ЭЛЕКТРОННЫЙ НАУЧНЫЙ ЖУРНАЛ «ДНЕВНИК НАУКИ»

УДК 519.233

DOI 10.51691/2541-8327_2023_12_32

АНАЛИЗ РАСПРЕДЕЛЕНИЯ ВЫСОТЫ МОРСКИХ ВОЛН. СРАВНЕНИЕ ОЦЕНОК И ПРИМЕНЕНИЕ КРИТЕРИЯ СОГЛАСИЯ ПИРСОНА

Облакова Т.В.

к.ф.-м.н., доцент кафедры,

Московский государственный технический университет имени Н. Э. Баумана,

Москва, Россия

Зубарев К.М.

Старший преподаватель кафедры

Московский государственный технический университет имени Н. Э. Баумана,

Москва, Россия

Яковлев Д.Ю.

студент

Московский государственный технический университет имени Н. Э. Баумана,

Москва, Россия

Аннотация. Настоящая работа посвящена исследованию статистического распределения высоты морских волн в открытом океане. Мы представляем подробный анализ выборки, включающий построение гистограммы, плотности распределения и эмпирической функции распределения с целью определения характера данных.

Сначала проводится статистический анализ, используя визуальные методы, такие как гистограмма, которая позволяет нам предварительно оценить форму распределения. Параллельно строится плотность распределения и эмпирическая функция распределения для получения более детального представления о структуре данных.

Далее, мы оцениваем параметры распределения высоты морских волн двумя методами: методом максимального правдоподобия (ММП) и методом моментов (ММ). Эта часть исследования направлена на выявление эффективности каждого метода в адекватном описании выборки.

Наконец, для подтверждения гипотезы о соответствии распределения Рэлея проводится статистический тест с использованием критерия хи-квадрат Пирсона. Результаты исследования имеют значение как для глубинного понимания природы морских волн, так и для развития эффективных методов их моделирования и прогнозирования в различных приложениях.

Ключевые слова: морские волны, статистическое распределение, метод максимального правдоподобия, метод моментов, критерий согласия Пирсона, эмпирическая функция распределения.

ANALYSIS OF THE DISTRIBUTION OF THE HEIGHT OF SEA WAVES: COMPARISON OF ESTIMATES AND APPLICATION OF THE PEARSON CONSENSUS CRITERION

Oblakova T.V.

Candidate of physical and mathematical Sciences, Associate Professor

Bauman Moscow state technical University,

Moscow, Russia

Zubarev K. M.

Senior Lecturer

Bauman Moscow state technical University,

Moscow, Russia

Yakovlev D.Y.

student

Bauman Moscow state technical University,

Moscow, Russia

Abstract. This work is devoted to the study of the statistical distribution of the height of sea waves in the open ocean. We present a detailed analysis of the sample, including the construction of a histogram, distribution density and empirical distribution function in order to determine the nature of the data.

First, a statistical analysis is performed using visual methods such as a histogram, which allows us to pre-evaluate the shape of the distribution. In parallel, the distribution density and the empirical distribution function are being constructed to obtain a more detailed understanding of the data structure.

Next, we estimate the parameters of the distribution of the height of sea waves by two methods: the maximum likelihood method (MMP) and the method of moments (MM). This part of the study is aimed at identifying the effectiveness of each method in an adequate description of the sample.

Finally, to confirm the hypothesis of the correspondence of the Rayleigh distribution, a statistical test using Pearson's chi-squared criterion is performed. The results of the study are important both for a deep understanding of the nature of sea waves, and for the development of effective methods for modeling and forecasting them in various applications.

Key words: sea waves, statistical distribution, maximum likelihood method, method of moments, Pearson chi-squared criterion, empirical distribution function.

Введение

Морские волны представляют собой явление, непрерывно волнующее умы исследователей в области океанографии и морских технологий. Высота волн играет ключевую роль в различных морских операциях, включая безопасность судоходства, проектирование морских инженерных конструкций и прогнозирование погоды [3; 7]. Однако, несмотря на важность этого параметра, статистический характер распределения высоты морских волн остается предметом активного исследования.

Цель нашего исследования заключается в определении статистического распределения высоты морских волн и эффективного метода оценки его параметров. Мы принимаем гипотезу о том, что данная выборка подчинена распределению Рэлея, которое широко используется для моделирования амплитуды случайных величин.

