# 4 Занятие четвертое. О непараметрическом оценивании, и о том, как бутстрэпить без знания распределения

## 4.1 Эмпирическая функция распределения

### 4.1.1 Непараметрическая модель

Мы привыкли рассматривать именно параметрическую модель $X_i \sim F_\theta$. С другой стороны, начиная исследование, мы часто не имеем никакой предварительной информации о распределении. В таком случае более подходящей является непараметрическая модель $X_i \sim F$, а оценки мы будем строить для $F(x)$, $\mathbf{E}X_1$ или другого показателя, связанного с моделью. При этом мы сохраняем определения состоятельности, несмещенности и асимптотической нормальности, просто не апеллируем в них к параметризации модели.

**Пример 1.** Так $\overline{X} = (X_1+...+X_n)/n$ будет несмещенной состоятельной оценкой $\mathbf{E}X_1$ и в непараметрической модели, а при $\mathbf{E}X_1^2 < \infty$ еще и асимптотически нормальной. Аналогично $S_0^2 = \sum_{i=1}^n (X_i - \overline{X})^2/(n-1)$ будет несмещенной состоятельной оценкой дисперсии, а при $\mathbf{E}X_1^4 < \infty$ асимптотически нормальной.

## 4.2 Как оценивать функции распределения и как с помощью этого строить непараметрические оценки?

Итак, давайте начнем с того, что оценим функцию распределения и плотность нашей выборки $X_i$.

### 4.2.1 ЭФР

Достаточно хорошей (несмещенной состоятельной и асимптотически нормальной) оценкой функции распределения в конкретной точке $x$ является эмпирическая функция распределения

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

**Вопрос 1.** Почему она при каждом $x$ обладает свойствами несмещенности, состоятельности и асимптотической нормальности?

**Вопрос 2.** Какой доверительный интервал для $F(x)$ можно построить при фиксированном $x$ на основе асимптотической нормальности ЭФР в этой точке?

**Вопрос 3.** Показать, что $\widehat{F}_n$ — ОМП для ф.р. в непараметрической модели (считая, что рассматриваемые распределения являются всеми дискретными распределениями с какими-либо значениями и вероятностями).
В R ЭФР может быть легко подсчитана по выборке с помощью команды ecdf(x). Обратите внимания, что при построении plot от этой функции, нужно будет написать plot(Vectorize(ecdf)).
В Python ЭФР автоматизирована в ECDF в statsmodels.distributions.empirical_distribution. У результата подсчета есть параметры x и y, построить график можно с помощью plot из matplotlib.pyplot.

Теорема Гливенко-Кантелли утверждает, что ЭФР состоятельна как оценка всей функции распределения, даже более того, $\widehat{F}_n \to F$ п.н. по равномерной норме. Теорема Колмогорова дает для непрерывных распределений утверждение

$$\mathbf{P}(\sqrt{n}||\widehat{F}_n - F|| > \varepsilon) \to 1 - K(\varepsilon),$$

где $K$ — функция распределения Колмогорова, а норма равномерная.

Тем самым, мы можем построить доверительную полосу для непрерывной функции распределения на основе ее ЭФР и теоремы Колмогорова. Кроме того, в случае непрерывной $F$ при каждом $n$ распределении статистики $||\widehat{F}_n - F||$ не зависит от $n$, а, значит, может быть заключена в доверительную полосу в случае небольших $n$.

Существует явная оценка сверху, не требующая непрерывности распределения: неравенство Дворецкого-Кифера-Вольфовица

$$\mathbf{P}(||\widehat{F}_n - F|| > \varepsilon) \leq 2e^{-2n\varepsilon^2},\ \varepsilon > 0.$$

При этом ф.р. Колмогорова близка к правой части неравенства при больших n и не очень больших $\varepsilon$, поэтому методы дают близкие результаты.

**Задача 1.** Смоделировать 100 выборок. Для каждой построить 95% доверительное множество для ф.р. распределения а) нормальной б) Коши с помощью неравенств Д.-К.-В и теоремы Колмогорова. Построить доверительный интервал 95% для функции распределения в каждой из точек $x$. Будет ли объединение этих интервалов давать 95% доверительное множество? В какой доле из выборок настоящая ф.р. не попадала в каждое из доверительных множеств?
Решение задачи на Python реализовано в ConfforCDF.py.

