# Курс: Статистические Методы Обработки Данных

## Лекция 2. Оценки Параметров Распределений

*Специальность: 1-53 01 02 – Автоматизированные системы обработки информации*

*УО «ГГУ им. Ф. Скорины»*

*Преподаватель: Бабич К.С, ст. преподаватель, 2016*

# Раздел 2. Оценки Параметров Распределений

## 4. Точечные оценки

Пусть дана выборка объема n.
Точечная оценка для математического ожидания (центра распределения) дается выражением:

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

Для дисперсии $D = \sigma^2$ такая оценка дается выражением:

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

# Раздел 2. Оценки Параметров Распределений

## 4. Точечные оценки

Величины $\overline{x}$, $S$ сами являются случайными величинами и, следовательно, они тоже могут иметь разброс, который характеризуется дисперсией.
Для $\overline{x}$:

$$S_{\overline{x}}^2 = \frac{S^2}{n}.$$

*Средне квадратичное отклонение среднего арифметического*

Для дисперсии:

$$D\left[S^2\right] = \frac{\mu_4}{n-1} - \frac{\sigma^4(n-3)}{n(n-1)},$$

где $\mu_4$ – четвертый алгебраический, а $\sigma$ – второй центральный моменты.

То есть:

$$D\left[S^2\right] = \sigma^4\left[\frac{\varepsilon}{n-1} - \frac{n-3}{n(n-1)}\right] = \frac{\sigma^4}{n-1}\left[\varepsilon - \frac{n-3}{n}\right].$$

# Раздел 2. Оценки Параметров Распределений

## 4. Точечные оценки

Четвертый центральный момент можно рассчитать (приближенно для больших n) следующим образом:

$$\mu_4^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4 .$$

С учетом предыдущего выражения при $n > 20$ с.к.о. будет задаваться выражением $(n-3 \approx n-1)$:

$$\sigma_\sigma = \sigma\frac{\sqrt{\varepsilon - 1}}{2\sqrt{n}} .$$

Тогда относительная ошибка для с.к.о.

$$\delta_\sigma = \frac{\sigma_\sigma}{\sigma} = \frac{\sqrt{\varepsilon - 1}}{2\sqrt{n}}$$

уже не зависит от самого с.к.о. $\sigma$.

# Раздел 2. Оценки Параметров Распределений

## 4. Точечные оценки

Таким образом точечные оценки:

| Для среднего значения | Дисперсия этой оценки | Для ц. мом. к-го пор. |
|---|---|---|
| $\overline{x} = \sum\limits_{i=1}^{n} \frac{x_i}{n}$ | $\sigma_x^2 = \frac{S^2}{n}$ | $\mu_k \approx \frac{1}{n} \sum\limits_{i=1}^{n} (x_i - \overline{x})^k$ <br> $\mu_4 \approx \frac{1}{n} \sum (x_i - \overline{x})^4$ |
| Для дисперсии $D = \sigma^2$ | Дисперсия дисперсии | |
| $S^2 = \sigma^2 = \frac{1}{n-1} \sum (x_i - \overline{x})^2$ | $\sigma_s^2 = S^4 \left[ \frac{\varepsilon}{n-1} - \frac{n-3}{n(n-1)} \right]$ | |

# Раздел 2. Оценки Параметров Распределений

## 4. Точечные оценки

Дисперсия для 4-го момента $\mu_4$ для различных законов выражается разными формулами. Так, относительные погрешности для нормального распределения для контрэксцесса $\kappa = \frac{1}{\sqrt{\varepsilon}}$

$$\delta(\kappa) = \sqrt{\frac{2}{3n}}, \quad (\varepsilon = 3, \kappa \approx 0,577)$$

для равномерного распределения

$$(\varepsilon = 1,8, \kappa \approx 0,74), \quad \delta(\kappa) = \sqrt{\frac{32}{315n}}$$

и т.д.

