Analiza estymatorów w regresji liniowej w kontekście selekcji zmiennych i redukcji błędu estymacji

Sylwia Patrijas

2025-03-31

Raport ten skupia się na analizie i porównaniu różnych metod **estymacji wektora współczynników regresji liniowej** β oraz procedur selekcji zmiennych pod kątem ich efektywności. Rozważanych będzie następujących sześć estymatorów: **estymator najmniejszych kwadratów**, **estymator Jamesa-Steina ściągający do zera**, **estymator Jamesa-Steina ściągający do wspólnej średniej** oraz **trzy estymatory "ucięte"**: dla procedury Bonferroniego, procedury Benjaminiego-Hochberga i klasyfikatora Bayesowskiego przy założeniu tej samej funkcji straty za błąd pierwszego i drugiego rodzaju.

Jako pierwsze, należy wygenerować **ortonormalną macierz planu** $X_{1000x1000}$, to znaczy taką, że $X^TX = I$. Zrobię to, korzystając z polecenia *randortho*, a następnie sprawdzę poprawność wygenerowanej macierzy, wyliczając X^TX.

Maksymalna różnica między elementami na przekątnej iloczynu macierzy X^TX a liczbą 1 to wartość **bardzo bliska zera**. Podobnie, maksymalna różnica między elementami poza przekątną tego iloczynu a liczbą 0 jest **bardzo zbliżona do zera**. Zatem, potwierdza to **poprawność wygenerowania macierzy ortonormalnej**, gdyż niewielkie różnice pojawiają się z powodu błędów numerycznych.

POJEDYNCZE DOŚWIADCZENIE

W pierwszej części tego raportu należy wygenerować **wektor współczynników regresji** jako ciąg niezależnych zmiennych losowych z rozkładu:

$$\beta_i \sim (1-\gamma)\delta_0 + \gamma\phi(0,\tau^2),$$

gdzie δ_0 jest rozkładem skupionym w 0, a $\phi(0,\tau^2)$ jest gęstością rozkładu normalnego $N(0,\tau^2)$. Należy rozważyć sześć przypadków:

- $\gamma = 0.01, \tau = 1.5\sqrt{2\log 1000}$,
- $\gamma = 0.05, \tau = 1.5\sqrt{2\log 1000}$,
- $\gamma = 0.1, \tau = 1.5\sqrt{2\log 1000}$,
- $\gamma = 0.01, \tau = 3\sqrt{2\log 1000}$,
- $\gamma = 0.05, \tau = 3\sqrt{2\log 1000}$,
- $\gamma = 0.1, \tau = 3\sqrt{2\log 1000}$.

Ponadto, dla każdego z tych przypadków należy wygenerować wektor odpowiedzi $Y = X\beta + \epsilon$, gdzie $\epsilon \sim N(0, I_{1000x1000})$ - zakładamy więc, że wariancja błędu jest znana ($\sigma^2 = 1$).

Zatem, zgodnie z zadaniem, dla każdej kombinacji wartości γ i τ wygeneruję wektor współczynników regresji β zgodnie z podanym rozkładem: z prawdopodobieństwem γ β_i pochodzi z rozkładu normalnego $N(0,\tau^2)$, a z prawdopodobieństwem $1-\gamma$ jest to 0. Następnie, w każdej z tych sytuacji wygeneruję także wektor odpowiedzi Y zgodnie ze wzorem $Y = X\beta + \epsilon$. W rezultacie, rozważam sześć sytuacji (dla różnych kombinacji γ i τ) oraz wygenerowane dla nich wektory β i Y. Jednak, przy stosowaniu różnych procedur w dalszej części raportu, nie wykorzystujemy owych informacji z procesu generującego dane.