Перед тем как перейти к оценке параметров, мы представляем детальный анализ выборки. С использованием гистограммы, плотности распределения и эмпирической функции распределения, мы стремимся выявить особенности и характеристики данных.

Основное внимание в нашем исследовании уделяется сравнительному анализу двух методов оценки параметров распределения высоты морских волн: метода максимального правдоподобия (ММП) и метода моментов (ММ). Этот выбор методологии обусловлен необходимостью не только точной оценки, но и сравнения их эффективности в конкретном контексте.

Наконец, для более надежной проверки нашей гипотезы, мы используем критерий согласия Пирсона. Полученные результаты не только обогащают теоретическое понимание распределения высоты морских волн, но также предоставляют практические рекомендации для оценки параметров и применения этих данных в морских приложениях [4].

Таким образом, наше исследование направлено на совершенствование методов анализа и моделирования морских волн, что может иметь важное значение для множества областей, связанных с морской деятельностью и безопасностью [3;7].

Распределение Рэлея

Распределение Рэлея — это статистическое распределение, используемое для моделирования амплитуды случайных величин [2]. Названо в честь британского инженера Лорда Рэлея, это распределение широко применяется в различных областях, включая теорию связи, физику и океанографию.

Функция плотности вероятности для распределения Рэлея задается формулой:

$$f(x,\sigma)=\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}},x\geq 0,\sigma>0 \tag{1}$$

где x – случайная величина, σ – параметр масштаба.

Математическое ожидание (среднее) случайной величины X с распределением Рэлея определяется следующим образом:

$$E[X]=\sqrt{\frac{\pi}{2}}\sigma \tag{2}$$

Математическое ожидание равно корню из половины числа π, умноженному на параметр масштаба σ. Это значение представляет собой "среднюю" амплитуду в распределении Рэлея.

Дисперсия случайной величины X в распределении Рэлея выражается формулой:

$$\mathrm{D}[\mathrm{X}]=\frac{4-\pi}{2}\sigma^2 \tag{3}$$

Дисперсия зависит от параметра масштаба σ и представляет собой меру разброса амплитуд в распределении. Чем больше σ, тем больше дисперсия.

Найдем оценку параметра масштабирования σ двумя методами и проверим их состоятельность для дальнейшего анализа оценок [2].

Оценка методом моментов

Приравнивая выборочное среднее $\overline{X}$ к его математическому ожиданию, получаем оценку параметра $\tilde{\sigma}$:

$$E[X]=\sqrt{\frac{\pi}{2}}\sigma=\overline{X}$$

$$\tilde{\sigma}=\frac{\overline{X}\sqrt{2}}{\sqrt{\pi}} \tag{4}$$

Покажем, что оценка является состоятельной:

$$M\tilde{\sigma} = M\,\frac{\overline{X}\sqrt{2}}{\sqrt{\pi}} = \frac{\sqrt{2}}{\sqrt{\pi}}M\overline{X} = \frac{\sqrt{2}}{\sqrt{\pi}}M\,\frac{1}{n}\sum_{k=1}^{n}X_k = \sum_{k=1}^{n}\frac{1}{n}\frac{\sqrt{2}}{\sqrt{\pi}}MX_k = \sigma$$

$$D\tilde{\sigma} = D\frac{\overline{X}\sqrt{2}}{\sqrt{\pi}} = D\frac{\sqrt{2}}{\sqrt{\pi}}\sum_{k=1}^{n}\frac{1}{n}X_k = \sum_{k=1}^{n}\frac{1}{n^2}\frac{2}{\pi}DX_k = \frac{2\left(2-\frac{\pi}{2}\right)\sigma^2}{\pi n} \to 0\, при\, n \to \infty$$

Оценка методом максимального правдоподобия

Функция правдоподобия для случайной выборки $X_1,\ldots,X_n$ закона Рэлея:

$$L\left(X_1,\ldots,X_n,\sigma\right) = \prod_{k=1}^{n}P_\sigma\left(\xi_k = X_k\right) = \prod_{k=1}^{n}\frac{X_k}{\sigma^2}e^{-\frac{X_k^2}{2\sigma^2}}$$

