Lecture 3. ***On point estimate***

I. Moment estimate properties

<u>*Definition 1:*</u> *Sequence of r.v. $\{X_n\}$ almost probably converge to r.v. X, if*

$$P\left\{\lim_{n\to\infty} X_n - X\right\} = 1.$$

<u>*Definition 2:*</u> *Sequence of r.v. $\{X_n\}$ converges in probability to r.v. X, if*

$$\lim_{n\to\infty} P\{|X_n - X| \le \varepsilon\} = 1.$$

<u>*Definition 3:*</u> *Sequence of r.v. $\{X_n\}$ converges in the distribution to r.v. X (or weakly converges), if for all points of continuity of the probability distribution function the equality holds*

$$\lim_{n\to\infty} F_{X_n}(x) = F_x(x) \quad \text{or} \quad F_{X_n}(x) \Longrightarrow F_x(x)$$

<u>*Remark 1:*</u> *Convergence almost probably implies convergence in probability!*

<u>*Remark 2:*</u> *Convergence in probability implies convergence in distribution!*

<u>**Moments properties:**</u>

The sample mean $\bar{X}$ is unbiased, consistent and asymptotically normal estimation for the theoretical mean (expectation of r.v.)

<u>**Theorem 6**</u>

1) If $E|X_1| < \infty$, then $E\bar{X} = EX_1 = a$
2) If $E|X_1| < \infty$, then $\bar{X} \overset{P}{\to} EX_1 = a,\ n \to \infty$
3) If $D|X_1| < \infty,\ DX_1 \ne 0$, then $\sqrt{n}(\bar{X} - EX_1) \Longrightarrow N(0,\ DX_1)$ ($\bar{X}$ is an asymptotically normal estimate for the true expectation EX_1; theoretical)

<u>*Proof:*</u>

According to properties of expectation:

1) $E\bar{X} = \frac{1}{n}(EX_1 + .. + EX_n) = \frac{1}{n} nEX_1 = EX_1 = a - unbiased$
2) and LLN: $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n) \overset{P}{\to} EX_1 = a$
3) $\sqrt{n}(\bar{X} - EX_1) = \frac{\sum_{i=1}^{n} X_i - nEX_1}{\sqrt{n}} = \frac{\mathrm{X}_1 - \mathrm{EX}_1 + \cdots + \mathrm{X}_n - \mathrm{EX}_n}{\sqrt{n}} = \ldots$

[according to (CLT) :

$\frac{X_1 - a + \cdots + X_n - a}{\sqrt{n}} \Rightarrow \mathcal{N}(0, \mathcal{D}X_1)$, where $\mathcal{D}X_1 = \sigma^2$ – true variance],

that is if $\mathrm{X_i} \in \mathcal{N}(a, \sigma^2)$ then $\mathrm{S_n} = \frac{1}{n}\sum \mathrm{X_i} \in \mathcal{N}\left(a, \frac{\sigma^2}{\mathrm{n}}\right)$,

and as result we obtain $\ldots = \dfrac{\mathrm{X}_1 - a + \cdots + \mathrm{X}_n - a}{\frac{\sigma}{\sqrt{\mathrm{n}}}} \in \mathcal{N}(0,1)$

Theorem 7

1. If $E|X_1|^k < \infty$, then $E\bar{X}^k = EX_1^k = m_k$
2. If $E|X_1|^k < \infty$, then $\bar{X}^k \overset{P}{\to} EX_1^k = m_k, n \to \infty$
3. If $DX_1^k < \infty$, and $DX_1^k \neq 0$ *then* $\sqrt{n}\left(\bar{X}^k - EX_1^k\right) \Rightarrow \mathcal{N}\left(0, DX_1^k\right)$

Theorem 8 ***Variance properties.***

Let $DX_1 < \infty$, then

1. Sample variance $s^2 = \frac{1}{n}\sum_1^n (X_i - \bar{X})^2$ and $s_0^2 = \frac{1}{n-1}\sum_1^n (X_i - \bar{X})^2$are consistent estimation for true variance: $s^2 \overset{P}{\to} DX_1 = \sigma^2$ and $s_0^2 \overset{P}{\to} DX_1 = \sigma^2$

2. Value s^2 – biased estimation of variance and s_0^2- unbiased one:
$$Es^2 = \frac{n-1}{n} DX_1 = \frac{n-1}{n}\sigma^2 \neq \sigma^2, \qquad Es_0^2 = DX_1 = \sigma^2$$

