

Paweł Matuszewski

Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie

John K. Kruschke, *Doing Bayesian Data Analysis: A Tutorial with R, Jags, and Stan*, Academic Press, San Diego, CA 2014, Kindle edition, ss. 776.

Od pewnego czasu w naukach społecznych rośnie zainteresowanie statystyką bayesowską i metodami symulacji Monte Carlo. W konsekwencji powstaje coraz silniejsze zapotrzebowanie na podręczniki, które przybliżyłyby te zagadnienia. Jedną z takich propozycji jest książka Johna K. Kruschke pt. *Doing Bayesian Data Analysis: A Tutorial with R, Jags, and Stan*.

Warto zwrócić uwagę, że opanowanie metod bayesowskich może być dla polskich naukowców prowadzących badania w ramach nauk humanistycznych i społecznych przedsięwzięciem dużo bardziej wymagającym niż zrozumienie klasycznej analizy statystycznej (tzw. Fisherowskiej). Związanych jest z tym kilka kwestii. Po pierwsze, kursy uniwersyteckie, które dotyczyłyby statystyki bayesowskiej i byłyby prowadzone na kierunkach humanistycznych lub społecznych, należą do rzadkości. Badacze często zmuszeni są do samodzielnych poszukiwań i trudno im uzyskać prawdziwie merytoryczne wsparcie od osób z większym doświadczeniem. Po drugie, autorowi nie są znane publikacje w języku polskim poświęcone statystyce bayesowskiej, które byłyby skierowane do socjologów, politologów, psychologów i przedstawicieli pokrewnych nauk. Z konieczności muszą oni poszukiwać odpowiedzi na swoje pytania w literaturze obcojęzycznej. Jest to też główny powód, dlaczego recenzowany jest podręcznik wydany w języku angielskim, a nie polskim. Po trzecie, właściwie nie istnieje dedykowane i proste w obsłudze oprogramowanie do analizy bayesowskiej. Najczęściej jest ona przeprowadzana w środowisku R przy użyciu pakietów rbugs (Bayesian Inference Using Gibbs Sampling), rjags (Just Another Gibbs Sampler) lub rstan. W związku z tym badacz, który chce opierać się na tych analizach, musi posiadać przynajmniej podstawowe umiejętności programowania (poza językiem R, często wykorzystuje się też język Python). Po czwarte, zrozumienie metod bayesowskich wymaga opanowania zagadnień matematycznych wykraczających poza poziom szkoły średniej, a więc poziom większości studentów i absolwentów kończących kierunki humanistyczne i społeczne (np. w sytuacji, gdy używane są zmienne ilościowe, niezbędna jest znajomość podstaw rachunku różniczkowego i całkowego). Autor, który podejmuje się napisać podręcznik wprowadzający do metod bayesowskich musi być świadomy powyższych przeszkód oraz wychodzić im naprzeciw. Recenzja została napisana z perspektywy badacza, który potrafi biegle czytać w języku angielskim, posiada dość zaawansowaną znajomość metod statystycznych i podstawowe umiejętności programowania.

W pierwszej kolejności warto omówić zakres wiedzy, który dostarcza książka Johna K. Kruschke. Podręcznik liczy 776 stron i jest podzielony na trzy części. W pierwszej z nich autor wyjaśnia podstawowe zagadnienia dotyczące analizy bayesowskiej i rachunku prawdopodobieństwa oraz dokonuje krótkiego wprowadzenia do języka R. Zgodnie z deklaracją zamieszczoną na samym początku książki („*This book explain show to actually do Bayesian data analysis, by real people (like you), for realistic data (like yours). The book starts at the basics, with elementary notions of probability and programming*”, [s. 1]), wszystkie zagadnienia wyjaśniane są bardzo dokładnie. Wydaje się, że nawet początkujący student nie powinien mieć problemu ze zrozumieniem, czym jest np. funkcja gęstości prawdopodobieństwa i jaka jest rola całki Lebesgue’a w wyznaczaniu prawdopodobieństwa wystąpienia danego zdarzenia. Należy jednak wyraźnie rozgraniczyć

zrozumienie, czym jest przykładowa funkcja i do czego służy od zrozumienia, jak ją obliczać. Tego autor nie wyjaśnia i prawdopodobnie jest to zabieg w pełni świadomy. Można powiedzieć, że Kruschke przekazuje wystarczająco dużo informacji, aby badacz wiedział, co i po co liczy. Umiejętność dokonywania szczegółowych obliczeń (np. całek) nie jest z kolei wymagana, ponieważ kwestią tą zajmuje się oprogramowanie komputerowe.