Логарифмируем:

$$\ln L\left(X_1,\ldots,X_n,\sigma\right) = \sum_{k=1}^{n}\ln\left(\frac{X_k}{\sigma^2}e^{-\frac{X_k^2}{2\sigma^2}}\right) = \sum_{k=1}^{n}\left(lnX_k - ln\sigma^2 - \frac{X_k^2}{2\sigma^2}\right)$$

$$\frac{\partial \ln L\left(X_1,\ldots,X_n,\sigma\right)}{\partial\sigma} = \sum_{k=1}^{n}\left(-\frac{2}{\sigma}+\frac{X_k^2}{\sigma^3}\right) \to \hat{\sigma} = \sqrt{\frac{1}{2n}\sum_{k=1}^{n}X_k^2} \qquad (5)$$

Проверим является ли полученная оценка несмещённой. Найдем чему будет равно математическое ожидание $M\hat{\sigma}$:

$$M\hat{\sigma} = M\sqrt{\frac{1}{2n}\sum_{k=1}^{n}X_k^2} = \sqrt{\frac{1}{2n}}M\sqrt{\sum_{k=1}^{n}X_k^2}$$

Для начала найдем как распределена величина X_k^2:

$$\eta = \xi^2$$

Найдем обратную функцию:

$$g(x) = x^2$$
$$g^{-1}(y) = \sqrt{y}$$

Посчитаем плотность распределения случайной η:

$$p_\eta(y) = p_\xi\left(g^{-1}(y)\right) * \left|(g^{-1}(y))'\right| = \frac{1}{2\sigma^2}e^{-\frac{y}{2\sigma^2}}$$

Случайная величина η будет иметь Гамма-распределение:

$$\eta \sim \gamma_{\frac{1}{2\sigma^2},1} \tag{6}$$

Тогда в силу того что $X_1,\ldots,X_n$ – одинаково, то сумма квадратов $X_1^2+X_2^2+\ldots+X_n^2 \sim \gamma_{\frac{1}{2\sigma^2},n}$.

Тогда

$$M\hat{\sigma}=\sqrt{\frac{1}{2n}}M\sqrt{\sum_{k=1}^{n}X_k^2}=\sqrt{\frac{1}{2n}}\int_0^\infty \frac{x^{n-\frac{1}{2}}}{2^n\sigma^{2n}\Gamma(n)}e^{-\frac{x}{2\sigma^2}}=\left\{\begin{matrix} t=\dfrac{x}{2\sigma^2} \\ dx=2\sigma^2 dt\end{matrix}\right\}=$$

$$\sqrt{\frac{1}{2n}}\int_0^\infty \frac{\left(2\sigma^2 t\right)^{n-\frac{1}{2}}}{2^n\sigma^{2n}\Gamma(n)}e^{-\frac{2\sigma^2 t}{2\sigma^2}}=\sqrt{\frac{1}{2n}}\frac{\sqrt{2}}{\Gamma(n)}\int_0^\infty t^{n-\frac{1}{2}}e^{-t}=\frac{\sigma\Gamma\left(n+\frac{1}{2}\right)}{\sqrt{n}\Gamma(n)}$$

Оценка $\hat{\sigma}=\sqrt{\frac{1}{2n}\sum_{k=1}^{n}X_k^2}$ является смещенной оценкой параметра σ.

Тогда $\hat{\sigma}=\dfrac{\sqrt{n}\Gamma(n)}{\Gamma\left(n+\frac{1}{2}\right)}\sqrt{\frac{1}{2n}\sum_{k=1}^{n}X_k^2}$ будет несмещенной оценкой σ.

Покажем, что оценка является состоятельной:

$$M\hat{\sigma}^2=M\frac{1}{2n}\sum_{k=1}^{n}X_k^2=\sum_{k=1}^{n}\frac{1}{2n}MX_k^2=\sigma^2$$

$$D\hat{\sigma}=M\hat{\sigma}^2-\left(M\hat{\sigma}\right)^2=\sigma^2-\left(\frac{\sigma\Gamma\left(n+\frac{1}{2}\right)}{\sqrt{n}\Gamma(n)}\right)^2\to 0\,при\,n\to\infty$$

Расчётная часть

В данной практической части нашего исследования мы обращаем внимание на конкретные данные — выборку высоты морских волн, полученную из измерений в Беринговом море у побережья Аляски. Данные были взяты с сайта NWS[10]. Наша цель заключается в проведении статистического анализа этих данных с последующим подтверждением или опровержением гипотезы о том, что высота морских волн подчиняется распределению Рэлея [5,6].

