Цель работы: построение гистограммы и эмпирической функции распределения.

Содержание работы

1. Для выборки объёма n из генеральной совокупности X реализовать в виде программы на ЭВМ
   - вычисление максимального значения $M_{\max}$ и минимального значения $M_{\min}$;
   - размаха R выборки;
   - вычисление оценок $\hat{\mu}$ и S^2 математического ожидания MX и дисперсии DX;
   - группировку значений выборки в $m = [\log_2 n] + 2$ интервала;
   - построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2;
   - построение на другой координатной плоскости графика эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2.
2. Провести вычисления и построить графики для выборки из индивидуального варианта.

1 Формулы для вычисления величин

$$\vec{x} = (x_1, ..., x_n)$$

1. **Максимальное значение выборки**

$$M_{\max} = \max\{x_1, .., x_n\}$$

2. **Минимальное значение выборки**

$$M_{\min} = \min\{x_1, .., x_n\}$$

3. **Размах выборки**

$$R = M_{\max} - M_{\min}$$

4. **Выборочное среднее (математическое ожидание)**

$$\hat{\mu}(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

5. **Состоятельная оценка дисперсии**

$$S^2(\vec{x}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

где $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

2 Определение эмпирической плотности и гистограммы

Эмпирической плотностью (отвечающей выборке $\vec{x}$) называют функцию

$$\hat{f}_n(x) = \begin{cases} \frac{n_i}{n\Delta}, x \in J_i, i = \overline{1;p} \\ 0, \text{ иначе} \end{cases}$$

где (J_i, n_i) – интервальный статистический ряд

Пусть $\vec{x}$ – выборка из генеральной совокупности X. Если объем n этой выборки велик, то значения x_i группируют не только в статистический ряд, но и в интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ (где $x_{(1)} = \min\{x_1, .., x_n\}$, $x_{(n)} = \max\{x_1, .., x_n\}$) делят на p равновеликих частей:

$$J_i = [a_i, a_{i+1}), i = \overline{1;p-1}$$

$$J_p = [a_p, a_{p+1}]$$

$$a_i = x_{(1)} + (i-1)\cdot\Delta, i = \overline{1;p+1}$$

$$\Delta = \frac{|J|}{p} = \frac{x_{(n)} - x_{(1)}}{p}$$

Интервальным статистическим рядом называют таблицу

J_1	...	J_i	...	J_p
n_1	...	n_i	...	n_p

Здесь n_i – количество элементов выборки $\vec{x}$, которые $\in J_i$

В нашем случае $p = m = [\log_2 n] + 2$

Гистограммой называют график эмпирической плотности.

3 Определение эмпирической функции распределения

Пусть $\vec{x} = (x_1, ..., x_n)$ – выборка из генеральной совокупности X. Обозначим $n(x, \vec{x})$ – чисор элементов вектора $\vec{x}$, которые имеют значения меньше x.

Эмпирической функцией распределения называют функцию $F_n : \mathbb{R} \to \mathbb{R}$, определенную условием $F_n(x) = \frac{n(x,\vec{x})}{n}$.

4 Текст программы

Листинг 1: Реализация

```
function main()
    X=[-5.05,-5.74,-6.39,-5.01,-4.94,-6.32,-4.73,-5.44,-4.79,-5.40,...
        -4.50,-3.43,-5.21,-5.22,-5.07,-5.51,-4.45,-5.24,-6.50,-4.99,...
        -5.42,-3.30,-5.06,-5.38,-6.48,-3.31,-5.56,-6.50,-5.22,-5.68,...
        -4.80,-6.67,-4.95,-5.32,-5.68,-6.32,-3.72,-4.59,-6.33,-5.03,...
        -4.49,-4.80,-6.04,-6.21,-3.60,-3.93,-5.89,-5.29,-7.41,-3.73,...
        -6.61,-4.34,-5.99,-5.24,-4.08,-4.68,-5.38,-6.38,-4.66,-3.67,...
        -4.61,-4.54,-4.51,-5.43,-6.47,-5.31,-4.30,-6.32,-5.82,-3.44,...
        -5.92,-4.76,-4.45,-3.52,-4.91,-5.65,-5.02,-5.00,-5.26,-4.98,...
        -6.16,-6.21,-4.42,-6.20,-5.84,-5.58,-5.34,-5.21,-5.78,-7.80,...
        -5.21,-4.79,-4.53,-4.78,-6.39,-7.04,-4.82,-5.53,-3.52,-6.24,...
        -3.58,-5.01,-5.79,-4.80,-6.04,-5.15,-7.03,-4.71,-4.38,-5.77,...
        -4.05,-5.76,-5.86,-6.45,-4.81,-5.68,-7.48,-3.97,-5.16,-3.48];

    %1)
    %a) Максимальное и минимальное значения
    Mmax = max(X);
    Mmin = min(X);
    fprintf("\na) Mmax (максимальное значение) = %f; Mmin (минимальное значенение) = %f", Mmax, Mmin);
```

```
%б) Размах
R = Mmax - Mmin;
fprintf("\nб) R (размах) = %f", R);

%в) Оценки
mu = sum(X) / length(X);
s2 = sum((X - mean(X)).^2) / (length(X) - 1);
fprintf("\nв)mu (оценка математического ожидания) = %f; s^2 (оценка дисперсии) = %f", mu, s2);

%г) Группировка значений выборки
% Нахождение количества интервалов
m = floor(log2(length(X))) + 2;
fprintf("\nг)Группировка значений выборки в m = [log2 n] + 2 интервала: m = %f\n", m);

% Разбиение выборки на интервалы от min до max, с помощью BinLimits
% объединяем только те значения, которые находятся в интервале от мин
% до макс
[counts, edges] = histcounts(X, m, 'BinLimits', [min(X), max(X)]);

for i = 1: length(counts)
    fprintf("[%f : %f] - %d\n", edges(i), edges(i + 1), counts(i));
end

%д)
hist = histogram();
hist.BinEdges = edges;
hist.BinCounts = counts / length(X) / ((max(X) - min(X)) / m);

hold on; % Продолжаем работать с той же системой

delta = R/m;
sigma = sqrt(s2);
Xn = min(X):delta/20:max(X);
Y = normpdf(Xn, mu, sigma);
plot(Xn, Y, 'blue');

%е)
figure;
[yy, xx] = ecdf(X);
stairs(xx, yy);

hold on;

delta = R/m;
Xn = min(X):delta/20:max(X);
Y = 1/2 * (1 + erf((Xn - mu) / sqrt(2*s2)));
```

```
    plot(Xn, Y, 'black');

end
```

5 Результаты расчетов для выборки из индивидуального варианта (вариант 22)

1. Максимальное значение выборки
2. Минимальное значение выборки
3. Размах выборки
4. Выборочное среднее (математическое ожидание)
5. Состоятельная оценка дисперсии
6. Группировка значений выборки в $m = [log_2 n] + 2$ интервала
7. Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2.

8. График эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2.