Homework I - Smirnov

Theoretical part

Problem I

> Let $X_1, \ldots, X_n$ be a sample from the Bernoulli distribution with parameter p. What is the distribution of $Y_1, \ldots Y_n$, where $Y_i = F(X_i)$, and $F(y)$ is the cumulative distribution function of the Bernoulli distribution?

Solution

If X_i are independent random variables with Bernoulli distribution with parameter p, $P(X_i = 1) = p,\ P(X_i = 0) = 1 - p$. $F(1) = 1, F(0) = 1 - p$, so with probability p $F(X_i) = 1$ and with probability $1 - p$ $F(X_i) = 1 - p$.

Problem II

> Let $F_n(y)$ be the empirical distribution corresponding to a sample $X_1, \ldots X_n$ of size n. Are the following functions empirical distribution for some sample? If yes, what sample it corresponds to?
> a) $F_n(y^3)$
> b) $(F_n(y))^3$

Solution

The answer is: a) - yes, and b) - no. a) corresponds to the sample $\{X_i^3\}$. b) is not an empirical distribution function, because it is not normalized to 1: $\sum_{i=1}^n \frac{1}{n^3} \neq 1$ for $n > 1$.

Problem III

> Let $X_1, \ldots, X_n$ be a sample from the Bernoulli distribution with parameter p. Check if X_1, X_1X_2, $X_1(1 - X_2)$ are
> a) Unbiased
> b) Consistent
> estimates for the parameters $p, p^2, p(1-p)$ respectively.

Solution

a) Unbiasedness

X_1 is unbiased because $\mathbb{E}[X_1] = p$
X_1X_2 is an unbiased estimate for p^2 because for independent random variables X_1, X_2
$\mathbb{E}[X_1X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2] = p \cdot p = p^2$
For the same reason (and linearity of $\mathbb{E}$) $X_1(1 - X_2)$ is an unbiased estimate for $p(1-p)$.

b) Consistency

All three estimates are not consistent, because they depend only on the first two terms of the sequence $X_1, \dots, X_n$, and for consistency we, by definition, need

$$\forall \varepsilon > 0 : \lim_{n\to\infty} P(|\hat{\theta}(X_1, \dots, X_n) - \theta| > \varepsilon) = 0$$

which in our case means

$$\forall \varepsilon > 0 : \lim_{n\to\infty} P(|X_1 - p| > \varepsilon) = 0 = P(|X_1 - p| > \varepsilon)$$

which is obviously not true seince $|X_1 - p|$ can't be lower than $\min\{p, 1-p\}$, a similar argument is valid for X_1X_2 and $X_1(1 - X_2)$.

Problem IV

> Empirical distribution function $\hat{F}_n(y)$ corresponds to a sample from the uniform distribution on $[0, a]$, where $a > 0$. What is the parameter $\theta(a)$ for which $\hat{F}_n(1)$ is unbiased? Is it also consistent?

Solution

The parameter in question is $\frac{1}{a}$. Indeed, the CDF of a uniform distribution on $[0, a]$ is linear on $[0, a]$, with slope being equal $\frac{1}{a}$ (having the form $F(x) = \frac{1}{a}x$, to be exact, the intercept is 0), so $F(1) = \frac{1}{a}$. The parameter is indeed unbiased and consistent, because $\hat{F}_n$ is exactly the mean of the binomial distribution with parameter $\frac{1}{a}$: we add $\frac{1}{n}$, if k-th sample out of n from the uniform distribution falls below 1, the probability of which is $\frac{1}{a}$, so the empitical distribution is

$$\hat{F}_n = \frac{1}{n}(\xi_1 + \xi_2 + \dots + \xi_n)$$

where $\xi_i = 1$ if $X_i < 1$ and 0 otherwise, and

$$\mathbb{E}[\hat{F}_n] = \frac{1}{n}(\mathbb{E}\xi_1 + \mathbb{E}\xi_2 + \dots + \mathbb{E}\xi_n) = \frac{np}{n} = p = \frac{1}{a}$$

which proves unbiasedness, and by the law of big numbers a sum of equally distributed independent random variables, divided by the number of variables, converges by probability to the mean of each random variable, which by definition means consistency, which finishes the proof.