Глава 4

Методы получения оценок. Метод моментов, метод максимального правдоподобия, метод спейсингов

4.1 Базовая часть

Сегодня мы рассмотрим несколько универсальных методов получения оценок, обладающих хорошими асимптотическими свойствами. Первый из них называется метод моментов.

4.1.1 Метод моментов

Одномерный случай

Итак, пусть у нас есть параметр $\theta \in \mathbb{R}$ и $X_i \sim F_\theta$. Как нам построить асимптотически нормальную, состоятельную оценку для θ? Мы уже знаем, что гладкие функции сохраняют асимптотическую нормальность и состоятельность, поэтому нам достаточно построить асимптотически нормальную и состоятельную оценку какой-то $g(\theta)$, а потом применить g^{-1} к обеим частям. Если g окажется взаимнооднозначной и гладкой, то мы получим то, чего хотели.

Но мы знаем хорошую оценку для $\mathbf{E}_\theta X_i = a(\theta)$ — это $\overline{X}$. Эта оценка состоятельна, а в случае $\mathbf{E}_\theta X_i^2 < \infty$ асимптотически нормальна. Следовательно, если a взаимнооднозначна и обратная к ней непрерывна, то $a^{-1}(\overline{X})$ будет состоятельной оценкой θ, а если обратная дифференцируема, то $a^{-1}(\overline{X})$ будет еще и асимптотически нормальной в силу соответствующей леммы.

Пример 1. Рассмотрим модель сдвига экспоненциального распределения, X_i имеют плотность $f(x,\theta) = e^{-(x-\theta)} I_{x>\theta}$. Тогда

$$\mathbf{E}X = \int_\theta^\infty x e^{-(x-\theta)} dx = \theta + \int_0^\infty x e^{-x} dx = \theta + \Gamma(2) = \theta + 1.$$

Отсюда $\widehat{\theta}(X_1, ..., X_n) + 1 = \overline{X}$ и оценка $\widehat{\theta} = \overline{X} - 1$ будет оценкой методом моментов параметра θ.

Пример 2. Пусть $X_i \sim \mathcal{N}(0, \theta^2)$. Тогда $\mathbf{E}_\theta X_1 = 0$, что нам не подходит. Как же быть? В таких случаях берут следующий момент — $\mathbf{E}_\theta X_1^2 = \theta^2$, а значит $\sqrt{\overline{X^2}}$ — состоятельная асимптотически нормальная оценка θ.
Такие оценки называются оценками методом моментов (ОММ).

Многомерный случай

Тот же метод работает и в случае, когда $\vec{\theta} = (\theta_1, ..., \theta_k)$ — вектор. В таком случае мы рассматриваем состоятельные, асимптотически нормальные оценки $\overline{X}, ..., \overline{X^k}$ для $a_1(\vec{\theta}) = \mathbf{E}_{\vec{\theta}} X_1, ..., a_k(\vec{\theta}) = \mathbf{E}_{\vec{\theta}} X_k$. Тогда

рассмотрим отображение $A : \mathbb{R}^k \to \mathbb{R}^k$, $A(\vec{\theta}) = (a_1(\theta), ..., a_k(\theta))$. Если найдется обратное отображение A^{-1}, причем оно будет гладким, то $A^{-1}(\overline{X}, ..., \overline{X^k})$ будет асимптотически нормальной состоятельной оценкой $\vec{\theta}$ в силу теоремы 1 и теоремы 2.
Технически удобно смотреть на получение этих оценок так — мы записываем систему уравнений $\overline{X^i} = a_i(\theta)$, $i \leq k$, и решаем ее относительно θ.

Определение 1. Оценка, полученная как решение системы $\overline{X^i} = a_i(\theta)$, $i \leq k$, называется оценкой методом моментов.

