ИНСТИТУТ
МАТЕМАТИКИ
МЕХАНИКИ
КОМПЬЮТЕРНЫХ
НАУК

ЮЖНЫЙ
ФЕДЕРАЛЬНЫЙ
УНИВЕРСИТЕТ

Parametric families.

Distributions connected with the Normal

Kurbatova Natalia Victorovna

nvkurbatova@sfedu.ru

Parametric families.

Lecture 3(II)

Definition 1.

$\mathbf{X}=\{X_i\}$ – *sample* $\in \mathcal{F}_\theta$ (*known family, unknown parameter* θ (*scalar of vector*)), $\theta \in \Theta$, Θ – *is the set of possible values*

for example:

$$\mathcal{F}_\theta = \begin{cases} P_\lambda, & \theta = \lambda > 0, & Poisson \\ B(p), & \theta = p \in (0,1), & Bernoulli \\ U(a,b), & \theta = a, b; \quad a < b, & Uniform \\ \mathcal{N}(a,\sigma^2), & \theta = a, \sigma; \; a \in R, \sigma > 0, & Normal \end{cases}$$

Statistics – an arbitrary Borel, measurable function – $\theta^* = \theta^*(X_1, .. X_n)$ is estimate of θ; θ^* – random value (as function of the sample $\mathbf{X}$) .

ИНСТИТУТ
МАТЕМАТИКИ
МЕХАНИКИ
КОМПЬЮТЕРНЫХ
НАУК

Point Estimate

Lecture 3(III) in details

ЮЖНЫЙ
ФЕДЕРАЛЬНЫЙ
УНИВЕРСИТЕТ

Definition 2. Statistics θ^* – unbiased $\theta^* = \theta^*(X_1, ..X_n)$ – estimation of the true parameter θ; if for $\forall\, \theta \in \Theta,\ E\theta^* = \theta,\ n - fixed$

- Unbiasedness – no error on average (after using)

Definition 3. Statistics θ^* – asymptotically unbiased estimation of θ; if $\forall\, \theta \in \Theta$ the convergence takes place: $E\theta^* \to \theta \quad if \quad n \to \infty$

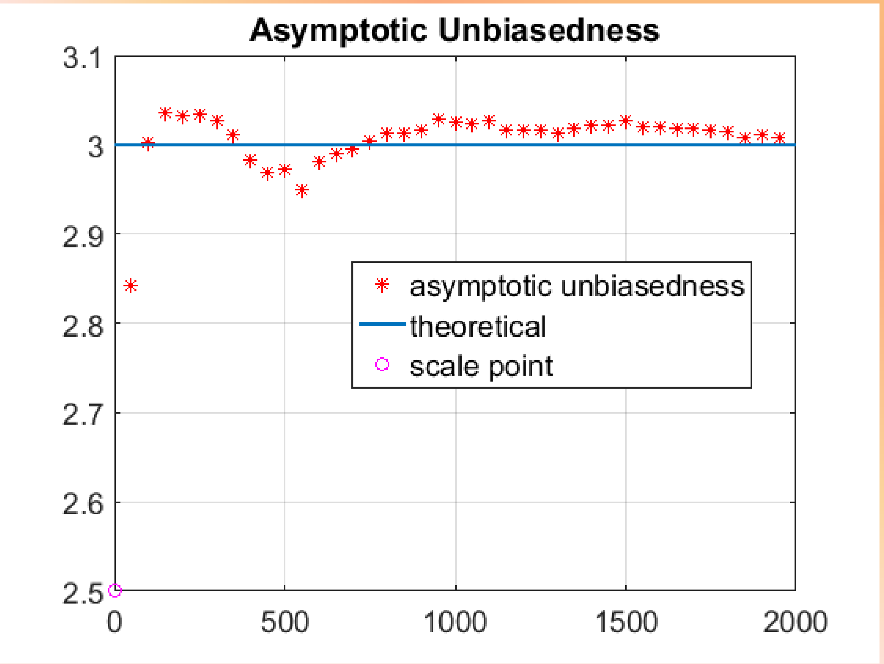
- Asymptotically unbiasedness – the difference between its mean and true parameter decrease with increasing of **sample size**

Definition 4. Statistics $\theta^* = \theta^*(X_1, ..X_n)$ – consistent estimation of θ, if for $\forall\, \theta \in \Theta$, $\theta^* \xrightarrow{P} \theta$, if $n \to \infty$

