Занятие третье. О параметрическом оценивании, о методах известных и не очень и о том, что такое бутстрэп

Вспоминаем умные слова

В рамках этого занятия мы довольно много будем повторять стандартный курс статистики. Рассматривая модель $X_i \sim F_\theta$, где $\theta \in \Theta$, мы называли оценкой функцию $T(X_1, ..., X_n)$, которая приближает параметр θ.

При этом важными свойствами для нас были:

1) Несмещенность $E_\theta T(X_1, ..., X_n) = \theta$,

2) Состоятельность: $T(X_1, ..., X_n) \xrightarrow{P_\theta} \theta$,

3) Асимптотическая нормальность: $\sqrt{n}(T(X_1, ..., X_n) - \theta) \xrightarrow{d} Z \sim \mathcal{N}(0, \sigma^2(\theta))$. Величину $\sigma^2(\theta)$ называют асимптотической дисперсией.

Свойство несмещенности нам важно, если мы повторяем опыт, тогда мы можем гарантировать себе, что у нас не будет систематического занижения или завышения оценки. Свойство состоятельности необходимо, если мы имеем возможность наращивать количество испытаний, свойство асимптотической нормальности уточняет состоятельность.

При этом мы привыкли рассматривать именно параметрическую модель $X_i \sim F_\theta$, где F — какое-то заданное семейство функций распределения, причем обычно довольно узкое — например, все нормальные распределения.

Итак, два базовых метода для построения оценок, которые вы рассматривали — метод моментов и метод максимального правдоподобия. Напомним как они устроены, обсудим их качества и добавим еще один метод, называемый методом спэйсингов.

Три базовых метода

Метод моментов базируется на следующих фактах:

а) ЗБЧ и ЦПТ. Выборочные средние $\overline{X^k} = \frac{X_1^k + ... + X_n^k}{n}$— хорошие оценки для $E_\theta X_1^k$, а именно состоятельные и при $E_\theta X_1^{2k} < \infty$ асимптотически нормальные с асимптотической дисперсией $D_\theta X_1^k$.

б) Дельта-метод или лемма об асимптотической нормальности. При применении непрерывной функции $f(x)$ к состоятельной оценке $\hat{\theta}$ функции $g(\theta)$ мы получаем состоятельную оценку для $f(g(\theta))$. В случае гладкости f и асимптотической нормальности $\hat{\theta}$, мы получим асимптотически нормальную оценку для $f(g(\theta))$, причем асимптотическая дисперсия увеличится в $(f'(g(\theta)))^2$ раз.

Этот факт мы сформулировали для одномерной функции f, аналогичный факт верен и в случае векторной функции, асимптотическая дисперсия при этом заменяется асимптотической матрицей ковариации $\Sigma(\theta)$, которая после применения f переходит в $J\Sigma J^t$, где J — матрица Якоби f в точке $g(\theta)$.

Метод моментов для оценки векторного параметра $\vec{\theta} = (\theta_1, ..., \theta_k)$ предлагает рассмотреть $\overline{X^1}$,...,$\overline{X^k}$ как оценки $\mu_1(\theta) = E_\theta X$,...,$\mu_k(\theta) = E_\theta X^k$ и применить к ним такое отображение f, которое переводит $(\mu_1(\theta), ..., \mu_k(\theta))$ в $\vec{\theta}$. Если это отображение окажется непрерывным, то $f(\overline{X}, ..., \overline{X^k})$ будет состоятельной для $\vec{\theta}$, а если гладким, то асимптотически нормальным (если, конечно $E_\theta X^{2k} < \infty$).

К сожалению, метод моментов не вполне удачен в плане асимптотической дисперсии, она зачастую бывает весьма большой, особенно в случае многомерных параметров.

Метод максимального правдоподобия исходит из простого соображения — мы должны искать такое θ, при котором появление нашей выборки особенно вероятно. Это равносильно максимизации совместного распределения (в случае дискретной выборки) или совместной плотности (в случае абсолютно-непрерывной плотности). Таким образом, метод максимального правдоподобия предписывает искать θ, такие что

$$L(x_1, ..., x_n; \theta) = f_\theta(x_1)...f_\theta(x_n) \to \max,$$

где f_θ — плотность или вероятность. Практическую зачастую удобнее искать аргумент максимума $\ln L(x_1, ..., x_n; \theta)$.

В некоторых случаях ОМП может быть неединственной или ее не будет совсем. Наиболее типичные случаи:

1) Если распределение X_i устроено так, что при разных параметрах θ величины принимают значения из разных множеств.
Вопрос 1. Что будет ОМП в случае $X_i \sim R[\theta, \theta+1]$?
2) Если плотность распределения X_i не гладко зависит от параметра
Вопрос 2. Что будет ОМП в случае X_i с плотностью $\exp(-|x-\theta|)/2$.
3) Если множество изменения параметра не является замкнутым
Вопрос 3. Что будет ОМП в случае X_i, с вероятностью $1/2$ имеющих $\mathcal{N}(0,1)$ распределение, а с вероятностью $1/2$ — $\mathcal{N}(\theta_1, \theta_2^2)$?
Подсказка Что будет, если $\theta_2 \to 0$?