## 4.3 Оценки плотности

### 4.3.1 Гистограмма

С той же целью полезно бывает оценить плотность $p(x)$, если мы знаем, что распределение абсолютно-непрерывно. Мы будем обозначать плотность $p$, поскольку $f$ будем использовать с другой целью.

Простейшей оценкой плотности является гистограмма, с которой вы уже знакомы. Гистограмма предлагает рассмотреть некоторое разбиение области значений нашей величины $(a_i, a_{i+1}]$ , $i = 1, 2, ....$ Тогда гистограмму на $(a_i, a_{i+1}]$ положим равной $N_i/N = \#\{x_j \in (a_i, a_{i+1}]\}/N$, то есть частоте попадания в полуинтервал наших наблюдений. В целом, гистограмма неплохая оценка, стремящаяся с ростом числа наблюдений и уменьшением ширины интервалов разбиения к истинному значению плотности, но она кусочно-постоянна.

### 4.3.2 Ядерная оценка плотности

Хорошим методом получения неперывной оценки для плотности является так называемая *ядерная оценка*. Назовем *ядерной* неотрицательную функцию $K(x)$, т.ч. $\int_{\mathbb{R}} K(x)dx = 1$, $\int_{\mathbb{R}} xK(x)dx = 0$, $\int_{\mathbb{R}} x^2K(x)dx > 0$ (то есть $K(x)$ —- некоторая плотность с нулевым матожиданием и конечной дисперсией). Назовем *ядерной оценкой* плотности

$$\widehat{p}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_n} K\left(\frac{x - X_i}{h_n}\right),$$

где $K$ — ядерная функция, а $h_n$ — некий параметр, называющийся шириной окна сглаживания.

Если $p$ — непрерывна, $h_n \to 0$, $nh_n \to \infty$, то $\widehat{p}_n(x)$ будет состоятельной для $p(x)$. При этом $\mathbf{E}(\widehat{p}_n(x) - p(x))^2$ есть $O(h_n^4) + O\left(\frac{1}{nh_n}\right) + O\left(\frac{1}{n}\right)$. Таким образом, наилучший порядок $h$ есть $O(n^{-1/5})$. Более точная формула для $h$, дающего минимум квадратичного отклонения

$$h^* = \left(\frac{\int K(x)^2 dx}{n\left(\int x^2K(x)dx\right)^2 \int (p''(x))^2 dx}\right)^{1/5}.$$

Эта формула не вполне удовлетворительна для применения, поскольку использует неизвестное нам $p''(x)$, но дает представить порядок $h^*$.
Функция $K$, как мы видим, не влияет на порядок сходимости. Часто берут в роли $K(x)$ равномерную на $[-1, 1]$ плотность (прямоугольное ядро), стандартную нормальную плотность (гауссово ядро) или $3(1 - x^2)I_{|x|\leq 1}/4$ (ядро Епанечникова).

В R ядерная оценка плотности встроена, например, в стандартный график qplot с геометрией geom=density, по умолчанию устанавливается гауссовкое ядро, но можно поменять его с помощью параметра kernel. В python она, например, входит в seaborne kdeplot и distplot.

Непосредственно ядерную оценку для плотности в R можно рассчитать с помощью функции density. Параметр kernel отвечает за вид ядра, параметр bw отвечает за ширину окна.

**Задача 2.** Оцените а) стандартную нормальную плотность на выборке размера 100. б) смесь из трех нормальных плотностей $\mathcal{N}(0,1)$, $\mathcal{N}(6,1)$, $\mathcal{N}(-3,1)$ с равными вероятностями. Что происходит при уменьшении ширины окна $h$?

Существуют и другие подходы к оценке плотности, например, оценить ее коэффициенты разложения в ряд Фурье. Мы не будем их касаться в рамках нашего курса.

#### 4.3.3 Кросс-валидация

Поговорим о том, как оценить качество полученной оценки для плотности. Хорошим показателем для нас являлась бы величина

$$\int_{\mathbb{R}} (\widehat{p}_n(x) - p(x))^2 dx = \int_{\mathbb{R}} \widehat{p}_n(x)^2 dx - 2\int_{\mathbb{R}} \widehat{p}_n(x)p(x)dx + \int_{\mathbb{R}} p(x)^2 dx,$$

которую мы не можем подсчитать, поскольку не знаем $p(x)$.