- Consistence – it means that the sequence of estimates tends to unknown parameter with increasing of the number of observations

Unbiasedness of Statistics

- MLM illustrates definition 1 for Family with Normal distribution (p.7, current presentation) .
- Interpretation of asymptotic unbiasedness (definition 2) is presented on the graph




```
clear      % AsymptoticUnbiasedness.m
X=random('norm',3,1,1,2000);
meani=zeros(1,40); n=zeros(1,40);
 for i=1:40
     rightpoint=50*i;
     meani(i)=mean(X(1:rightpoint));
     n(i)=rightpoint;
 end
  plot(n(1:end),meani(1:end),'r*'); hold on;
  plot([0;n(end)],[3;3],'linewidth',1.5);
  plot(0,2.5,'mO');
set(gca,'fontsize',14)
lg=legend('asymptotic unbiasedness',...
              'theoretical','scale point')
    set(lg,'fontsize',14); grid on;
    title('Asymptotic Unbiasedness')
```

How to get a point estimate ?

1. Moment method .

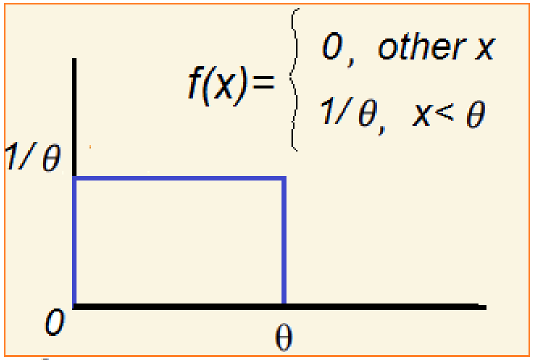
The main idea: each moment of r. v. X_1 – is some function h of θ; substituting the sample analogue of the moment in the inverse function h^{-1} with respect to θ instead of the true value, we get an estimate θ* of the true value θ.

Property of MM estimation:

Let $\theta^* = h^{-1}\left(\overline{g(X)}\right)$ – MM estimate of , h^{-1} – continuous function then θ^* – *consistent estimate.*

Interpretation: The MM estimate is taken as an estimate of a random parameter value, at which the true point coincides with the moment of sampling

Example : $X_1, \ldots X_n$ – sample $\in$ uniform distribution $U(0,\theta)$, $\theta > 0$.



Determine θ_1^* and θ_k^*(using the first and k – th moments):

a) θ_1^*: $g(y) = y$; for uniform distributed random variable f.e. X_1

$$EX_1 = \frac{\theta}{2}, \text{so } \theta = 2EX_1, \theta_1 = 2\overline{X}$$

$$b)\ \theta_k^*: EX_1^k = \int_0^\theta y^k \frac{1}{\theta} dy = \frac{\theta^k}{k+1}; \quad \theta = \sqrt[k]{(k+1)EX_1^k} \Rightarrow \theta_k^* = \sqrt[k]{(k+1)\overline{X_1^k}}$$

Here you can watch a short explanation about properties of uniform distribution:

https://www.youtube.com/watch?app=desktop&v=UC-CBUSQXAo

2. Maximum likelihood method (MLM)

2. Maximum likeli-hood method (MLM)

the most likely ~ the most probably

MLM – another approach to construct estimate of unknown distribution's parameters using sample $(X_1, \dots X_n)$.

The main idea: as the most plausible parameter value will be taken the value θ, maximizing probability of obtaining the sample $(X_1, \dots, X_n)$

$$P(X_1 \in (y, y+dy) = f_\theta(y)dy \quad (\text{noted } f(y,\theta) = f_\theta(y))$$

Given the nature of random variable, we proposed the following kind of density function:

$$f(y,\theta) = \begin{cases} f(y,\theta), & if\ \mathcal{F}_\theta - \text{absolutely continuous} \\ P_\theta(X_1 = y), & if\ \mathcal{F}_\theta - \text{descrete} \end{cases}$$

Here $\mathcal{F}_\theta$ – distribution family.