В случае достаточно гладко зависящих от θ распределений, замкнутого множества изменения параметра и независимости носителя от параметра, ОМП оказываются довольно хорошими оценками — состоятельными, асимптотически нормальными, да еще и с наилучшей возможной асимптотической дисперсией среди оценок с непрерывной асимптотической дисперсией.
Вопрос 4. Рассмотрим $\hat{\theta}_n = \overline{X}(1 + I_{|\overline{X}|>n^{2/3}})/2$, $X_i \sim \mathcal{N}(0,1)$. Показать, что эта оценка асимптотически нормальна и ее асимптотическая дисперсия равна дисперсии ОМП при $\theta \neq 0$ и меньше ее при $\theta = 0$.

Обе рассмотренных оценки являются эквивариантными, то есть если $\hat{\theta}$ — ОМП или ОММ для θ, то $f(\hat{\theta})$ — ОМП или ОММ для $f(\theta)$. Это удобное свойство, которое, к сожалению, плохо сочетается с несмещенностью.

Упражнение 1. Построить алгоритм, численно вычисляющий по однопараметрической функции плотности ОММ и ОМП.
Для решения уравнений пригодится функция uniroot, для интегрирования — функция integrate (лучше интегрировать не по всей прямой, а по конечному отрезку), а для минимизации — nlm или optimize. Функция nlm подходит и для векторного случая, также минимизацию можно осуществлять optim, а вот uniroot в многомерном случае заменяется на polyroot из библиотеки rootsolve)
Реализованная оценка ОМП с помощью optim находится в пакете stats4, эта функция носит название mle, ее аргументом является логарифмическая функция правдоподобия с обратным знаком.

Метод спейсингов (method of maximal spacing) предлагает для оценивания параметра θ семейства распределений с плотностями, сосредоточенными на отрезке $[a,b]$, рассмотреть

$$S(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} D_i(\theta),$$

где $D_i(\theta) = P_\theta(X \in [x_{(i)}, x_{(i+1)}]) = F_\theta(x_{(i+1)}) - F_\theta(x_{(i)})$, $x_{(0)} = a$, $x_{(n+1)} = b$, где $x_{(i)}$ — вариационный ряд. Тогда θ, максимизирующее S, называют оценкой методом спейсингов.

Логика довольно проста — при правильном θ $Y_i = F_\theta(X_{(i)})$ есть вариационный ряд равномерного распределения. Тогда $Y_i - Y_{i-1}$ одинаково распределены. Но $\max_{y_1+...+y_{n+1}}(y_1...y_{n+1})$ достигается при $y_i = 1/(n+1)$. Следовательно, максимизация S будет связана с выравниванием $D_i(\theta)$, что соответствует искомому θ. Эта оценка состоятельна, а в случае регулярных оценок асимптотически эффективна.
Вопрос 5. Какая оценка методом спейсингов для равномерного распределения $R[\theta_1, \theta_2]$.
Упражнение 2. $F_{\theta_1,\theta_2}(x) = 1 - e^{-(x-\theta_1)^{\theta_2}}$, $x > \theta_1$. Построить численно оценку методом спейсингов. Сравнить ее с ОМП.

Бутстрап или как прикрывать свои недостатки

Построив оценку вида $g(\hat{F}_n)$ или $g(\hat{f}_n)$ для какого-то функционала g, мы никак не гарантируем несмещенности. Но если мы часто повторяем эксперимент, нам может быть важно снизить смещение. Поэтому мы бы хотели научиться исправлять или снижать смещенность оценки.
Идея метода бутстрапа заключается в том, что если оценка $\hat{\theta}(X_1, ..., X_n)$ близка к настоящему параметру θ, а семейство распределений непрерывно зависит от параметра, то распределение $F_{\hat{\theta}}$ будет похоже на F_θ. Таким образом, мы можем взять выборку Y_1,...,Y_n из $F_{\hat{\theta}}$, и подсчитать по ней $\hat{\theta}(Y_1, ..., Y_n)$. Тогда смещение $\hat{\theta}(Y_1, ..., Y_n) - \hat{\theta}(X_1, ..., X_n)$ будет близко к смещению $\hat{\theta}(X_1, ..., X_n) - \theta$. Таким образом, метод бутстрапа предписывает взять оценку $2\hat{\theta}(X_1, ..., X_n) - \hat{\theta}(Y_1, ..., Y_n)$.

