

Sprawozdanie 2

Bartosz Michalak

8 grudnia 2025

1 Zadanie 1: Wpływ rozkładu a priori na estymację

1.1 Opis problemu

Wnioskowanie bayesowskie opiera się na aktualizacji wiedzy a priori (przed doświadczeniem) o dane z próby, tworząc rozkład a posteriori. Rozważamy trzy nieinformujące rozkłady a priori dla parametru $\vartheta \in (0, 1)$:

1. **Rozkład Laplace'a** (płaski): Beta(1, 1). Zakłada, że każda wartość parametru jest jednakowo prawdopodobna.
2. **Rozkład Jeffreysa**: Beta(1/2, 1/2). Jest niezmienniczy ze względu na reparametryzację modelu.
3. **Rozkład MDIP** (Maximum Data Information Prior): Maksymalizuje ilość informacji w danych.

Celem jest wyznaczenie wartości oczekiwanej oraz mody rozkładu a posteriori dla zadanych n i x .

1.2 Metodyka obliczeń

Jeśli rozkład a priori to Beta(α, β), to rozkład a posteriori jest postaci Beta(α_n, β_n), gdzie parametry kształtu aktualizujemy o dane z próby:

$$\alpha_n = \alpha + x, \quad \beta_n = \beta + n - x$$

Dla tak określonego rozkładu a posteriori wyznaczamy analitycznie:

- **Wartość oczekiwana**:

$$E[\vartheta|x] = \frac{\alpha_n}{\alpha_n + \beta_n}$$

- **Mode** (dla $\alpha_n > 1, \beta_n > 1$):

$$\text{Mode} = \frac{\alpha_n - 1}{\alpha_n + \beta_n - 2}$$

Do wyrażenia rozkładu **MDIP** potrzebujemy znać ujemną entropię rozkładu zmiennej X :

$$\kappa(\vartheta) = E[\ln(p_\vartheta(X))]$$

gdzie w tym przypadku p_ϑ jest funkcją masy zmiennej X . Rozkład ten nie ma zwartego wzoru na entropię, ale możemy natomiast zapisać z definicji:

$$\kappa(\vartheta) = \sum_{k=0}^n p_\vartheta(k) \ln(p_\vartheta(k))$$

gdzie $p_\vartheta(k) = \binom{n}{k} \vartheta^k (1 - \vartheta)^{n-k}$. Wartość tę obliczamy numerycznie i przypominamy sobie, że rozkład a priori MDIP ma gęstość $\pi_{MDIP}(\vartheta)$ spełniającą zależność:

