし「親」) であることを意味している。 X の全ての親を $PA(X)$ で表す。また X から Y に矢印を順に伝って行けると、 X は Y の (間接的な) 原因であり、 Y は X の (間接的な) 結果である。なおグラフは非巡回、すなわちいかなる変数もそれ自身の原因ではないとする。^{*2}

マルコフ条件

因果グラフで結ばれた V について、さらにその確率分布 $P(V)$ を考える。因果グラフ理論では、このような因果的構造と確率分布の間に、以下のような対応が成立すると主張される

$$\text{任意の確率変数 } V \perp_P V \text{ の結果でない全ての変数 } | PA(V)$$

つまり、変数の直接原因 (「親」) は、その変数を、その結果以外の全ての変数から無関係にする。よってある変数の状態を知りたいとき、その直接の原因となっている変数の状態を知ってしまえば、結果以外の変数の状況を知っている必要はなくなる。これをマルコフ条件 (Markov condition) という。

マルコフ条件は、次の条件と同値である。

$$P(V) = \prod_{V \in V} P(V|PA(V))$$

つまり、 V 全体の確率 (同時確率分布) は、「各要素の親のもとでの条件付確率」を全変数にわたって掛け合わせたものに等しくなる。ここから、全体の確率 $P(V)$ を得るためには、まず最下層の変数の親のもとでの条件付確率を求め、その親の親の条件付確率を掛け合わせ、さらにそのまた親の・・・と繰り返していき、最終的に最も上位の親の無条件確率を掛け合わせれば良いことになる。これを分解 (factorization) という。

介入

変数 X に介入してその値を \tilde{x} としたときの Y の確率分布を、 $P(Y|do(X = \tilde{x}))$ と表す。一般にこれは、変数 X が x であると観測されたときの条件付分布 $P(Y|X = x)$ とは異なることに注意せよ。例えば X の値を「道路が濡れている (1)・いない (0)」、 Y を「雨が降る (1)・降らない (0)」としたら、明らかに $P(Y|X = 1) \neq P(Y|X = 0)$ だが、 $P(Y|do(X = 1)) = P(Y|do(X = 0))$ 、つまり道路に水をまいても降水確率は変わらないだろう。ここから、介入の結果 $P(\bullet|do(\bullet))$ を予測するためには、確率分布以外の道具立てが必要になることが分かる。

^{*1} より厳密には、因果グラフはノードの集合 V と矢印の集合 $E \subset V \times V$ の組として定義される。

^{*2} これを Direct Acyclic Graph あるいは DAG という。

介入結果を計算する方法として、Pearl の do 計算 (do calculus) あるいは Spirtes らの操作定理 (manipulation theorem) がある。これらは以下の手順で介入結果を予測する。なお前提として、因果グラフ \mathcal{G} とマルコフ条件を満たす確率分布 $P(\mathbf{V})$ が与えられているとする。

1. 介入対象の変数 (ここでは X とする) に流れ込む矢印をすべて \mathcal{G} から削除し、新しいグラフを得る。
2. 新しいグラフにもとづき、確率分布 $P(\mathbf{V})$ を $\prod P(V|PA(V))$ へと分解する。このとき介入により X は原因を持たないので、最も上位に $P(X)$ の形で現れるはずである。
3. X の分布を介入に合わせて設定する。例えば $do(X = 1)$ であれば、 $P(X)$ は $X = 1$ のときだけ 1、それ以外は 0 となるようにする。
4. 3 で得られた介入後の X の分布を 2 の分解に代入して、介入後の確率分布 $P(\mathbf{V}|do(X = 1))$ を得る。

授業では簡単な例を用いてこれを解説する。

4 参考文献

- Take a Risk : 林岳彦の研究メモ、なぜ無作為化なのか : 『因果推論の根本問題』とその解法
<http://takehiko-i-hayashi.hatenablog.com/entry/2013/11/21/221514>
- 大塚淳. (2010). ベイズネットから見た因果と確率. 科学基礎論研究, 38(1), 39-47.
https://junotkja.files.wordpress.com/2015/10/kisoron_otsuka2010.pdf