Definition. Likelihood function (LF) is

$$f(x_1, x_2, \dots, x_n, \theta) = f\ (X_1,\theta) \cdot f(X_2,\theta) \cdot \dots \cdot f(X_n,\theta) =$$

$$= \prod_{i=1}^{n} f(X_i,\theta) \quad \text{and (LLF) Logarithmic likelihood function}$$

$$- is\ L(X_1, X_2, \dots, X_n, \theta) = \ln\big(f(X_1, X_2, \dots X_n, \theta)\big) = \sum_{i=1}^{n} \ln f(X_i,\theta)$$

ИНСТИТУТ МАТЕМАТИКИ МЕХАНИКИ КОМПЬЮТЕРНЫХ НАУК

MLM for Poisson distribution. Example

ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

Let $X_1, \dots, X_n \in P_\lambda$, Poisson family, $\lambda > 0$.

Find $\hat{\lambda}$: Based on density function for Poisson family distribution P_λ:

$$\boxed{f_\lambda(y) = P(X_1 = y) = \frac{\lambda^y}{y!} e^{-\lambda}; \quad y = 0, 1, 2, \dots}$$

We will determine likelihood function

$$f(X_1, X_2, \dots, X_n, \lambda) = \prod_{i=1}^{n} \frac{\lambda^{X_i}}{X_i!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n X_i!} e^{-\lambda n} = \frac{\lambda^{n\bar{X}}}{\prod_{i=1}^n X_i!} e^{-\lambda n}; \quad \lambda > 0,$$

but easier to use L : $\underline{\mathrm{L}(\mathrm{X}_1, \mathrm{X}_2, \ldots \mathrm{X}_\mathrm{n}, \lambda) = \ln f(X_1, \ldots X_n, \lambda)} =$

$$= \ln\left(\frac{\lambda^{n\bar{X}}}{\prod X_i!} e^{-n\lambda}\right) = \underline{n\bar{X} \ln \lambda - \ln \prod X_i! - n\,\lambda};$$

partial derivative:

$$\frac{\partial}{\partial \lambda} L(\mathrm{X}_1, \mathrm{X}_2, \ldots \mathrm{X}_\mathrm{n}, \lambda) = \frac{n\bar{X}}{\lambda} - n = 0, \quad \hat{\lambda} = \bar{X}, \quad \hat{\lambda} - \textit{ maximal value}$$

MLM for Normal distribution. Example

Let sample $X_1,..,X_n \in \mathcal{N}(a,\sigma^2), a \in R, \sigma > 0; a,\sigma$ – unknown

$$\boxed{f(y,a,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(y-a)^2}{2\sigma^2}};} \quad - N(a,\sigma^2)$$

$$LF: f(X_1,X_2,..X_n,a,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x_i-a)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{\frac{-\sum_{i=1}^{n}(x_i-a)^2}{2\sigma^2}}$$

$$LLF: L(X_1,X_2,..X_n,a,\sigma^2) = \ln f(X_1,X_2,\dots,X_n,a,\sigma^2) =$$

$$= -\ln(2\pi)^{\frac{n}{2}} - \frac{n}{2}\ln(\sigma^2) - \frac{\sum_{i=1}^{n}(X_i-a)^2}{2\sigma^2}$$

Find extreme points:

$$\begin{cases} \dfrac{\partial L}{\partial a} = \dfrac{2\sum_{i=1}^{n}(X_i-a)}{2\sigma^2} = \dfrac{n\bar{X}-na}{\sigma^2} = 0 \\ \dfrac{\partial L}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{\sum_{i=1}^{n}(X_i-a)^2}{2\sigma^4} = 0 \end{cases}$$

$$LM\ estimations\ : n\bar{X} - na = 0\,;\ -\sigma^2 + \frac{1}{n}\sum_{i=1}^{n}(X_i-a)^2 = 0$$

$\boxed{\hat{a} = \bar{X},\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S^2}$ – identical to the first and the second empirical moments.

Gamma function properties

Recall the basic properties of the gamma function:

Remark:

1) $\Gamma(\alpha, \lambda) = \int_0^\infty x^{\lambda-1} e^{-\alpha x}\, dx = \Gamma(\lambda)/\alpha^{\lambda}$,

2) $\Gamma(\alpha, \lambda) = \int_0^\infty x^{\lambda-1} e^{-x/\alpha}\, dx = \alpha^{\lambda}\Gamma(\lambda)$

3) $\Gamma(1) = 1, \ (\alpha = 1)$

4) $\Gamma(\lambda + 1) = \lambda\Gamma(\lambda), \ (\alpha = 1)$

5) $\Gamma(1/2) = \sqrt{\pi}$, parameters α – scale, λ – shape.

The gamma function provides the relationship of different distributions!