ESTYMATOR NAJMNIEJSZYCH KWADRATÓW

Jako pierwszy rozważymy w tej sytuacji **estymator najmniejszych kwadratów** $\hat{\beta}^{LS}$ dla wektora β. Wiemy, iż wyraża się on następującym wzorem:

$$\hat{\beta}^{LS} = (X'X)^{-1}X'Y.$$

Jeśli przyjmiemy (jak podano w treści zadania), że $X'X = I$, to otrzymujemy: $\hat{\beta}^{LS} = X'Y$. Z kolei rozkład tego estymatora wygląda następująco:

$$\hat{\beta}^{LS} \sim N(\beta, \sigma^2(X'X)^{-1}).$$

Przyjmując znów z treści zadania, że $X'X = I$, a $\sigma^2 = 1$, dostajemy: $\hat{\beta}^{LS} \sim N(\beta, I)$. Wiemy również, że owy estymator najmniejszych kwadratów $\hat{\beta}^{LS}$ jest w tej sytuacji równy estymatorowi największej wiarogodności $\hat{\beta}_{MLE}$.

Znając już te fakty, napiszę funkcję, która dla podanych macierzy X i wektora Y zwraca wyliczony zgodnie z podanym wzorem **estymator najmniejszych kwadratów** dla wektora β.

ESTYMATOR JAMESA-STEINA ŚCIĄGAJĄCY DO ZERA

Estymator Jamesa-Steina ściągający do zera dla wektora β to estymator postaci: $\hat{\beta}^c = c\hat{\beta}_{MLE}$. Owa wartość c jest znajdowana w taki sposób, aby zminimalizować błąd średniokwadratowy estymatora $\hat{\beta}^c$, w rezultacie czego dostajemy: $c_{opt} = argminMSE(\hat{\beta}^c) = \frac{||\beta||^2}{||\beta||^2+\sigma^2 n}$. Jednak taki estymator nie może być wykorzystany w praktyce, gdyż wartość β nie jest znana. W związku z tym, zauważamy, że: $E||\hat{\beta}||^2 = \sum_{i=1}^n E(\hat{\beta}_i) = \sum_{i=1}^n Var(\hat{\beta}_i) + [E(\hat{\beta}_i)]^2 = \sum_{i=1}^n \sigma^2 + \beta_i^2 = n\sigma^2 + ||\beta||^2$. Wynika z tego, iż $c_{opt} = 1 - \frac{n\sigma^2}{E||\hat{\beta}||^2}$. Zastępując $E||\hat{\beta}||^2$ przez $||\hat{\beta}||^2$ oraz zauważając, że $\frac{\sigma^2}{||\hat{\beta}||^2} \sim Inv\chi^2(n)$, więc $E\frac{\sigma^2}{||\hat{\beta}||^2} = \frac{1}{n-2}$, dostajemy:

$$\hat{\beta}_{JS} = c_{JS}\hat{\beta}_{MLE}, c_{JS} = 1 - \frac{(n-2)\sigma^2}{||\hat{\beta}_{MLE}||^2},$$

pamiętając, iż tutaj $\sigma^2 = 1$. Znając już owy wzór, napiszę funkcję, która dla podanego wektora $\hat{\beta}_{MLE}$ wyznacza omawiany **estymator Jamesa-Steina ściągający do zera**.

ESTYMATOR JAMESA-STEINA ŚCIĄGAJĄCY DO WSPÓLNEJ ŚREDNIEJ

Estymator Jamesa-Steina ściągający do wspólnej średniej dla wektora β to estymator postaci: $\hat{\beta}^d = (1-d)\hat{\beta}_{MLE} + d\bar{\hat{\beta}}_{MLE}$. Owa wartość d jest znajdowana w taki sposób, aby zminimalizować błąd średniokwadratowy estymatora $\hat{\beta}^d$, w rezultacie czego dostajemy: $d_{opt} = argminMSE(\hat{\beta}^d) = \frac{\sigma^2}{var(\beta)+\sigma^2}$. Jednak taki estymator nie może być wykorzystany w praktyce, gdyż wartość β nie jest znana. Postępując podobnie jak w przypadku poprzedniego estymatora, dostajemy następujący wynik:

$$\hat{\beta}_{JS} = (1 - d_{JS})\hat{\beta}_{MLE} + d_{JS}\bar{\hat{\beta}}_{MLE}, d_{JS} = \frac{(n-3)}{(n-1)}\frac{\sigma^2}{var(\hat{\beta}_{MLE})},$$

pamiętając, iż tutaj $\sigma^2 = 1$. Znając już owy wzór, napiszę funkcję, która dla podanego wektora $\hat{\beta}_{MLE}$ wyznacza omawiany **estymator Jamesa-Steina ściągający do wspólnej średniej**.

USTALENIE ISTOTNYCH ZMIENNYCH

W tej części raportu należy **ustalić, które zmienne są istotne**, stosując następujące procedury:

- procedura Bonferroniego,
- procedura Benjaminiego-Hochberga,
- klasyfikator Bayesowski przy założeniu tej samej funkcji straty za błąd pierwszego i drugiego rodzaju.

Procedura Bonferroniego:

Procedura Bonferroniego polega na tym, iż odrzucamy hipotezę H_i, gdy $p_i \leq \frac{\alpha}{n}$, gdzie p_i jest p-wartością dla H_i, a n to liczba hipotez. Zakładamy, iż $\alpha = 0.05$. Napiszę zatem funkcję, która dla podanych p-wartości zwraca informację o tym, które zmienne są istotne na podstawie rozważanej procedury.

P-wartości natomiast potrzebne do zastosowania tej funkcji wyznacza się, rozważając test istotności współczynnika β_i o hipotezach: $H_0 : \beta_i = 0, H_1 : \beta_i \neq 0$. W takiej sytuacji statystyka testowa wygląda następująco: $T_i = \frac{\hat{\beta}_i}{s(\hat{\beta}_i)}$ i przy prawdziwości hipotezy zerowej ma ona rozkład Studenta z $n - p$ stopniami swobody. Jednak w rozważanej sytuacji wiemy, iż $\sigma = 1$, a $X'X = I$, stąd statystyka testowa to $T_i = \frac{\hat{\beta}_i}{1} = \hat{\beta}_i$ i przy hipotezie H_0 ma ona rozkład normalny $N(0, 1)$. Stąd też szukane p-wartości obliczane są wzorem: $p_i = 2 \cdot \left(1 - \Phi\left(|\hat{\beta}_i|\right)\right)$.

Procedura Benjaminiego-Hochberga:

Z kolei w **procedurze Benjaminiego-Hochberga** należy na początku posortować p-wartości w kolejności rosnącej: $p_{(1)} \leq ... \leq p_{(n)}$, gdzie $H_{(1)}, ..., H_{(n)}$ są odpowiadającymi im hipotezami. Następnie, niech i_0 oznacza największy indeks taki, że $p_{(i)} \leq \frac{i}{n}\alpha$. Wówczas, procedura ta mówi o tym, że odrzucamy hipotezy $H_{(1)}, ..., H_{(i_0)}$. Napiszę zatem funkcję, która dla podanych p-wartości (wyliczonych w taki sam sposób jak dla poprzedniej metody) zwraca informację o tym, które zmienne są istotne na podstawie rozważanej procedury.

Klasyfikator Bayesowski przy założeniu tej samej funkcji straty za błąd pierwszego i drugiego rodzaju:

Funkcja straty C polega na tym, iż:

- nie ma straty, gdy podejmujemy prawdziwą decyzję,
- ponosimy pewne koszty w przypadku błędnych decyzji: C_0 za błąd I rodzaju i C_1 za błąd drugiego rodzaju.