$$\pi_{MDIP}(\vartheta) \propto \exp(\kappa(\vartheta))$$

```
bayesian_binomial_post_summary <- function(
  n,
  x,
  prior = c("Laplace", "Jeffreys", "MDIP")
) {

  results <- data.frame(
    Prior = character(),
    Expected_Value = numeric(),
    Mode = numeric(),
    stringsAsFactors = FALSE
  )

  # (a) Laplace (Beta(1, 1))
  if ("Laplace" %in% prior) {
    alpha_L <- x + 1
    beta_L <- n - x + 1
    E_L <- alpha_L / (alpha_L + beta_L)
    Mode_L <- ifelse(
      x > 0 && n - x > 0, (alpha_L - 1) /
      (alpha_L + beta_L - 2), ifelse(x == 0, 0, 1)
    )
    results[nrow(results) + 1, ] <- list("Laplace", E_L, Mode_L)
  }

  # (b) Jeffreys (Beta(1/2, 1/2))
  if ("Jeffreys" %in% prior) {
    alpha_J <- x + 0.5
    beta_J <- n - x + 0.5
    E_J <- alpha_J / (alpha_J + beta_J)

    # Moda (gęstość dąży do nieskończoności na brzegu, gdy x=0 lub x=n)
    if (x == 0) {
      Mode_J <- 0
    } else if (x == n) {
      Mode_J <- 1
    } else {
      Mode_J <- (alpha_J - 1) / (alpha_J + beta_J - 2)
    }
    results[nrow(results) + 1, ] <- list("Jeffreys", E_J, Mode_J)
  }

  # (c) MDIP
  if ("MDIP" %in% prior) {
    dmdip.binom <- function(n, theta) {
      pmf <- outer(
        seq(theta, 1 - theta, length.out = n),
        rep(1:n, times = n),
        FUN = function(theta, k) {
          dbeta(k, n - k + 1, theta)
        })
      pmf / sum(pmf)
    }
  }
}
```

```

    0:n,
    theta,
    Vectorize(\(x, theta) dbinom(x, size = n, prob = theta))
)
pmf[pmf == 0] <- 1e-15
entropy <- colSums(pmf * log(pmf), na.rm = TRUE)
exp(entropy)
}

E_M <- integrate(\(theta) theta * dbinom(x, n, theta) *
  dmdip.binom(n, theta), 0, 1)$value /
integrate(\(theta) dbinom(x, n, theta) *
  dmdip.binom(n, theta), 0, 1)$value

Mode_M <- optimize(
  \(\theta) log(dbinom(x, n, theta)) + log(dmdip.binom(n, theta)),
  interval = c(0, 1),
  maximum = TRUE
)$maximum

results[nrow(results) + 1, ] <- list("MDIP", E_M, Mode_M)
}
results$Expected_Value <- as.numeric(results$Expected_Value)
results$Mode <- as.numeric(results$Mode)

return(results)
}

```

Poniżej przedstawiono wyniki estymacji dla małej próby ($n = 10, x = 4$).

```

n <- 10; x <- 4
cat(sprintf("Wyniki dla n = %d i x = %d:\n", n, x))

## Wyniki dla n = 10 i x = 4:

print(bayesian_binomial_post_summary(n, x, prior =
c("Laplace", "Jeffreys", "MDIP")))

##      Prior Expected_Value      Mode
## 1 Laplace     0.4166667 0.4000000
## 2 Jeffreys   0.4090909 0.3888889
## 3 MDIP       0.4078548 0.3876068

```

```

n <- 10; x <- 5
cat(sprintf("Wyniki dla n = %d i x = %d:\n", n, x))

## Wyniki dla n = 10 i x = 5:

print(bayesian_binomial_post_summary(n, x, prior =
c("Laplace", "Jeffreys", "MDIP")))

```

```

##      Prior Expected_Value Mode
## 1 Laplace      0.5  0.5
## 2 Jeffreys    0.5  0.5
## 3 MDIP        0.5  0.5

n <- 10; x <- 6
cat(sprintf("Wyniki dla n = %d i x = %d:\n", n, x))

## Wyniki dla n = 10 i x = 6:

print(bayesian_binomial_post_summary(n, x, prior =
c("Laplace", "Jeffreys", "MDIP")))

##      Prior Expected_Value     Mode
## 1 Laplace      0.5833333 0.6000000
## 2 Jeffreys    0.5909091 0.6111111
## 3 MDIP        0.5921452 0.6123932

n <- 1000; x <- 523
cat(sprintf("Wyniki dla n = %d i x = %d:\n", n, x))

## Wyniki dla n = 1000 i x = 523:

print(bayesian_binomial_post_summary(n, x, prior =
c("Laplace", "Jeffreys", "MDIP")))

##      Prior Expected_Value     Mode
## 1 Laplace      0.5229541 0.5230000
## 2 Jeffreys    0.5229770 0.5230230
## 3 MDIP        0.5229680 0.5230231

```

1.3 Wnioski

Analizując powyższe wyniki, można zauważyc, że przy małej liczebności próby wybór rozkładu a priori ma zauważalny wpływ na jakość estymacji. Różnice między estymatorami wartości oczekiwanej i mody są bardziej widoczne dla $n = 10$ niż dla $n = 1000$. W przypadku większej próby, estymatory z różnych rozkładów a priori zbliżają się do siebie, co jest zgodne z intuicją, że przy dużej ilości danych wpływ rozkładu a priori maleje.