3. If $0 < D(X_1 - EX_1)^2 < \infty$, then s^2 and s_0^2 – asymptotically normal estimation of the true variance: $\sqrt{n}(s^2 - DX_1) \Rightarrow \mathcal{N}(0, D(X_1 - EX_1)^2)$

Proof:

1. $s^2 = \overline{X^2} - (\bar{X})^2 \overset{P}{\to} EX_1^2 - (EX_1)^2 = \sigma^2.\ (theorem\ 7.1)$
$$\frac{n}{n-1} \mapsto 1,\ so \quad s_0^2 = \frac{n}{n-1}s^2 \overset{P}{\to} \sigma^2$$

2. $ES^2 = E\left(\overline{X^2} - \bar{X}^2\right) =^{prop\ .E}= E\overline{X^2} - E(\bar{X}^2) = theor.7 = EX_1^2 - E(\bar{X}^2)$= [т. к. $D(\bar{X}) = E\bar{X}^2 - (E\bar{X})^2$] $= EX_1^2 - ((E\bar{X})^2 + D(\bar{X})) = EX_1^2 - (EX_1)^2 - D\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \sigma^2 - \frac{1}{n^2}nDX_1 = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2,$
$$ES_0^2 = \frac{n}{n-1}ES^2 = \sigma^2$$

3. $Introduce\ new\ variables\ Y_i = X_i - a,\ such\ that:$
$$DY_1 = DX_1 = \sigma^2, \qquad EY_1 = 0$$

$Sample\ variance\ s^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - a - (\bar{X} - a)\right)^2 = \overline{Y^2} - (\bar{Y})^2,$

$$so\ \sqrt{n}(s^2 - \sigma^2) = \sqrt{n}\left(\overline{Y^2} - (\bar{Y})^2 - \sigma^2\right) = \sqrt{n}\left(\overline{Y^2} - EY_1^2\right) - \sqrt{n}(\bar{Y})^2$$
$$= (theorem\ 7\ )$$
$$= \left[\frac{\sum_{i=1}^n Y_i^2 - nEY_1^2}{\sqrt{n}} - \bar{Y}\sqrt{n}\bar{Y}\right] \Rightarrow^{\{tends\ to\}} \mathcal{N}(0, DY_1^2)$$
$$= \mathcal{N}(0, D(X_1 - a)^2),$$
$$because{:}\ if\ \bar{Y} \overset{P}{\to} EY_1 = 0,\ \text{then}\ \bar{Y}\sqrt{n}\bar{Y} \to 0$$

II. Parametric families. Point Estimate

Parametric families of distribution

Definition 1.
X=$\{X_i\} - sample \in$
$\mathcal{F}_\theta\big(\, known\ family, unknown\ parameter\ \theta\ (\ scalar\ of\ vector)\big)$,
$\theta \in \Theta\ for\ example$:

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

$$\Theta - is\ the\ set\ of\ possible\ values.$$

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

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

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

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

Interpretation:

- Unbiasedness - no error on average (after using)
- Asymptotically unbiasedness – the difference between its mean and true parameter decrease with increasing of **sample size**
- Consistence – it means that the sequence of estimates tends to unknown parameter with increasing of the number of observations

Before we proved

Theorem 9.

- estimation for the **true mean** is the sample average ($a = EX_1 \leftarrow a^* = \bar{X}$), and a^* - is consistent and unbiased;
- for **true variance** two estimations exist:
  - $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$
  - $S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Both variances are consistent, but S^2 – biased and asymptotically unbiased, and S_0^2 – unbiased.

There are some standard methods for obtaining point estimates:

III. Point estimate construction.

1. Moment method (MM)

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.

$Let\ X_1, .. X_n$ – the sample $\in \mathcal{F}_\theta, \theta \in \Theta \subseteq R$. Function $g(y)$ such that $Eg(X_1) = h(\theta)$ $(g: R \to R)$ $and\ h^{-1}$ – exists, $\theta = h^{-1}\big(Eg(X_1)\big)$, $\theta^* = h^{-1}\big(\overline{g(X)}\big) = h^{-1}(\frac{1}{n}\sum_{i=1}^{n} g(X_i))$

$$Let\ g(y) = y^k\ (k = 1, ..), (\text{– general choice})\ EX_1^k = h(\theta),$$

$$\theta = h^{-1}\big(EX_1^k\big), \quad \theta^* = h^{-1}\big(\overline{X_1^k}\big) = h^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^k\right)$$

<u>Property of MM estimation:</u>

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

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

<u>Example</u> : $X_1, .. X_n$ – sample $\in$ uniform distribution $U(0, \theta), \theta > 0$.
Determine θ_1^* and θ_k^*(using the first and k – th moments):
$a)\ \theta_1^*: g(y) = y$; for uniform distributed random variable f.e. X_1

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

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

2. Maximum likelihood method (MLM)

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

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

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

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

Both functions are random for a fixed θ.