Мы хотим минимизировать эту величину, выбирая $\widehat{p}_n$, а значит минимизировать сумму первых двух слагаемых.

**Определение 1.** *Оценкой кросс-валидацией* для оценки плотности $\widehat{p}_n$ называют

$$\widehat{J}(h) = \int_{\mathbb{R}} \widehat{p}_n(x)^2 dx - \frac{2}{n}\sum_{i=1}^{n} \widehat{p}_{n,i}(X_i),$$

где $\widehat{p}_{n,i}$ — оценка, построенная по выборке с исключенным $X_i$.

**Вопрос 4.** Показать, что $\mathbf{E}\widehat{J}(h) = \mathbf{E}\left(\int_{\mathbb{R}} \widehat{p}_n(x)^2 dx - 2\int_{\mathbb{R}} \widehat{p}_n(x)p(x)dx\right)$.
Тем самым, маленькие значения $\widehat{J}(h)$ указывают на то, что $\widehat{p}_n(x)$ близко к $p(x)$.

Для ядерной оценки оценку кросс-валидации можно найти по формуле

$$\widehat{J}(h) = \frac{1}{hn^2}\sum_i\sum_j K^*\left(\frac{X_i - X_j}{h}\right) + \frac{2}{nh}K(0) + O\left(\frac{1}{n^2}\right).$$

$K^* = K * K - 2K$, $*$ — свертка. С помощью оценки кросс-валидации удобно измерять качество приближения. В частности, с ее помощью можно откалибровать ширину $h$.

Осуществить выбор $h$ с помощью кросс-валидации в R можно, указав bw=bw.ucv(x).

### 4.4 Оценивание параметров распределения

#### 4.4.1 Естественные оценки параметра

Естественно оценивать параметр $f(F)$ функции распределения, где $f$ – некий функционал, величиной $f(\widehat{F}_n)$. Этот параметр будем называть естественной оценкой $f(F)$.

В качестве альтернативы можно представлять наш параметр как $f(p)$, где $p$ — плотность распределения. Скажем, с помощью оценки для плотности можно строить оценки $\int g(x)\widehat{p}_n(x)dx$ для функционалов вида $\int g(x)p(x)dx$. Например, для математического ожидания мы получим оценку

$$\int_{\mathbb{R}} x\widehat{p}_n(x)dx = \overline{X}.$$

**Вопрос 5.** Почему верно последнее тождество?

Из сходимости $\widehat{F}_n \to F$ при п.в. $\omega$ вытекает сходимость при тех же $\omega$ $f(\widehat{F}_n) \to f(F)$ для любого непрерывного (по равномерной норме) функционала $f$. Вполне естественно, таким образом, исследуя функционалы $f(F)$, оценивать их $f(\widehat{F}_n)$.

**Пример 2.** Для оценки математического ожидания $\mathbf{E}X_1 = \int_{\mathbb{R}} xdF(x)$ и дисперсии $\mathbf{D}X_1 = \int_{\mathbb{R}} x^2 dF(x) - (\int_{\mathbb{R}} xdF(x))^2$ возьмем

$$\widehat{\theta}_1 = \int_{\mathbb{R}} xd\widehat{F}_n(x),\ \widehat{\theta}_2 = \int_{\mathbb{R}} x^2 d\widehat{F}_n(x) - \widehat{\theta}_1^2.$$

**Вопрос 6.** Будут ли эти функционалы непрерывными по равномерной норме?

**Вопрос 7.** Почему $\widehat{\theta}_1 = \overline{X}$, $\widehat{\theta}_2 = S^2$?

**Задача 3.** Построить оценку для асимметрии: $\mathbf{E}(X - \mathbf{E}X)^3(\mathbf{D}X)^{-3/2}$.

#### 4.4.2 Асимптотическая нормальность оценок

Можно показать асимптотическую нормальность естественных оценок. Подробно это разобрано в факультативной части. Здесь же мы приведем несколько конкретных приложений. Будем использовать так называемую функцию влияния $I_f(x)$, с помощью которой можно получить асимптотическую дисперсию

$$\sigma^2(F) = \int_{\mathbb{R}} I_f^2(x) dF(x).$$

1. Для функционала $f(F) = \int_{\mathbb{R}} a(x)dF(x)$

   $$I_f(x) = a(x) - \int_{\mathbb{R}} a(u)dF(u),$$

   если $F$ имеет конечный момент $\int_{\mathbb{R}} a^2(x)dF(x)$.