Celem jest **minimalizacja ryzyka** - wartości oczekiwanej funkcji straty. **Klasyfikator Bayesowski** mówi o tym, że Γ_1, czyli obszar odrzucenia H_0, to: $\Gamma_1 = \{x : \frac{f_1(x)}{f_0(x)} \geq \frac{C_0}{C_1}\frac{P(H_0)}{P(H_1)}\}$, gdzie f_0 i f_1 to funkcje gęstości prawdopodobieństwa, które opisują rozkład zmiennej losowej x pod warunkiem, że rzeczywista hipoteza to, odpowiednio, H_0 lub H_1. W przypadku tego zadania $C_0 = C_1$.

Rozważamy zatem tutaj następujące hipotezy: $H_0 : \beta_i = 0, H_1 : \beta_i \neq 0$. Należy zauważyć, iż $\hat{\beta}^{MLE} = X'Y = X'(X\beta + \epsilon)$. Przy **prawdziwości hipotezy zerowej** $\beta_i = 0$, więc $\hat{\beta}_i^{MLE} = (X'\epsilon)_i \sim N(0, 1)$. Z kolei przy **prawdziwości hipotezy alternatywnej** $\beta_i \neq 0$, z czego wnioskujemy, że $\beta_i \sim N(0, \tau^2)$. W sytuacji tej $\hat{\beta}_i^{MLE} = \beta_i + (X'\epsilon)_i$, gdzie $\beta_i \sim N(0, \tau^2)$, a $(X'\epsilon)_i \sim N(0, 1)$. Wynika z tego (wykorzystując niezależność), iż wówczas $\hat{\beta}_i^{MLE} \sim N(0, \tau^2 + 1)$.

Zauważamy, iż $P(H_0) = 1 - \gamma$ oraz $P(H_1) = \gamma$. Ponadto, wyliczamy:

$$\frac{f_1(x)}{f_0(x)} = \frac{\frac{1}{\sqrt{2\pi(\tau^2+1)}} e^{\frac{-x^2}{2(\tau^2+1)}}}{\frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}} = \frac{1}{\sqrt{\tau^2+1}} e^{\frac{(x\tau)^2}{2(\tau^2+1)}},$$

a następnie sprawdzamy, które x należą do obszaru odrzucenia H_0:

$$\frac{1}{\sqrt{\tau^2+1}} e^{\frac{(x\tau)^2}{2(\tau^2+1)}} \geq \frac{1-\gamma}{\gamma}.$$

W rezultacie przeprowadzonych obliczeń dostajemy następujący warunek:

$$x^2 \geq \log(\sqrt{\tau^2+1} \cdot \frac{1-\gamma}{\gamma}) \cdot \frac{2(\tau^2+1)}{\tau^2}.$$

Na tej podstawie napiszę funkcję, która dla podanych wartości γ i τ wskazuje istotne zmienne wykorzystując wyznaczony warunek.

Porównanie wyników:

Przedstawię teraz tabelę ukazującą **liczbę istotnych zmiennych** uzyskanych przy wykorzystaniu wyżej omówionych procedur w sześciu rozważanych sytuacjach (dla różnych kombinacji γ i τ).

Table 1: Liczba istotnych zmiennych

gamma	tau	Bonferroni	BH	k. Bayesowski
0.01	5.575	4	5	5
0.05	5.575	31	38	38
0.10	5.575	65	86	86
0.01	11.151	6	6	6
0.05	11.151	37	46	41
0.10	11.151	59	68	68

Na podstawie przedstawionej tabeli możemy wyciągnąć wniosek, iż **zwiększanie γ prowadzi do wzrostu liczby istotnych zmiennych** w każdej z procedur. Dzieje się tak, gdyż γ reprezentuje wagę, jaką przypisujemy rozkładowi normalnemu w procesie generowania wektora współczynników regresji β. Im wyższa wartość γ, tym większa szansa, że współczynniki β_i będą różne od zera, co powoduje wykrycie większej liczby zmiennych jako istotne. Podobnie, **zwiększanie τ powoduje wzrost liczby istotnych zmiennych**, choć efekt ten jest zauważalny przy mniejszych wartościach γ. Jest tak dlatego, że wartość τ wpływa na zmienność współczynników β - większa wartość powoduje większą zmienność współczynników, co z kolei prowadzi do wykrywania większej liczby zmiennych jako istotnych.