Пример 3. Пусть X_i имеют гамма-распределение $\Gamma(\theta_1, \theta_2)$ с плотностью $\theta_1^{\theta_2} x^{\theta_2 - 1} e^{-\theta_1 x} I_{x>0} / \Gamma(\theta_2)$. Тогда

$$\mathbf{E}_\theta X = \frac{\theta_1^{\theta_2}}{\Gamma(\theta_2)} \int_0^\infty x \cdot x^{\theta_2 - 1} e^{-\theta_1 x} dx = \frac{\Gamma(\theta_2 + 1)}{\theta_1 \Gamma(\theta_2)} = \frac{\theta_2}{\theta_1},$$

$$\mathbf{E}_\theta X^2 = \frac{\theta_1^{\theta_2}}{\Gamma(\theta_2)} \int_0^\infty x^2 \cdot x^{\theta_2 - 1} e^{-\theta_1 x} dx = \frac{\Gamma(\theta_2 + 2)}{\theta_1^2 \Gamma(\theta_2)} = \frac{\theta_2(\theta_2 + 1)}{\theta_1^2}.$$

Отсюда

$$A_2 - (A_1)^2 = \frac{\widehat{\theta}_2}{\widehat{\theta}_1^2} = \frac{A_1}{\widehat{\theta}_1},$$

и, следовательно,

$$\widehat{\theta}_1 = \frac{A_1}{A_2 - A_1^2} = \frac{\overline{X}}{S^2},\ \widehat{\theta}_2 = A_1 \widehat{\theta}_1 = \frac{\overline{X}^2}{S^2},$$

где

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2.$$

Иногда удобнее сразу записать уравнения в терминах центрированных моментов, например, для случая двух параметров - систему уравнений

$$\overline{X} = \mathbf{E}_\theta X,\ S^2 = D_\theta X.$$

Эта система получается из исходных вычитанием из второго уравнения квадрата первого.

Свойства оценок

Какими свойствами будут обладать ОММ?
Можем ли мы рассчитывать на их несмещенность? Оценки $\overline{X^i}$ будут несмещенными оценками $a_m(\theta)$, однако, никаких причин ожидать, что функции f_i от $\overline{X^i}$ будут несмещенными оценками $\theta_i = A^{-1}(a_1(\theta), ..., a_m(\theta))$. Исключением является ситуация, когда A – линейное отображение.

С другой стороны, состоятельность у них будет, если A^{-1} непрерывно (в силу ЗБЧ и того, что непрерывные функции сохраняют сходимость по вероятности), а асимптотическая нормальность всегда, когда у X_i конечны $2k$ моментов и A^{-1} – гладкое отображение.

В тех случаях, когда как в примере 2, неудобно брать уравнения именно на основе *первых* k моментов (например, для нормальных $\mathcal{N}(0, \theta^2)$ величин первое уравнение даст нам $\overline{X} = 0$, в котором θ не фигирирует), часть уравнений можно отбросить, а вместо них взять следующие уравнения $\overline{X^l} = a_l(\widehat{\theta})$, $l > k$. Состоятельность (а в случае гладкого обратного отображения и асимптотическая нормальность) при этом сохранится.

Более общим вариантом метода является использование так называемых пробных функций $g_1(X)$, ..., $g_k(X)$. Выбирая некоторые функции $g_1(X)$,...,$g_k(X)$, мы рассматриваем для нахождения θ систему уравнений

$$\overline{g_1(X)} = \mathbf{E}_\theta g_1(X), ..., \overline{g_k(X)} = \mathbf{E}_\theta g_k(X).$$

К сожалению, асимптотическая дисперсия у ОММ не слишком хороша. Немудрено, ведь уже при трехпараметрическом семействе мы будем базироваться на оценке A_3, среднем величин X_i^3, на которые очень сильно воздействуют большие X_i (например, если на 100 наблюдений из $[0,1]$ придется одно наблюдение 5, $\overline{X^3}$ будет не менее 1.25.

4.1.2 Оценка максимального правдоподобия

Основная идея

Основой метода максимума правдоподобия является следующее соображение. Пусть наша выборка приняла значение $\vec{x} = (x_1, ..., x_n)$. Тогда при каждом θ для дискретных выборок мы можем посчитать вероятность попадания нашей выборки в $\vec{x}$, для абсолютно-непрерывных — плотность в точке $\vec{x}$, т.е. "удельную вероятность" попадания в $\vec{x}$. Чем больше это значение, тем более "вероятна" x при данном θ. Вот и выберем то θ, при котором данная выборка наиболее вероятна. Иначе говоря, возьмем

$$\forall x_1, ..., x_n \ \widehat{\theta}(x_1, ..., x_n): \ f_{X_1,...,X_n}(x_1, ..., x_n; \widehat{\theta}(x_1, ..., x_n)) \geq f_{X_1,...,X_n}(x_1, ..., x_n; \tilde{\theta}), \ \forall \tilde{\theta} \in \Theta.$$