Ponadto można zauważyc, że dla $x < \frac{n}{2}$ moda jest mniejsza niż wartość oczekiwana, co wynika z asymetrii rozkładu Beta w tych przypadkach. Analogicznie, dla $x > \frac{n}{2}$, moda jest większa niż wartość oczekiwana.

2 Zadanie 2: Aproksymacja BCTG

2.1 Opis metodyki

Twierdzenie Bernsteina-von Misesa (BCTG) pozwala aproksymować rozkład a posteriori rozkładem normalnym dla dużych prób. W badaniu porównujemy dwie wersje tej aproksymacji:

1. **BCTG 1 (Aproksymacja momentowa):** Rozkład a posteriori przybliżamy rozkładem normalnym $N(\mu_n, \delta_n^2)$, gdzie parametry odpowiadają dokładnej wartości oczekiwanej i wariancji rozkładu a posteriori Beta:

$$\mu_n = \frac{\alpha + x}{\alpha + \beta + n}, \quad \delta_n^2 = \frac{(\alpha + x)(\beta + n - x)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)}$$

2. **BCTG 2 (Aproksymacja modalna):** Rozkład a posteriori przybliżamy rozkładem normalnym $N(m_n, v_n^2)$, gdzie m_n jest modą rozkładu a posteriori, a v_n^2 jest odwrotnością ujemnej drugiej pochodnej logarytmu gęstości w punkcie mody (związane z informacją Fishera):