- For **discrete case**, when $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ are values (outcomes) of random variables $\boldsymbol{X} = (X_1, X_2, \dots X_n)$, X_i – independent then probability to obtain $\boldsymbol{X}$ depends on θ and equal to
$$f(\boldsymbol{X}, \theta) = \prod_{i=1}^{n} f(X_i, \theta) = P_\theta(X_1 = x_1) \cdot \dots \cdot P_\theta(X_n = x_n) =$$
$$= P_\theta(X_1 = x_1,\ X_2 = x_2, \dots,\ X_n = x_n)$$

- For **absolutely continuous case,** this function is proportional to the probability of getting "almost" to the point $(x_1, x_2, \dots, x_n)$, namely is cube with the sides $dx_1, dx_2, \dots, dx_n$ around the point.

Definition. Likelihood estimation $\hat{\theta}$ of θ is a value of θ such that the maximum of the functions $f(X_1, \dots X_n, \theta)$ or $L(X_1, \dots, X_n, \theta)$ is achieved.

Remark: maximum points for both functions are the same because of monotonicity of the function ln x (function values in extreme points are different).

Example 1.

$$Let\ X_1, \dots, X_n \in P_\lambda, \qquad \lambda > 0.\ \ Find\ \hat{\lambda}.$$

Based on density function for Poisson family distribution P_λ:

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

We will determine likelihood function

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

$$\text{likelihood function f} - \text{differentiated function, but easier to use L}$$

$$\mathrm{L}(\mathrm{X_1, X_2, .. X_n}, \lambda) = \ln f(X_1, .. X_n, \lambda) = \ln\left(\frac{\lambda^{n\bar{X}}}{\prod X_i!} e^{-n\lambda}\right) = n\bar{X} ln\lambda - ln \prod X_i! - n\,\lambda;$$

partial derivative:

$$\frac{\partial}{\partial\lambda} L(\mathrm{X_1, X_2, .. X_n}, \lambda) = \frac{n\bar{X}}{\lambda} - n = 0$$

$$\hat{\lambda} = \bar{X}, \qquad \text{where } \hat{\lambda} - \ maximal\ value, why?$$

Example 2.

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

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

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

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

$$= \ln f(X_1, X_2, \dots, X_n, a, \sigma^2) = -\ln(2\pi)^{\frac{n}{2}} - \frac{n}{2}\ln(\sigma^2) - \frac{\sum_{i=1}^{n}(X_i - a)^2}{2\sigma^2}$$

Find extreme points:

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

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

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

IV. Estimate Comparison.

1. Mean square approach (MSA)

There are examples where the estimates according to MM and LM are different.

Let's introduce "distance" between true parameter θ and its estimation $\hat{\theta}$ as expectation of deviation $E(\hat{\theta} - \theta)^2$ for each permissible θ.

$Let\ X_1, .. X_n -$ sample $\in \mathcal{F}_\theta, \theta \in \Theta$.

Definition. θ_1^* better than θ_2^* in sense of MSA, if for any $\theta \in \Theta$,

$$E(\theta_1^* - \theta)^2 \le E(\theta_2^* - \theta)^2$$

2. Asymptotically normal estimate (ANE)

MSA creates problems for calculation moment of $\theta^* - \theta$, because of density functions non-linearity leads to methodological difficulties, often.

Statement. $Let\ X_1, X_2, \dots X_n -$ sample $\in \mathcal{F}_\theta, \theta \in \Theta$.

Estimation θ^* is called ANE of θ with variance $\sigma^2(\theta)$,

if for $\forall\ \theta \in \Theta$ there is a weak convergence:

$$\sqrt{n}(\theta^* - \theta) \Rightarrow \mathcal{N}(0, \sigma^2(\theta)) \text{ or the same } \sqrt{n}\frac{\theta^* - \theta}{\sigma(\theta)} \Rightarrow \mathcal{N}(0,1), n \to \infty$$

Remark. This statement important for estimates sequence,
to determine confidence interval, or statistical tests.

Definition

It is said that there is weak convergence $\xi_n \to F$, if for $\forall$ $(\cdot)$ $x -$ point of function continuity take a place convergence $P(\xi_n < x) \to F(x)$ as $n \to \infty$