Выборка, полученная с указанного сайта, представляет собой ценный набор данных, который может способствовать лучшему пониманию динамики морских волн в данной области [1]. Анализ этой выборки позволит установить, насколько хорошо распределение Рэлея описывает наблюдаемые высоты волн.

Мы начнем с визуального представления данных [9], построив гистограмму и полигон относительных частот (Рис.2). Эти методы позволят нам оценить форму распределения и выделить особенности, которые могут быть ключевыми для дальнейшего анализа.

	Границы интервала	Середина интервала	Частота	Относительная частота p
1	(0.14, 1.218)	0.679	78	0.104
2	(1.218, 2.296)	1.757	162	0.216
3	(2.296, 3.374)	2.835	167	0.222667
4	(3.374, 4.452)	3.913	163	0.217333
5	(4.452, 5.53)	4.991	90	0.12
6	(5.53, 6.608)	6.069	53	0.070667
7	(6.608, 7.686)	7.147	21	0.028
8	(7.686, 8.764)	8.225	11	0.014667
9	(8.764, 9.842)	9.303	4	0.005333
10	(9.842, 10.92)	10.381	1	0.001333

Рис. 1 –Таблица с посчитанными относительными частотами. Авторская разработка.

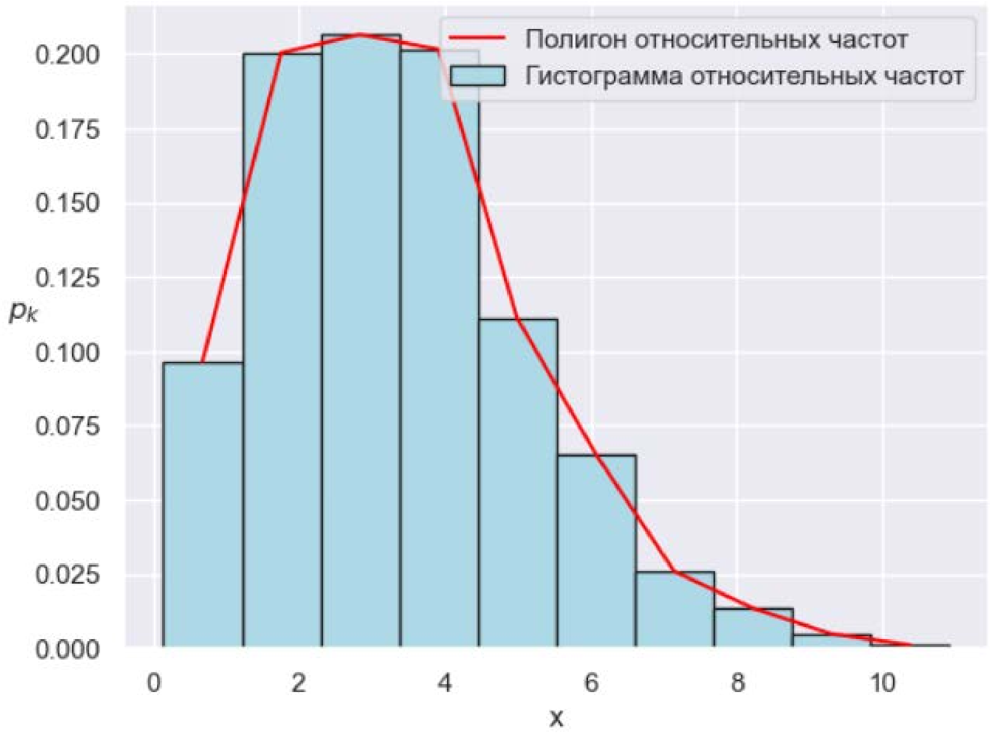

Рис. 2 – График гистограммы относительных частот. Авторская разработка.

