Modele i Wnioskowanie Statystyczne
Laboratorium 3

Sprawozdanie

Igor Markiewicz

Zadanie 1

Rzut pinezką można zamodelować używając próby Bernoulliego :

$$f_X(k) = \begin{cases} p & \text{gdy } k = 1 \\ p - 1 & \text{gdy } k = 0 \end{cases}$$

Jeśli za rozkład priori parametru p przyjmiemy rozkład beta (co jest bardzo wygodne ze względu na duże mozliwości dostosowania funkcji i nośnika na przedziale $[0, 1]$), to ze względu na sprzężenie modelu Bernouliego względem tego rozkładu mamy określony rozkład a posteriori w nstępujący sposób :

$$f(\alpha, \beta) \longrightarrow f\left(\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i\right)$$

gdzie $f(\alpha, \beta)(p)$:

$$f(\alpha, \beta)(p) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1 - p)^{\beta-1}$$

Arbitralnie przyjęto parametry rozkładu a priori $\alpha = 2$, $\beta = 2$, co zapewnia symetryczy rozkład z maksimum dla $p = 0,5$. Obliczono jednocześnie podstawowe parametry rozkłdu, celem ich porównania :

$$E(X) = \frac{\alpha}{\alpha + \beta}$$

$$Var(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Następnie wykonano dwa eksperymenty :

- 20 rzutów pinezką $X_1 =< 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1 >$

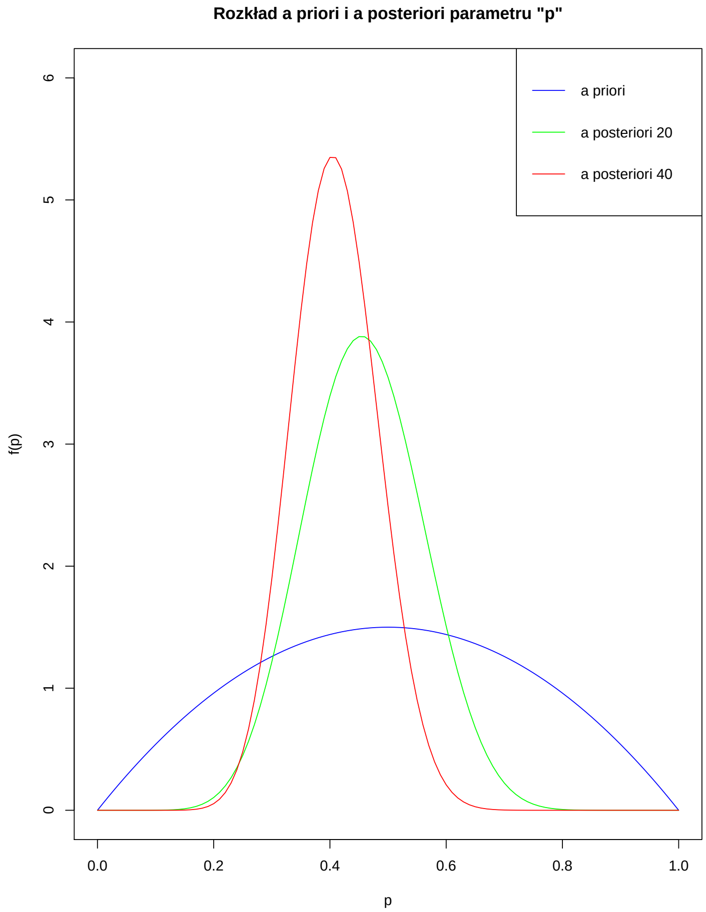
$$E(X) \approx 0,46$$

$$Var(X) \approx 0,0099$$

- 40 rzutów pinezką $X_2 = X_1 ^\frown X_1'$
  $X_1' =< 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0 >$

$$E(X) \approx 0,41$$

$$Var(X) \approx 0,0054$$



Rys. 1: Wykresy funkcji gęstości prawdopodobieństwa

Wnioski :
Możemy zaboserwować że wraz ze wzrostem liczby pomiarów, rośnie pewność oczekiwanego wyniku (zmniejsza się wariancja), a wartość oczekiwana jest modyfikowana w ten sposób, aby być ”blisko” empirycznych estymat p. Dla 20 pomiarów otrymujemy $p = 0,45$, zaś dla 40 mamy $p = 0,4$