Gamma distribution

Probability density function:

$$\Gamma \sim \gamma(\alpha,\lambda): \quad f_{\gamma(\alpha,\lambda)}(x) = \begin{cases} \dfrac{\alpha^{\lambda}}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Estimate of distribution parameters by the sample:

$$\hat{\alpha} = \left(\frac{\bar{x}}{S^2}\right), \qquad \hat{\lambda} = \left(\frac{\bar{x}}{S}\right)^2$$

Expectation and variance of Gamma distribution:

$$E\,\gamma(\alpha,\lambda) = \frac{\lambda}{\alpha}, \qquad D\,\gamma(\alpha,\lambda) = \frac{\lambda}{\alpha^2}$$

Properties of Gamma distribution.

Lemma 1.

The property of stability by summation:

Let $X_1, \dots, X_n$ are independent and ξ_i have Gamma distribution $\Gamma_{\alpha,\lambda_i}$, $i = 1, \dots, n$, then their *sum* has Gamma distribution with parameters α, $\lambda = \lambda_1 + \cdots + \lambda_n$: $\Gamma_{\alpha,\lambda}$.

Remark.

The square of r. v. with standard normal distribution has Gamma distribution. $\Gamma(1/2, 1/2)$

It would be great to confirm the remark graphically!

Generate sample in ML:

```
Y=random('gamma',0.5,0.5,1,2000);
```

Lemma 2

If ξ has standard normal distribution,
then ξ^2 has Gamma distribution $\Gamma_{1/2, 1/2}$.

Proof:

Find the derivative of distribution function of r.v. ξ^2, let's show that this resulting function – is the density function:

$$y \le 0: \qquad F_{\xi^2}(y) = P(\xi^2 \overset{?}{<} y) = 0 \to f_{\xi^2}(y) = 0$$

$$y > 0: \quad F_{\xi^2}(y) = P(-\sqrt{y} < \xi < \sqrt{y}) = F_\xi(\sqrt{y}) - F_\xi(-\sqrt{y}),$$

$$f_{\xi^2}(y) = \left(F_{\xi^2}(y)\right)' = F'_\xi(\sqrt{y})\frac{1}{2\sqrt{y}} + F'_\xi(-\sqrt{y})\frac{1}{2\sqrt{y}} =$$

Normal st. dens. fun

$$[f_\xi(\sqrt{y}) + f_\xi(-\sqrt{y})]\frac{1}{2\sqrt{y}} = (\text{symm } f_\xi) = \frac{f_\xi(\sqrt{y})}{\sqrt{y}} = f_\xi^N = \frac{1}{\sqrt{2\pi y}} e^{\frac{-y}{2}}$$

Continue proof, taking into account both semi-intervals for F_{ξ^2}

$$\int_{-\infty}^{\infty} f_{\xi^2}(y)\,dy = \int_{-\infty}^{\infty} \left(1/2\right)^{1/2} \frac{y^{1/2-1} e^{-y/2}}{\Gamma(1/2)} dy = \left(1/2\right)^{1/2} \frac{\Gamma(1/2, 1/2)}{\Gamma(1/2)} =$$

$$= [\,remarks\ (1{,}4)] = \frac{\left(1/2\right)^{1/2}\Gamma(1/2)}{\Gamma(1/2)\left(1/2\right)^{1/2}} = 1$$

so $f_{\xi^2}(y)$ – density function!

Distribution χ 2 (Pearson)

Lemma 3: If $\xi_1, \dots, \xi_k$ – are independent and have standard normal distribution, then r. v. $\chi^2 = \xi_1^2 + \dots + \xi_k^2$ has χ^2 or Pearson distribution – $\Gamma\left({}^1/_2, {}^k/_2\right)$.

Definition: Distribution of k squares of independent r. v. with St. Norm. Distr. is called chi-square (χ_k^2) with k-degrees of freedom and denoted H_k. According to lemma 3 H_k is the same as $\Gamma\left({}^1/_2, {}^k/_2\right)$, so distribution function for family distributions H_k depends on k and equal

$$f(y) = \begin{cases} \dfrac{1}{2^{\frac{k}{2}}\Gamma\left({}^k/_2\right)} y^{\frac{k}{2}-1} e^{-\frac{y}{2}}, & y > 0 \\ 0, & y \le 0 \end{cases}$$

Remarks:

Stability of χ^2 with respect to summation follows from the stability of Γ- distribution.