Для многих распределений эту относительную погрешность (с точностью 8-10%) можно приближенно описать:

$$\delta(\kappa) = \sqrt[n]{\frac{(\varepsilon^2 - 1)^4}{\sqrt{29n}}}\ .$$

# Раздел 2. Оценки Параметров Распределений

# 4.(Часть 2) - Интервальные оценки

Точечные оценки параметров случайных величин не позволяют судить о степени близости выборочных значений к оцениваемому параметру. Более содержательны процедуры оценивания параметров, связанные с построением интервала с известной степенью доверительности. Для того, чтобы перейти к тому как найти эти доверительные интервалы, рассмотрим понятие квантиля распределения. Квантилем случайной величины $x$ уровня значимости $\alpha$ называется величина $x_\alpha$, для которой имеем:

$$P(x_\alpha) = P_{\text{пов}}(x \le x_\alpha) = \int_{-\infty}^{x_\alpha} p(x)\,dx = 1 - \alpha$$

или

$$P_{\text{пов}}(x > x_\alpha) = \int_{x_\alpha}^{\infty} p(x)\,dx = \alpha$$

Точка $x_\alpha$ называется $100 \times \alpha$-процентной точкой распределения с плотностью распределения $p(x)$.

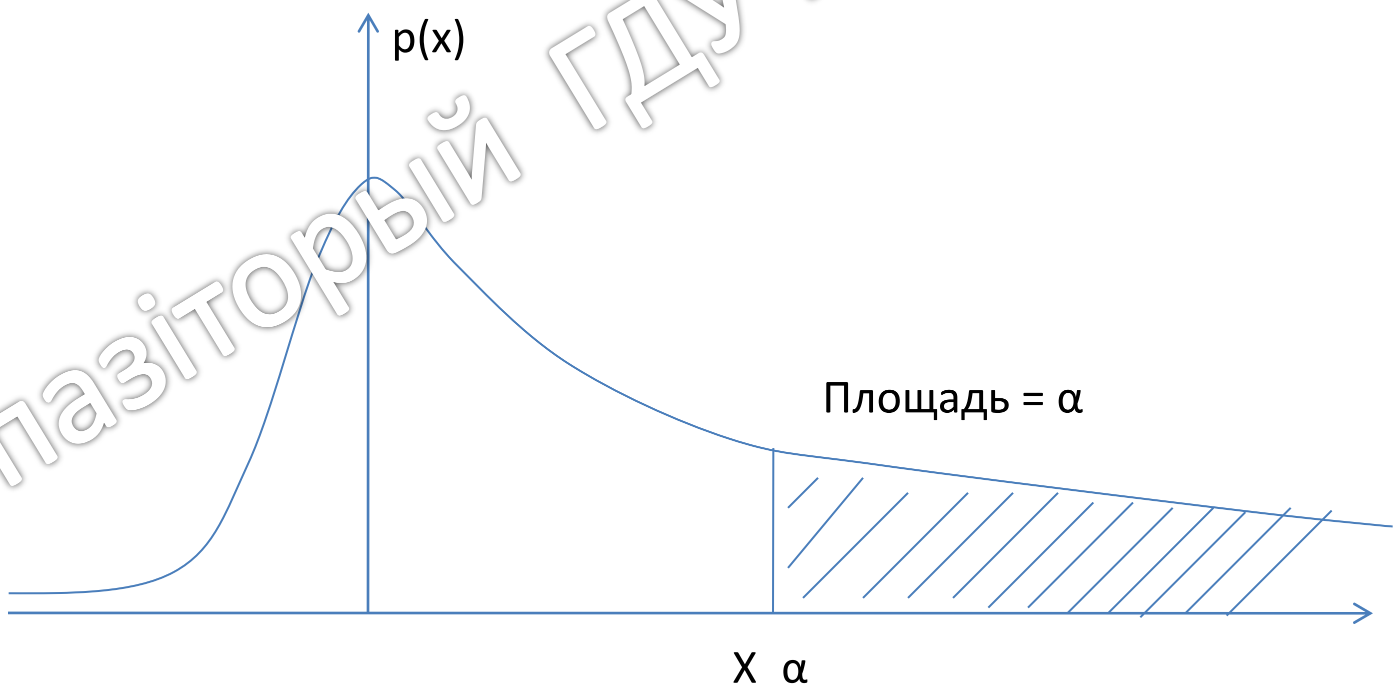
# Раздел 2. Оценки Параметров Распределений

# 4.(Часть 2) - Интервальные оценки

$$P\left(x_{\alpha}\right) = P_{\text{пов}}\left(x \le x_{\alpha}\right) = \int\limits_{-\infty}^{x_{\alpha}} p\left(x\right) dx = 1 - \alpha$$

или

$$P_{\text{пов}}\left(x > x_{\alpha}\right) = \int\limits_{x_{\alpha}}^{\infty} p\left(x\right) dx = \alpha$$



# Раздел 2. Оценки Параметров Распределений

## 4.(Часть 2) - Интервальные оценки

Доверительный интервал для м.о. $\mu_1$.