Zadanie 2

Rozkład posteriori :

$$f_{post}(\lambda;\boldsymbol{t}) = \frac{L_n(\boldsymbol{t};\lambda) f_{prior}(\lambda)}{\int_\Theta L_n(\boldsymbol{t};\lambda) f_{prior}(\lambda)\mathrm{d}\lambda}$$

Funkcja wiarygodności dla rozkładu wykładniczego (przy założeniu niezależności zmiennych):

$$L_n(\boldsymbol{t};\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda n \bar{\boldsymbol{t}}}$$

Rozkład gamma :

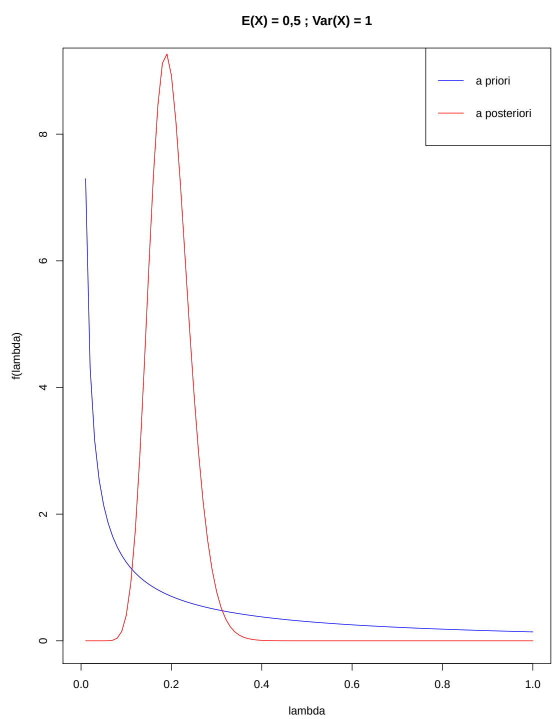
$$f_X(\lambda) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{\alpha-1} e^{-\frac{\lambda}{\beta}}$$

Wyliczenie parametrów rozkładu gamma :

$$\begin{cases} E(X) = \alpha\beta \\ Var(X) = \alpha\beta^2 \end{cases}$$

$$\begin{cases} \beta = \frac{Var(X)}{E(X)} \\ \alpha = \frac{E^2(X)}{Var(X)} \end{cases}$$

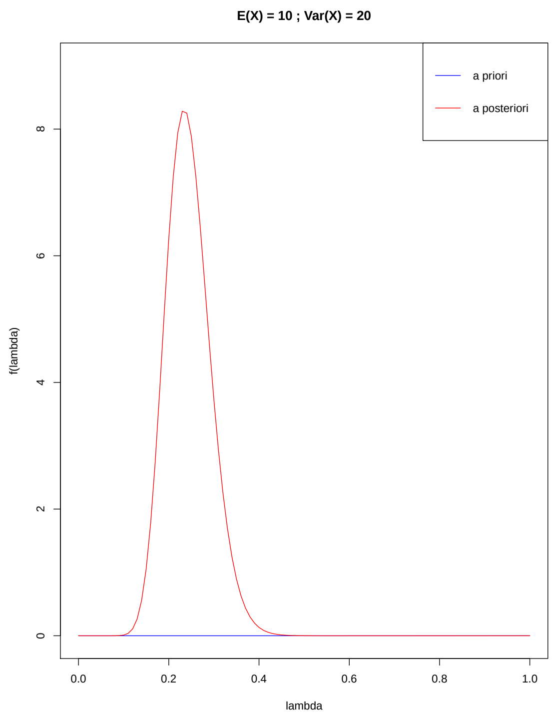
a)



Rys. 2: Wykresy funkcji gęstości prawdopodobieństwa

$$\arg\max\left(f_{post}(\lambda)\right) = 0,19$$
$$E(X) = \frac{1}{0,19} \approx 5,26$$

b)