Определение 2. Функцией правдоподобия выборки $X_1, \ldots, X_n$ в дискретном случае называют совместное распределение

$$L(x_1, \ldots, x_n; \theta) = \mathbf{P}_\theta(X_1 = x_1, \ldots, X_n = x_n),$$

в абсолютно-непрерывном случае – совместную плотность

$$L(x_1, \ldots, x_n; \theta) = f_{X_1,...,X_n}(x_1, \ldots, x_n).$$

В действительности, нас устроит чтобы меры $\mathbf{P}_\theta$ были при всех θ были абсолютно-непрерывны около некоторой общей меры μ, правдоподобием при этом будет плотность (то есть производная Радона-Никодима) меры $\mathbf{P}_\theta$ относительно этой меры. Дискретный случай – это частный случай этой схемы при дискретной мере μ на счетном множестве (множестве значений нашей выборки), а абсолютно-непрерывный – при μ – мере Лебега.

Определение 3. Оценкой максимального правдоподобия (ОМП) $\widehat{\theta}(X_1, \ldots, X_n)$ называют функцию $\widehat{\theta}(x_1, \ldots, x_n)$, т.ч.

$$L(x_1, \ldots, x_n; \widehat{\theta}(x_1, \ldots, x_n) \geq L(x_1, \ldots, x_n; \theta), \quad \forall x_i, \ \forall \theta.$$

Иначе говоря, мы выбираем точку максимума функции правдоподобия по переменной θ. Здесь θ может быть как скалярной, так и векторной величиной. Такая оценка $\widehat{\theta}(X_1, ..., X_n)$ называется оценкой максимального правдоподобия (ОМП).

Пример 4. Для схемы Бернулли с параметром θ функция правдоподобия имеет вид $\theta^{x_1+...+x_n}(1-\theta)^{n-x_1-...-x_n}$,

$$\ln L(x_1, ..., x_n; \theta) = (x_1 + ... + x_n) \ln \theta + (n - x_1 - ... - x_n) \ln(1 - \theta)$$

при $x_1, ..., x_n \in \{0, 1\}$, 0 иначе. При $x_1, ..., x_n \in \{0, 1\}$ дифференцируем полученную функцию по θ и видим, что подозрительными на экстремум при $\theta \in [0, 1]$ являются точки 0, 1 и решение уравнения

$$\frac{x_1 + ... + x_n}{\theta} = \frac{n - x_1 - ... - x_n}{1 - \theta},$$

то есть $\theta = \overline{x}$. При этом, в точке $\overline{x}$ у функции $\ln L$ максимум, следовательно, ОМП будет $\overline{X}$.
Здесь x_i — это конкретная реализация случайных величин, то есть обычные числа, а X_i — сами случайные величины. Мы находим максимум при каждой реализации и полученную функцию $\widehat{\theta}$ рассматриваем как функцию от случайных величин.

Пример 5. Рассмотрим $X_i \sim \mathcal{N}(\theta_1, \theta_2^2)$. Тогда

$$L(x_1, ..., x_n; \theta) = \frac{1}{(2\pi\theta_2^2)^{n/2}} \exp\left(-\sum_{i=1}^{n} \frac{(x_i-\theta_1)^2}{2\theta_2^2}\right),$$

$$\ln L(x_1, ..., x_n; \theta) = -n\ln(2\pi)/2 - n\ln\theta_2 - \sum_{i=1}^{n} \frac{(x_i-\theta_1)^2}{2\theta_2^2}.$$

Приравнивая частные производные $\ln L$ по θ_1 и θ_2 нулю, имеем уравнения

$$\sum_{i=1}^{n} \frac{x_i - \widehat{\theta}_1}{\widehat{\theta}_2^2} = 0,\ -\frac{n}{\widehat{\theta}_2} + \sum_{i=1}^{n} \frac{(x_i-\widehat{\theta}_1)^2}{\widehat{\theta}_2^3} = 0.$$

Из первого уравнения $\widehat{\theta}_1 = \overline{X}$, из второго $\widehat{\theta}_2 = S^2$. Это и будут ОМП для θ_1, θ_2, если в этой точке действительно достигается максимум. Это вытекает из того, что

$$L(\vec{x}, \theta_1, \theta_2) \leq L(\vec{x}, \widehat{\theta}_1, \theta_2) \leq L(\vec{x}, \widehat{\theta}_1, \widehat{\theta}_2.$$