2. Для функционала квантили $f(F) = F^{-1}(q)$

   $$I_f(x) = \begin{cases} \frac{q-1}{p(F^{-1}(q))}, & x \le F^{-1}(q), \\ \frac{q}{p(F^{-1}(q))}, & x > F^{-1}(q), \end{cases}$$

   где $F$ предполагается абсолютно-непрерывной и монотонно возрастающей в точке $F^{-1}(q)$.

3. Для функционала $g(f_1(F), \ldots, f_m(F))$

   $$I_g(x) = \sum_{i=1}^{m} \frac{\partial g}{\partial f_i}(f_1(F), \ldots, f_m(F)) I_{f_i}(x).$$

Те же формулы верны и в случае многомерного распределения.

Отсюда можно построить доверительный интервал для $f(F)$, оценив $\sigma^2(F)$ с помощью ЭФР $\widehat{F}_n$ или оценки плотности.

**Вопрос 8.** Какой вид будет иметь оценка для ковариации $cov(X, Y)$ и как выглядит для нее доверительный интервал?

### 4.5 Исправление смещения с помощью bootstrap

Если мы захотим получать несмещенные оценки, то нам, как и прежде, будет полезна процедура бутстрепинга. Однако, раньше мы брали выборку из распределения с оцениваемым параметром, то теперь мы будем брать выборку из распределения с оцениваемой функцией распределения. Если в качестве оценки для функции распределения используется эмпирическая функция распределения, то это равносильно рассмотрению выборок $X^*_{i,1}, ..., X^*_{i,n}$, $i = 1, ..., M$, взятых из нашей выборки с возвращением. Это может показаться странным — зачем брать с возвращением элементы из нашей выборки? Тем не менее метод действительно вполне осмыслен, если функционал $f$ неперерывен.

Таким образом, если мы рассматриваем статистику $f(\widehat{F}_n)$ в качестве оценки $f(F)$, мы можем брать выборки из распределения $\widehat{F}_n$ и на основе этих выборок изучать распределение $f(\widehat{F}_n)$, ожидая, что оно

близко к распределению $f(F)$. В частности, мы можем исследовать смещение $f(\widehat{F}_n)$ по сравнению с $f(F)$, приближая его смещением $f(\tilde{F}_n) - f(\widehat{F}_n)$, $\tilde{F}_n$ – ЭФР выборки из $\widehat{F}_n$. Аналогичным образом мы можем изучать и другие параметры распределения $f(\widehat{F}_n)$, например, дисперсию. Для бутстрэпа в пакете bootstrap в R есть одноименная функция, хотя реализация этого алгоритма совсем проста и не требует специальных функций.

**Задача 4.** Сгенерировать выборку $R[0,1]$ размера 50 и оценить дисперсию оценок а) $\overline{X}$, б) $MED$.

**Задача 5.** Исправить смещение статистики $S^2$ как оценки дисперсии с помощью bootstrap по выборке размера 100 из распределения $\mathcal{N}(0,1)$.

### 4.6 Доверительные интервалы с помощью bootstrap

Предположим, что $\widehat{\theta} = f(\widehat{F}_n)$ — оценка какого-то параметра распределения $f(F)$. Обсудим как построить доверительные интервалы на основе bootstrap.

1. percentile-интервал строится тем же методом, что и прежде — генерируется $m$ выборок из функции распределения $\widehat{F}_n$ (то есть $m$ выборок того же размера с возвращением), ранжируется значение нашей оценки $\widehat{\theta}^*$ на этих выборках и в качестве левой границы доверительного интервала берется $[\alpha m]$-е по возрастанию из значений $\widehat{\theta}^*$, а в качестве правой границы — $[(1-\alpha)m]$-е. Он также достаточно требователен к данным и оценке.

2. pivotal-интервал так же строится аналогично параметрической версии — генерируется $m$ выборок из функции распределения $\widehat{F}_n$, ранжируется значение нашей оценки $\widehat{\theta}^*$ на этих выборках и в качестве левой границы доверительного интервала берется $[\alpha m]$-е по возрастанию из значений $2\widehat{\theta} - \widehat{\theta}^*$, а в качестве правой границы — $[(1-\alpha)m]$-е.

