Экспоненциальные семейства распределения и эффективные оценки для них

$\{F(x,\theta);\ \theta\in\Theta\}$

$f(x,\theta)$ - плотность

Если семейство регулярно и $f(x.\theta)=\exp(A(\theta)B(x)+C(\theta)+D(x))$, то его называют экспоненциальным

$U(X,\theta)=\frac{\partial}{\partial\theta}\ln L(X,\theta)=\frac{\partial}{\partial\theta}\sum\limits_{i=1}^{n}(A(\theta)B(X_i)+C(\theta)+D(X_i))=A'(\theta)\sum\limits_{i=1}^{n}B(X_i)+nC'(\theta)=nA'(\theta)(T^*-(-\frac{C'(\theta)}{A'(\theta)}))=U(X,\theta)$

$T^*=\frac{1}{n}\sum\limits_{i=1}^{n}B(X_i)$, $\tau(\theta)=-\frac{C'(\theta)}{A'(\theta)}$, $a(\theta)=\frac{1}{nA'(\theta)}$

Получаем условие Рао-Крамера

$T^*-\tau(\theta)=a(\theta)U(X,\theta)$

Следовательно, T^* является эффективной оценкой для $\tau(\theta)=-\frac{C'(\theta)}{A'(\theta)}$

Найдём дисперсию этой оценки

$\tau(\theta)=\int\limits_{\mathbb{R}^n}T^*(x)L_n(x,\theta)\lambda(dx)$

$\tau'(\theta)=\int\limits_{\mathbb{R}^n}T^*(x)U(x,\theta)L_n(x,\theta)\lambda(dx)$

$\tau'(\theta)=E_\theta(T^*U(X,\theta))=cov(T^*,U(X,\theta))=cov(T^*,\frac{1}{a}(T^*-\tau(\theta)))=\frac{D_\theta T^*}{a}$

Отсюда $D_\theta T^*=a(\theta)\tau'(\theta)=\frac{1}{nA'(\theta)}\cdot(-\frac{C'(\theta)}{A'(\theta)})'$ - формула для дисперсии

Пример 1 - $N(\alpha,\theta^2)$

$f(x,\theta)=\frac{1}{\sqrt{2\pi}\theta}\exp(-\frac{1}{2\theta^2}(x-\alpha)^2)=\exp(-\frac{1}{2\theta^2}(x-\alpha)^2-\ln\theta-\ln\sqrt{2\pi})$

$A(\theta)=-\frac{1}{2\theta^2}$, $B(x)=(x-\alpha)^2$, $C(\theta)=-\ln\theta$, $D(x)=-\ln\sqrt{2\pi}$

$\tau(\theta)=-\frac{(-\ln\theta)'}{(-\frac{1}{2\theta^2})'}=\frac{\frac{1}{\theta}}{\frac{1}{\theta^3}}=\theta^2$

Эффективное оценивание возможно для θ^2

$T^*(X)=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\alpha)^2$

$D_\theta T^*=a(\theta)\tau'(\theta)=\frac{1}{n\frac{1}{\theta^3}}2\theta=\frac{2}{n}\theta^4$ - наилучшее значение

Пример 2 - $Bi(k,\theta)$ - x - число успехов среди k испытаний

$f(x,\theta)=C_k^x\theta^x(1-\theta)^{k-x}=\exp(x\ln\theta+(k-x)\ln(1-\theta)+\ln C_k^x)=\exp((\ln\theta-\ln(1-\theta))x+k\ln(1-\theta)+\ln C_k^x)$

$\tau(\theta)=-\frac{C'(\theta)}{A'(\theta)}=\frac{-\frac{k}{1-\theta}}{\frac{1}{\theta}+\frac{1}{1-\theta}}=k\theta$