Ponadto, możemy zauważyć, że **procedura Bonferroniego** generuje najmniejszą liczbę zmiennych uznanych za istotne. Jest ona jedną z bardziej restrykcyjnych procedur, która stosuje bardzo surowe kryteria odrzucenia hipotez zerowych. Natomiast dwie pozostałe metody - **procedura Benjaminiego-Hochberga** i **klasyfikator Bayesowski** - wykrywają więcej zmiennych jako istotne, a ich wyniki są bardzo do siebie zbliżone. Jest tak dlatego, że procedury te są mniej restrykcyjne.

ESTYMATORY "UCIĘTE"

W tej części raportu, dla każdej wyżej omówionej procedury, należy wyznaczyć **"ucięte" estymatory** wektora β, które konstruuje się w następujący sposób:

$$\hat{\beta}_i^{uc} = \begin{cases} \hat{\beta}_i^{LS}, & \text{jeżeli odrzucono } H_{0i} : \beta_i = 0; \\ 0, & \text{w przeciwnym wypadku.} \end{cases}.$$

Zatem, tam, gdzie zgodnie z procedurą wielokrotnego testowania $\beta_i \neq 0$, estymujemy β_i za pomocą estymatora największej wiarogodności, a tam, gdzie zgodnie z tą procedurą $\beta_i = 0$, estymujemy β_i za pomocą 0.

Owe estymatory będę wyznaczać, modyfikując wektor $\hat{\beta}_{MLE}$ w taki sposób, by miał on zera w miejscach odpowiadających zmiennym, które poszczególna procedura uznała za nieistotne.

PORÓWNANIE ESTYMATORÓW POD KĄTEM BŁĘDU KWADRATOWEGO

W tej części porównamy wszystkie sześć omówionych estymatorów pod kątem **błędu kwadratowego** w sześciu rozważanych sytuacjach (dla różnych kombinacji γ i τ).

Błąd kwadratowy to suma kwadratów różnic między oszacowanymi wartościami $\hat{\beta}$ a rzeczywistymi wartościami β. Mierzy on błąd dla jednej realizacji eksperymentu i wylicza się go następującym wzorem:

$$SE = ||\hat{\beta} - \beta||^2.$$

Przedstawię teraz zatem tabelę uzyskanych wyników w rozważanej sytuacjii.

Table 2: Porównanie błędu kwadratowego

gamma	tau	est. LS	est. JS śc. do zera	est. JS śc. do średniej	est. ucięty (Bonf)	est. ucięty (BH)	est. ucięty (Bayes)
0.01	5.575	956.573	171.353	171.298	30.353	21.721	21.721
0.05	5.575	1010.389	660.656	660.941	1213.413	1122.647	1122.647
0.10	5.575	985.468	813.013	812.619	485.473	307.751	307.751
0.01	11.151	1012.689	436.775	437.473	1025.528	1025.528	1025.528
0.05	11.151	967.262	858.322	859.250	119.267	133.410	114.287
0.10	11.151	938.411	869.386	869.128	1087.508	969.203	969.203

Analizując ową tabelę, należy pamiętać, iż są to wyniki jedynie dla pojedynczego doświadczenia, więc losowość odgrywa tutaj dużą rolę. Widzimy jednak, że **estymator najmniejszych kwadratów** osiąga jedne z najwyższych błędów kwadratowych. Jest tak dlatego, że nie stosuje on żadnej formy regularizacji, a ponieważ niektóre współczynniki β są zerowe, dopasowuje on szum, co prowadzi do dużego błędu. Widzimy również, że **estymatory Jamesa-Steina** działają dużo lepiej - znacząco zmniejszają błąd kwadratowy. Zmniejszają one błąd kwadratowy, wprowadzając pewne obciążenie i redukując wariancję. Następnie, możemy zauważyć, iż **estymatory "ucięte"** w wielu rozważanych sytuacjach mają najmniejszy błąd kwadratowy. Działają one lepiej, gdyż usuwają zbędny szum, zamiast tylko go redukować. Ponadto, użycie korekty Bonferroniego powoduje wyższy błąd kwadratowy, gdyż procedura ta nadmiernie "ucina" zmienne - przesadnie upraszcza model, ignorując część istotnych zmiennych. Dokładniejsze wnioski będzie można wyciągnąć w dalszej części raportu, przy powtórzeniu owego doświadczenia 1000 razy.

