Raport 3

Magdalena Potok

2023-11-15

Zadanie 1

Przedział ufności dla średniej $\bar{X}$ w modelu normalnym $N(\mu, \sigma^2)$ o znanej wariancji σ^2 na poziomie ufności $1-\alpha$ wygląda następująco:

$$[\bar{X} - z^* \frac{\sigma}{\sqrt{n}} \ ; \ \bar{X} + z^* \frac{\sigma}{\sqrt{n}}]$$

gdzie $\bar{X}$ to średnia próbkowa zmiennych niezależnych $X_1, X_2, ..., X_n$, czyli nieobciążony estymator średniej oczekiwanej, a z^* to kwantyl z rozkładu normalnego rzędu $1 - \frac{\alpha}{2}$.

Zadanie 2

W tym zadaniu wygeneruję $n \in \{20, 50, 100\}$ obserwacji z rozkładu:

(a) normalnego o parametrach: $\mu = 0, \sigma \in \{1, 2, 3\}$,
(b) logistycznego o parametrach: $\mu = 0, \sigma \in \{1, 2, 3\}$,
(c) Cauchy'ego o parametrach: $\mu = 0, \sigma \in \{1, 2, 3\}$,
(d) wykładniczego z parametrem $\lambda \in \{1, 1/2, 1/3\}$,
(e) chi-kwadrat z ν stopniami swobody: $\nu \in \{1, 2, 3\}$.
Na podstawie tych rozkładów wyznaczę przedziały ufności średniej na poziomie ufności 0.95 omówionego w zadaniu 1. oraz jego długość. Doświadczenie powtórzę 10 000 razy, aby oszacować prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności oraz jego długość. Wyniki przedstawię w tabeli.

n		N(0,1)	N(0,2)	N(0,3)	L(0,1)	L(0,2)	L(0,3)	C(0,1)	C(0,2)	C(0,3)
20	dł	0.877	1.753	2.630	0.877	1.753	2.630	0.877	1.753	2.630
20	prawd	0.951	0.949	0.951	0.729	0.726	0.720	0.264	0.266	0.257
50	dł	0.554	1.109	1.663	0.554	1.109	1.663	0.554	1.109	1.663
50	prawd	0.952	0.951	0.950	0.725	0.723	0.721	0.170	0.172	0.173
100	dł	0.392	0.784	1.176	0.392	0.784	1.176	0.392	0.784	1.176
100	prawd	0.950	0.950	0.947	0.722	0.722	0.720	0.122	0.129	0.123

n		Exp(1)	Exp(1/2)	Exp(1/3)	Chi(1)	Chi(2)	Chi(3)
20	dł	0.877	1.753	2.630	1.240	1.753	2.147
20	prawd	0.951	0.945	0.951	0.955	0.953	0.954
50	dł	0.554	1.109	1.663	0.784	1.109	1.358
50	prawd	0.948	0.954	0.956	0.949	0.950	0.951
100	dł	0.392	0.784	1.176	0.554	0.784	0.960
100	prawd	0.952	0.949	0.946	0.950	0.949	0.951

Z tabelki można zauważyć, że w przypadku każdego z rozkładów długość przedziałów ufności dla średniej się zmniejsza dla coraz większego n. Oznacza to, że z większą ilością próby jesteśmy w stanie precyzyjniej wskazać ten przedział.
Dla rozkładu normalnego prawdopodobieństwo w każdym z przypadków wyniosło ok. 0.95, czyli dokłądnie $1-\alpha$, jest to zgodne z naszymi oczekiwaniami, bo właśnie tyle wynosi pewność, że rzeczywisty parametr dla rozkłądu normalnego znajduje się w przedziale ufności o poziomie istotności $\alpha = 0.05$. Można również zauważyć, że dla rozkładu normalnego, logistycznego oraz Cauchy'ego wzrost skali σ powoduje wydłużenie przedziału ufności, co jest logiczne, ponieważ ten parametr odpowiada za rozrzut obserwacji, a co za tym idzie, mniej precyzyjniej jesteśmy w stanie oszacować średnią naszej próby.
Prawdopodobieństwo, że przedział ufności pokryje prawdziwą wartość średniej próby dla rozkładu logistycznego wynosi, dla każdej wielkości próby, około 0.722, jest to mniej niż dla rozkładu normalnego. Wynika to z asymetrycznego charakteru rozkładu logistycznego.
Dla rozkładu Cauchy'ego to prawdopodobieństwo wychodzi najniższe ze wszystkich, wynika to z tego, że ten rozkład charakteryzuje się szerokimi ogonami. Sprawia to, że estymacja przedziału ufności jest trudniejsza. Można też zauważyć, że ze wzrostem próby to prawdopodobieństwo maleje.
W przypadku rozkładów wykładniczego oraz chi-kwadrat zmniejszenie parametru λ i wzrost parametru ν również powoduje zwiększenie długości przedziałów. Wzrost próby, zgodnie z nasyzmi oczekiwaniami, zwęża przedziały ufności. Dla każdego z przypadku tych rozkładów prawodpodobieństwo zawarcia średniej w przedziale wynosi około 0.95, tak jak dla rozkładu normalnego.

