# Lection 1. The Introduction to Nonparametric Statistics

## Motivation

In first course of statistics we consider parametric statistical models. We consider a sample — a vector of $n$ i.i.d. random variables with some unknown cumulitive distribution function (c.d.f.). In parametric model we have an initial assumption that the distribution belongs to some known parametric family. Sometimes this assumption is natural. Let's consider four typical cases:

- Finite Case.
  For example, if $X$ takes values in the set $\{0, 1\}$, then $X$ has Bernoulli distribution with some parameter $\theta = \mathbf{P}(X = 1)$. If the set $A$ of possible values of $X$ is finite, the distribution of $X$ can be parameterized by $|A| - 1$ parameters.

- On Physical Grounds.
  Consider, for example, a small particle moving in a fluid and let $(X, Y, Z)$ be its coordinates. From physical grounds we can say that the motion of our particle is described by the Wiener process. Therefore, the increments of $X$, $Y$, $Z$ are i.i.d. r.v. with gaussian distribution $\mathcal{N}(\theta_1, \theta_2^2)$, where $\theta_1 = 0$ if there's no drift.

- On Probabilistic Grounds.
  For example, if we measure some parameters with noise, we usually can approximate our observations by gaussian random variable (due to the Central Limit Theorem). Similarly, if $X_i$ is $\sum_{j=1}^{m} Y_{i,j}$, $Y_{i,j}$ — Bernoulli r.v. with a small probability $p$, we can approximate $X_i$ by Poisson distribution.

- On Previous Trials.
  Imagine that you start to sell a new drug and you want to predict your income. If it's not your first experiment, you have some information about the distribution of the sales. Sometimes this knowledge can be represented as a functional form of distribution.

In situations described above the parametrization is natural. But in general case we have no reason to restrict a set of possible distributions to some parametric family. So, we need to use a general nonparametric model:

$$X_1, ..., X_n \sim F, \ \ F \in \mathcal{F},$$

where $\mathcal{F}$ is some nonparametric family of distribution, for example, all distributions or all distribution with a finite mean or all continuos distributions.

## Three Problems of Nonparametric Statistics

In parametric model we consider three kinds of problems: point estimation of $\theta$, confidence estimation of $\theta$, hypothesis testing. In nonparametric model instead of $\theta$ we use some parameter of distribution, for example, the mean. In other words, we deal with $f(F)$, where $f : \mathcal{F} \to \mathbb{R}^k$ is some functional on $\mathcal{F}$. So, the three kinds of problems of nonparametric statistics are:

1. to find an estimator $\hat{\theta}(X_1, ..., X_n)$ for $f(F)$, where $f : \mathcal{F} \to \mathbb{R}^k$ is a given functional.
2. to find a set $S = S(X_1, ..., X_n) \subset \mathbb{R}^k$ such that for every $F \in \mathcal{F}$ $\mathbf{P}_F(f(F) \in S(X_1, ..., X_n)) = 1 - \alpha$ for given $\alpha \in (0, 1)$, $f : \mathcal{F} \to \mathbb{R}^k$.
3. to test a hypothesis about $F$ — to find a decision rule $\delta : \mathbb{R}^n \to \{0, 1\}$ for hypothesises $H_0 : F \in \mathcal{F}_1$, $H_1 : F \in \mathcal{F}_2$, where $\mathcal{F}_1 + \mathcal{F}_2 = \mathcal{F}$.

## Point Estimation. Definitions

Let $X_1, ..., X_n \sim F$, $F \in \mathcal{F}$, $f : \mathcal{F} \to \mathbb{R}^k$.

**Definition 1.** An *estimator* of $f(F)$ is a measurable function $\hat{\theta} : \mathbb{R}^n \to \mathbb{R}^k$.

**Definition 2.** An estimator is called *unbiased* if $\mathbf{E}_F\hat{\theta}(X_1, ..., X_n) = f(F)$ for every $F \in \mathcal{F}$.

**Definition 3.** A sequence of estimators $\hat{\theta}_n : \mathbb{R}^n \to \mathbb{R}^k$, $n \in \mathbb{Z}$, is called *consistent* if $\mathbf{P}_F(|\hat{\theta}_n(X_1, ..., X_n) - f(F)| > \varepsilon) \to 0$, $n \to \infty$, for every $F \in \mathcal{F}$, $\varepsilon > 0$.

**Definition 4.** A sequence of estimators $\hat{\theta}_n$ is called *asymptotically normal* if

$$\sqrt{n}\frac{\hat{\theta}_n(X_1, ..., X_n) - f(F)}{\sigma(F)} \xrightarrow{d} Z \sim \mathcal{N}(0,1), \ n \to \infty,$$

for some $\sigma : \mathcal{F} \to \mathbb{R}^+$. The function $\sigma$ is called asymptotic variance of $\widehat{\theta}$.

**Example 1.** Let $f(F) = \int_{\mathbb{R}} g(x)dF(x)$ be a mean of $g(X_1)$, $\mathcal{F}_k$ be the set of distributions with finite $k$-th moment of $g(X_1)$. Then the estimator $\overline{g(X)} = (g(X_1) + ... + g(X_n))/n$ for $\mathbf{E}_F g(X)$ is unbiased in $\mathcal{F}_1$ due to linearity of expectation, consistent in $\mathcal{F}_1$ due to law of large numbers and asymptotically normal in $\mathcal{F}_2$ due to central limit theorem.

