

MWS Ćwiczenie 3

Estymacja bayesowska

Tomasz Korzeniowski, 265753

8 grudnia 2017

1 Zadanie 1

Rzucona pinezka upada ostrzem do dołu lub do góry. Eksperyment polegający na tym, że pinezka spadnie ostrzem go góry można opisać schematem Bernoulliego. Rozkładem parametru p tego schematu jest $Beta(\alpha, \beta)$. Nie chcemy faworyzować strony na jaką spadnie pinezka, zatem potrzebujemy rozkładu, który będzie rozłożony dość równomiernie. Przyjmiemy rozkład $Beta(2, 2)$, który to będzie rozkładem a priori. Do przeprowadzenia eksperymentu wykorzystamy poniższą próbę losową powstałą z dwóch tur dwudziestu kolejnych rzutów pinezką.

	x_i																			
Pierwsza tura	0	1	0	0	1	1	1	1	0	1	0	1	1	0	0	1	0	1	1	1
Druga tura	1	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	1	0	1	0

Tabela 1: Wartości rzutów pinezką w kolejnych turach.

$$x_i = \begin{cases} 1, & \text{pinezka spadła ostrzem do góry} \\ 0, & \text{w p.p.} \end{cases}$$

Po wykonaniu pierwszej tury dwudziestu rzutów wyznaczmy rozkład a posteriori $Beta(\alpha_{post}, \beta_{post})$ oraz estymator parametru \hat{p} . Następnie powtórzmy eksperyment i wyznaczmy te same statystyki dla obu tur rzutów pinezką.

$$\alpha_{post} = \alpha + \sum_{i=1}^N x_i \quad (1)$$

$$\beta_{post} = \beta + n - \sum_{i=1}^N x_i \quad (2)$$

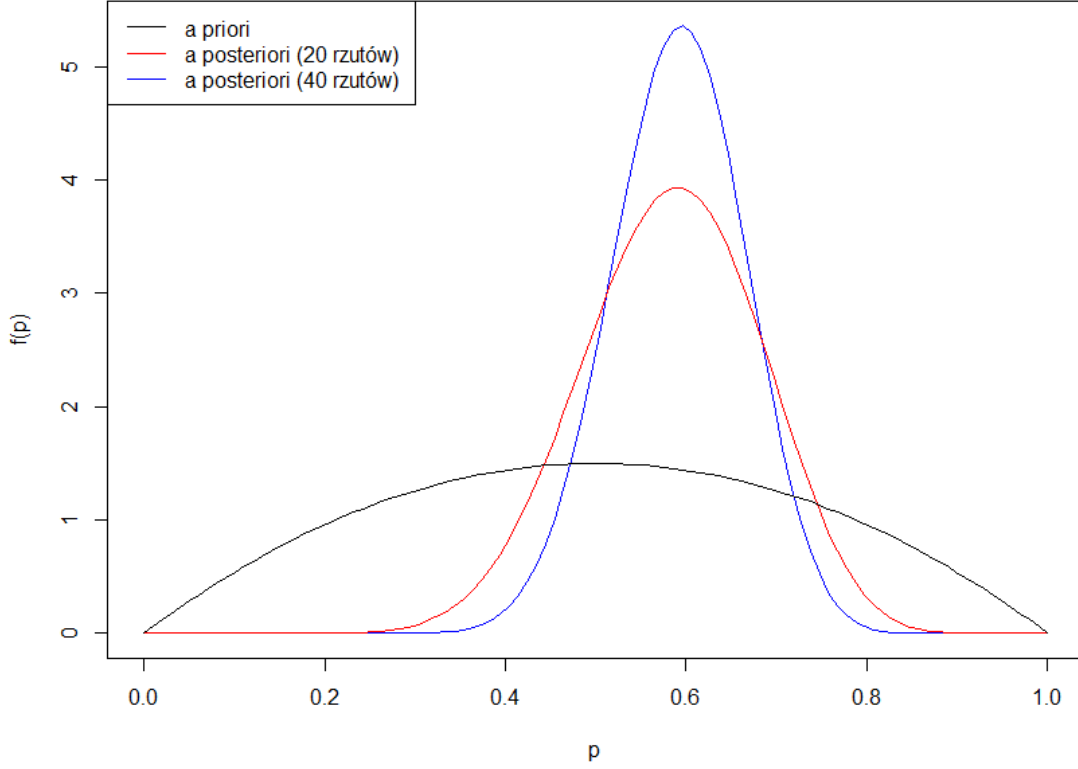
$$\hat{p} = \frac{\alpha_{post}}{\alpha_{post} + \beta_{post}} \quad (3)$$