Zadanie 3

Przedział ufności dla średniej w modelu normalnym o nieznanej wariancji na poziomie ufności $1-\alpha$ konstruujemy przy użyciu przedziału ufności opartego na rozkładzie t-Studenta. Wygląda on tak:

$$[\bar{X} - t^* \frac{s}{\sqrt{n}} \ ; \ \bar{X} + t^* \frac{s}{\sqrt{n}}]$$

gdzie $\bar{X}$ to średnia próbkowa niezazleżnych zmiennych $X_1, X_2, ..., X_n$ pochodzących z rozkładu normalnego $N(\mu, \sigma^2)$, s^2 to próbkowa wariancja, czyli s to nieobciążony estymator σ, a t^* to kwantyl rozkładu t-Studenta z n-1 stopniami swobody i rzędu $1 - \frac{\alpha}{2}$. Przedział ten bazuje na założeniu, że próbka ma rozkład normalny, a wariancja jest nieznana. Zastosowanie rozkładu t-Studenta uwzględnia większą niepewność związanąn z estymacją wariancji.

Zadanie 4

W tym zadaniu powtórzę eksperyment z zadania 2., tym razem oszacuję prawodpodobieństwo pokrycia nieznanej średniej przez przedział ufności z zadania 3. na poziomie ufności 0.95 oraz jego długość.

n		N(0,1)	N(0,2)	N(0,3)	L(0,1)	L(0,2)	L(0,3)	C(0,1)	C(0,2)	C(0,3)
20	dł	0.928	1.850	2.777	1.666	3.334	4.990	29.508	42.492	79.740
20	prawd	0.955	0.952	0.953	0.955	0.949	0.951	0.978	0.980	0.980
50	dł	0.565	1.130	1.694	1.021	2.051	3.073	29.465	154.889	64.288
50	prawd	0.950	0.950	0.953	0.950	0.952	0.950	0.979	0.979	0.980
100	dł	0.396	0.791	1.187	0.717	1.433	2.152	22.762	58.414	57.176
100	prawd	0.948	0.951	0.951	0.951	0.951	0.949	0.980	0.979	0.980

n		Exp(1)	Exp(1/2)	Exp(1/3)	Chi(1)	Chi(2)	Chi(3)
20	dł	0.895	1.803	2.694	1.228	1.799	2.214
20	prawd	0.919	0.920	0.922	0.891	0.921	0.927
50	dł	0.558	1.117	1.673	0.780	1.119	1.372
50	prawd	0.935	0.938	0.935	0.925	0.937	0.942
100	dł	0.394	0.785	1.177	0.552	0.786	0.964

n		Exp(1)	Exp(1/2)	Exp(1/3)	Chi(1)	Chi(2)	Chi(3)
100	prawd	0.943	0.940	0.944	0.934	0.943	0.944

Porównując do wyników z zadania 2. wszystkie przedziału ufności są szersze (mają większą długość). Wynika to z tego, że te przedziały ufności uwzględniają większą niepewność związaną z estymacją wariancji. Najbardziej wyróżniającymi wynikami są długości przedziałów dla rozkładu Cauchy'ego, wynika to z tego, że bardzo ciężko oszacować wariancje tego rozkładu na podstawie próby. Wszystkie prawdopodobieństwa zbliżyły się do wartości 0.95, czyli dla rozkładu logistycznego oraz Cauchy'ego znacząco te wartości się zwiększyły. Wynika to ze zwiększenia się tych przedziałów, im większe one są, tym większe prawdopodobieństwo, że średnia się w nich znajdzie.

