Введение в математическую статистику

Основная задача М.С. заключ. в том, чтобы на основании стат. наблюдений предложить адекватную М. модель

Основные задачи М.С

1) П.О. пар-ов распределения (точечные оценки)

2) Интервальные О., Д.И (доверительные интервалы)

3) Стат. проверка стат. гипотез

(ГС) Генеральная совокупность и выборка (В)

Пусть имеется $X = \{X_1, X_2, \dots, X_N\}$
Из этого мн. осущ. выборка с возвращ. n эл-тов вида $x = \{x_1, x_2, \dots, x_n\}$. Тогда исходное мн. наз. Г.С., мн. x наз. выборкой объёмом n из этой Г.С.

Зам

Понятие Г.С. не сводится к конечным мн-вам. В общем случае, это некий генератор значений СВ сл с конкретным распред.

Упоряд. по возр. В наз. вариационным рядом

$(x_1) \le (x_2) \le \ldots \le (x_n)$

Эмпирический з-н распред.

X	x_1	x_2	$\ldots$	x_k
n_i	n_1	n_2	$\ldots$	n_k
W_i	W_1	W_2	$\ldots$	W_k

k – т.к. м/б повторяющиеся эл-ты

x_i – варианта

n_i – частота варианты; $\sum_{i=1}^{k} n_i = n$

$W_i = \frac{n_i}{n}$ – относ. ч.в.; $\sum_{i=1}^{k} W_i = 1$

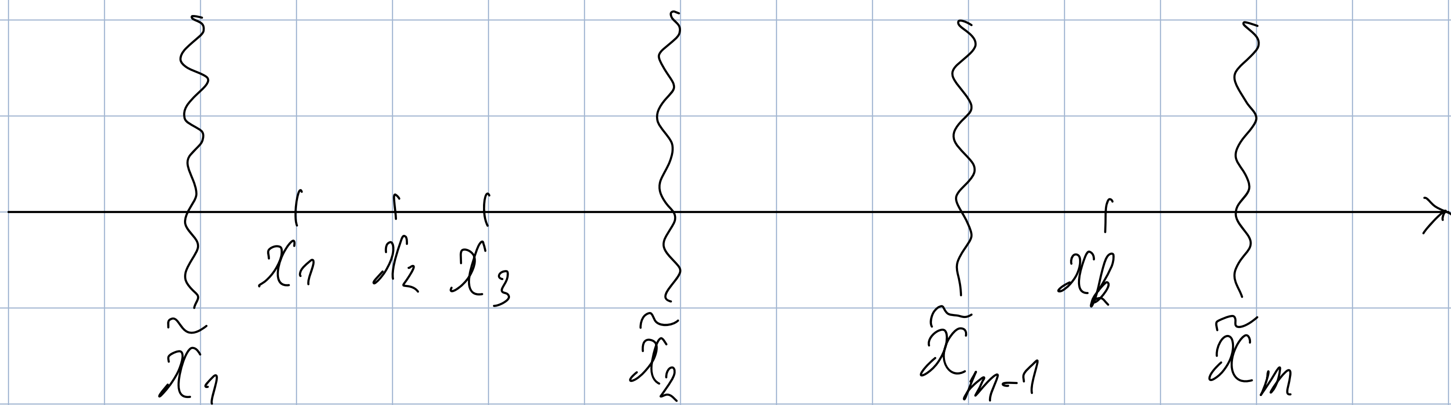
Эмп. з-н не явл. С, он сам СВ

Эмпирическая ф. распред.

$$\hat{F}(x)=\sum_{x_i \le x} w_i$$

Э. ф. распред. обл. всеми св-вами теоретической ф. распред

Интервальные з-ны распред.