Wyznaczone wartości przedstawia tabela 2.

	po 20 rzutach	po 40 rzutach
α_{post}	14	26
β_{post}	10	18
\hat{p}	0.5833333	0.5909091

Tabela 2: Wartości parametrów rozkładu a posteriori i \hat{p} .

Do wyznaczenia rozkładu $Beta$ została wykorzystana funkcja $dbeta$ z parametrami α_{post} i β_{post} wyznaczonymi powyżej. Na jej podstawie powstał wykres 1. Widać, że początkowy dobór rozkładu a priori nie odgrywa zbyt dużej roli. Jest on ważny w sytuacji gdy mamy niewielką próbę losową, lecz jego pierwotne parametry nie wpływają na ostateczny wynik (wraz ze wzrostem liczności próby losowej). Bardziej zbliżonym do rzeczywistości rozkładem opisującym rzuty pinezką jest rozkład $Beta(26,18)$, a parametrem schematu Bernoulliego $p = 0.6$.



Wykres 1: Rozkład a priori i a posteriori dla eksperymentu rzutu pinezką.

2 Zadanie 2

Czas oczekiwania na obsługę w pewnej kolejce jest modelowany rozkładem $Exp(\lambda)$. Na podstawie znanego rozkładu a priori wyznaczamy numerycznie funkcje gęstości rozkładów a posteriori tychże rozkładów. Wiadomo, że rozkład a priori dla tej kolejki to rozkład $f_{prior}(\lambda) = Gamma(\alpha, \beta)$. Znamy średnią i wariancję tego rozkładu, a na ich podstawie możemy wyznaczyć potrzebne parametry rozkładu

$$\begin{aligned} EX &= \alpha\beta & \beta &= \frac{VX}{EX} \\ & \implies & \alpha &= \frac{EX}{\beta} \\ VX &= \alpha\beta^2 \end{aligned} \quad (4)$$

Ponadto wiemy, że średni czas oczekiwania dla losowo wybranych 20 osób to 5.1 minuty. Możemy przystąpić do numerycznego wyznaczania rozkładu a posteriori na podstawie poniższego wzoru

$$f_{post}(\lambda, x) = \frac{\mathcal{L}_n(x; \lambda) f_{prior}(\lambda)}{\int_0^\infty \mathcal{L}_n(x; \lambda) f_{prior}(\lambda) d\lambda} \quad (5)$$

gdzie

$$\mathcal{L}_n(x; \lambda) = f_\lambda(x_1) \times \dots \times f_\lambda(x_n) = \lambda e^{-\lambda x_1} \cdot \dots \cdot \lambda e^{-\lambda x_n} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i} = \lambda^n e^{-\lambda n\bar{x}} \quad (6)$$

Należy zwrócić uwagę, że wzór (6) jest prawdziwy jedynie w sytuacji, gdy pomiary czasów pobytu w kolejce są niezależne. Oznacza to, że pomiary były dokonywane np. w różnych dniach o różnych godzinach, tak, żeby czas oczekiwania jednej z osób nie wpływał na czas oczekiwania innej (branej pod uwagę do liczenia średniej).