Zadanie 5

Przedział ufności wariancji w modelu normalnym o znanej średniej na poziomie ufności $1-\alpha$ kontruujemy w następujący sposób:

$$[\frac{n}{\chi^2_{\alpha/2}}s^2\ ;\ \frac{n}{\chi^2_{1-\alpha/2}}s^2]$$

gdzie s^2 to próbkowa wariancja, $\chi^2_{\alpha/2}$ to kwantyl rzędu $\frac{\alpha}{2}$ z rozkładu chi-kwadrat o n stopniach swobody, a $\chi^2_{1-\alpha/2}$ to kwantyl tego samego rozkładu rzędu $1-\frac{\alpha}{2}$. Przedział ten zapewnia nam pewność, że prawdziwa wariancja mieści się w nim z prawdopodobieństwem $1-\alpha$.

Zadanie 6

W tym zadaniu powtórzyłam eksperyment z zadania 2., na jego podstawie oszacuję prawdopodobieństwo pokrycia nieznanej wariancji przez przedziały ufnośći z zadania 5. na poziomie ufności 0.95 oraz jego długość.

n		N(0,1)	N(0,2)	N(0,3)	L(0,1)	L(0,2)	L(0,3)	C(0,1)	C(0,2)	C(0,3)
20	dł	1.495	5.977	13.517	1.666	19.781	44.544	258374.4	1212337.783	778695.437
20	prawd	0.950	0.948	0.952	0.955	0.072	0.071	0.0141	0.012	0.0141
50	dł	0.843	3.371	7.608	2.783	11.111	24.996	301296.206	2179795.807	6421994.201
50	prawd	0.950	0.953	0.949	0.001	0.001	0.001	0	0.000	0
100	dł	0.576	2.296	5.175	1.889	7.588	17.073	256338446.377	1498823.363	2832641198.324
100	prawd	0.948	0.951	0.946	0.000	0.000	0.000	0	0.000	0

n		Exp(1)	Exp(1/2)	Exp(1/3)	Chi(1)	Chi(2)	Chi(3)
20	dł	1.502	5.968	13.422	3.043	5.936	8.913
20	prawd	0.744	0.744	0.747	0.596	0.748	0.811
50	dł	0.847	3.400	7.617	1.690	3.405	5.092
50	prawd	0.703	0.701	0.702	0.578	0.703	0.776
100	dł	0.577	2.299	5.159	1.144	2.287	3.451
100	prawd	0.693	0.688	0.686	0.569	0.695	0.762

To co można z tabelki wyczytać, to fak, że w rozkładach normalnym, logistycznym oraz Cauchy'ego dla każdej wielkości próby, wzrost σ powoduje zwiększenie przedziału ufności dla wariancji. W rozkładzie normalnym prawdopodobieństwo utrzymuje się tak, jak w poprzednich przedziałach ufności, czyli ok. 0.95. Dla rozkładu logistycznego wraz z zwiększaniem się parametru σ to prawdopodobieństwo maleje oraz wraz z większaniem się rozmiaru próbki, dla n = 50 i n = 100 to prawdopodobieństwo jest znikome. W przypadku rozkładu Cauchy'ego ma on najniższe prawdopodobieństwo z całej tabelki, dla n = 50 i n = 100 wynosi po prostu 0, a dla n = 20 jest bardzo małe, bliskie 0. Długość przedizału dla rozkładów wykładniczego oraz chi-kwadrat również zależy od ich parametrów. W przypadku wykładniczego zmniejszenie się parametru λ powoduje zwiększanie długości przedziału, a dla chi-kwadrat zwiększanie parametru ν daje ten sam efekt, czyli szerszy przedział. Zwiększanie próby w obu tych przedziałach wpływa na prawdopodobieństwo tego, czy prawdziwa wariancja należy do przeedziału ufności w taki sposób, że to prawdopodobieństwo maleje. Ale nie jest to tak niski wynik, jak w przypadku rozkładu logistycznego lub Cauchy'ego, bo plasuje się w okolicach 0.7, co jest satysfakcjonującym wynikiem.

Zadanie 7

Przedział ufności dla wariancji w modelu normalnym o nieznanej średniej na poziomie ufności $1-\alpha$ wygląda następująco:

$$[\frac{n-1}{\chi^2_{\alpha/2}}s^2\ ;\ \frac{n-1}{\chi^2_{1-\alpha/2}}s^2]$$

gdzie s^2 to próbkowa wariancja, $\chi^2_{\alpha/2}$ to kwantyl rzędu $\frac{\alpha}{2}$ z rozkładu chi-kwadrat o n stopniach swobody, a $\chi^2_{1-\alpha/2}$ to kwantyl tego samego rozkładu rzędu $1-\frac{\alpha}{2}$. Przedział ten zapewnia nam pewność, że prawdziwa wariancja mieści się w nim z prawdopodobieństwem $1-\alpha$.