$$\left[\overline{x} - \frac{S\, t_{N,\alpha/2}}{\sqrt{n}} \le \mu_1 \le \overline{x} + \frac{S\, t_{N,\alpha/2}}{\sqrt{n}}\right], \quad N = n - 1$$

$t_{N,\alpha}$ определяется соотношением:

$$\int_{t_{n,\alpha}}^{\infty} p(t)\, dt = \alpha \ ,$$

где

$$p(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)} \left[1 + \frac{t^2}{n}\right]^{-\frac{n+1}{2}} \ -$$

$t$-распределение с $n$ степенями свободы (распределение Стьюдента). Коэффициенты $t_{N,\alpha/2}$ носят название коэффициентов Стьюдента.

*У.С. Госсет (1876 – 1937)*

# Раздел 2. Оценки Параметров Распределений

# 4.(Часть 2) - Интервальные оценки

Для дисперсии $\sigma$

$$\left[\frac{nS^2}{\chi^2_{N;\alpha/2}} \le \sigma_x < \frac{nS^2}{\chi^2_{N;-\alpha/2}}\right], \quad N = n - 1$$

$\chi^2_{n;\,\alpha}$ - квантиль распределения $\chi^2$.

$$\int_{\chi^2_{n;\alpha}}^{\infty} P(\chi^2)d\chi^2 = \alpha$$

и

$$P(\chi^2) = \frac{(\chi^2)^{(\frac{n}{2}-1)}e^{-\frac{\chi^2}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}, \quad (\chi^2 > 0) -$$

распределение $\chi^2$.
Значения $\mu_1$ и $\sigma$ лежат в интервале с доверительной вероятностью $1 - \alpha$.

# Раздел 2. Оценки Параметров Распределений

## Выбор доверительных интервалов

Сделаем несколько замечаний относительно доверительных интервалов.

Какую вероятность необходимо выбрать в зависимости от числа измерений n. Дело в том, что при очень малом $(n \leq 30)$ величины $\overline{x}, S$, входящие в формулы доверительных интервалов будут иметь сами, как случайные величины большой расброс, поэтому оценки с помощью квантилей будут иметь лишь небольшую доверительную вероятность $1-\alpha$. На этот вопрос можно ответить следующим образом:

# Раздел 2. Оценки Параметров Распределений

## Выбор доверительных интервалов

1) Пусть имеем $n$ значений случайной величины $x$: $x_1$, $x_2$, …, $x_n$, которые распределены в порядке возрастания $x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n$.

2) Разобьём весь интервал вероятностей $[0,1]$ на $n+1$ интервалов с равными значениями вероятности т.е. вероятности попадания в каждый интервал $]-\infty, x_1[$ $[x_1, x_2[$ … $[x_n, +\infty[$ одинакова т.е $x_1, \ldots, x_n$ – $\frac{1}{n+1}100\%$ - квантили.

3) Если отбросим два крайних слева и справа интервала, оценка величины может быть определена с доверительной вероятностью

$$P_D \leq \frac{n-1}{n+1}\ .$$

# Раздел 2. Оценки Параметров Распределений

## Выбор доверительных интервалов

4) На практике отбрасывается еще $n_{sp}$ с каждого края, тогда

$$P_D \le \frac{(n - 1 - 2n_{sp})}{n + 1} \ .$$

Тогда имеем:

$$n \ge (1 + P_D + 2n_{sp}) \,/\, (1 - P_D) \ , \quad n_{sp} = 1$$

| $P_D$ | 0,8 | 0,9 | 0,95 | 0,98 | 0,99 | 0,995 | 0,997 |
|---|---|---|---|---|---|---|---|
| n | 20 | 40 | 80 | 200 | 400 | 800 | 1333 |

# Раздел 2. Оценки Параметров Распределений

## Выбор доверительных интервалов

| $P_D$ | 0,8 | 0,9 | 0,95 | 0,98 | 0,99 | 0,995 | 0,997 |
|---|---|---|---|---|---|---|---|
| n | 20 | 40 | 80 | 200 | 400 | 800 | 1333 |

Таким образом, по экспериментальным данным легко найти доверительный интервал лишь с доверительной вероятностью 0,95 ($n \approx 80$), а более высокое $P_D$ практически трудно допустимо.
Отметим, что выше указанные рассуждения говорят лишь о верхней границе т.е. взяв $n = 80$ мы получим, что доверительная вероятность не может быть большей, чем 0.95, (однако нет оснований утверждать, что обязательно она должна быть равна 0.95).