Теория вероятностей и математическая статистика
Лектор А.А. Лобузов
Семестр 6 Лекция 3

Непрерывная статистическая модель

Если $\mathbf{X} = (X_1, X_2, ..., X_N)$ − случайная выборка, полученная при измерении непрерывной случайной величины ξ, то при обработке выборки удобно рассматривать **группированную выборку (или интервальный вариационный ряд).**

Интервальный ряд (группированная выборка):

$(a_{i-1}, a_i]$	$[a_0, a_1]$	$(a_1, a_2]$	…	$(a_{m-1}, a_m]$
n_i	n_1	n_2	…	n_m
w_i	w_1	w_2	…	w_m

$\sum_{i=1}^{m} n_i = N$; $w_i = \frac{n_i}{N}$, $\sum_{i=1}^{m} w_i = 1$;

$x_i^* = \frac{a_{i-1} + a_i}{2}$ – середина интервала $(a_{i-1}, a_i]$.

Эмпирическая функция распределения

$$F_N(x, \mathbf{x}) = \frac{1}{N}\sum_{k=1}^{N} I_{(-\infty, x]}(x_k) = \frac{1}{N}\sum_{k=1}^{N} I_{(x_k \le x]}$$

$F_N(x, \mathbf{x}) = F_N(x, x_1, x_2, ..., x_N)$ − зависит от выборки.

$$F_N(x, \mathbf{x}) = \begin{cases} 0, x < x_{(1)}; \\ \frac{k}{N}, x_{(k)} \le x < x_{(k+1)}, 1 \le k \le N; \\ 1, x \ge x_{(N)}. \end{cases}$$

Пример.

Упорядоченная выборка (вариационный ряд) объёмом 100

2.0464	2.3134	2.5010	2.5563	2.5656	2.6724	2.7031	2.7974	2.9178	2.9682
3.0492	3.0904	3.0918	3.0930	3.1071	3.1451	3.1703	3.2509	3.2605	3.2754
3.3322	3.3774	3.4208	3.4281	3.4705	3.4964	3.6107	3.6401	3.6603	3.6841
3.7139	3.7337	3.7610	3.7699	3.7894	3.8320	3.8590	3.8772	3.8817	3.9139
3.9467	3.9566	3.9661	3.9822	3.9824	4.0000	4.0363	4.0534	4.0792	4.1025
4.1066	4.1128	4.1572	4.1908	4.1974	4.2141	4.2316	4.2589	4.2650	4.2833
4.2946	4.3423	4.3509	4.3854	4.3993	4.4570	4.4759	4.5120	4.5121	4.5140
4.5201	4.5322	4.5530	4.5609	4.6017	4.6094	4.6210	4.6407	4.6429	4.6532
4.6579	4.7031	4.7191	4.7341	4.7722	4.8061	4.8468	4.9601	4.9741	4.9855
5.0702	5.0718	5.0718	5.1286	5.1612	5.2736	5.2992	5.4612	5.5232	5.9649

