Университет ИТМО

Практическое задание №1 по дисциплине

«Теория информации и информационных систем»

Исследование статистических и информационных характеристик сигналов

Цель работы

Изучение методики расчёта статистических характеристик и энтропии сигналов при передаче информации в информационных системах.

Теоретические основы

Построение гистограмм широко используется при экспериментальном исследовании статистических характеристик случайных величин.

Гистограмма является приближенной оценкой плотности вероятности случайной величины. При известной плотности вероятности $p(s)$ случайной величины s несложно определить элемент вероятности $p(s)ds$, т.е. вероятность попадания случайной величины в интервал $[s; s+ds]$.

Для дискретной случайной величины, которая принимает значения s_i, $i=1,\ldots,M$, вероятность этих значений обозначается как $P(s_i)$. Знание вероятностей $P(s_i)$ позволяет найти статистические моменты случайной величины, в частности, математическое ожидание

$$\langle s \rangle = \sum_{i=1}^{M} s_i P(s_i) \tag{1}$$

и дисперсию

$$\sigma_s^2 = \langle s_i^2 \rangle - \langle s \rangle^2, \tag{2}$$

где угловые скобки обозначают усреднение по ансамблю значений случайной величины.

Если рассматривать случайную величину как формируемую источником дискретных сообщений с алфавитом S, $s_i \in S$, то можно вычислить энтропию источника по формуле Шеннона

$$H(S) = -\sum_{i=1}^{M} P(s_i) \log P(s_i) \tag{3}$$

и тем самым определить среднее количество информации, которое содержит одно полученное сообщение (одно значение s_i).

Применительно к сигналам как носителям информации под значениями случайной величины s_i понимают значения сигнала, попадающие в i-ый интервал $[s_i; s_i + \Delta s]$ квантования сигнала по уровню, где Δs – ширина интервала (шага) квантования.

Построение гистограммы значений сигнала осуществляется следующим образом: область изменения величины сигнала разбивается на определённое количество интервалов M, после чего для каждого интервала подсчитывается количество значений сигнала (рассматриваемых как значения случайной величины) попадающих в каждый интервал.

Пример сигнала и его гистограммы представлен на Рис. 1. На Рис. 2 показана нормализованная гистограмма.

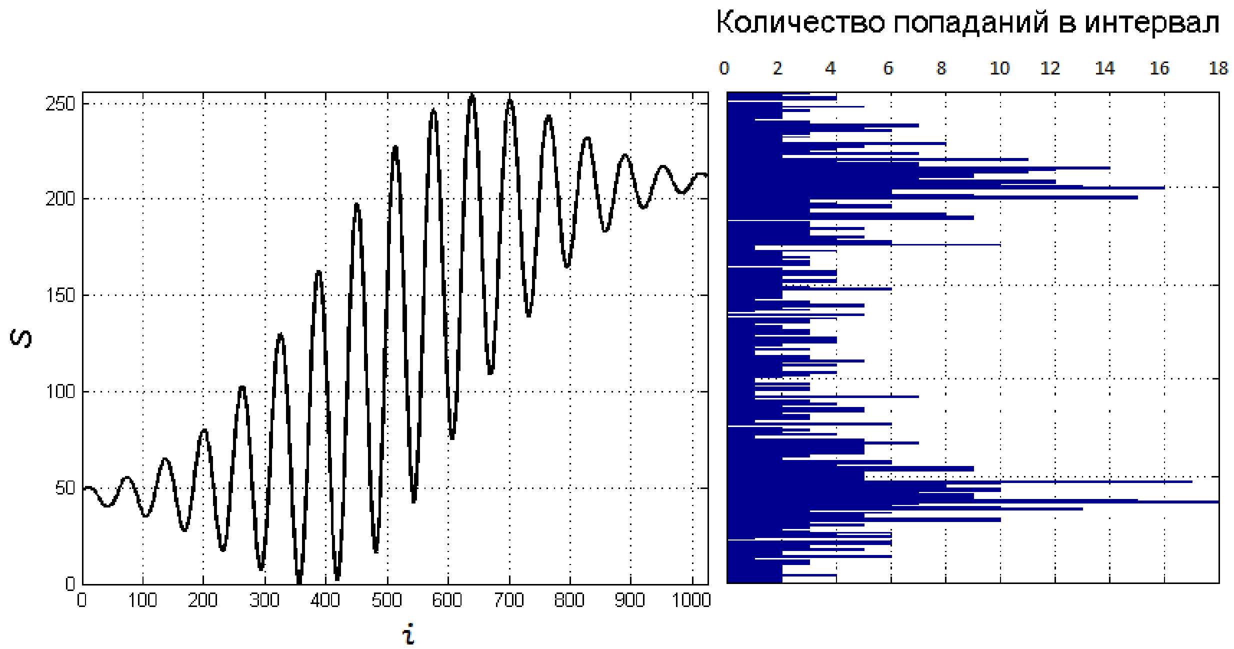

Рис. 1. График сигнала и гистограмма значений сигнала.