Далее, мы приступим к оценке параметров распределения высоты морских волн с использованием метода максимального правдоподобия и метода моментов [2]. Этот этап анализа позволит нам сравнить эффективность двух методов и выбрать тот, который наилучшим образом соответствует особенностям наших данных.

Посчитаем выборочное среднее $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = 3.337 \tag{7}$$

Найдём исправленную выборочную дисперсию S^2:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = 3.269 \tag{8}$$

Сравним дисперсии для оценок ММП и ММ при размере выборке n=750:

$$D\hat{\sigma} = \sigma^2 - \left(\frac{\sigma\Gamma\left(n+\frac{1}{2}\right)}{\sqrt{n}\Gamma(n)}\right)^2 = 0.00033327$$

$$D\tilde{\sigma} = \frac{2\left(2 - \frac{\pi}{2}\right)\sigma^2}{\pi n} = 0.00036431$$

Оценка, посчитанная ММП более эффективная, чем оценка, посчитанная ММ [8].

Посчитаем оценку параметра σ:

$$\hat{\sigma} = \sqrt{\frac{1}{2n}\sum_{k=1}^{n} X_k^2} = 2.683825$$

После успешной оценки параметра σ для распределения Рэлея, мы переходим к визуализации результатов с использованием графиков плотности (Рис.3), эмпирической функции распределения и теоретической функции распределения (Рис.4). Эта часть анализа направлена на более наглядное представление соответствия между оцененной моделью и фактическими данными [9].

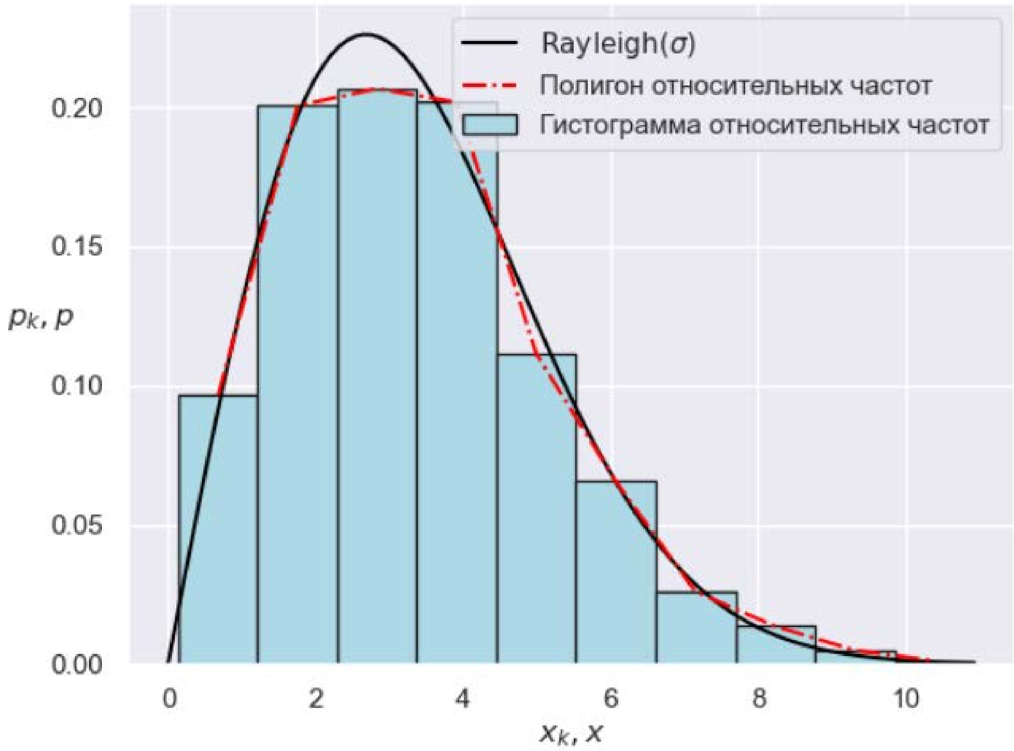

Рис. 3 – Совмещенные графики гистограммы относительных частот и плотности распределения Рэлея с параметром $\hat{\sigma}$. Авторская разработка.