Некоторые замечания

- Зачастую легче рассматривать $\ln L$ – максимум этой функции достигается там же, но произведение в определении функции правдоподобия превращается в более удобную сумму
- Обращайте внимание на носитель плотности или распределения. Носитель плотности, зависящий от параметра, приведет к тому, что носитель правдоподобия как функции θ при заданных $x_1, \ldots, x_n$ будет зависеть от $x_1, \ldots, x_n$.
- Не забудьте доказать, что найденная вами точка – точка максимума, а не просто критическая точка.
- Если точка максимума не единственна, то все такие точки называются ОМП. Найти ОМП означает найти множество всех таких максимумов или показать, что их нет.
- Свойства ОМП значительно сложнее получить, чем ОММ, однако, позже мы увидим, что ОМП является асимптотически нормальной в широком классе моделей. О недостатках ОМП вы можете прочитать в факультативе.

4.1.3 Оценка методом спейсингов

Формулировка

Менее известный метод спейсингов является удачной альтернативой методу максимального правдоподобия в случае, если наблюдения абсолютно-непрерывны и одномерны (параметр при этом может быть векторным).

Определение 4. Пусть

$$D(x_1, \ldots, x_n; \theta) = F(x_{(1)}; \theta)(F(x_{(2)}; \theta) - F(x_{(1)}; \theta)) \cdots (1 - F(x_{(n)}; \theta)).$$

Тогда оценкой методом спейсингов (ОМС) называют функцию $\widehat{\theta}(x_1, \ldots, x_n)$, т.ч.

$$D(x_1, \ldots, x_n; \widehat{\theta}(x_1, \ldots, x_n) \geq D(x_1, \ldots, x_n; \theta), \quad \forall x_i,\ \forall\theta.$$

Метод базируется на трех тезисах:

- Если параметр определен верно, то $F(X_i;\theta) \sim R[0,1]$;
- Если $Y_{(i)}$ – вариационный ряд из $R[0,1]$ величин, то $\mathbf{E}Y_{(i)} = i/(n+1)$.
- Максимум $a_1 \cdots a_{n+1}$, где $a_1 + \cdots + a_{n+1} = 1$, $a_i \geq 0$, достигается при $a_1 = \ldots = a_{n+1} = 1/(n+1)$.

Тем самым, при верном параметре величины $F(x_{(i)};\theta)$ будут "близки" к $i/(n+1)$, а расстояния между ними к $1/(n+1)$ и произведение этих расстояний будет близко к единице.

Примеры

Пример 6. Пусть $X_i \sim R[0,\theta]$. Тогда

$$D(x_1,\ldots,x_n;\theta) = \frac{x_{(1)}(x_{(2)} - x_{(1)})\cdots(\theta - x_{(n)})}{\theta^{n+1}} I_{0<x_{(1)}<x_{(n)}<\theta}.$$

Максимизация D равносильна максимизации $\ln(\theta - x_{(n)} - (n+1)\ln\theta$. Дифференцируя, приходим к

$$\frac{1}{\theta - x_{(n)}} - \frac{(n+1)}{\theta} = 0.$$

Это уравнение имеет единственное решение $\theta = (n+1)x_{(n)}/n$. Нетрудно убедиться, что это действительно максимум. Эта оценка в отличие от ОМП является несмещенной.

Некоторые замечания

- Зачастую легче рассматривать $\ln D$ – максимум этой функции достигается там же, но произведение превращается в более удобную сумму
- Не забудьте доказать, что найденная вами точка – точка максимума, а не просто критическая точка.
- У ОМС гораздо меньше проблем с неограниченностью плотности и недостижимостью максимума.
- В тех условиях, в которых ОМП является асимптотически нормальной и эффективной, ОМС также обладает теми же свойствами.