3. Можно действовать более хитрым путем. Сперва с помощью bootstrap оценим величиной $\widehat{\sigma}$ стандартное отклонение $\widehat{\theta}$. Теперь начнем новый bootstrap для построения доверительного интервала. Пусть $X^*_{i,1}, ..., X^*_{i,n}$ — одна из выборок с возвращением из нашей $(X_1, ..., X_n)$. Найдем $\widehat{\theta}^*_i = f(\widehat{F}^*_{n,i})$ и оценим еще одним bootstrap'ом ее стандартное отклонение $\widehat{\sigma}^*_i$. Построим

$$Y_i = \frac{\widehat{\theta}^*_i - \widehat{\theta}}{\widehat{\sigma}^*_i}$$

   для всех выборок $X^*_{i,\cdot}$, $i \le m$, выберем среди них $[\alpha m]$-е и $[(1-\alpha)m]$-е по возрастанию, а затем в качестве доверительного интервала возьмем

$$\left(\widehat{\theta} - (\widehat{\theta} - Y_{([(1-\alpha)m])})\widehat{\sigma}, \widehat{\theta} - (\widehat{\theta} - Y_{([\alpha m])})\widehat{\sigma}\right)$$

   Рассмотрим теперь величину $(f(\widehat{F}_n) - f(F))/S(f)$. Построим для нее bootstrap-интервал ширины 0.95, откуда получим доверительный интервал для $f(F)$. Этот интервал называется studentized pivotal.

   Важно то, что в том случае необходимо оценивать $\sigma$ заново на каждом шагу. Из-за этого количество итераций бутстрэп растет квадратично, зато и скорость сходимости к $1-\alpha$ повышается соответственно.

Можно сократить вторую часть процедуры (оценку стандартного отклонения) с помощью дельта-метода.

**Задача 6.** Рассматривая $f(F) = F^{-1}(1/3)$, найти для $R[0,1]$ с помощью $I_f(x)$ доверительный интервал для $f(F)$ а) непосредственно из дельта-метода б) с помощью стьюдентовского бутстрэпа.

Оказывается, что указанные интервалы являются асимптотическими уровня $1-\alpha$, если наши оценки асимптотически нормальны.

**Вопрос 9.** Как при больших $n$ ведет себе 60% pivotal доверительный интервал для параметра "супремум возможных значений случайной величины" на основе $\max X_i$, где $X_i \sim R[0,1]$, $i \le n$.?

**Задача 7.** Испытайте метод pivotal для интервалов для а) среднего $R[0,1]$, б) дисперсии $\exp(1)$. Рассмотренные интервалы не точные, их уровень доверия близок к $1-\alpha$ с ростом $n$ (хотя и не равен ему). Стьюдентовский интервал имеет более высокую скорость сходимости к $1-\alpha$, равную $O(1/n)$, обычный pivotal интервал имеет скорость $O(1/\sqrt{n})$.

### 4.6.1 Сглаженный бутстрэп

Недостатков непараметрического бутстрэпа является то, что возможно непрерывное распределение $F$ подменяется дискретным распределением $\widehat{F}_n$. В качестве альтернативы в абсолютно-непрерывном случае можно рассмотреть оценку $\widehat{p}_n$ для плотности $p$ и генерировать бутстрэп-выборки на ее основе.

Оказывается, что моделировать бутстрэп-выборки из ядерной оценки плотности можно следующим образом:

1. Выбираем $R \sim R\{1,\dots,n\}$;
2. Выберем величину $Y$ с плотностью $K(x)$;
3. Положим $X_i^* = X_R + hY$.

**Вопрос 10.** Почему данная процедура дает величину с плотностью $\widehat{p}_n$?

При этом мы повышаем дисперсию наших наблюдений, что можно снизить с помощью замены третьего шага на

$$X_i^* = \overline{X} + (X_R - \overline{X} + hY)/(1 + h^2\sigma_K^2/S^2)^{1/2}.$$

На основе этих данных мы строим $\widehat{\theta}_i^*$ и далее действуем обычным путем для построения интервала. В случае исправленного третьего шага этот метод называют усеченным сглаженным бутстрэпом.