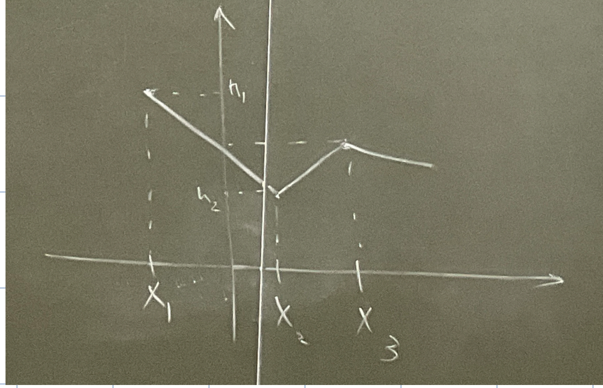


$[\tilde{x}_i; \tilde{x}_{i+1})$	$[\tilde{x}_1; \tilde{x}_2)$	...	$[\tilde{x}_{m-1}; \tilde{x}_m]$
n_i	n_1	...	n_{m-1}
w_i	w_1	...	w_{m-1}

Полигон частот и отн. частот. В

Пол. ч. наз. ломанная, кот. послед. соед. (·)ки плос-ти с коорд. $(x_i; n_i)$

Пол. отн. ч. наз... $(x_i; w_i)$



Пол. ч.

В случае И. распред. пол. ч. выстраивается по (•) $\left(\frac{\tilde{x}_i + \tilde{x}_{i+1}}{2}; n_i\right)$

В ... пол. отн. ч. ... (•) $\left(\frac{\tilde{x}_i + \tilde{x}_{i+1}}{2}; w_i\right)$

(Г.) Гистограмма частот и отн. частот В

Г. строится только в случ. И. распред.

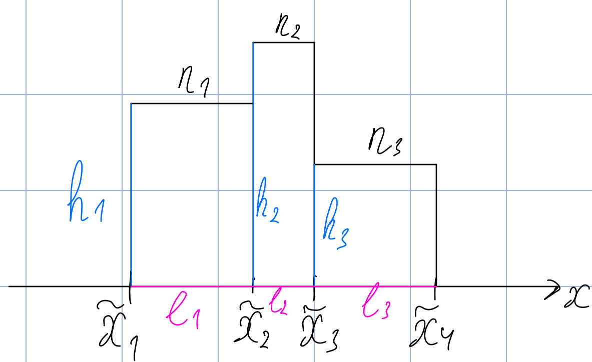
Алгоритм построения Г. частот

1) На ось абсцисс выносятся все знач-ия И-лов, т. е (интервалов)
$\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_m$

2) На каждом отр., как на основании, достраивается , высота которого $h_i=\frac{n_i}{l_i}$; l_i – длина i-го И-ла

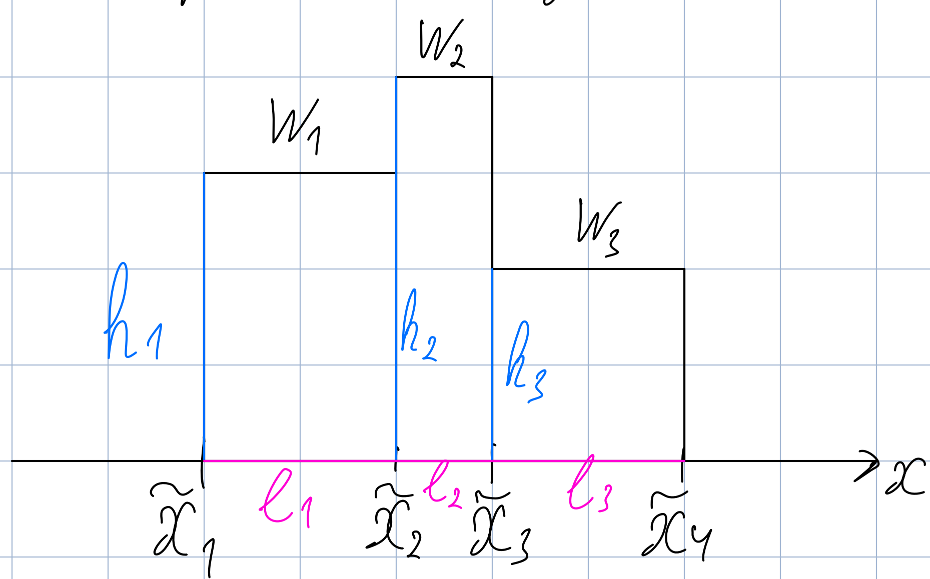
3) Многоуг., кот. возникает, наз. Г. ч-т

4) Если в п.2 высоты прямоуг. строятся по фор. $h_i=\frac{W_i}{l_i}$, где l_i – длина i-го отрезка, то возникающая фигура наз. Г. отн. частот