Метод бутстрапа называется в честь ремешков на обуви барона Мюнгхаузена, ухватившись за которые, он вытянул себя из болота. Вот и этот метод позволяет нам улучшить оценку с помощью самой этой оценки.

Упражнение 3. С помощью бутстрэппинга исправить смещение оценки S^2 для выборки размера 20 из $\mathcal{N}(0,1)$.

Использовать метод бутстрэпа можно не только для оценивания смещения, но и, например, для оценки квадратичного смещения. Так мы могли бы взять $(\hat{\theta}(Y_1,...,Y_n) - \hat{\theta}(X_1,...,X_n))^2$ в качестве оценки дисперсии $\hat{\theta}(X_1,...,X_n)$.

Упражнение 4. Пусть известно $\overline{X}$ из выборки. Как с его помощью бутстрэппингом оценить дисперсию известного распределения? Применить метод для $R[0,1]$.

Этот метод имеет несколько неоспоримых плюсов — он прост в использовании и не требует вычислений, применим даже к весьма громоздким моделям. С другой стороны, мы не можем явным образом оценить его погрешность, а в случае, если оценка $\hat{\theta}$ значимо промахнулась мимо θ, рискуем неправильно изменить оценку.

Вопрос 6. Предположим, что мы оценили среднее в модели $R[0,\theta]$ с помощью $\overline{X}$. Теперь мы берем новую выборку из $R[0,2\overline{X}]$ и оцениваем с помощью ее среднего θ. Какую дисперсию будет иметь эта новая оценка?

Доверительные беседы о доверительных интервалах

Будем рассматривать одномерные параметры $\theta \in \mathbb{R}$.

Напомним, что доверительным интервалом называют пару статистик $\hat{\theta}_1$, $\hat{\theta}_2$, таких, что при всех θ выполнено (не)равенство

$$P_\theta(\theta \in (\hat{\theta}_1,\hat{\theta}_2))(\leq) = 1-\alpha,$$

где α — заданное число, называемое уровнем доверия.

Построение точных доверительных интервалов достаточно трудоемко, а вот асимптотические доверительные интервалы (т.е. те, где равенство заменено на равенство в пределе) можно строить на основе любой асимптотически нормальной оценки. Так если $\hat{\theta}$ — ОМП для θ в регулярной модели, то ее асимптотическая дисперсия есть $1/I(\theta)$ – информация Фишера, которую можно состоятельно оценить величиной $I(\hat{\theta})$. Из асимптотической нормальности ОМП

$$\sqrt{n}\frac{\hat{\theta}-\theta}{I^{-1/2}(\theta)} \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

Но, в силу состоятельности $I(\hat{\theta})$ как оценки $I(\theta)$

$$\sqrt{n}\frac{\hat{\theta}-\theta}{I^{-1/2}(\hat{\theta})} = \sqrt{n}\frac{\hat{\theta}-\theta}{I^{-1/2}(\theta)}\sqrt{\frac{I(\hat{\theta})}{I(\theta)}} \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

Тогда

$$P\left(\sqrt{n}\frac{\hat{\theta}-\theta}{I^{-1/2}(\hat{\theta})} \leq x\right) \to \Phi(x),$$

откуда мы можем построить доверительный интервал вида

$$\left(\hat{\theta} - \frac{z_\beta I^{-1/2}(\hat{\theta})}{\sqrt{n}}, \hat{\theta} - \frac{z_\gamma I(\hat{\theta})^{-1/2}}{\sqrt{n}}\right),$$

$\beta - \gamma = 1-\alpha$.

Может показаться, что для оценки методом моментов доверительный интервал строится сложнее. Но здесь нам на помощь приходит дельта-метод. Асимптотическую дисперсию оценки $\overline{X^k}$ как оценки $\mu_k(\theta)$ мы знаем — это $D_\theta X_k = \mu_{2k}(\theta) - \mu_k^2(\theta)$. Следовательно, если ОММ для θ есть $f(\overline{X^k})$, то ее асимптотическая дисперсия $\sigma^2(\theta) = (\mu_{2k}(\theta) - \mu_k^2(\theta))(f'(\mu_k(\theta)))^2$, т.е. доверительный интервал будет иметь вид

$$\left(f(\overline{X^k}) - \frac{z_\beta \sigma(f(\overline{X^k}))}{\sqrt{n}}, f(\overline{X^k}) - \frac{z_\gamma \sigma(f(\overline{X^k}))}{\sqrt{n}} \right)$$