$$m_n = \frac{\alpha + x - 1}{\alpha + \beta + n - 2}$$

$$v_n^2 = \frac{(\alpha + x - 1)(\beta + n - x - 1)}{(\alpha + \beta + n - 2)^3}$$

```
aproksymacja_BCTG <- function(n, x, alfa, beta) {
  alfa_n <- alfa + x
  beta_n <- beta + n - x

  mu_n_v1 <- alfa_n / (alfa_n + beta_n)
  delta_n_sq_v1 <- (alfa_n * beta_n) /
    ((alfa_n + beta_n)^2 * (alfa_n + beta_n + 1))

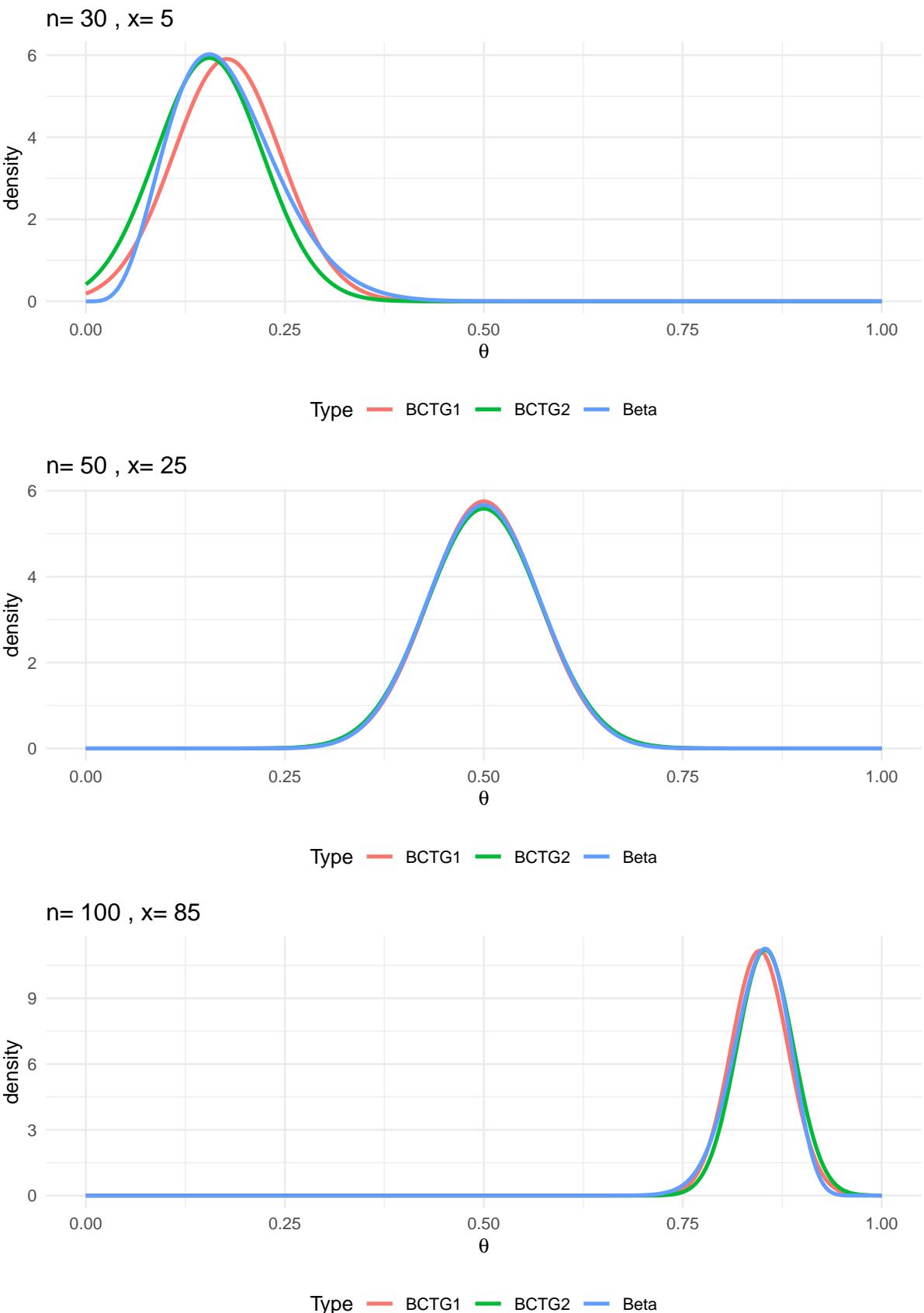
  if (alfa_n > 1 && beta_n > 1) {
    m_n_v2 <- (alfa_n - 1) / (alfa_n + beta_n - 2)
    v_n_sq_v2 <- ((alfa_n - 1) * (beta_n - 1)) / (alfa_n + beta_n - 2)^3
  } else {
    m_n_v2 <- NA
    v_n_sq_v2 <- NA
  }

  return(list(
    mu1 = mu_n_v1,
    sd1 = sqrt(delta_n_sq_v1),
    mu2 = m_n_v2,
    sd2 = sqrt(v_n_sq_v2)
  ))
}
```

2.2 Wizualizacja i Wnioski

Na podstawie wygenerowanych wykresów gęstości można sformułować następujące wnioski dotyczące jakości aproksymacji BCTG:

- **Weryfikacja twierdzenia Bernsteina-von Misesa:** Wykresy wizualnie potwierdzają treść twierdzenia. Dla małej liczby próby ($n = 30$) widoczne są wyraźne rozbieżności między rozkładem dokładnym a aproksymacjami normalnymi. Wraz ze wzrostem liczby próby (do $n = 100$), różnice te zanikają, a wszystkie trzy krzywe zaczynają się coraz bardziej pokrywać. Świadczy to o asymptotycznej zbieżności rozkładu a posteriori do rozkładu normalnego.



Rysunek 1: Porównanie gęstości dokładnej (Beta) z aproksymacjami normalnymi (BCTG) dla różnych wielkości próby.

