Занятие четвертое. О непараметрическом оценивании, о том как хорошо оценивать и как оценивать еще лучше, о том, как бутстрэпить без знания распределения

Общие слова

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 $F(x)$, EX_1 или другого показателя, связанного с моделью. При этом все мы сохраняем определения состоятельности, несмещенности и асимптотическое нормальность, просто не апеллируем в них к параметризации модели

Пример 1. Так $\overline{X} = (X_1 + ... + X_n)/n$ будет несмещенной состоятельной оценкой EX_1 и в непараметрической модели, а при $EX_1^2 < \infty$ 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 $EX_1^4 < \infty$ асимптотически нормальной.

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

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

Вопрос 1. Почему она обладает всеми свойствами 1)-3)?
Вопрос 2. Показать, что $\hat{F}_n$ — ОМП для ф.р. в непараметрической модели.
Теорема Гливенко-Кантелли утверждает, что оценка состоятельна как оценка всей функции распределения, даже более, $\hat{F}_n \to F$ п.н. по равномерной норме. Отсюда можно- сделать вывод, что $f(\hat{F}_n) \to f(F)$ для любого непрерывного (по равномерной норме) функционала f от ф.р. F сходится п.н. Кроме того, полезна также явная оценка, так называемое неравенство Дворецкого-Кифера-Вольфовица

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

Упражнение 1. Построить доверительное множество для ф.р. распределения Коши с помощью неравенства Д.-К.-В.
Вполне естественно, таким образом, исследуя функционалы $f(F)$, оценивать их $f(\hat{F}_n)$.
Пример 2. Для оценки математического ожидания $EX_1 = \int_{\mathbb{R}} x dF(x)$, а дисперсии $DX_1 = \int_{\mathbb{R}} x^2 dF(x) - (\int_{\mathbb{R}} x dF(x))^2$ возьмем

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

Вопрос 3. Почему $\hat{\theta}_1 = \overline{X}$, $\hat{\theta}_2 = S^2$?
Упражнение 2. Построить оценку для асимметрии: $E(X - EX)^3 DX^{-3/2}$.

Как оценить плотность?

С той же целью полезно бывает оценить плотность. Простейшей оценкой является гистограмма, с которой вы и так хорошо знакомы. В целом, гистограмма неплохая оценка, стремящаяся с ростом числа наблюдений и уменьшением ширины интервалов разбиения к истинному значению плотности, но она кусочно-постоянна. Хорошим методом получения непрерывной оценки для плотности является так называемая *ядерная оценка.* Назовем ядерной неотрицательную функцию $K(x)$, т.ч. $\int_{\mathbb{R}} K(x)dx = 1$, $\int_{\mathbb{R}} xK(x)dx = 0$, $\int_{\mathbb{R}} x^2 K(x)dx > 0$. В роли $K(x)$ сгодится любая плотность. Назовем ядерной оценкой плотности

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

где K — ядерная функция, а h_n — ширина окна сглаживания. Если f — непрерывна, $h_n \to 0$, $nh_n \to \infty$, то $\hat{f}_n(x)$ сходится по вероятности к $f(x)$. При этом $E(\hat{f}_n(x) - f(x))^2$ 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

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

Эта формула не вполне удовлетворительна для применения, поскольку использует неизвестное нам $f''(x)$, 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 $3(1-x^2)I_{|x|\le 1}/4$ (ядро Епанчикова).

Упражнение 3. 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?

Оценкой кросс-валидацией для оценки плотности $\hat{f}_n$ называют

$$\hat{J}(h) = \int \hat{f}_n^2 dx - \frac{2}{n}\hat{f}_{n,i}(X_i),$$

где $\hat{f}_{n,i}$ — оценка, построенная по выборке с исключенным X_i. Для ядерной оценки ее можно найти по формуле

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

$K^* = K*K - 2K$, $*$ — свертка. Это несмещенная оценка кросс-валидации $-\int(\hat{f}_n(x) - f(x))^2 dx$, расстояния между плотностью и оценкой в L^2. С помощью оценки кросс-валидации удобно измерять качество приближения. В частности, с ее помощью можно откалибровать ширину h.