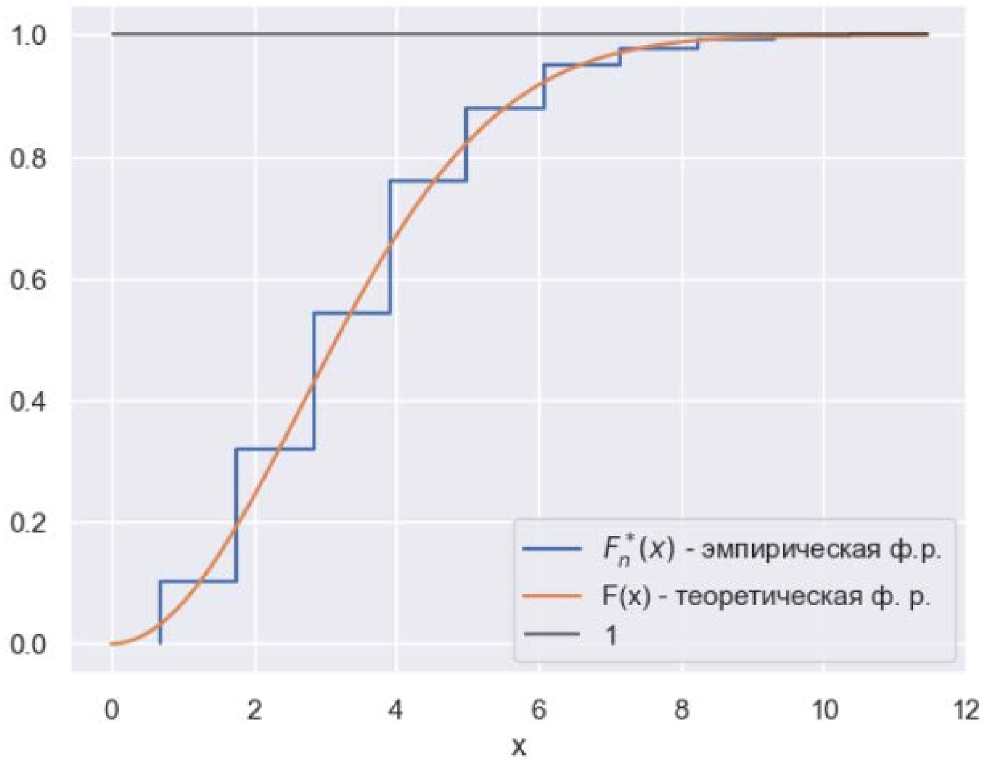

Рис. 4 – График сравнения эмпирической и теоретической функций распределения. Авторская разработка.

Графики подтверждают, что теоретическая модель плотности распределения Рэлея с оцененным параметром хорошо соответствует фактическим данным.

На первом графике (Рис.3) мы видим, что форма гистограммы данных схожа с теоретической кривой плотности. Это свидетельствует о том, что модель распределения Рэлея хорошо аппроксимирует наши измерения.

График эмпирической функции распределения (Рис.4) поддерживает этот вывод, показывая, что накопленная доля значений в выборке совпадает с теоретической функцией распределения.

И, наконец, мы проверим гипотезу о соответствии распределения Рэлея нашей выборке, применяя критерий хи-квадрат Пирсона. Результаты этого шага будут служить ключевым компонентом наших заключений.

Формула критерия согласия Пирсона имеет вид:

$$\chi^2 = \sum_{i=1}^{m} \frac{\left(\nu_i - np_i\right)^2}{np_i}$$

Вычисленные эмпирические и теоретические частоты представлены на рисунке 5.

	1	2	3	4	5	6	7	8	9	10
Интервал	(0.14, 1.218)	(1.218, 2.296)	(2.296, 3.374)	(3.374, 4.452)	(4.452, 5.53)	(5.53, 6.608)	(6.608, 7.686)	(7.686, 8.764)	(8.764, 9.842)	(9.842, 10.92)
Эмпирические частоты	78	162	167	163	90	53	21	11	4	1
Теоретические частоты	72.371994	156.447865	179.853832	150.838324	99.697347	53.574305	23.776277	8.793465	2.72558	1.921012

Рис. 5 – Эмпирические и теоретические частоты. Авторская разработка.