Ostatecznie możemy wyznaczyć estymator właściwego rozkładu

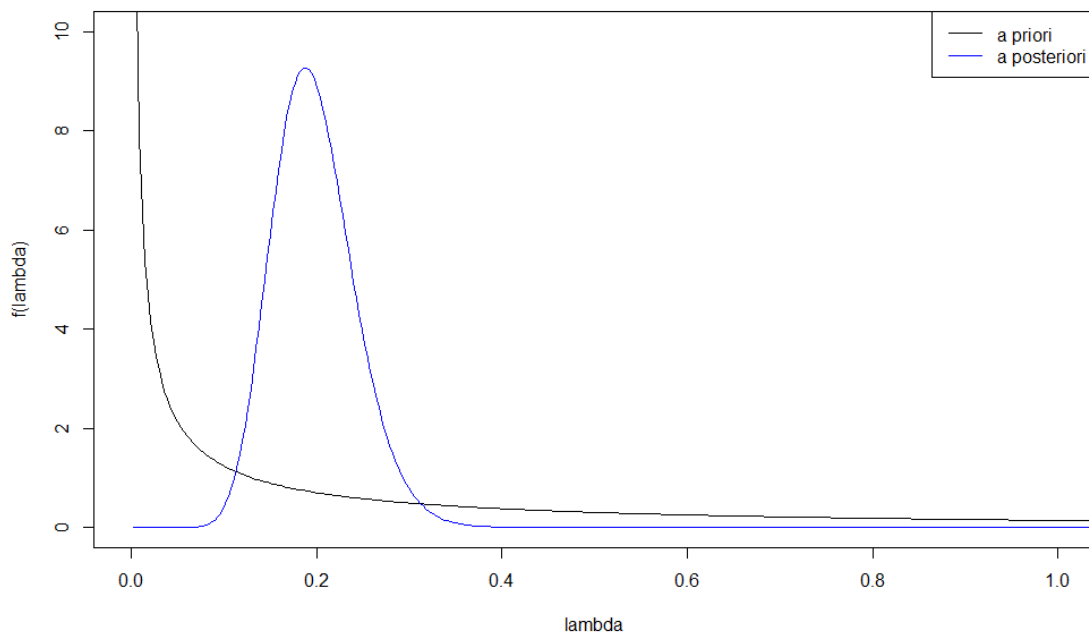
$$\hat{\lambda} = \int_0^{\infty} \lambda f_{post}(\lambda) d\lambda \quad (7)$$

W poniższej tabeli znajdują się wyniki dla zadanych rozkładów a priori

$\mathbb{E}X$	$\mathbb{V}X$	α	β	$\hat{\lambda}$
0.5	1	0.25	2	0.1974948
10	20	5	2	0.2437802

Tabela 3: Wartości parametrów rozkładu a priori i $\hat{\lambda}$.

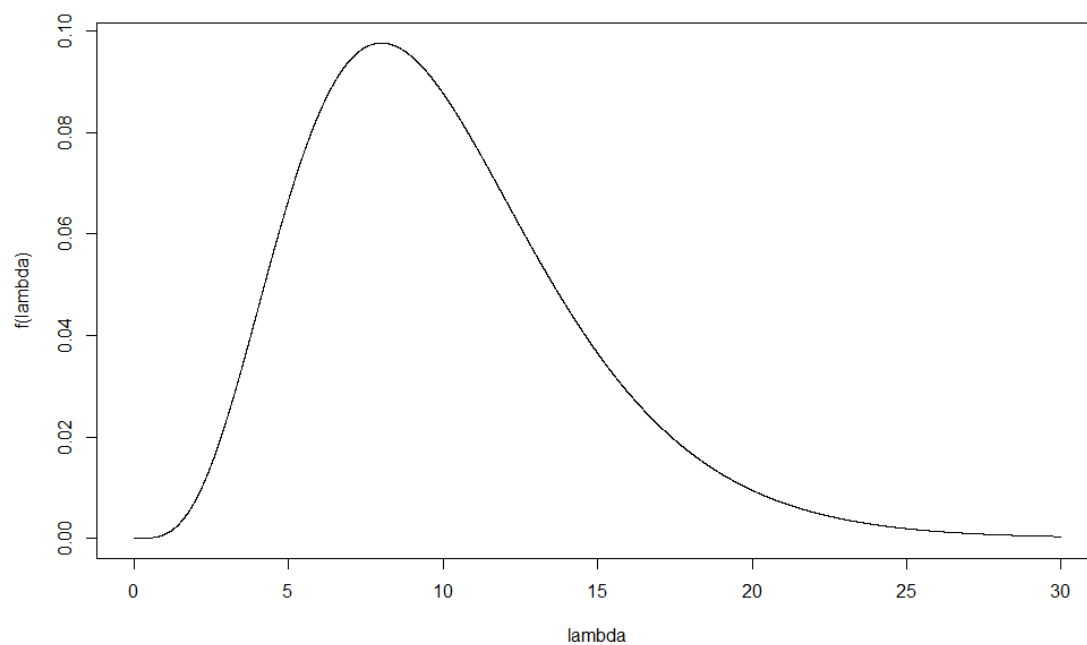
Ponownie można zauważyć, że rozkład a priori nie jest głównym czynnikiem wpływającym na rozkład a posteriori. Mimo dość różnych parametrów α obu rozkładów, estymatory λ są do siebie zbliżone. Rozkłady a priori i a posteriori można zaliczyć do rodziny rozkładów wykładniczych.