С помощью оценки для плотности можно строить оценки $\int g(x)\hat{f}_n(x)dx$ для функционалов вида $\int g(x)f(x)dx$. Например, для математического ожидания мы получим оценку

$$\int_{\mathbb{R}} x\hat{f}_n(x)dx = \overline{X}$$

Вопрос 4. Почему верно последнее тождество?

Бутстреп или как прикрывать свои недостатки в непараметрическом случае

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 $X_1^*, ..., X_n^*$, взятых из нашего распределения с возвращением.

Таким образом, если мы рассматриваем статистику $f(\hat{F}_n)$ в качестве оценки $f(F)$, мы можем брать выборки из распределения $\hat{F}_n$ и на основе этих выборок изучать распределение $f(\hat{F}_n)$, ожидая, что оно близко к распределению $f(F)$. В частности, мы можем исследовать смещение $f(\hat{F}_n)$ по сравнению с $f(F)$, приближая его смещением $f(\tilde{F}_n) - f(\hat{F}_n)$, $\tilde{F}_n$ – ЭФР выборки из $\hat{F}_n$. Аналогичным образом мы можем изучать и другие параметры распределения $f(\hat{F}_n)$, например, дисперсию.

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

Упражнение 5. Оценить ф.р. статистики $\overline{X}$ по выборке размера 100 из распределения $\mathcal{N}(0,1)$.

Предположим, что $\hat{\theta} = f(\hat{F}_n)$ — оценка для какого-то параметра распределения $f(F)$. Тогда мы можем оценить ее среднее квадратичное отклонение $S(F)$. Рассматривая величину $(f(\hat{F}_n) - f(F))/S(F)$, мы можем оценить построить для нее bootstrap-интервал ширины 0.95, откуда получим доверительный

интервал для $f(F)$, так называемый Стьюдентовский pivotal интервал.
Можно было делать аналогичную операцию без использования $S(F)$, строя интервал для $f(\hat{F}_n) - f(F)$ напрямую. Такой интервал называется pivotal
Упражнение 6. 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, но не равен ему. Стьюдентовский интервал имеет более высокую скорость сходимости к $1-\alpha$, равную $O(1/n)$, обычный pivotal интервал имеет скорость $O(1/\sqrt{n})$.

Функции влияния и Дельта-метод в непараметрическом случае

Пусть f — функционал от ф.р. Рассмотрим $L_{f,F}(x) = \lim_{p\to 0}(f((1-p)F + p\delta_x) - f(F))/p$, где δ_x — ф.р. константы x. Иначе говоря, мы рассматриваем насколько изменится функционал, если в данные добавлять некоторый процент данных со значением x. $L_{f,F}$ называется функцией влияния.
Вопрос 5. Пусть $f(F) = \int_{\mathbb{R}} a(x)dF(x)$. Найти $L_{f,F}$.
Оказывается, что верна такая теорема

$$\sqrt{n}\frac{T(\hat{F}_n) - T(F)}{\tau} \to Z \sim \mathcal{N}(0,1),$$

где $\tau^2 = \int_{\mathbb{R}} L_{T,F}(x)^2 dF(x)$. Кроме того,

$$\sqrt{n}\frac{T(\hat{F}_n) - T(F)}{\hat{\tau}} \to Z \sim \mathcal{N}(0,1),$$

где $\hat{\tau} = \frac{1}{n}\sum_{i=1}^n L^2_{T,\hat{F}_n}(X_i)$. Это некоторый аналог Дельта-метода — мы видим, как меняется дисперсия величины при применении функционала T.
Нетрудно заметить, что этот метод упрощает построение Стьюдентовских интервалов, поскольку позволяет сократить расходы на подсчет дисперсии.
Упражнение 7. Рассматривая $f(F) = F(1/3) - F(1/4)$, найти для $R[0,1]$ с помощью функции влияния доверительный интервал для $f(F)$ а) обычный б) бутстрэповский стьюдентовский.