PORÓWNANIE PROCEDUR TESTOWANIA POD KĄTEM SUMY LICZBY BŁĘDÓW

W tej części należy porównać omawiane procedury testowania pod kątem **sumy błędów pierwszego i drugiego rodzaju**. **Błąd pierwszego rodzaju** to błąd polegający na odrzuceniu hipotezy zerowej w sytuacji, gdy jest ona prawdziwa. Z kolei **błąd drugiego rodzaju** to błąd polegający na przyjęciu hipotezy zerowej w sytuacji, gdy hipoteza alternatywna jest prawdziwa. Obliczę zatem sumę tych błędów w rozważanych sześciu sytuacjach i przedstawię wyniki w tabeli.

Table 3: Porównanie sumy błędów

gamma	tau	Bonferroni	BH	k. Bayesowski
0.01	5.575	7	6	6
0.05	5.575	40	33	33
0.10	5.575	56	41	41
0.01	11.151	5	5	5
0.05	11.151	14	15	14
0.10	11.151	26	17	17

Ponownie należy pamiętać, iż są to jedynie wyniki uzyskane w pojedynczym doświadczeniu. Widzimy jednak, że **procedura Bonferroniego** w prawie każdej sytuacji ma najwyższą sumę błędów. Metoda ta stosuje mocne poprawki na wielokrotne testowanie, więc odrzuca mniej hipotez, z czego wynika jej duża liczba błędów drugiego rodzaju. Możemy także zauważyć, iż **klasyfikator Bayesowski** osiąga najmniejsze sumy błędów, co może wynikać z tego, że korzysta on z dodatkowej wiedzy o rozkładzie β. Bardzo podobne, niskie wyniki uzyskała też **procedura Benjaminiego-Hochberga** - jest ona mniej restrykcyjna niż procedura Bonferroniego i dopuszcza więcej istotnych zmiennych, dzięki czemu lepiej równoważy ona liczbę błędów pierwszego i drugiego rodzaju. Widzimy również, że im większa wartość γ, tym większa liczba popełnionych błędów, gdyż wówczas występuje więcej niezerowych współczynników β_i. Z kolei im większa wartość τ, tym

mniejsza liczba błędów, gdyż wówczas łatwiej odróżnić niezerowe β_i od szumu, więc procedury mają mniej błędów.

POWTÓRZENIE DOŚWIADCZENIA 1000 RAZY

W tej części raportu przeprowadzone doświadczenie zostanie powtórzone **1000 razy** dla każdej kombinacji γ i τ. Wygeneruję zatem 1000 razy dane analogicznie jak w pierwszej części raportu, a następnie napiszę funkcję, która wyznacza dla nich wszystkie estymatory i stosuje rozważane procedury tak, jak poprzednio.

PORÓWNANIE ESTYMATORÓW POD KĄTEM BŁĘDU ŚREDNIOKWADRATOWEGO

Błąd średniokwadratowy, czyli **MSE**, to wartość oczekiwana omawianego wcześniej błędu kwadratowego:

$$MSE = E(SE).$$

Obliczę go jako średnią uzyskanych błędów kwadratowych dla każdej kombinacji γ i τ ze wszystkich wykonanych powtórzeń. Przedstawię teraz tabelę z uzyskanymi wynikami.