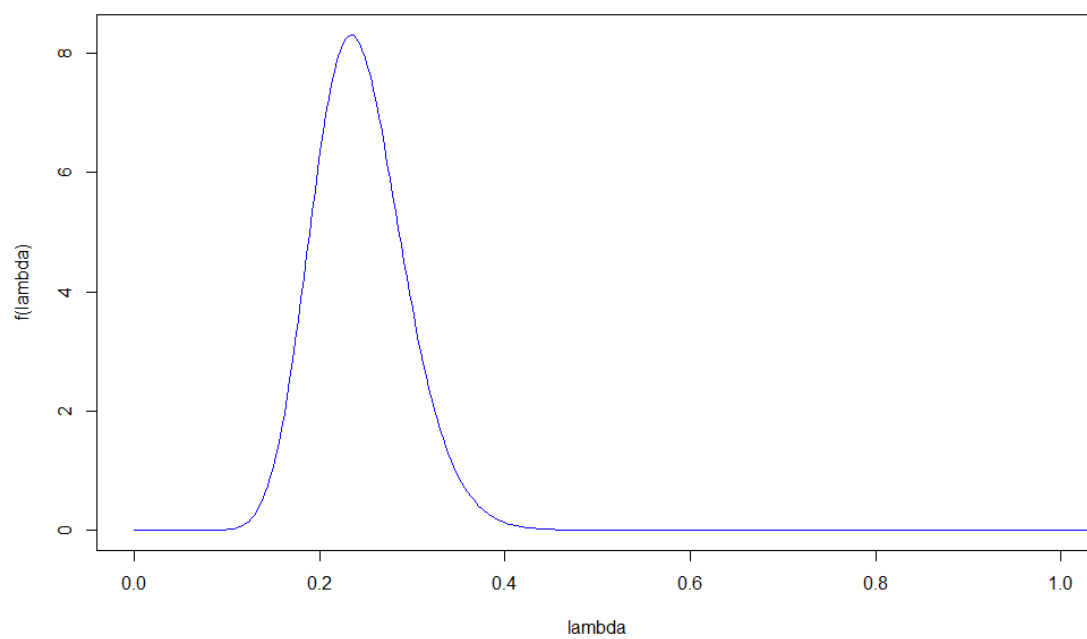
Wykres 2: Gęstość rozkładu a priori i a posteriori dla $\mathbb{E}X = 0.5$ i $\mathbb{V}X = 1$.

Do wyznaczenia rozkładu a posteriori została wykorzystana funkcja *integrate* do policzenia całki w mianowniku (5), a także do wyznaczenia estymatora $\hat{\lambda}$ w (7). Przyjmuje ona funkcję do scałkowania (napisaną samodzielnie), wraz z przedziałem całkowania. Ponadto wykorzystano funkcję *dgamma* do generowania rozkładu a priori o zadanych parametrach α i β .

Ze względu na różną skalę rozkładu a priori i a posteriori dla rozkładu $Gamma(5, 2)$, ich wykresy zostały przedstawione oddzielnie. W przeciwnym przypadku, można by mylnie stwierdzić, że rozkład a priori jest jednostajny.



Wykres 3: Gęstość rozkładu a priori dla $\mathbb{E}X = 10$ i $\mathbb{V}X = 20$.