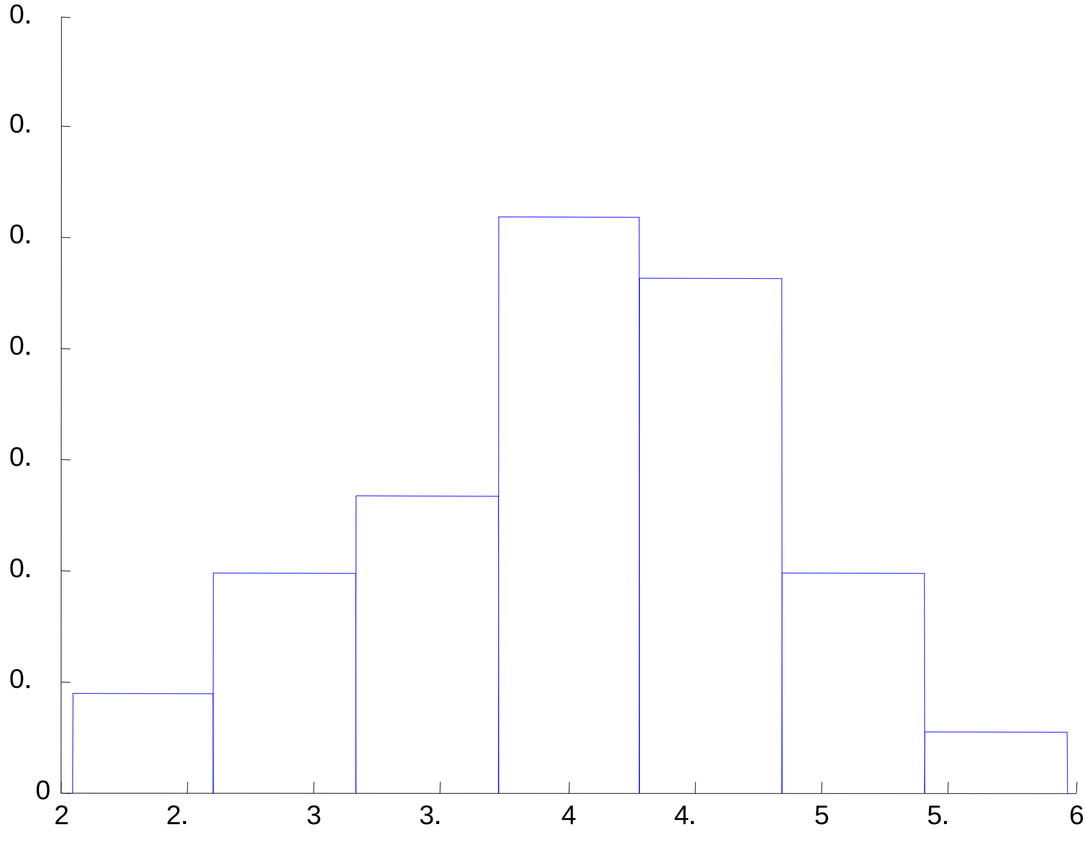
Интервальный вариационный ряд

Число интервалов определяется по формуле Стерджеса

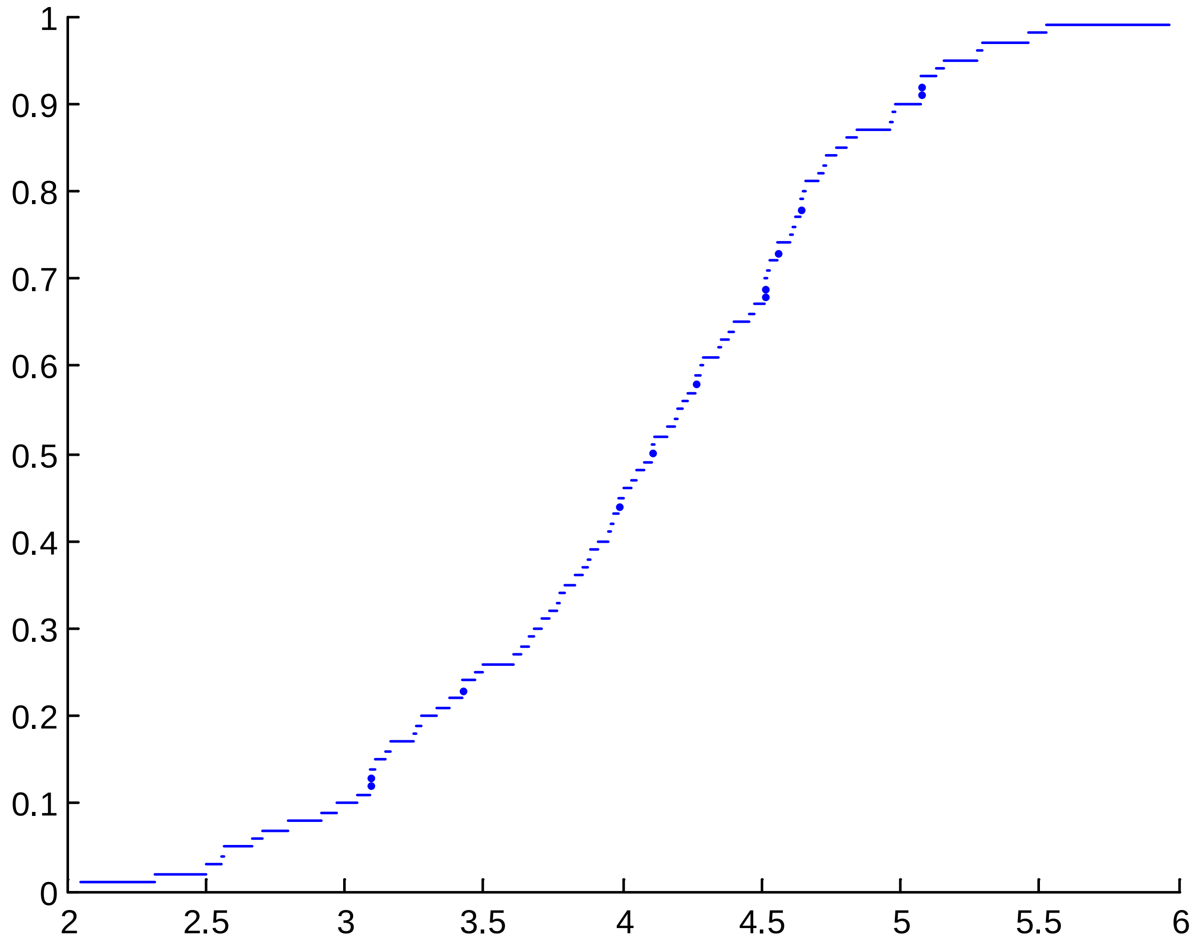
$m = 1 + [\log_2 N] = 1 + [\log_2 100] = 7$