Сглаженная оценка при правильном выборе $h$ более удачна, но проседает при неудачно выбранном $h$.

**Задача 8.** Для данных из двумерного нормального распределения с единичной дисперсией у обеих компонент и ковариацией $1/2$ получить доверительные pivotal интервалы а) обычным б) сглаженным в) усеченным сглаженным бутстрэпом.

## 5 Ответы на вопросы

**Ответ 1.** Несмещенность вытекает из линейности матожидания

$$\mathbf{E}\frac{1}{n}\sum_{i=1}^{n} I_{X_i \le x} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E} I_{X_i \le x} = \mathbf{P}(X_1 \le x) = F_X(x).$$

Состоятельность — прямое следствие ЗБЧ, а асимпотическая нормальность — ЦПТ, примененных к $I_{X_i \le x}$.

**Ответ 2.** В силу ЦПТ

$$\sqrt{n}(\widehat{F}_n(x) - F(x)) \xrightarrow{d} Z \sim \mathcal{N}(0, F(x)(1-F(x))).$$

Отсюда

$$\frac{\sqrt{n}(\widehat{F}_n(x) - F(x))}{\sqrt{\widehat{F}_n(x)(1-\widehat{F}_n(x))}} \overset{Z}{\sim} \mathcal{N}(0,1).$$

Следовательно, получаем доверительный интервал

$$F(x) \in \left(\widehat{F}_n(x) - \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\widehat{F}_n(x)(1-\widehat{F}_n(x))}, \widehat{F}_n(x) - \frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{\widehat{F}_n(x)(1-\widehat{F}_n(x))}\right).$$

**Ответ 3.** Если значения нашей случайной величины $y_1, ..., y_m$ с вероятностями $p_1, ..., p_m$, а выборка $x_1, ..., x_n$, то а) если какой-то из $x_i$ не встречается среди $y_i$, то правдоподобие 0.
б) иначе правдоподобие принимает вид $L = p_1^{N_1}...p_m^{N_m}$, где $N_1$,...,$N_m$ — число $x_i$, равных $y_1, ..., y_m$ соответственно. Но тогда будем рассматривать параметры $p_1, ..., p_{m-1}$ (и $p_m = 1 - p_1 - ... - p_{m-1}$):

$$\ln L(p_1, ..., p_{m-1}) = \sum_{i=1}^{m-1} N_i \ln p_i + N_m \ln(1 - p_1 - ... - p_{m-1}),\ \frac{\partial}{\partial p_i} \ln L = \frac{N_i}{p_i} - \frac{N_m}{p_m},$$

То, что найденная точка максимум, можно показать напрямую или рассматривая пару $p_i, p_j$, фиксируя их сумму.

**Ответ 4.** Первые слагаемые в обоих выражениях совпадают. Рассмотрим матожидание левой части второго из них

$$\mathbf{E}\frac{2}{n}\sum_{i=1}^{n} \widehat{f}_{n,i}(X_i) = \frac{2}{n}\sum_{i=1}^{n} \mathbf{E}\widehat{f}_{n,i}(X_i).$$

В силу линейности матожидания

$$\mathbf{E}\widehat{f}_{n,i}(X_i) = \frac{1}{n-1}\sum_{j\neq i}\frac{1}{h_n}\mathbf{E}K\left(\frac{X_i - X_j}{h_n}\right) = \frac{1}{h_n}\int_{\mathbb{R}^2} K\left(\frac{x-y}{h_n}\right) f(x)f(y)dydx.$$

С другой стороны,

$$\mathbf{E}\int_{\mathbb{R}} \widehat{f}_n(x)f(x)dx = \int_{\mathbb{R}} \frac{1}{n}\frac{1}{h_n}\sum_{j=1}^{n}\mathbf{E}K\left(\frac{x - X_j}{h_n}\right)dx = \frac{1}{h_n}\int_{\mathbb{R}} K\left(\frac{x-y}{h_n}\right) f(y)f(x)dydx.$$

Подставляя полученные выражения в правую и левую части искомого тождества, получаем требуемое

**Ответ 5.**

$$\int_{\mathbb{R}} x\widehat{f}_n(x)dx = \frac{1}{n}\frac{1}{h_n}\sum_{i=1}^{n}\int_{\mathbb{R}} xK\left(\frac{x - X_i}{h_n}\right)dx = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}}(y\cdot h_n + X_i)K(y)dy = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

где в последнем тождестве мы воспользовались тем, что $\int_{\mathbb{R}} yK(y)dy = 0$, $\int_{\mathbb{R}} K(y)dy = 1$.