Wykres 4: Gęstość rozkładu a posteriori dla $\mathbb{E}X = 10$ i $\mathbb{V}X = 20$.

3 Zadanie 3

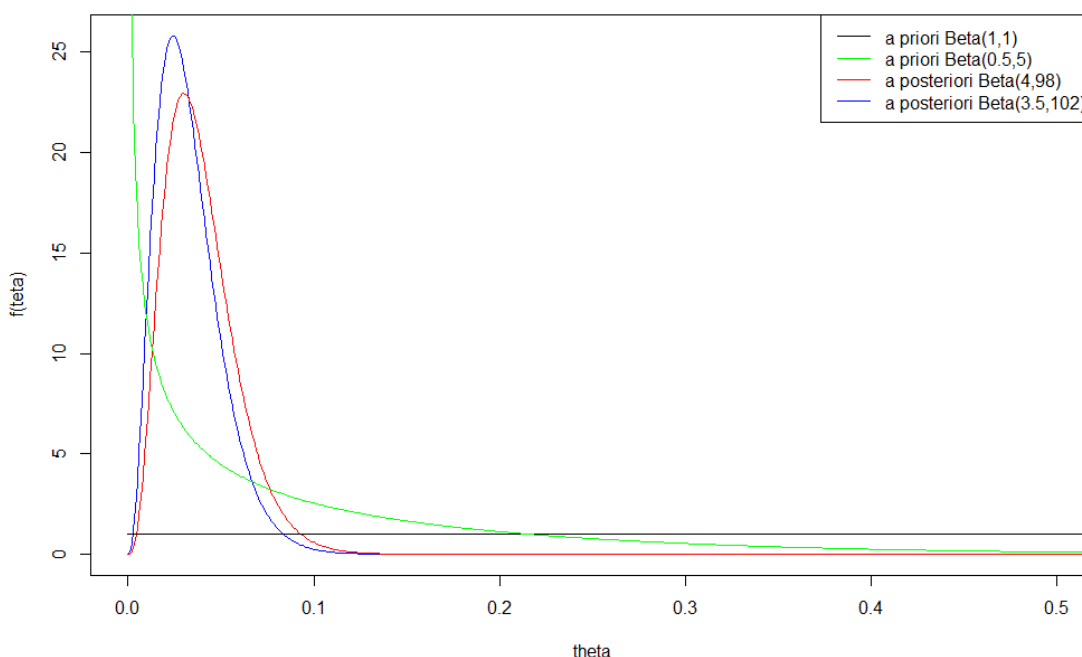
Rozpatrujemy linię produkcyjną, której 100 wyprodukowanych urządzeń zostało poddanych testom. Przyjmujemy, że linia produkuje x wadliwych urządzeń, równą ostatniej liczbie indeksu autora sprawozdania. Frakcję wadliwych urządzeń wytwarzanych przez linię θ można policzyć ze wzoru

$$\theta = \frac{\text{liczba wadliwych urządzeń}}{\text{liczba wyprodukowanych urządzeń}} = \frac{3}{100} \quad (8)$$

Podobnie jak w przypadku zadania pierwszego, rozkład a priori to rozkład $Beta(\alpha, \beta)$. Zgodnie ze wzorami (1) - (3) wyznaczmy parametry rozkładu a posteriori i estymatora $\hat{\theta}$.

α_{prior}	β_{prior}	α_{post}	β_{post}	$\hat{\theta}$
1	1	4	98	0.03921569
0.5	5	3.5	102	0.03317536

Tabela 4: Wartości parametrów rozkładu a priori i $\hat{\lambda}$.



Wykres 5: Rozkłady a priori i a posteriori dla danych rozkładów Beta.

Na podstawie wykresu można podejrzewać, że rozkład $Beta(3.5, 102)$ lepiej opisuje frakcję wytwarzanych wadliwych urządzeń. Wartość estymowanego parametru θ jest bliższa założonej, a także wykres bardziej skupiony wokół tej właśnie wartości. Ponownie, dobór parametrów rozkładu a priori, nie odgrywa kluczowej roli.