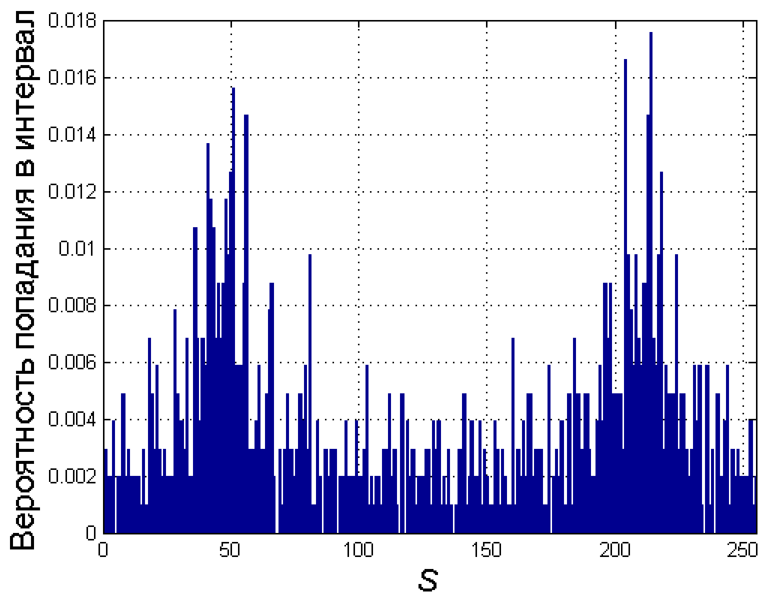

Рис. 2. Нормализованная гистограмма значений сигнала.

Высота каждого "столбика" гистограммы выражает оценку вероятности попадания значений сигнала в интервал $[s_i; s_i + \Delta s]$ согласно формуле

$$P(s_i) = K_i / K, \tag{4}$$

где K_i – количество попаданий в i-ый интервал квантования, $i = 1,\ldots,M$, при общем количестве K анализируемых значений в последовательности отсчетов сигнала.

Важно отметить, что точность оценки вероятности по формуле (4) возрастает с увеличением количества (объема выборки) K анализируемых значений сигнала. Поскольку выборка отсчетов сигнала формируется при дискретизации сигнала в области независимой переменной x с шагом дискретизации Δx, уменьшение шага дискретизации способствует повышению точности оценок вероятности значений сигнала за счет увеличения объема выборки K.

Пусть n – нормально распределённый аддитивный шум с нулевым средним значением, статистически независимый от сигнала s. Представим сигнал, искажённый таким шумом, следующим выражением:

$$\xi_i = s_i + n_i, \quad i = 1,\ldots,M. \tag{5}$$

Вероятности значений принятого сигнала (5), идеального сигнала и шума обозначим, соответственно, как $P_\xi(\xi)$, $P_s(s)$ и $P_n(n)$, а их плотности вероятности – как $p_\xi(\xi)$, $p_s(s)$ и $p_n(n)$.

Известно, что *плотность вероятности* суммы двух статистически независимых случайных величин определяется свёрткой плотностей вероятностей этих величин. С учетом принятых обозначений можно записать выражение для плотности вероятности значений принятого сигнала:

$$p_\xi(\xi) = p_s(s) * p_n(n), \tag{6}$$

где $*$ – символ операции свертки.

Используя соотношение (6) и формулу (3) для энтропии, можно показать, что *совместная энтропия* двух статистически независимых случайных величин равна сумме энтропий этих величин. Однако *энтропия суммы* двух статистически независимых случайных величин, вообще говоря, не равна их совместной энтропии и определяется соотношением:

$$H(\Xi) \le H(S,N) = H(S) + H(N), \tag{7}$$

где для рассматриваемого случая $H(\Xi)$ – энтропия принятого сигнала (5), $H(S,N)$ – совместная энтропия сигнала и шума, $H(S)$ – энтропия идеального сигнала, $H(N)$ – энтропия шума. (8)

Порядок выполнения работы

1. Смоделировать сигнал согласно выданному индивидуальному заданию.
2. Построить оценку плотности распределения вероятности значений сигнала в форме гистограммы сигнала.
3. Оценить математическое ожидание, дисперсию и энтропию значений сигнала.
4. К смоделированному сигналу добавить реализацию нормального шума с нулевым средним. Построить гистограмму шума.
5. Для зашумленного сигнала (п. 4) выполнить пп. 2 – 3.
6. Построить расчетную гистограмму зашумленного сигнала, используя соотношение (6) и оценки плотности вероятности, полученные по пп. 2, 4, и сравнить ее с экспериментальной гистограммой.
7. Оценить совместную энтропию и энтропию суммы сигнала и шума, учитывая соотношение (7), при различном количестве отсчётов в сигнале.
8. Оценить совместную энтропию через совместную гистограмму при различном количестве отсчётов в сигнале. Сравнить результаты с результатами пункта (8).
9. Проанализировать результаты, сформулировать выводы, составить отчёт по работе.