### Empirical Cumulative Distribution Function

Let's begin with an estimation of c.d.f. $F(x)$. Since $F(x) = \mathbf{E}I_{X\le x}$ the estimator

$$\widehat{F}_n(x; X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i \le x}$$

is a natural estimator for $F(x)$. This estimator is called the empirical cumulative distribution function (e.c.d.f.).

1. $\widehat{F}_n(x; \vec{X})$ is an unbiased estimator since

$$\mathbf{E}\widehat{F}_n(x; \vec{X}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}I_{X_i \le x} = \mathbf{P}(X_1 \le x) = F(x).$$

2. $\widehat{F}_n(x; \vec{X})$ is a consistent estimator since

$$\frac{\sum_{i=1}^n I_{X_i \le x}}{n} \xrightarrow{P} \mathbf{E}I_{X_i \le x} = \mathbf{P}(X_i \le x) = F(x), \ n \to \infty.$$

3. $\widehat{F}_n(x; \vec{X})$ is an asymptotically normal estimator since

$$\sqrt{n}\left(\frac{\sum_{i=1}^n I_{X_i \le x}}{n} - F(x)\right) = \frac{\sum_{i=1}^n I_{X_i \le x} - nF(x)}{\sqrt{n}} \xrightarrow{d} Z \sim \mathcal{N}(0, F(x)(1 - F(x)), \ n \to \infty.$$

4. By Glivenko-Cantelly theorem

$$\sup_x |\widehat{F}_n(x; \vec{X}) - F(x)| \xrightarrow{a.s.} 0, \ n \to \infty.$$

5. By 4) for a.s. $X_1, ..., X_n, ...$ we have

$$\widehat{F}(x; \vec{X}) \xrightarrow{d} F(x), \ n \to \infty.$$

By properties 1)-3) $\widehat{F}_n(x)$ is a nice estimator for $F(x)$ as $x$ is fixed. However, it's not enough for our purposes. Properties 4)-5) shows that $\widehat{F}_n(x)$ is a nice estimator for $F(x)$ in $\mathcal{F}$.

Moreover, we will prove that

$$\sqrt{n}(\widehat{F}_n(x; \vec{X}) - F(x)) \xrightarrow{d} Y(x),$$

where $Y(x)$ is some gaussian stochastic process. This result generalizes property 3).

**Consistent Estimators and Weakly Continuous Functionals**

We see that $\widehat{F}_n(x;\vec{X})$ is an excellent estimator for $F(x)$. Therefore, it's reasonable to estimate $f(F)$ by $f(\hat{F}_n)$, where $f$ is some functional $f:\mathcal{F}\to\mathbb{R}^k$.

**Definition 5.** A functional $f$ is called *weakly continuous* if $f(F_n)\to f(F)$ for every sequence $F_n$ of c.d.f. such that $F_n\xrightarrow{d}F$, $n\to\infty$.

Therefore, if $f$ is a weakly continuous functional, then $f(\hat{F}_n)$ is a consistent estimator of $f(F)$.

**Example 2.** The functional $f_1(F)=\int_{\mathbb{R}} x dF(x)=\mathbf{E}_F X$ is not weakly continuous. Really, consider a sequence of c.d.f.

$$F_n(x)=\left(1-\frac{1}{n}\right)F(x)+\frac{1}{n}I_{x\geq n}.$$

Then $F_n\xrightarrow{d}F$ as $n\to\infty$ but

$$f_1(F_n)=\left(1-\frac{1}{n}\right)\mathbf{E}_F X+\frac{1}{n}\cdot n=\left(1-\frac{1}{n}\right)\mathbf{E}_F X+1\to f_1(F)+1.$$

It's natural since $X_n\xrightarrow{d}X$ doesn't mean that $\mathbf{E}X_n\to\mathbf{E}X$. However, for every bounded function $g:\mathbb{R}\to\mathbb{R}$ the functional $f(F)=\int_{\mathbb{R}} g(x)dF(x)$ is weakly continuous. Therefore,

$$f(\widehat{F}_n)=\int_{\mathbb{R}} g(x)d\widehat{F}_n(x)=\sum_{i=1}^{n} g(X_i)\frac{1}{n}=\overline{g(X)}$$

is a consistent estimator for $f(F)$.

**Example 3.** The median $f_2(F)=x_{1/2}=F^{-1}(1/2)=\inf\{x: F(x)\geq 1/2\}$ isn't a weakly continuous functional. Really, consider

$$F_n(x)=\begin{cases}0, & x<0,\\ \frac{1}{2}-\frac{1}{n}+\frac{2x}{n}, & x\in[0,1)\\ 1, & x\geq 1.\end{cases},\quad F(x)=\begin{cases}0, & x<0,\\ \frac{1}{2}, & x\in[0,1),\\ 1, & x\geq 1.\end{cases}$$

Then $F_n\xrightarrow{d}F$, $f_2(F_n)=1/2$, $f_2(F)=0$.