$n=n_1+n_2+n_3$

Плош. Г. ч-т равна объёму выборки (n)



При пост. Г. отн. ч-т, площ = 1

Зам

Г. отн. ч-т явл. стат. аналогом плот. распред. вероят-ей НСВ

Функ. хар-ки ГС и В

$X = \{X_1, \dots, X_N\}$

Средневыборочным зн. ГС наз. ср. арифм. всех её эл-тов

$$\overline{X}_{ген} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

Д. ГС наз. выр. вида:

$$D_{ген} = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{X}_{ген}\right)^2$$

$$\sigma_{ген} = \sqrt{D_{ген}}$$

$x = \{x_1, \dots, x_n\}$

Средневыборочным зн. В наз. ср. арифм. всех её эл-тов

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

Д. В наз. выр. вида:

$$D_{выб} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}_{выб}\right)^2$$

$$\sigma_{выб} = \sqrt{D_{выб}}$$

Т.О. пар-ов распред. СВ, кот. порождается ГС

ξ	0	1
p	0,6	0,4

Предположим, что ГС порождает СВ, обл. некоторым пар. Θ (тэта)

Т.О. $\hat{\Theta}$ пар Θ распред СВ наз. фор. след. вида:

$\hat{\Theta} = \hat{\Theta}(x_1, \ldots, x_n)$, в кот. н/подставить зн-ия В, чтобы получить приближ. зн. искомого пар. Θ

$$\hat{\Theta} = \hat{\Theta}(x_1, \ldots, x_n) \approx \Theta$$

Св-ва Т.О.:

1) Смещенность и не - Т.О.

Т.О. наз. несмещённой, если её М.О. совпадает с оцениваемым пар Θ

$$M\hat{\Theta} = \Theta$$

Иначе О. наз. смещённой и выр. $M\hat{\Theta} - \Theta$ наз. смещением О.

Среднекв. ошибкой О. наз. М.О. от квадрата откл. О. от самого пар., т.е. $M(\hat{\Theta} - \Theta)^2$

Для смещ. О.:

$$M(\hat\theta-\theta)^2 = M(\hat\theta^2 - 2\theta\hat\theta+\theta^2) = M(\hat\theta^2) - 2\theta\cdot M\hat\theta+\theta^2 =$$

$$= \underbrace{M(\hat\theta^2)-(M\hat\theta)^2}+\underbrace{(M\hat\theta)^2-2\theta\cdot M\hat\theta+\theta^2} = D\hat\theta + M(\hat\theta-\theta)^2$$

Для несм. О.:

$$M(\hat\theta-\theta)^2 = D\hat\theta$$

1) $\bar x_{выб} = \dfrac{x_1+x_2+\dots+x_n}{n}$

X_1 ||

$$M(\bar x_{выб}) = \frac{Mx_1+Mx_2+\dots+Mx_n}{n} = \bar X_{ген}$$

x_i	x_1	x_2	…	x_n
p_i	p_1	p_2	…	p_n

$\bar x_{выб} \longrightarrow$ несм. О $\bar X_{ген}$

2)

$$M D_{выб.} = \frac{n-1}{n} D_{ген}$$

$D_{выб} \longrightarrow$ смещ. О $D_{ген.}$

$$M\left(\frac{n}{n-1} D_{выб.}\right) = D_{ген}$$

Дистр.

Дисп. $\to$ несм. О. $D_{ген}$

- Если $n < 30 \to D_{выб}$ и $D_{испр.}$ различают
- Если $n > 30 \to D_{выб} \approx D_{испр.}$, т.е. не различают

2) Эффективность Т. О.

Т. О. наз. **эффективной**, если среднекв. ошибка этой О. явл. наим. среди всех возможных.

$$M(\hat{\theta}-\theta)^2 \le M(\tilde{\theta}-\theta)^2; \forall \tilde{\theta}$$

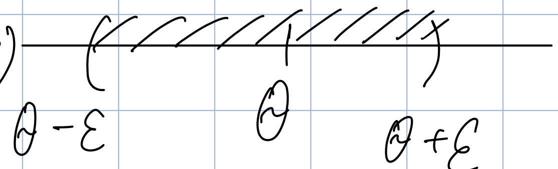
$\bar{x}_{выб} \to$ Э.О. $\bar{X}_{ген}$

$D_{выб} \to$ Э.О. $D_{ген}$

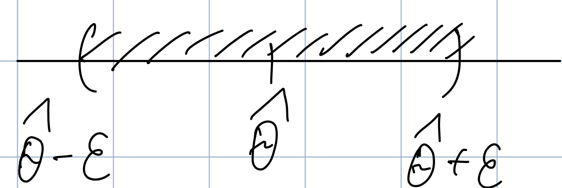
3) Состоятельность О.