**Ответ 6.** Нет, из равномерной сходимости $F_n$ к $F$ не следует, что

$$\int_{\mathbb{R}} a(x)dF_n(x) \to \int_{\mathbb{R}} a(x)dF(x),$$

если функция $a$ не является ограниченной. Например, для $a(x) = x$ можно взять $F_n = (1 - 1/n)F(x) + nI_{x\geq n}$. Тогда

$$\int_{\mathbb{R}} xdF_n(x) = \left(1 - \frac{1}{n}\right)\int_{\mathbb{R}} xdF(x) + \frac{1}{n}\cdot n \to \int_{\mathbb{R}} xdF(x) + 1.$$

**Ответ 7.** Если $x_1, ..., x_n$ (значения элементов выборки) различны, то $\widehat{F}_n(x)$ — ф.р. дискретной случайной величины $\widehat{X}$ со значениями $x_1, ..., x_n$ и вероятностями $1/n$. Тогда

$$\mathbf{E}\widehat{X} = x_1\frac{1}{n} + x_2\frac{1}{n} + ... + x_n\frac{1}{n} = \overline{x},\ \mathbf{D}\widehat{X} = (x_1 - \overline{x})^2\frac{1}{n} + ... + (x_n - \overline{x})^2\frac{1}{n} = S^2.$$

Если же какие-то значения совпадают, то слагаемые, соответствующие одинаковых $x_i$, объединятся.

**Ответ 8.** Поскольку $\mathrm{cov}(X, Y)$ есть

$$\int_{\mathbb{R}^2} xydF(x, y) - \int_{\mathbb{R}^2} xdF(x, y)\int_{\mathbb{R}^2} ydF(x, y).$$

Тогда естественная оценка есть

$$\int_{\mathbb{R}^2} xyd\widehat{F}_n(x,y) - \int_{\mathbb{R}^2} xd\widehat{F}_n(x,y) \int_{\mathbb{R}^2} yd\widehat{F}_n(x,y) = \overline{XY} - \overline{X}\ \overline{Y}.$$

Ее функция влияния имеет вид

$$I(x,y) = xy - \int_{\mathbb{R}^2} uvdF(u,v) - (x - \int_{\mathbb{R}^2} udF(u,v)) \int_{\mathbb{R}^2} vdF(u,v) - (y - \int_{\mathbb{R}^2} vdF(u,v)) \int_{\mathbb{R}^2} udF(u,v) = (x - \mathbf{E}X)(y - \mathbf{E}Y) - \mathrm{cov}(X,Y).$$

При этом

$$\sigma^2(F) = \mathbf{E}(X - \mathbf{E}X)^2(Y - \mathbf{E}Y)^2 - (\mathrm{cov}(X,Y))^2.$$

Чтобы построить доверительный интервал мы можем оценить дисперсию величиной

$$\sigma^2(\widehat{F}) = \overline{(X - \overline{X})^2(Y - \overline{Y})^2} - (\overline{(X - \overline{X})(Y - \overline{Y})})^2,$$

откуда получаем интервал

$$\overline{(X - \overline{X})(Y - \overline{Y})} \pm z_{1-\alpha/2}\sigma(\widehat{F})n^{-1/2}.$$

**Ответ 9.** Пусть $X_1, ..., X_n \sim R[0,1]$, $A = \max X_i$. Тогда в выборке $X_1^*, ..., X_n^*$ с возвращением из $X_1, ..., X_n$ тот же максимум останется с вероятностью

$$1 - \left(1 - \frac{1}{n}\right)^n \to 1 - e^{-1},\ n \to \infty.$$

Таким образом, bootstrap в 63% случаев даст смещение 0. Значит 60% доверительный интервал для $\theta$ на основе максимума может оказаться нулевой длины.

**Ответ 10.** Почему данная процедура дает величину с плотностью $\widehat{p}_n$?

Формула ядерной оценки плотности сооветствует формуле свертке дискретной величины с $R\{X_1, \ldots, X_n\}$ распределением и велиины $hY$.