Note that $H_2 = \Gamma_{1/2,1} = EXP(1/2)$, Exponential distribution.

ИНСТИТУТ МАТЕМАТИКИ МЕХАНИКИ КОМПЬЮТЕРНЫХ НАУК

Distribution χ 2 properties (1-3)

ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

P 1:

If r. v. $\chi^2 \in H_k$ and $\psi^2 \in H_m$, then the sum $\chi^2+\psi^2 \in H_{m+k}$.

Let $\xi_1, \xi_2, \dots$ – are independent , $\in N(0,1)$, then $\xi_1^2+\dots+\xi_k^2$ has the same distribution as χ^2, analogically ψ^2 and $\xi_k^2+\dots+\xi_{k+m}^2$, so the sum $\xi_1^2+\dots+\xi_{k+m}^2 \in H_{m+k}$, is proven.

P 2: ∃

If $\chi^2 \in H_k \Rightarrow E\chi^2 = k,\ D\chi^2 = 2k$

$\xi_1, \xi_2, \dots$ – are independent with NSD, then

$E\xi_1^2 = 1, D\xi_1^2 = E\xi_1^4 - (E\xi_1^2)^2 = [\dots] = 3 - 1 = 2,$

[*based on theory prob. property*:

$E\xi^{2k} = (2k-1)!! = (2k-1)(2k-3) * \dots * 3 * 1,\ E\xi_1^4 = 3$]

so $E\chi^2 = E(\xi_1^2+\dots+\xi_k^2) = k;$

$D\chi^2 = D(\xi_1^2+\dots+\xi_k^2) = kD(\xi_1^2) = 2k,$ is proven.

(1-3)

Try to guess approach for proof, the main idea!

P 3:

Let $\chi_n^2 \in H_n$, then if $n\to\infty$ $\frac{\chi_n^2}{n} \xrightarrow{P} 1,\ \frac{\chi_n^2-n}{\sqrt{2n}} \Rightarrow N(0,1)$.

For any n, χ_n^2 has the same distribution as $\xi_1^2+\dots+\xi_n^2$, $\xi_i \in N(0,1)$ and independent. According to LLN and CLT (central limit theorem), have

$$(\xi_1^2+\dots+\xi_n^2)/n \xrightarrow{P} E\xi_1^2 = 1,\quad \text{and}\quad \frac{\xi_1^2+\dots+\xi_n^2-n}{\sqrt{2n}} = \frac{\xi_1^2+\dots+\xi_n^2-nE\xi_1^2}{\sqrt{nD\xi_1^2}} \Rightarrow N(0,1)\ \square$$

Distribution χ 2 properties (4-5)

P 4: Let $\chi_n^2 \in H_n$, then if $n \to \infty$, the following weak convergence takes place $\sqrt{2\chi_n^2} - \sqrt{2n-1} \Rightarrow N(0,1)$

Therefore for large n $\exists$ the approximation for distribution function

$$H_n(x) = P(\chi_n^2 < x) \text{ and } H_n(x) \approx \Phi_{0,1}(\sqrt{2x} - \sqrt{2n-1})$$

- $\sqrt{2n} - \sqrt{2n-1} \to 0,\ n \to \infty$ (clear);

$$\sqrt{2\chi_n^2} - \sqrt{2n} = \frac{2}{1+\sqrt{\chi_n^2/n}} \frac{\chi_n^2 - n}{\sqrt{2n}} \Rightarrow N(0,1)$$

because based on P3 $\frac{2}{1+\sqrt{\chi_n^2/n}} \xrightarrow{p} 1$ and $\frac{\chi_n^2 - n}{\sqrt{2n}} \Rightarrow N(0,1)$

- $P(\chi_n^2 < x) = P(\sqrt{2\chi_n^2} - \sqrt{2n-1} < \sqrt{2x} - \sqrt{2n-1})$ □

P 5: If r. v. $\xi_1, \dots, \xi_k$ – Independent and belong to $N(a, \sigma^2)$,

then $\chi_k^2 = \sum_{i=1}^{k} \left(\frac{\xi_i - a}{\sigma}\right)^2 \in H_k$ $\boxed{\text{clear}}$

ИНСТИТУТ
МАТЕМАТИКИ
МЕХАНИКИ
КОМПЬЮТЕРНЫХ
НАУК

Examples of real characteristics submitting this law

ЮЖНЫЙ
ФЕДЕРАЛЬНЫЙ
УНИВЕРСИТЕТ

- **The normalized sample variance**

- **Measure of deviation of a hypothetical distribution from a theoretical one**

Student Distribution (StD)