Т. О. наз. **состоятельной**, если по мере увелич. объёма В. эта О. сходится по вероят-ти к самому пар., т.е. $\forall \varepsilon > 0$ выполн.:

След.: $\lim\limits_{n\to\infty} P(|\hat{\theta}-\theta| < \varepsilon) = 1 \iff$ 1)



2)



$\bar{x}_{выб}$ – сост. О. $\bar{X}_{ген}$

$D_{выб}$ – сост. О. $D_{ген}$.

И.О пар-ов распред. Построение Д.И.

Пусть ГС порождает СВ, кот. хар-ется неким пар θ.

Пусть также для этого пар сущ. несмещ. $\hat{\theta}$.

Если $\exists \varepsilon > 0$ для кот с вероят γ выполн усл.:

$$P\left(|\hat{\theta} - \theta| < \varepsilon\right) = \gamma$$

$$P(\hat{\theta} - \varepsilon < \theta < \hat{\theta} + \varepsilon) = \gamma$$

Т. И. $(\hat{\theta} - \varepsilon; \hat{\theta} + \varepsilon)$ наз. дов. интер. θ.

ε – точность О.

γ – дов. вероят. или надежность О.

Зам. 1:

Не верно говорить, что пар. θ попадает в Д.И.

Правильно говорить, что Д.И накрывает пар. θ

Зам. 2:

Сами границы Д.И. явл. СВами

Зад. 1.

„О расчёте ДИ М.О. н/р Г.С. с известным станд. откл.:

Дано: ГС – норм. распред. с пар $(a;\sigma)$;

σ – задана;

$X=\{x_1, x_2, \dots, x_n\}$;

γ.

Найти:

ε – ?

$(\overline{x}_{выб}-\varepsilon;\ \overline{x}_{выб}+\varepsilon)$ – ?

① $\overline{x}_{выб}=\dfrac{x_1+x_2+\dots+x_n}{n}$; x_i – н/р. с пар. $(a;\sigma) \longrightarrow \overline{x}_{выб}$ – н/р

$M_{\overline{x}_{выб}}=a$; $D_{\overline{x}_{выб}}=\dfrac{\sigma^2}{n}$

$\overline{x}_{выб}$ – н/р с пар. $\left(a;\dfrac{\sigma}{\sqrt{n}}\right)$

② $P\left(|\overline{x}_{выб}-a|<\varepsilon\right)=2\cdot\varphi\left(\dfrac{\varepsilon}{\frac{\sigma}{\sqrt{n}}}\right)=\gamma \xrightarrow{\text{по т. Лапласса}} \dfrac{\varepsilon}{\frac{\sigma}{\sqrt{n}}}=t^* \rightarrow$

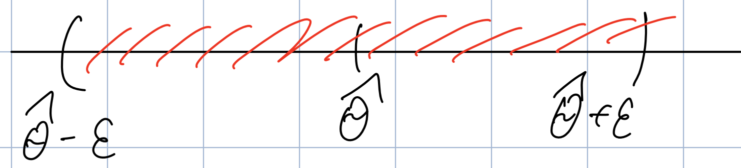
$$\varphi\left(\dfrac{\varepsilon}{\frac{\sigma}{\sqrt{n}}}\right)=\dfrac{\gamma}{2}$$

$\rightarrow \varepsilon=\dfrac{t^*\cdot\sigma}{\sqrt{n}}$

$(\overline{x}_{выб}-\varepsilon;\ \overline{x}_{выб}+\varepsilon)$ – Д.И.

Зам.:

$$n=\left(\frac{\sigma\cdot t^*}{\varepsilon}\right)^2$$



Зад. 2.

