***Wnioskowanie bayesowskie***

***wprowadzenie w miarę bezbolesne :)***

Łukasz Kwiatkowski

Katedra Ekonometrii i Badań Operacyjnych

**Plan wykładu**

1. Prawdopodobieństwo – *revisited*
2. Dwa podejścia do wnioskowania statystycznego
3. Statystyka bayesowska – wyróżniki
4. Wnioskowanie bayesowskie – to „lepsiejsze”
5. Bayesowski model statystyczny – krok po kroku

**Prawdopodobieństwo – *revisited***

* Pytanie 1: Jakie jest prawdopodobieństwo, że jutro między 7:00 a 9:00 spadnie deszcz?

**Prawdopodobieństwo – *revisited***

* Pytanie 2: Mam w ręce monetę 5 zł. Jakie są szanse, że jeśli nią rzucę, to wypadnie orzeł?

**Prawdopodobieństwo – *revisited***

* Interpretacje prawdopodobieństwa:
	+ „**klasyczna**” – prawdopodobieństwo „obiektywne”, częstościowe (definicja Andrieja Kołmogorowa, doprecyzowująca tę zaproponowaną wcześniej przez Richarda von Misesa), rozumiane jako **idealizacja częstości względnej**; zwane także „**fizycznym**”, ale p-stwo **NIE jest jakąś cechą fizyczną**, immanentnie związaną z i określającą jakiś „obiekt” (np. monetę, urnę czy kurs walutowy), zdarzenie czy zdanie logiczne (hipotezę)
	+ **bayesowska** – prawdopodobieństwo jest miarą subiektywnego stopnia przekonania/pewności co do zajścia jakiegoś zdarzenia (losowego, tj. obarczonego niepewnością) czy prawdziwości pewnej hipotezy

🡪Prawdopodobieństwo jest jak … – każdy ma swoje ;) ALE nie oznacza to jakiegoś dogmatyzmu (*Moje prawdopodobieństwo jest „najmojsze”, i za Chiny Ludowe go nie zmienię!*)

🡪 P-stwo ma 2 swoje źródła: wychodzimy od **p-stwa *a priori*** (czyli przed doświadczeniem, eksperymentem, przed wglądem w dane, w empirię), a następnie modyfikujemy je wraz z napływającymi danymi = aktualizacja wiedzy („mojej” wiedzy, rozumianej tu jako „moje” wyobrażenie o kształtowaniu się niepewności odnośnie danego zdarzenia/hipotezy itp.)

🡪 Statystyka bayesowska oferuje w pełni formalny, zgodny z podstawowymi zasadami rachunku prawdopodobieństwa (wychodząc od definicji wg. Pierre Simone de Laplace), wewnętrznie spójny (koherentny) – choć też nie pozbawiony pewnych paradoksów – sposób modyfikacji mojej wiedzy *a priori* w *a posteriori* (tj. po doświadczeniu, po wglądzie w dane/obserwacje), czyli sposób wyznaczania „mojej” końcowej wiedzy, poprzez połączenie wiedzy *a priori* z informacją płynącą z obserwacji (co jest zgodne z neurofizjologią procesu uczenia się :)

**Prawdopodobieństwo – *revisited***

* Powtórzmy:
	+ P-stwo jest matematycznym sposobem wyrażania „mojego” (zatem subiektywnego) stopnia przekonania (a tym samym stopnia *niepewności*)

🡪 Na gruncie statystyki bayesowskiej prawdopodobieństwo ma interpretację subiektywistyczną

* + P-stwo ulega modyfikacji (w wyniku procesu „uczenia się”, aktualizacji wiedzy) – od wiedzy/informacji wstępnej (tzw. *a priori*, czyli przed wglądem w dane/obserwacje) do wiedzy końcowej (tj. *a posteriori*, czyli uwzględnieniu informacji zawartej w danych; ang. *evidence*)

🡪 W modelowaniu parametrycznym (np. regresji) owa „wiedza”/niepewność dotyczyć będzie *parametrów modelu*

* + Matematyczny dowód istnienia i jednoznaczności p-stwa jako miary subiektywnego stopnia przekonania nie jest trywialny, ALE JEST :) (XX w.: Morris H. DeGroot, Bruno de Finetti, Leonard Savage)

**Dwa podejścia do wnioskowania statystycznego**

* Dwa wiodące – istnieją jeszcze inne koncepcje…
* Te dwie wzmiankowane w tym miejscu bazują, odpowiednio, na „klasycznej” i bayesowskiej interpretacji prawdopodobieństwa:
	+ **Statystyka „klasyczna”** – inaczej: niebayesowska, częstościowa (ang. *frequentist statistics/inference, frequentist approach to statistics*):
		- Jerzy Spława Neyman (1894-1984) – przedziały ufności
		- Ronald Fisher (1890-1962) – MNW
	+ **Statystyka bayesowska** (ang. *Bayesian statistics/inference, Bayesian approach to statistics*):
		- **Thomas** **Bayes** (1702-1760) – angielski matematyk i duchowny; twórcą jako tako nie jest, ale autorem sławetnego **wzoru Bayesa** (na p-stwo warunkowe; opublikowanego pośmiertnie):
		- Pierre Simone de Laplace (1749-1827) – pierwsza formalna definicja p-stwa

🡪 Chronologicznie podejście bayesowskie jest bardziej „klasyczne” niż podejście „klasyczne” ;)