Definition

Let $\xi_0, \xi_1, \ldots, \xi_k \in N(0,1)$ and independent, the distribution of the random variable

$$t_k = \frac{\xi_0}{\sqrt{\frac{\xi_1^2 + \cdots + \xi_k^2}{k}}}$$

is called Student distribution with k degrees of freedom and denoted T_k or the same r. v.:

$$t_k = \frac{\xi}{\sqrt{\chi_n^2 / k}}, \xi \in N(0,1), \chi_k^2 \in H_k$$

with density function:

$$f_k(y) = \frac{\Gamma((k+1)/2)}{\sqrt{\pi k}\,\Gamma(k/2)} \left(1 + \frac{y^2}{k}\right)^{-(k+1)/2}$$

ИНСТИТУТ МАТЕМАТИКИ МЕХАНИКИ КОМПЬЮТЕРНЫХ НАУК

Student Distribution Properties

ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

P 6: Student density function $f_k(y)$ is symmetric: if random variable $t_k \in T_k$, then $-t_k \in T_k$.

P 7: Student distribution slightly (weakly) converges to $N(0,1)$ if $n \to \infty$.

Proof: According to P3, $\frac{\chi_n^2}{n} \xrightarrow{P} 1$, if $n \to \infty$ and $t_k = \frac{\xi}{\sqrt{\chi_k^2/k}} \Rightarrow \xi \in N(0,1)$, is proven.

Remark: For large k one can use normal density function for approximation StD!

Remark: For k= 1 Student distribution is Cauchy distribution with density function:

$$f_1(y) = \frac{1}{\pi}(1 + y^2)^{-1}$$

Proof: *follows* from $\Gamma\left({}^1/_2\right) = \sqrt{\pi}$ and $\Gamma(1) = 1$.

Property. Brain Storm

P.8 There are moments of order $m<k$ for Student distribution and there are no moments of order $m \geq k$.

Brain Storm:

All existing moments of odd order are equal to zero!

Why?

Examples of real characteristics submitting Student law

- **The normalized of the deviation the mean value of the sample taken from a normally distributed General Population and theoretical one**

- **A measure of the deviation of the mean value of two independent samples taken from a normally distributed General Population**

Fisher distribution

The Fisher distribution is also closely related to the normal one, it is often called the **distribution of the variance ratio!**

Let $\chi_k^2 \in H_k$, $\psi_n^2 \in H_n$ both r. v. independent. Distribution of r. v.

$$f_{k,n} = \frac{\chi_k^2/k}{\psi_n^2/n} = \frac{n}{k}\frac{\chi_k^2}{\psi_n^2}$$

Definition

is called Fisher's distribution.

P 9:

If r. v. $f_{k,n} \in F_{k,n} \to \frac{1}{f_{k,n}} \in F_{n,k}$.

For any x : $F_{k,n}(x) = P\left(f_{k,n} < x\right) =$

$= P\left(\frac{1}{f_{k,n}} > \frac{1}{x}\right) = 1 - F_{n,k}\left(\frac{1}{x}\right).$ $\square$

P 10:

The Fisher distribution $F_{k,n}$ weakly converges to the degenerate distribution at the point c = 1 for any tendency k,$n \to \infty$.

P 11:

Let r.v. $t_k \in T_k$, then $t_k^2 \in F_{1,k}$.

Proof follows from definition of t_k, and $f_{1,k}$

ИНСТИТУТ МАТЕМАТИКИ МЕХАНИКИ КОМПЬЮТЕРНЫХ НАУК

Examples of real characteristics submitting Fisher law

ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

- **Ratio of variances of the two independent samples taken from a normally distributed General Population**

- **A multidimensional analogue of Student statistics describing the difference of two sample vector averages constructed from two independent samples taken from a multidimensional normal population**

Expectation (theoretical):

$$E\,F(m_1, m_2) = \frac{m_2}{m_2 - 2}, \quad \exists\ m > 2$$

Variance: (theoretical):

$$D\,F(m_1, m_2) = \frac{2{m_2}^2(m_1 + m_2 - 2)}{m_1(m_2 - 2)^2(m_2 - 4)}, \quad \exists\ m_2 > 4$$

THANKS FOR YOUR ATTENTION!
BE HEALTHY!