Результат вычисленного критерия:

$$\chi^2 = \sum_{i=1}^{m} \frac{\left(\nu_i - np_i\right)^2}{np_i} = 5.39862$$

И найдем критическое значение χ^2 с 9 степенями свободы и уровнем доверия 0.99:

$$\chi^2_{1-\alpha}(9) = 12.24214$$

В итоге получаем что наше значение не превосходит критическое значение, для того чтобы отвергнуть гипотезу о распределении Рэлея.

Таким образом, практическая часть нашего исследования направлена на конкретные данные, собранные на морском участке, и на их статистический анализ с целью проверки гипотезы о распределении высоты морских волн по закону Рэлея.

Выводы

Результаты статистического анализа выборки позволяют утверждать, что распределение Рэлея хорошо соответствует наблюдаемым высотам волн, что поддерживает нашу теоретическую гипотезу.

Сравнение методов оценки параметров распределения подчеркнуло важность выбора метода, учитывая конкретные характеристики данных. Мы выявили, что метод максимального правдоподобия может быть предпочтителен в условиях, когда данные хорошо соответствуют предположениям о распределении.

Применение критерия согласия Пирсона подтвердило, что распределение Рэлея является статистически значимым для нашей выборки высоты морских волн. Эти результаты имеют важное значение для морских инженеров, океанографов и других специалистов, чья деятельность напрямую связана с динамикой океана.

Библиографический список

1. Бенин, А. В. Планирование эксперимента / А. В. Бенин, В. В. Гарбарук; А. В. Бенин, В. В. Гарбарук ; Федеральное агентство ж.-д. трансп., Федеральное гос. образовательное учреждение высш. проф. образования "Петербургский гос. ун-т путей сообщ.". – Санкт-Петербург : ПГУПС, 2010. – 89 с. – ISBN 978-5-7641-0260-3. – EDN QJYKOB.
2. Боровков, А. А. Математическая статистика : учебник / А. А. Боровков; А. А. Боровков. – Изд. 4-е, стер.. – Санкт-Петербург [и др.] : Лань, 2010. – 703 с. – (Знание. Уверенность. Успех!). – ISBN 978-5-8114-1013-2. – EDN QJWFVV.
3. Запевалов, А. С. Влияние групповой структуры морских волн на бортовую качку судна / А. С. Запевалов // Нанотехнологии: наука и производство. – 2023. – № 5. – С. 48-53. – EDN ZHRKAI.
4. Ларичев, О. И. Вербальный анализ решений / О. И. Ларичев ; О.И. Ларичев ; Рос. акад. наук, Ин-т систем. анализа. – Москва : Наука, 2006. – 181 с. – ISBN 5-02-033979-2. – EDN QJPVZD.
5. Математические и инженерные примеры законов распределний случайных величин в ЦОС Nomotex / Т. В. Облакова, К. М. Зубарев, А. А. Сальникова, Д. С. Шинаков // Дневник науки. – 2022. – № 12(72). – DOI 10.51691/2541-8327_2022_12_29. – EDN VLCPIK.
6. Методика графического представления данных об успеваемости в Цифровой образовательной среде NOMOTEX / Ю. И. Димитриенко, Е. А. Губарева, К. М. Зубарев [и др.] // Дневник науки. – 2022. – № 12(72). – DOI 10.51691/2541-8327_2022_12_36. – EDN TAXWIB.
7. Морозов, Н. Ф. Сейсмические барьеры для защиты от поверхностных и головных волн: множественные рассеиватели и метаматериалы / Н. Ф. Морозов, В. А. Братов, С. В. Кузнецов // Известия Российской академии наук. Механика твердого тела. – 2021. – № 6. – С. 33-44. – DOI 10.31857/S057232992106009X. – EDN CVSSVK.

8. Мочалов, И. А. Оценивание параметров модели по нечетким случайным данным / И. А. Мочалов, М. С. Хрисат // Информационные технологии. – 2014. – № 2. – С. 14-22. – EDN RVVXLL.
9. Орлов, А. И. Прикладная статистика : учебник / А. И. Орлов ; А. И. Орлов. – Москва : Экзамен, 2006. – 671 с. – ISBN 5-472-01122-1. – EDN QJPVDP.
10. National Data Buoy Center: [Электронный ресурс]. URL: https://www.weather.gov/. (Дата обращения: 08.11.2023).

Оригинальность 95%