"О расчёте ДИ М.О. н/р. Г.С. с неизвестным. станд. откл."

Дано: ГС – норм. распред. с пар-ми $(a;\sigma)$;

$X=\{x_1,x_2,\dots,x_n\}$;

γ.

Найти:

ε – ?

$(\overline{x}_{выб}-\varepsilon;\ \overline{x}_{выб}+\varepsilon)$ – ?

Решение:

1) $\overline{x}_{выб}$ – н/р. с пар $\left(a;\frac{\sigma}{\sqrt{n}}\right)$;

$$\xi=\frac{\overline{x}_{выб}-a}{\frac{\sigma}{\sqrt{n}}}$$ – н/р с пар $(0;1)$

2) $$V=\frac{(n-1)\cdot D_{исп}}{D_{ген}}$$ – χ^2 распред. с $(n-1)$ ст. св.

3) $$t=\frac{\xi}{\sqrt{\frac{1}{n-1}V}}$$ – t-распред. с $(n-1)$ ст. св.

$$t = \frac{\overline{x}_{выб} - a}{\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{\cancel{(n-1)} D_{испр.}}{\cancel{(n-1)} D_{ген}}}} = \frac{(\overline{x}_{выб} - a) \cdot \sqrt{n}}{\sigma_{исп.}}$$

4) $P\left(|\overline{x}_{выб} - a| < \varepsilon\right) = \gamma$

$$P\left(\frac{|\overline{x}_{выб} - a| \sqrt{n}}{\sigma_{испр.}} < \frac{\varepsilon \sqrt{n}}{\sigma_{испр.}}\right) = \gamma$$

$$P\left(|t| < \frac{\varepsilon \sqrt{n}}{\sigma_{испр.}}\right) = \gamma$$

по табл
t- распр.

$$\frac{\varepsilon \sqrt{n}}{\sigma_{испр.}} = t^{**}$$

$$\varepsilon = \frac{\sigma_{испр.} \cdot t^{**}}{\sqrt{n}}$$

$$\left(\overline{x}_{выб} - \varepsilon ; \overline{x}_{выб} + \varepsilon\right)$$

Зад. 3.

„Рассчёт ДИ. средне. кв. откл. н/р СВ"

Дано: ГС – норм. распред. с пар $(a; \sigma)$;

$X = \{x_1, x_2, \ldots, x_n\}$;

γ.

Найти:

ε - ?

$(\sigma_{испр.}-\varepsilon; \sigma_{испр.}-\varepsilon)$ - ?

Решение:

$$P(|\sigma_{испр.}-\sigma|<\varepsilon)=\gamma \longrightarrow |\sigma_{испр.}-\sigma|<\varepsilon \longrightarrow \sigma_{испр.}\left|1-\frac{\sigma}{\sigma_{испр.}}\right|<\frac{\varepsilon}{\sigma_{испр.}}$$

$$\frac{\varepsilon}{\sigma_{испр}}=q: \quad -q<1-\frac{\sigma}{\sigma_{испр}}<q$$

$$-1-q<\frac{\sigma}{\sigma_{испр}}<q-1$$

$$1+q>\frac{\sigma}{\sigma_{испр}}>1-q \qquad a>b>c>0$$

$$\frac{1}{1+q}<\frac{\sigma_{испр.}}{\sigma}<\frac{1}{1-q} \qquad \frac{1}{a}<\frac{1}{b}<\frac{1}{c}$$

$$\frac{\sqrt{n-1}}{1+q}<\frac{\sqrt{n-1}\,\sigma_{испр.}}{\sigma}<\frac{\sqrt{n-1}}{1-q}$$

2) $\frac{\sqrt{n-1}\,\sigma_{испр}}{\sigma}$ имеет χ распред. с $(n-1)$ ст. св.

$$P\left(\frac{\sqrt{n-1}}{1+q}<\chi<\frac{\sqrt{n-1}}{1-q}\right)=\gamma \xrightarrow[\chi\text{-распред}]{\text{по табл.}} q=\frac{\varepsilon}{\sigma_{испр.}} \longrightarrow$$

$$\rightarrow \varepsilon=q\cdot\sigma_{испр.}; \qquad (\sigma_{испр}-\varepsilon; \sigma_{испр}+\varepsilon)$$