(2.0464, 2.6062]	(2.6062, 3.1660]	(3.1660, 3.7257]	(3.7257, 4.2855]	(4.2855, 4.8453]	(4.8453, 5.4051]	(5.4051, 5.9649]
5	11	15	29	26	11	3
0.05	0.11	0.15	0.29	0.26	0.11	0.03

Гистограмма относительных частот



Эмпирическая функция распределения



Характеристики выборки

Выборочное среднее:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{m} x_i^* \cdot n_i = \sum_{i=1}^{m} x_i^* \cdot w_i \text{ .}$$

Выборочная дисперсия:

$$D_B = \sum_{i=1}^{m}(x_i^* - \overline{x})^2 \cdot w_i = \sum_{i=1}^{m}(x_i^*)^2 \cdot w_i - (\sum_{i=1}^{m} x_i^* \cdot w_i)^2 = \overline{x^2} - (\overline{x})^2 \text{ .}$$

Выборочная дисперсия с поправкой Шеппарда

$$s_B^2 = \sum_{i=1}^{m}(x_i^* - \overline{x})^2 \cdot w_i - \frac{h^2}{12}, \text{где } h = (a_m - a_0)/m \text{ .}$$

Выборочный момент k-ого порядка:

$$\bar{\mu}_k = \overline{x^k} = \sum_{i=1}^{m} (x_i^*)^k \cdot w_i, \; \bar{\mu}_1 = \bar{x}.$$

Выборочный центральный момент k-ого порядка:

$$\bar{\mu}_k^0 = \sum_{i=1}^{m} (x_i^* - \bar{x})^k \cdot w_i, \; \bar{\mu}_1^0 = 0, \; \bar{\mu}_2^0 = D_B = \bar{\mu}_2 - (\bar{\mu}_1)^2.$$

Выборочное среднее квадратичное отклонение:

$$\bar{\sigma} = \sqrt{D_B}.$$

Выборочная мода

$$\bar{M}_0 = a_{k-1} + h \frac{w_k - w_{k-1}}{2w_k - w_{k-1} - w_{k+1}}$$

a_{k-1} – левая граница модального интервала (a_{k-1}, a_k) (интервала, имеющего наибольшую частоту);

w_k – относительная частота на модальном интервале;

w_{k-1}, w_{k+1} – относительные частоты интервалов слева и справа от модального интервала.

Выборочная медиана:

$$\bar{M}_e = a_{k-1} + \frac{h}{w_k}\left(\frac{1}{2} - \sum_{i=1}^{k-1} w_i\right), \text{ если } \sum_{i=1}^{k-1} w_i < \frac{1}{2} < \sum_{i=1}^{k} w_i;$$

$$\bar{M}_e = a_k, \text{ если } \sum_{i=1}^{k} w_i = \frac{1}{2}.$$

Выборочный коэффициент асимметрии:

$$\bar{a}_s = \bar{\gamma}_1 = \frac{\bar{\mu}_3^0}{\bar{\sigma}^3}.$$

Выборочный коэффициент эксцесса:

$$\bar{\varepsilon} = \bar{\gamma}_2 = \frac{\bar{\mu}_4^0}{\bar{\sigma}^4} - 3.$$