- **Porównanie aproksymacji w przypadku asymetrii ($n = 30, x = 5$):** Dla małej próby i parametru ϑ bliskiego 0, rozkład a posteriori jest silnie prawoskoły. Widać tu wyraźną różnicę między metodami:
 - **BCTG 2:** Ponieważ parametry tej aproksymacji (m_n, v_n^2) oparte są na modzie, jej wierzchołek idealnie pokrywa się ze szczytem dokładnego rozkładu Beta (niebieska linia). Jest to metoda preferowana, gdy zależy nam na precyzji wokół estymatora MAP.
 - **BCTG 1:** Ta aproksymacja jest wycentrowana na wartości oczekiwanej (μ_n). W rozkładach prawoskołnych średnia leży na prawo od mody, co na wykresie objawia się przesunięciem czerwonej krzywej względem szczytu rozkładu dokładnego.
- **Przypadek symetryczny ($n = 50, x = 25$):** Gdy estymowana frakcja wynosi 0.5, rozkład a posteriori jest symetryczny. W takim przypadku średnia pokrywa się z modeą, w rezultacie czego obie wersje aproksymacji (BCTG 1 i BCTG 2) dają niemal identyczne wyniki i bardzo dobrze przybliżają rozkład dokładny.

3 Zadanie 3: Testowanie Hipotez i Czynnik Bayesa

3.1 Opis metodyki

Weryfikujemy hipotezę zerową $H_0 : \vartheta \leq \vartheta_0$ przeciwko hipotezie alternatywnej $H_1 : \vartheta > \vartheta_0$.

Do podjęcia decyzji wykorzystujemy **czynnik Bayesa**. Do oceny siły dowodów na rzecz hipotezy alternatywnej (H_1), posługujemy się wskaźnikiem B_{10} , zdefiniowanym jako:

$$B_{10} = \frac{P(\vartheta \in \Theta_1 | x)\pi_0}{P(\vartheta \in \Theta_0 | x)\pi_1}$$

gdzie:

- $P(\vartheta \in \Theta_1 | x)$ oraz $P(\vartheta \in \Theta_0 | x)$ to prawdopodobieństwa a posteriori odpowiednio dla hipotezy alternatywnej i zerowej.
- π_1 oraz π_0 to prawdopodobieństwa a priori tych hipotez.

```

zadanie_3 <- function(
  n,
  alpha_prior,
  beta_prior,
  x,
  theta0,
  lambda
) {
  alpha_n <- alpha_prior + x
  beta_n <- beta_prior + n - x

  # =====
  # 1. ROZKŁAD DOKŁADNY (Podejście numeryczne)
  # =====
  # (a) ESTYMATOR MAP (Maximum A Posteriori)
  # Definicja 1: arg max h(theta | x)
  # Definiujemy funkcję gęstości i szukamy jej maksimum.
  funkcja_gestosci_a_posteriori <- function(theta) {
    
```

```

    dbeta(theta, shape1 = alpha_n, shape2 = beta_n)
}

# optimize() szuka maksimum w przedziale (0, 1)
wynik_optymalizacji <- optimize(f = funkcja_gestosci_a_posteriori,
                                 interval = c(0, 1),
                                 maximum = TRUE)

map_exact <- wynik_optymalizacji$maximum

# (b) OBSZAR NAJWIĘKSZEJ GĘSTOŚCI (HPD)
# Szukamy przedziału, gdzie gęstość na krańcach
# jest równa (dbeta(L) == dbeta(U))
calc_hpd_custom <- function(a, b, conf_level) {
  obj_func <- function(p) {
    lower_q <- qbeta(p, a, b)
    upper_q <- qbeta(p + conf_level, a, b)
    return(dbeta(lower_q, a, b) - dbeta(upper_q, a, b))
  }
  # uniroot szuka p takiego, że różnica gęstości wynosi 0
  res <- uniroot(obj_func, interval = c(1e-10, 1 - conf_level - 1e-10))
  p_best <- res$root
  return(c(qbeta(p_best, a, b), qbeta(p_best + conf_level, a, b)))
}

hpд_exact <- calc_hpd_custom(alpha_n, beta_n, 1 - lambda)

# (c) CZYNNIK BAYESA (Bayes Factor)
# Hipoteza H0: theta <= theta0 (Całka z gęstości od 0 do theta0)
# Hipoteza H1: theta > theta0 (Całka z gęstości od theta0 do 1)
prob_H0_prior <- pbeta(theta0, alpha_prior, beta_prior)
odds_prior <- prob_H0_prior / (1 - prob_H0_prior)

prob_H0_post <- pbeta(theta0, alpha_n, beta_n)
odds_post <- prob_H0_post / (1 - prob_H0_post)

bf_exact <- odds_prior / odds_post

# =====
# 2. APROKSYMACJE BCTG 1 (wartość oczekiwana)
# =====
mu_n <- alpha_n / (alpha_n + beta_n)
d_n_sq <- (alpha_n * beta_n) / ((alpha_n + beta_n)^2 * (alpha_n + beta_n + 1))
sd_n <- sqrt(d_n_sq)

map_bctg1 <- mu_n
z_crit <- qnorm(1 - lambda / 2)
hpд_bctg1 <- c(mu_n - z_crit * sd_n, mu_n + z_crit * sd_n)
prob_H0_bctg1 <- pnorm(theta0, mean = mu_n, sd = sd_n)
bf_bctg1 <- odds_prior / (prob_H0_bctg1 / (1 - prob_H0_bctg1))