🡪 Odnowiciele bayesizmu: Harold Jeffreys, Bruno de Finetti, Leonard Savage, Arnold Zellner

**Statystyka bayesowska – wyróżniki**

* Podejście bayesowskie do statystyki odróżnia się od ujęcia wnioskowania „klasycznego” w co najmniej kilku aspektach, wśród których za fundamentalne można uznać:
	+ Od strony epistemologii (czyli filozofii poznania): subiektywną – w odróżnieniu od częstościowej – interpretację p-stwa

🡪 p-stwo jako miara „mojego” stopnia niepewności – najbardziej adekwatną metodą do wyrazu/ujęcia tej niepewności okazują się rozkłady prawdopodobieństwa

* + Od strony „technicznej”: wszystkie elementy modelu statystycznego (tj. zarówno dane, jak i – UWAGA – parametry!) są zmiennymi losowymi

🡪 dla tych, których nie dowierzają w to, co słyszą powtórzmy: na gruncie bayesowskim PARAMETRY są ZMIENNYMI LOSOWYMI

🡪 przypomnijmy: na gruncie „klasycznym” parametry są… (?)

* Analogia pomiędzy parametrami (bayesowskiego) modelu statystycznego a położeniem elektronu (fizyka kwantowa)
* Tak więc na gruncie bayesowskim parametry modelu mają swoje… rozkłady prawdopodobieństwa

🡪 Nie ma więc potrzeby konstruowania czegoś takiego jak estymator (=funkcja danych, która ma „ustrzelić” nieznaną, stałą wartość parametru), choć na gruncie bayesowskim także (o ile chcemy) możemy wyznaczać oceny punktowe parametrów (czyli punktowe wskazania jego położenia) – jako wartość oczekiwana/mediana/modalna *rozkładu* – ale już nie estymatora, tylko bezpośrednio parametru (wydaje się bardziej naturalne :)

**Wnioskowanie bayesowskie – to „lepsiejsze” ;)**

* Zasadniczo, statystyka bayesowska:
	+ daje po prostu pełny obraz kształtowania się **niepewności** co do możliwych wartości parametrów (poprzez bezpośrednio *ich* – a nie estymatora – rozkłady prawdopodobieństwa), a o to właśnie chodzi w statystyce (🡨 nauka o niepewności, a nie – by tak rzec – o ocenach punktowych)
	+ wewnętrznie spójne (najbardziej ze wszystkich podejść do wnioskowania statystycznego, choć w dalszym ciągu nie jest wolne od pewnych paradoksów)
	+ bardzo wymierna i już teraz namacalna dla Państwa korzyść z WB: w końcu naturalnie brzmiące interpretacje przedziałów „ufności” (zwanych w statystyce bayesowskiej *przedziałami wiarygodności*; ang. *credible intervals*)
* Choć „lepsze”, to jednak jego zwolennikom/twórcom/propagatorom nie szczędzono batów na przestrzeni dziejów. Te przeplatająco-ścierające się dzieje statystyki bayesowskiej i „klasycznej” to istna „Moda na sukces”...

🡪 Sharon B. McGrayne, (2011), *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*, Yale University Press

**Bayesowski model statystyczny – krok po kroku**

* Patrz materiał „***Podstawowe pojęcia***.pdf” (Moodle, w sekcji „Materiały pomocnicze”)
* We wnioskowaniu statystycznym (każdym parametrycznym, nie tylko bayesowskim) mamy zasadniczo dwa obiekty: dostępne **dane** (obserwacje) i **parametry** (pomijamy prognozowanie, gdzie mamy jeszcze *obserwacje niedostępne*):
	+ **Dane** – **wektor obserwacji dostępnych**:

- dla zakładamy, że zostały „wygenerowane” zgodnie z pewnymi założeniami (np. KMNRL), przyjmując dla nich tzw. ***model próbkowy*** (ang. *sampling model*) albo inaczej ***rozkład próbkowy***, czyli rozkład obserwacji, przy założeniu konkretnych (choć o nieznanych wartościach) parametrów, – przykładowo w KMNRL:

- – **notacja rodzajowa**: funkcja gęstości / masy rozkładu prawdopodobieństwa zmiennej

- ten „mechanizm powstawania obserwacji” – jeśli spojrzeć na niego tak, jakby argumentem były parametry, a nie obserwacje – to zarazem **funkcja wiarygodności** (ang. *likelihood function*), :

- Uwaga: jako wektor losowy (wektorowa zmienna losowa, wektor zmiennych losowych) ORAZ jako *realizacja* tego wektora (notacyjnie tego tu nie rozróżniamy)

**Bayesowski model statystyczny – krok po kroku**

* Patrz materiał „***Podstawowe pojęcia***.pdf” (Moodle, w sekcji „Materiały pomocnicze”)
* We wnioskowaniu statystycznym (każdym parametrycznym, nie tylko bayesowskim) mamy zasadniczo dwa obiekty: dostępne **dane** (obserwacje) i **parametry** (pomijamy prognozowanie, gdzie mamy jeszcze *obserwacje niedostępne*):
	+ **Parametry** – **wektor** (wszystkich) **parametrów** danego modelu:

- przypomnijmy, że „tu”: – wektor losowy (a nie nieznanych stałych, jak w ujęciu częstościowym)