Zadanie 8

Powtórzę ponownie eksepryment z zadania 2., na jego podstawie oszacuję prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności z zadania 7. na poziomie ufności 0.95 oraz jego długość.

n		N(0,1)	N(0,2)	N(0,3)	L(0,1)	L(0,2)	L(0,3)	C(0,1)	C(0,2)	C(0,3)
20	dł	1.642	6.560	14.724	5.366	21.678	48.375	3918454.354	4240383.167	6660768.789
20	prawd	0.947	0.950	0.949	0.064	0.060	0.060	0.013	0.011	0.013
50	dł	0.874	3.496	7.859	2.873	11.448	25.848	3077160.857	558212.854	2522094.874
50	prawd	0.949	0.945	0.949	0.001	0.001	0.001	0.000	0.000	0.000
100	dł	0.586	2.340	5.273	1.924	7.688	17.307	235411.065	1439585.419	8342983.658
100	prawd	0.951	0.952	0.953	0.000	0.000	0.000	0.000	0.000	0.000

n		Exp(1)	Exp(1/2)	Exp(1/3)	Chi(1)	Chi(2)	Chi(3)
20	dł	1.646	6.544	14.835	3.299	6.543	9.766
20	prawd	0.762	0.765	0.761	0.620	0.771	0.829
50	dł	0.877	3.481	7.886	1.736	3.511	5.267
50	prawd	0.716	0.710	0.709	0.579	0.716	0.779
100	dł	0.584	2.341	5.234	1.163	2.337	3.484
100	prawd	0.693	0.688	0.705	0.562	0.694	0.768

Porównując do poprzedniego zadania, gdzie konstruowaliśmy przedziały ufności wariancji przy znanej średniej, to tym razem średnie długości przedziałów wyszły większe. Wynika to z niepewności związanej z estymacją wariancji. Zwiększanie próbki, w każdym z rozkładów, powoduje zwężenie przedziału ufności. Takie same wnioski, co do zmian parametru i wpływu jego na długość przedziało, co w zadaniu 6. To samo można powiedzieć o prawdopodobieństwie, w rozkładzie normalnym mamy oczekiwaną wartość 0.95, w rozkładzie logistycznym dla większych n to prawdopodobieństwo jest bardzo małe, bliskie 0. W rozkładach chi-kwadrat oraz wykłądniczym plasuje się w okolicach 0.7 dla każdej wielkości próby. Natomiast najmniejsze, znikome prawdopodobieństwo jest dla rozkładu Cauchy'ego, czyli dokładnie jest tak samo, jak w przypadku, gdy znaliśmy średnią.

Zadanie 9

Podam asymptotyczny przedział ufności dla proporcji na poziomie ufności $1-\alpha$.

$$[\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ; \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$$

gdzie $\hat{p}$ to estymator proporcji, $z_{1-\alpha/2}$ to kwantyl rozkłądu normalnego rzędu $1-\frac{\alpha}{2}$, n to wielkość próby.

Zadanie 10

Powtórzę eksperyment z zadania 2. dla podpunktów a, b i c. Na jego podstawie oszacuje prawdopodobieństwo pokrycia neiznanej proporcji dodatnich obserwacji przez przedział ufności z zadania 9. na poziomie ufności 0.95 oraz jego długość.

n		N(0,1)	N(0,2)	N(0,3)	L(0,1)	L(0,2)	L(0,3)	C(0,1)	C(0,2)	C(0,3)
20	dł	0.427	0.427	0.427	0.427	0.427	0.427	0.427	0.427	0.427
20	prawd	0.959	0.960	0.960	0.956	0.960	0.962	0.958	0.959	0.960
50	dł	0.274	0.274	0.274	0.274	0.274	0.274	0.274	0.274	0.274
50	prawd	0.933	0.936	0.938	0.935	0.935	0.934	0.939	0.935	0.933
100	dł	0.195	0.195	0.195	0.195	0.195	0.195	0.195	0.195	0.195
100	prawd	0.943	0.946	0.943	0.944	0.945	0.942	0.942	0.944	0.946

Jak spojrzymy na długości przedziałów dla zadanego n, to można zauważyć, że wszystkie długości są takie same. Oznacza to, że w przypadku estymacji asymptotycznej proporcji długość przedziału ufności jest stała, niezależnie od wybranego rozkładu. Wzrost n zmniejsza nam ten przedział, czyli zwiększa precyzję estymacji. Prawdopodobieństwo w każdym przypadku jest podobne i plasuje się w okolicach 0.95.