Условный экстремум

В некоторых задачах вам понадобится исследовать функцию на условный экстремум. Для этого напомним как устроено такого рода исследование:
Пусть $f(x_1,...,x_m)$ — интересующая нас функция (гладкая), $\phi_i(x_1,...,x_m) = 0$, $i = 1,...,k$ — дополнительные условия (гладкие), наложенные на переменную $\vec{x}$ (то есть мы рассматриваем только $x \in D = \{\vec{x} : \phi_i(\vec{x}) = 0, i = 1,...,k\}$). Предположим что функция f достигает максимума на D во внутренней точке D $\hat{\vec{x}}$. Тогда найдутся такие $\widehat{\lambda}_1,...,\widehat{\lambda}_k$, что все частные производные функции

$$g(x_1,..,x_m,\lambda_1,...,\lambda_k) = f(\vec{x}) + \lambda_1\phi_1(\vec{x}) + ... + \lambda_k\phi_k(\vec{x})$$

равны 0 при $\vec{x} = \hat{\vec{x}}$, $\vec{\lambda} = \hat{\vec{\lambda}}$.

Пример 7. Найдем минимумом $f(x,y) = x^2+y^2$ при $x+y = 1$. Для этого составим функцию $g(x,y,\lambda) = x^2 + y^2 - \lambda(x+y-1)$. Продифференцируем ее по x, y, λ и приравняем производные 0:

$$\frac{\partial}{\partial x}g(x,y,\lambda) = 2x - \lambda = 0,\ \ \frac{\partial}{\partial y}g(x,y,\lambda) = 2y - \lambda = 0,\ \ \frac{\partial}{\partial \lambda}g(x,y,\lambda) = x + y - 1 = 0.$$

Единственное решение $x = y = 1/2$, $\lambda = 1$ очевидно является минимумом (при движении по прямой $x + y = 1$ в одну из двух "бесконечностей" функция неограничено возрастает, значит не достигнуть инфимума там она не может, а других претендентов на минимум нет). Тем самым мы получаем геометрически очевидный ответ — минимум достигается в точке $(1/2, 1/2)$.

4.2 Факультатив

4.2.1 Плюсы и минусы ОМП и ОМС

- В широком классе моделей мы покажем, что ОМП асимптотически нормальна и при этом имеет наименьшую асимптотическую дисперсию среди непрерывных асимптотических дисперсий. ОМС при этом также асимптотически нормальна и обладает той же асимптотической дисперсией.

- ОМП функционально инвариантна – ОМП для $g(\theta)$ есть $g(\widehat{\theta})$, где $\widehat{\theta}$ – ОМП для θ. ОМС обладает тем же свойством.

- ОМП не зависит от потенциальных возможных реализаций экспериментов, отличной от реализовавшихся. Мы сравниваем правдоподобия именно в той точке, которая выпала в выборке, не взирая на все остальные возможные выпадающие точки. Это проиллюстрировано примером ниже. С ОМС это не совсем так, хотя в примере ниже ОМС также будет работать без всяких проблем.

- При этом ОМП может не существовать, может не быть единственна, может сходиться к величине отличной от параметра. Наиболее частой плачевной ситуацией для ОМП является случай неограниченных плотностей. ОМС не столь страдает от неограниченности плотности, она реже бывает неединственной, она состоятельна в гораздо более широких условиях.

- Увы, ОМС в такой формулировке работает только в непрерывном случае и только если данные одномерные. ОМП накладывает в этом смысле куда меньше условий.

Пример 8. Предположим, что физик измерил с помощью вольтметра с диапазоном от 2 до 5 вольт напряжение на n приборах и отослал вам для исследования. Спустя сутки он обнаружил, что вольтметр сломался и на напряжениях свыше 4 вольт не работает, всегда при таких напряжениях показывая 4 вольта. Он сообщил об этом нам, добавив, что не знает, сломался ли к моменту проведения наших измерений вольтметр или еще нет, но, к счастью, все показания в выборке от 2 до 3 вольт, так что это неважно. Однако, если, скажем, мы захотим оценить некий параметр методом моментов, то нам понадобятся теоретически средние a_k, которые зависят от того, зашкаливающий вольтметр мы имеем или нет. Таким образом, нам придется заставить знакомого переделать свои измерения и прислать их заново. С другой стороны, для метода максимального правдоподобия совершенно неважно, может ли зашкаливать вольтметр, поскольку при в нашей точке, в которой нет значений 4 и выше, плотности в случае наличия и отсутствия зашкаливания идентичны. Таким образом, ОМП не интересуется общим поведением функции L при всех x_i, θ. Она зависит только от поведения L в данной фиксированной точке x.

Пример 9.