Table 4: Porównanie błędu średniokwadratowego

gamma	tau	est. LS	est. JS śc. do zera	est. JS śc. do średniej	est. ucięty (Bonf)	est. ucięty (BH)	est. ucięty (Bayes)
0.01	5.575	997.939	225.055	225.780	36.241	32.780	32.309
0.05	5.575	999.049	603.731	604.112	177.973	132.013	131.599
0.10	5.575	999.814	753.186	753.402	351.771	234.040	232.507
0.01	11.151	998.369	512.286	512.699	25.726	25.709	24.379
0.05	11.151	1000.076	855.104	855.261	124.302	104.429	101.282
0.10	11.151	1000.466	923.872	923.955	251.923	193.281	188.270

Widzimy, iż **estymator najmniejszych kwadratów** osiąga najwyższy błąd średniokwadratowy we wszystkich przypadkach. Dzieje się tak, gdyż w sytuacjach, gdy wiele współczynników jest równe zeru, nie wykorzystuje on tej informacji - nie zawiera żadnych mechanizmów regularizacji. Następnie, widzimy, iż **oba estymatory Jamesa-Steina** mają zbliżone do siebie, znacznie niższe błędy średniokwadratowe. Stosują one kurczenie współczynników w kierunku zera lub wspólnej średniej, co pomaga zmniejszyć wariancję estymatora. W modelach, gdzie wiele współczynników jest równe zero, estymatory Jamesa-Steina lepiej radzą sobie z identyfikacją i kurczeniem tych współczynników. Znamy również twierdzenie, że dla $n \geq 3$ (w tym przypadku $n = 1000$) owe estymatory są zawsze lepsze pod względem błędu średniokwadratowego od estymatora największej wiarogodności, co potwierdza wyciągnięte wnioski. Ponadto, zauważamy, że **estymatory "ucięte"** jeszcze bardziej obniżają błąd średniokwadratowy, jednak różnią się między sobą skutecznością. **Procedura Bonferroniego** jest najbardziej restrykcyjna i odrzuca najmniej hipotez zerowych, co skutkuje większą liczbą błędów drugiego rodzaju i wyższym MSE. **Procedura Benjaminiego-Hochberga** ma niższy błąd średniokwadratowy niż poprzednia metoda, gdyż odrzuca ona więcej hipotez zerowych i pozwala na wykrycie większej liczby niezerowych współczynników. Z kolei **estymator Bayesowski** ma najniższe MSE, co wynika z optymalnego balansu między błędami pierwszego i drugiego rodzaju. Warto także zauważyć, iż **większe wartości** γ powodują większą wartość MSE, gdyż wówczas więcej współczynników β_i jest niezerowych, co utrudnia poprawne estymowanie. Ponadto, gdy **wartość** τ **rośnie**, to błąd średniokwadratowy dla estymatora najmniejszych kwadratów i estymatorów Jamesa-Steina również rośnie, gdyż wtedy współczynniki β_i są większe. Natomiast pozostałe trzy "ucięte" estymatory mają wówczas mniejsze MSE, gdyż procedury selekcji zmiennych mogą łatwiej wykryć istotne zmienne i usunąć tylko te, które rzeczywiście są bliskie 0.

Oszacowanie teoretyczne:

Wiemy, iż dla **estymatora największej wiarogodności** błąd średniokwadratowy powinien wynosić $n\sigma^2$, czyli w naszym przypadku $1000 \cdot 1 = 1000$. Z kolei dla **estymatora Jamesa-Steina ściągającego do zera** owe MSE powinno być równe $\frac{n\sigma^2||\beta||^2}{||\beta||^2+\sigma^2 n}$, a dla **estymatora Jamesa-Steina ściągającego do wspólnej średniej** liczba ta to: $n\sigma^2 - \frac{\sigma^4(n-1)}{(var(\beta)+\sigma^2)}$. Wynika z tego, iż dla $n \geq 3$ esymatory Jamesa-Steina zawsze osiągają niższy błąd średniokwadratowy niż estymator największej wiarogodności. Obliczę teraz teoretyczne oszacowanie dla tych trzech estymatorów, wykorzystując wygenerowane wektory β, aby sprawdzić poprawność estymacji. Dla każdej kombinacji γ i τ wyznaczę średnią uzyskanych teoretycznych oszacowań MSE i przedstawię wyniki w tabeli.