Rys. 3: Wykresy funkcji gęstości prawdopodobieństwa

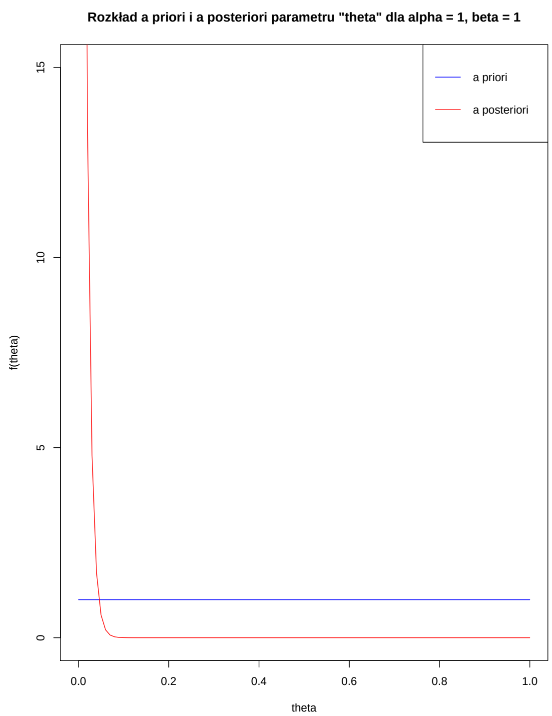
$$\arg\max\left(f_{post}(\lambda)\right) = 0,23$$
$$E(X) = \frac{1}{0,23} \approx 4,35$$

Wnioski :
W obydwu przypadkach rozkłady a piori i a posteriori różnią się znacząco. Ponad to w podpunkcie **b)**, rozkład a priori przyjmuje dla zadanych parametrów tak małe wartości, że w porównaniu z rozkładem a posteriori przedstawia linię prostą. Zauważamy również że w drugim przypadku wariancja rozkładu jest większa oraz obliczona wartość oczekiwana jest gorsza w stodunku do $5,1$ min.

Zadanie 3
Podobnie jak w **Zadanie 1**, można zamodelować podane zjawisko stosując próbę Bernoulliego z parametrem :

$$\theta = \frac{\text{liczba urządzeń wadliwych = ostatnia cyfra indeksu}}{\text{liczba wyprodukowanych urządzeń}} = 0$$

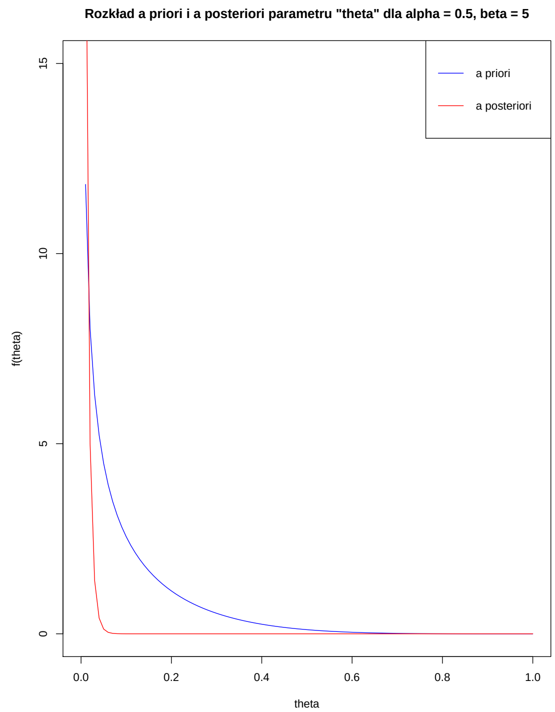
oznaczającym prawdopodobieństwo popsucia się urządzenia w pojedyńczej próbie. Można podobnie wykorzystać jako rozkład a posteriori rozkład beta. **a)**



Rys. 4: Wykresy funkcji gęstości prawdopodobieństwa a priori i a posteriori

$$E(X) = 0,0098$$
$$Var(X) = 9,43 \cdot 10^{-5}$$
$$\arg\max\left(f_{post}(\theta)\right) = 0$$

b)



Rys. 5: Wykresy funkcji gęstości prawdopodobieństwa a priori i a posteriori

$$E(X) = 0,0047$$
$$Var(X) = 4,43 \cdot 10^{-5}$$
$$\arg\max\left(f_{post}(\theta)\right) = 0$$

Wnioski :

Dla $\alpha = \beta = 1$, otrzymujemy wartość stałą równą 1 na przedziale $[0, 1]$ co odpowiada rozkładowi jednostajnemu. Dlatego też pierwszy rozkład ma większą wariancję niż drugi i bardziej odstającą wartość oczekiwaną. W obydwu przypadkach największą wartość funkcji gęstości prawdopodobieństwa otrzymujemy dla $\hat{\theta} = 0$.