Пример 10. Для $R(\theta, \theta+1)$ функция правдоподобия будет иметь вид $L(x_1, ..., x_n; \theta) = \prod_{i=1}^n I_{x_i \in [\theta, \theta+1]}$. Эта функция равна 1 при $\max(x_i) - 1 \le \theta \le \min(x_i)$ и 0 иначе, максимум у нее достигается на целом отрезке. Здесь ОМП не единственна.

Пример 11. Напротив, рассмотрим следующую модель: пусть плотность $f_{x,\theta}$ имеет вид

$$p_1 \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x-a_1)^2}{2\sigma_1^2}\right) + (1-p_1)\frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(x-a_2)^2}{2\sigma_2^2}\right),$$

зависящую от 5 параметров $a_1, a_2, \sigma_1, \sigma_2, p_1$. Содержательно модель можно описать так: мы имеем монету с вероятностью p_1, если она выпадает на орла, то мы выбираем одно нормальное распределение, а иначе второе. Тогда возьмем выборку $X_1, ..., X_n$, какие-нибудь $p_1 \in (0,1)$, a_2, σ_2 и $a_1 = X_1$. Тогда функция правдоподобия удовлетворяет неравенству

$$L(x_1, ..., x_n; a_1, a_2, \sigma_1, \sigma_2, p_1) \geq p_1 \frac{1}{\sqrt{2\pi}\sigma_1} (1 - p_1)^n \prod_{i=2}^{n} f_{a_2, \sigma_2}(x_i),$$

где f_{a_2,σ_2} — нормальная плотность с параметрами a_2, σ_2^2. Действительно, в каждой из плотностей мы оставили лишь одно из наших двух слагаемых — в первой первое, а в остальных второе. Полученная нижняя оценка для L уходит в бесконечность с убыванием σ_1 к 0, значит $\sup L = \infty$ и ОМП не существует.

Интересно, что при этом один из локальных максимумов при этом будет ОМП. ОМС в этой модели ведет себя вполне прилично и является состоятельной.

4.2.2 Непараметрическая ОМП

Если мне ничего неизвестно о распределении X_i, то строить правдоподобие в этом случае не вполне выходит. В наиболее общем случае (если я рассматриваю вообще все распределения), то они не являются абсолютно-непрерывными около какой-то одной меры. Более того, плотность в заданных точках можно устремить в бесконечность, правильно подбирая функцию. В связи с этим в непараметрическом случае используют другое понятие правдоподобия:

Определение 5. Непараметрическим правдоподобием называют

$$L_{NPE}(x_1, \ldots, x_n; F) = (F(x_1) - F(x_1 - 0))(F(x_2) - F(x_2 - 0)) \cdots (F(x_n) - F(x_n - 0)),$$

где в качестве параметра рассматривается ф.р. F.

Определение 6. Непараметрической оценкой максимального правдоподобия (НОМП) для параметра $f(F)$ в классе ф.р. $\mathcal{F}$ называется оценка $f(\widehat{F}_{MLE})$, где

$$\widehat{F}_{MLE}: \; L_{NPE}(x_1, \ldots, x_n; \widehat{F}_{MLE}) \geq L_{NPE}(x_1, \ldots, x_n; F), \quad \forall F \in \mathcal{F}.$$

Пример 12. В классе всех распределений НОМП будет естественная оценка $f(\widehat{F}_n)$, где $\widehat{F}_n$ – ЭФР. Действительно, пусть F – ф.р. с атомами вероятности $p_1, \ldots, p_n$ в точках $x_1, \ldots, x_n$ (для простоты предположим, что x_i различны), то

$$L_{NPE}(x_1, \ldots, x_n; F) = p_1 \cdots p_n.$$

При этом $p_1 + \cdots + p_n \leq 1$, поскольку это вероятности. При этом в силу неравенства о среднем геометрическим и среднем арифметическим

$$\sqrt[n]{p_1 \cdots p_n} \leq \frac{p_1 + \cdots + p_n}{n} \leq \frac{1}{n},$$

равенство в котором достигается при $p_1 = \ldots = p_n$, максимум L_{NPE} достигается при $p_1 = \cdots = p_n = 1/n$. Тем самым, максимум правдоподобия будет при F_{MLE} равной ЭФР.

Этот подход позволяет хорошо искать оценки, используя имеющиеся ограничения на распределения, хотя ограничен, по существу, дискретными распределениями.