Table 5: Porównanie błędu średniokwadratowego - teoretycznie

gamma	tau	est. LS	est. JS śc. do zera	est. JS śc. do średniej
0.01	5.575	1000	223.505	224.283
0.05	5.575	1000	601.546	601.956
0.10	5.575	1000	752.394	752.647
0.01	11.151	1000	511.653	512.131
0.05	11.151	1000	854.458	854.602
0.10	11.151	1000	923.706	923.780

Porównując te wyniki z tabelą estymowanych wartości, widzimy, iż **są one bardzo zbliżone**, co potwierdza poprawność przeprowadzonych estymacji.

PORÓWNANIE PROCEDUR TESTOWANIA POD KĄTEM WARTOŚCI OCZEKIWANEJ SUMY LICZBY BŁĘDÓW

W tej części należy porównać analizowane procedury pod kątem **wartości oczekiwanej sumy liczby błędów pierwszego i drugiego rodzaju**. Wyliczę zatem sumę liczby błędów dla wszystkich wygenerowanych danych analogicznie jak w przypadku pojedynczego powtórzenia, a następnie dla poszczególnych kombinacji wartości γ i τ obliczę średnią uzyskanych wyników i ukażę je w tabeli.

Table 6: Porównanie wartości oczekiwanej sumy błędów

gamma	tau	Bonferroni	BH	k. Bayesowski
0.01	5.575	5.275	5.044	5.001
0.05	5.575	26.460	22.777	22.731
0.10	5.575	52.606	42.587	42.322
0.01	11.151	2.857	2.866	2.769
0.05	11.151	14.158	12.738	12.466
0.10	11.151	28.533	24.189	23.612

Na przedstawionej tabeli widzimy, iż zastosowanie **korekty Bonferroniego** w praktycznie wszystkich przypadkach daje największą wartość oczekiwaną sumy liczby błędów. Jest tak dlatego, iż metoda ta zbyt surowo kontroluje błędy pierwszego rodzaju kosztem wielu błędów drugiego rodzaju. Dalej, **korekta Benjaminiego-Hochberga** osiąga niższe średnie wartości sumy liczby popełnionych błędów, gdyż mniej surowo kontroluje ona błędy pierwszego rodzaju, w związku z czym jest bardziej zrównoważona. Natomiast dla **klasyfikatora Bayesowskiego** obserwujemy najniższe wyniki - optymalizuje on kompromis między błędami pierwszego i drugiego rodzaju. Uwzględnia on prawdopodobieństwa i minimalizuje błąd globalny. Warto również zauważyć, że **im większa wartość** γ, tym większe oczekiwane sumy liczby popełnionych

błędów - występuje wówczas więcej istotnych współczynników β, więc trudniej poprawnie identyfikować zmienne. Z kolei wraz **ze wzrostem wartości** τ omawiane wyniki maleją - wtedy wartości niezerowych współczynników β są bardziej wyraźne, co ułatwia ich wykrycie.

PODSUMOWANIE

Raport ten pozwolił porównać działanie sześciu estymatorów wektora współczynników β w rozważanej sytuacji. Pokazał on, że **estymatory Jamesa-Steina** znacznie obniżają błąd średniokwadratowy w stosunku do estymatora największej wiarogodności. Także **estymatory "ucięte"** znacząco obniżają uzyskany błąd średniokwadratowy - są one asymptotycznie optymalne, jeżeli sprawdzone jest założenie rzadkości. Porównane zostały także procedury testowania - **procedura Bonferroniego** osiąga największe średnie liczby popełnionych błędów, a **klasyfikator Bayesowski** pod tym względem jest najbardziej optymalny.