$T^*=\frac{1}{n}\sum\limits_{i=1}^{n}X_i=\overline{X}$

$D_\theta T^*=\frac{1}{n}DX_1=\frac{k\theta(1-\theta)}{n}$

Нужна эффективная оценка для θ

$\tilde{T}=\frac{1}{kn}\sum\limits_{i=1}^{n}X_i$

$D_\theta\tilde{T}=\frac{1}{n}DX_1=\frac{\theta(1-\theta)}{kn}$

Для экспоненциальных семейств всегда можно найти эффективную оценку (и только для них)

Теорема

Если регулярная статистическая модель допускает эффективное оценивание хотя бы одной функции от параметра, то она является экспоненциальной моделью

Доказательство

Пусть X_1 - выборка объёма 1

И пусть $T^*(X_1)$ - эффективная оценка некоторой функции $\tau(\theta)$

Тогда для неё выполнено условие Рао-Крамера

$T^*(X_1)-\tau(\theta)=a_1(\theta)U(X_1,\theta)$ (U - вклад одного элемента выборки)

$T^*(X_1) - \tau(\theta) = a_1(\theta)\frac{\frac{\partial}{\partial\theta}f(X_1,\theta)}{f(X_1,\theta)}$

Для почти всех x верно следующее:

$T^*(x) - \tau(\theta) = a_1(\theta)\frac{\frac{\partial}{\partial\theta}f(x,\theta)}{f(x,\theta)}$

Делим на a_1, интегрируем по θ и получаем ($\ln D(x)$ вылезает после взятия интеграла):

$\int \frac{1}{a_1(\theta)}(T^*(x) - \tau(\theta))d\theta = \ln f(x,\theta) - \ln D(x)$

$A(\theta) = \int \frac{d\theta}{a_1(\theta)}$, $B(x) = T^*(x)$, $C(\theta) = -\int \frac{\tau(\theta)}{a_1(\theta)}$

Потенциализируем и в итоге:

$\exp(A(\theta)B(x) + C(\theta) + D(x)) = f(x,\theta)$ - QED

Как получать оптимальную оценку для других распределений?

Критерий оптимальности Бхаттачария

$U = \frac{\partial}{\partial\theta}\ln L(X,\theta) = \frac{L'_\theta(X,\theta)}{L(X,\theta)}$

В критерии Бхаттачария кроме выписанной дроби используются ещё дроби вида $\frac{L^{(2)}_\theta(X,\theta)}{L(X,\theta)}, \ldots, \frac{L^{(n)}_\theta(X,\theta)}{L(X,\theta)}$

Ищется оценка, для которой $T - \tau(\theta)$ является линейной комбинацией указанных дробей с коэффицентами, зависящими от параметра

Неравенство и условие Бхаттачария

Пусть $T(X)$ - несмещённая оценка для функции от параметра θ, $E_\theta T(X) = \tau(\theta)$

Тогда её дисперсия удовлетворяет условию $D_\theta T \geq \sum\limits_{i,j=1}^{r} c_{i,j}a_i a_j$, где $c_{i,j} = E_\theta(\frac{L^{(i)}}{L}\cdot\frac{L^{(j)}}{L})$, где коэффиценты определяются системой уравнений

$\tau^{(i)}(\theta) = \sum\limits_{j=1}^{r} c_{i,j}(\theta)a_j(\theta)$, $i = 1, \ldots, r$

Знак в равенстве неравенства Бхаттачария достигается тогда и только тогда, когда $T - \tau(\theta) = \sum\limits_{i=1}^{r} a_i\frac{L^{(i)}}{L}$

Замечание

Если матрица $(c_{i,j})^r_{i,j=1}$ является невырожденной, то неравенство Бхаттачария примет вид $D_\theta \geq \sum\limits_{i,j=1}^{r} c^{i,j}\tau^{(i)}\tau^{(j)}$, где $c^{i,j}$ - элементы обратной матрицы

$a = c^{-1}\tau'$

$(\tau')^T(c^{-1})^T c c^{-1}\tau = (\tau')^T(c^{-1})^T\tau = (\tau')^T c^{-1}\tau$

Это обобщение неравенства Рао-Крамера