Druga część książki stanowi łagodne wprowadzenie do wnioskowania opartego na teoremacie Bayesa. Autor pokazuje w bardzo przejrzysty sposób, odwołując się do przykładów, jak zastosować metody bayesowskie, gdy posiadamy dane typu logicznego (boolowskiego). Równolegle wprowadzane są bardziej szczegółowe zagadnienia z zakresu analizy statystycznej. Autor wyjaśnia, na czym polega symulacja Monte Carlo, opisuje algorytm Metropolis, próbnik Gibbsa oraz zastosowanie czynnika Bayesa. Szczególnie warte odnotowania są przekonujące porównania metod bayesowskich z tradycyjnymi testami opartymi na statystyce Fisherowskiej. Pod koniec tej części znajduje się dodatkowo wprowadzenie do pakietu rstan.

Ostatnia część książki poświęcona jest bayesowskim odpowiednikom różnego rodzaju „tradycyjnych” testów. Kruschke pokazuje, w jaki sposób przeprowadzić test t , analizę regresji liniowej (prostą i wielokrotną), jednoczynnikową i wieloczynnikową analizę wariancji, analizę regresji logistycznej, test chi-kwadrat oraz jak obliczyć wielkość próby. W porównaniu do dwóch poprzednich części tym razem od czytelnika w znacznie większym stopniu wymaga się posiadania dość dużej wiedzy wstępnej. Główny akcent położony jest bowiem na pokazanie, jak robić tego typu analizy za pomocą metod bayesowskich a nie na to, na czym dokładnie te techniki polegają. Wydaje mi się, że czytelnik, który wcześniej nie spotkał się np. z analizą regresji liniowej, może mieć problem z jej zrozumieniem. Z kolei badacze posiadający dużą wiedzę i doświadczenie w tym zakresie uzyskają wszystkie niezbędne informacje, aby je z powodzeniem stosować w ujęciu bayesowskim.

Przed podręcznikiem do statystyki stoi zadanie nie tylko przekazania wiedzy, ale i wykształcenia umiejętności wykorzystania tej wiedzy w praktyce. Książka Johna Kruschke ma pod tym względem wiele do zaoferowania. Po pierwsze, moim zdaniem, jedną z ważniejszych zalet jest załączony plik z gotowymi schematami, które umożliwiają przeprowadzanie analiz za pomocą pakietu rjags lub rstan. Nie trzeba zatem od samego początku tworzyć programów (co może być poważną barierą dla kogoś, kto nigdy wcześniej nie programował), a jedynie dostosowywać gotowe skrypty do własnych potrzeb. Po drugie, na końcu każdego rozdziału są pytania i zadania (odpowiedzi można znaleźć w Internecie). Na uwagę zasługuje to, że z jednej strony, nawiązują one do omawianych treści, a z drugiej, wymagają od czytelnika samodzielnego myślenia i stanowią wyzwanie, dzięki któremu można utrwalić sobie dany materiał. Ważny też jest dobrze dostosowany poziom trudności tych zadań, ponieważ zarówno zbyt niski, jak i zbyt wysoki może działać demobilizująco. W tym przypadku można powiedzieć, że każde zadanie jest możliwe do rozwiązania w oparciu o treści znajdujące się w książce, choć są oczywiście zadania łatwe i trudniejsze. Po trzecie, autor prowadzi internetowego bloga, do którego nie raz się odwołuje. Daje to szansę pogłębiania przedstawionych w książce zagadnień. Na koniec warto wspomnieć o sposobie przekazywania treści. Styl pisania Kruschke’go jest bardzo bezpośredni i przejrzysty, dzięki czemu dość łatwo jest zrozumieć nawet bardzo trudne i nowe zagadnienia. Dodatkowo autor posługuje się prostymi przykładami, co pozwala od razu stworzyć powiązania między materiałem teoretycznym a tym, jak go wykorzystać w praktyce.

Książka jest niewątpliwie bardzo dobrym podręcznikiem wprowadzającym do metod bayesowskich. John Kruschke porusza wszystkie najważniejsze zagadnienia i dostarcza

czytelnikowi niezbędnych informacji, aby mógł samodzielnie przeprowadzać swoje własne analizy. Książka ma przede wszystkim wymiar praktyczny w takim sensie, że pokazuje, jak stosować metody bayesowskie i jak interpretować wyniki. Natomiast może ona nie spełnić oczekiwań tych badaczy, którzy chcieliby zrozumieć matematyczne podstawy metod bayesowskich. Autor w znacznym stopniu redukuje te wątki do niezbędnego minimum, a obliczenia wykonują wcześniej przygotowane skrypty. Oczywiście, nie należy tego traktować w kategoriach zalet lub wad. Zależy to od potrzeb czytelnika.