```

```

# =====
# 3. APROKSYMACJE BCTG 2 (moda)
# =====

m_n <- (alpha_n - 1) / (alpha_n + beta_n - 2)
v_n_sq <- ((alpha_n - 1) * (beta_n - 1)) / ((alpha_n + beta_n - 2)^3)
sd_v <- sqrt(v_n_sq)
map_bctg2 <- m_n
hp_d_bctg2 <- c(m_n - z_crit * sd_v, m_n + z_crit * sd_v)
prob_H0_bctg2 <- pnorm(theta0, mean = m_n, sd = sd_v)
bf_bctg2 <- odds_prior / (prob_H0_bctg2 / (1 - prob_H0_bctg2))

results <- data.frame(
  Metoda = c("Dokładna", "BCTG 1", "BCTG 2"),
  MAP_Estymator = c(map_exact, map_bctg1, map_bctg2),
  HPD_Dol = c(hpd_exact[1], hpd_bctg1[1], hpd_bctg2[1]),
  HPD_Gora = c(hpd_exact[2], hpd_bctg1[2], hpd_bctg2[2]),
  BayesFactor_10 = c(bf_exact, bf_bctg1, bf_bctg2)
)

print(results)
return(invisible(results))
}

```

3.2 Wyniki obliczeń

```

out <- rozwiaz_zadanie_3_zgodnie_z_definicja(
  n = 10, alpha_prior = 0.5, beta_prior = 0.5,
  x = 4, theta0 = 0.5, lambda = 0.05)

##      Metoda MAP_Estymator    HPD_Dol   HPD_Gora BayesFactor_10
## 1 Dokładna     0.3888882 0.14257620 0.6838786     0.3602247
## 2 BCTG 1       0.4090909 0.13090975 0.6872721     0.3530329
## 3 BCTG 2       0.3888889 0.07039603 0.7073817     0.3281311

out <- rozwiaz_zadanie_3_zgodnie_z_definicja(
  n = 10, alpha_prior = 1, beta_prior = 1,
  x = 4, theta0 = 0.5, lambda = 0.05)

##      Metoda MAP_Estymator    HPD_Dol   HPD_Gora BayesFactor_10
## 1 Dokładna     0.3999980 0.15857762 0.6817731     0.3781965
## 2 BCTG 1       0.4166667 0.14867012 0.6846632     0.3719553
## 3 BCTG 2       0.4000000 0.09636369 0.7036363     0.3500788

out <- rozwiaz_zadanie_3_zgodnie_z_definicja(
  n = 10, alpha_prior = 0.5, beta_prior = 0.5,
  x = 5, theta0 = 0.5, lambda = 0.05)

##      Metoda MAP_Estymator    HPD_Dol   HPD_Gora BayesFactor_10