Свои варианты для построения доверительного множества дает и бутстрэп. Такого рода методов несколько, но мы рассмотрим наиболее простой — pivotal интервал.
Для этого рассмотрим оценку $\hat{\theta}$ для θ. Будем брать выборки из $F_{\hat{\theta}}$ и строить на основе них оценки $\hat{\theta}_1$,...,$\hat{\theta}_m$. Рассмотрим $\Delta_i = \hat{\theta}_i - \hat{\theta}$. Мы ожидаем, что это выборка из величин, близких к $\hat{\theta} - \theta$. Упорядочим Δ_i и выберем те из них Δ_-, Δ_+, которые стоят на местах $[\gamma m]$ и $[\beta m]$ по возрастанию. Тогда

$$(\hat{\theta} - \Delta_+, \hat{\theta} - \Delta_-)$$

и будет нашим интервалом.
Упражнение 5. Сравнить на выборках размера 50 для а) $\mathcal{N}(\theta, 1)$, б) $R[0, \theta]$ доверительные интервалы на основе ОММ, ОМП, бутстрэпа с помощью $\overline{X}$.

Доверительные множества или то же, но на плоскости и в пространстве

Для векторного параметра $\theta = (\theta_1, ..., \theta_k)$ нам уже не очень помогают доверительные интервалы для каждой θ_i, поскольку знание вероятностей накрытия каждой из θ_1,...,θ_k своим интервалом не дает нам возможность выписать вероятность того, что все параметры одновременно попадут в соответствующий параллелипипед. Поэтому рассматривают доверительное множество $A(X_1, ..., X_n) \subset \mathbb{R}^k$, которое накроет мой параметр с вероятностью $1 - \alpha$.

Для многомерной нормальной выборки $(X_1, ..., X_k) \sim \mathcal{N}(\vec{\mu}, \Sigma)$, мы можем утверждать, что $(\vec{X} - \vec{\mu})^t \Sigma^{-1} (\vec{X} - \vec{\mu})$ будет вектором с распределением χ_k^2, поскольку $A^{-1}(\vec{X} - \vec{\mu}) \sim \mathcal{N}(0, E)$, где $A^t A = \Sigma$. Но тогда мы можем сказать, что выборка попадает в эллипсоид с центром $\vec{\mu}$

$$(\vec{X} - \vec{\mu})^t \Sigma^{-1} (\vec{X} - \vec{\mu}) \leq x$$

с вероятностью $F_{\chi_k^2}(x)$.

Таким образом мы можем строить доверительные множества на основе векторных асимптотически нормальных оценок. Для ОМП в роли асимптотической дисперсии $\Sigma(\theta)$ будет выступать $I^{-1}(\theta)$, где $I(\theta)$ — информационная матрица Фишера с элементами

$$I_{i,j}(\theta) = cov\left(\frac{\partial}{\partial \theta_i} \ln f_\theta(X), \frac{\partial}{\partial \theta_j} \ln f_\theta(X) \right).$$

$$\sqrt{n}(\hat{\theta} - \vec{\theta}) \xrightarrow{d} \mathcal{N}(0, \Sigma(\theta)),$$

откуда имеем доверительный эллипсоид для $\vec{\theta}$, заданный соотношением

$$(\hat{\theta} - \vec{\theta})^t I(\hat{\theta})(\hat{\theta} - \vec{\theta}) \leq y_{1-\alpha},$$

где y — квантиль χ_k^2.

Аналогично строится доверительное множество на основе ОММ, только для подсчета матрицы ковариации придется использовать многомерный Дельта-метод и ЦПТ для векторов. Более конкретно, если $\Sigma_1(\theta)$ — матрица ковариации вектора $X, X^2, ..., X^k$, g — отображение, такое что $g(\mu_1(\theta), ..., \mu_k(\theta)) = \theta$, а $J(\theta)$ — его матрица Якоби в точке $\mu_1(\theta), ..., \mu_k(\theta)$, то для $\hat{\theta} = g(\overline{X}, \overline{X^2}, ..., \overline{X^k})$ выполнено соотношение

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} Z \sim \mathcal{N}(0, J^t(\hat{\theta}) \Sigma_1(\hat{\theta}) J(\hat{\theta})).$$

Соответственно, доверительный эллипсоид имеет вид

$$(\hat{\theta} - \theta)^t \Sigma^{-1}(\hat{\theta})(\hat{\theta} - \theta) \leq y_{1-\alpha}/n,$$

где $\Sigma = J \Sigma_1 J^t$.

Упражнение 6. Построить асимптотический доверительный эллипс на основе ОММ и ОМП для $X_i \sim \Gamma(\theta_1, \theta_2)$ на выборке размера 50 при $\theta_1 = 2$, $\theta_2 = 2$. Обратную матрицу к A можно построить с помощью функции solve(A). Доверительные эллипсы нормального распределения удобно рисовать с помощью функции ellipse пакета mixtools. Так функция ellipse(mu,Sigma,alpha = .05, npoints = 50, newplot = TRUE, type = "l") строит доверительный эллипс для нормального $\mathcal{N}(mu, Sigma^2)$ распределения, n точек уровня α.