```

```

## 1 Dokładna      0.5 0.2235287 0.7764713      1
## 2   BCTG 1      0.5 0.2171036 0.7828964      1
## 3   BCTG 2      0.5 0.1733393 0.8266607      1

```

3.3 Wnioski

Analiza wyników uzyskanych dla małej liczebności próby ($n = 10$) pozwala na sformułowanie następujących wniosków dotyczących weryfikacji hipotez i jakości aproksymacji:

1. **Interpretacja Czynnika Bayesa (B_{10}) dla $x = 4$:** W przypadkach, gdzie liczba sukcesów wynosiła $x = 4$ (przy $n = 10$), estymator MAP oscylował wokół wartości 0.4, co jest poniżej progu testowanej hipotezy $H_1 : \vartheta > 0.5$. Uzyskane wartości czynnika Bayesa B_{10} wyniosły około 0.33 – 0.38 (zależnie od przyjętego rozkładu a priori i metody).
 - Ponieważ B_{10} jest za małe, możemy stwierdzić, że dane nie dostarczają dowodów na rzecz hipotezy alternatywnej (H_1).
 - Wartość $B_{10} \approx 0.36$ oznacza, że po uwzględnieniu danych, szanse na prawdziwość hipotezy H_1 zmalały w stosunku do szans a priori. Innymi słowy, dane około 3-krotnie bardziej wspierają hipotezę zerową ($H_0 : \vartheta \leq 0.5$) niż alternatywną (gdyż $B_{01} = 1/B_{10} \approx 2.7$).
2. **Przypadek graniczny ($x = 5$):** Trzeci wynik przedstawia sytuację, w której liczba sukcesów wynosi dokładnie połowę liczebności próby ($x = 5, n = 10$).
 - W tym przypadku estymator MAP wynosi dokładnie 0.5, a czynnik Bayesa $B_{10} = 1$ dla wszystkich metod.
 - Oznacza to, że dane są całkowicie "neutralne" względem testowanych hipotez. Rozkład a posteriori jest symetryczny względem $\vartheta_0 = 0.5$, więc prawdopodobieństwo a posteriori obu hipotez jest jednakowe. Dane nie zmieniły naszych początkowych szans.
 - Warto zauważyć, że przedziały największej gęstości a posteriori (HPD) są symetryczne co jest zgodne z intuicją.
3. **Porównanie metod aproksymacji (BCTG vs Dokładna):** Mimo małej liczebności próby ($n = 10$), aproksymacje BCTG działają zaskakująco dobrze, choć widać subtelne różnice:
 - **Estymacja MAP:** Metoda **BCTG 2 (Modalna)** zazwyczaj daje wyniki bliższe (lub identyczne) metodzie dokładnej niż BCTG 1. Widać to wyraźnie dla a priori Laplace'a ($x = 4$), gdzie BCTG 2 i metoda dokładna dają MAP = 0.400, podczas gdy BCTG 1 (oparta na średniej) daje 0.417. Wynika to z faktu, że przy asymetrycznym rozkładzie średnia różni się od mody.
 - **Czynnik Bayesa:** Wszystkie trzy metody prowadzą do tych samych wniosków dedykacyjnych ($B_{10} < 1$ dla $x = 4$ oraz $B_{10} = 1$ dla $x = 5$). Różnice w wartościach liczbowych B_{10} między metodą dokładną a aproksymacjami są na tyle małe, że nie wpływają na interpretację merytoryczną wyniku.
4. **Wpływ rozkładu a priori:** Porównując wyniki dla a priori Jeffreysa ($\alpha = \beta = 0.5$) i Laplace'a ($\alpha = \beta = 1$) przy $x = 4$, różnice są minimalne. Rozkład Laplace'a (płaski) nieznacznie mocniej przesuwa estymatory w stronę centrum (0.5), co skutkuje minimalnie wyższym B_{10} (0.378 vs 0.360), ale nie zmienia to ogólnego wniosku o braku podstaw do przyjęcia H_1 .