Математическая статистика. Теория

Александр Сергеев

1 Введение

Есть генеральная совокупность. Надо выбрать часть генеральной совокупности – выборку. По выборке хотим сделать вывод о всей совокупности
В исследовании есть следующие этапы

1. Сбор данных
2. Препроцессинг (чистка)
3. Построение модели и анализ
4. Интерпретация

Определение
Репрезентативность – свойство выборки, означающее, что по выборке можно судить по всей совокупности

2 Простейшая модель выборки. Эмпирическая функция распределения

Простейшая модель выборки
$X_1, \dots, X_n$ – i.i.d.
F – функция распределения (теоретическая, мы ее не знаем)
$x_1, \dots, x_n$ – реализация выборки

Глобальная цель – оценить из реализации $x_1, \ldots, x_n$ теоретическую функцию F

Определение (эмпирическая функция выборки)
$\mu_n(t) = \sum\limits_{i=1}^{n} \mathbb{1}(X_i \le t)$
$F_n(t) = \dfrac{\mu_n(t)}{n}$ – эмпирическая функция распределения
Замечание
Заметим, что $\mathbb{1}(X_i \le t) \sim Bin(F(t))$
Тогда $\mu_n(t) \sim Bin(n, F(t))$
$P(F_n(x) = \dfrac{k}{n}) = P(\mu_n(x) = k) = \dbinom{n}{k} F_n^k(x)(1 - F_n(x))^{n-k}$
Отсюда $E(\mu_n(x)) = nF_n(x)$
$E(F_n(t)) = F(t)$ – несмещенность
$\operatorname{Var} F_n(t) = \dfrac{F(t)(1 - F(t))}{n}$
По ЦПТ $\dfrac{\mu_n(t) - nF(t)}{\sqrt{F(t)(1 - F(t))n}} = \sqrt{n}\dfrac{F_n(t) - F(t)}{\sqrt{F(t)(1 - F(t))}} \xrightarrow{d} N(0, 1)$ – асимптотическая нормальность
Теорема Гливенко – Кантелли
$\sup\limits_{t \in \mathbb{R}} |F_n(t) - F(t)| \xrightarrow[n \to \infty]{a.s.} 0$
Теорема Колмагорова
$D_n = \sup\limits_x |F_n(x) - F(x)|, F \in C(\mathbb{R}), t \ge 0 \Rightarrow P(\sqrt{n} D_n \le t) \to K(t) = \sum\limits_{j=-\infty}^{\infty} e^{-2j^2t^2}$ – функция распределения Колмагорова
Теорема Смирнова
$X_1, \ldots, X_n, Y_1, \ldots, Y_n$ – независимы
Обе распределены на $F \in C(\mathbb{R})$
$D_{m,n} = \sum\limits_x |F_n(x) - F_m(x)|$
Тогда $P(\sqrt{\dfrac{nm}{n+m}} D_{m,n} \le t) \xrightarrow{m,n \to \infty} K(t)$

3 Выборочные моменты

$\alpha_k = EX_1^k$ – k-ый теоретический момент
$\beta_k = E(X_1 - EX_1)^k$ – k-ый теоретический момент
$\overline{g(X)} := \dfrac{1}{n}\sum\limits_{k=1}^{n} g(X_k), g(\bullet) : \mathbb{R} \to \mathbb{R}$
$\widehat{\alpha}_k = \overline{X^k} = \dfrac{1}{n}\sum\limits_{j=1}^{n} X_j^k$ – k-ый выборочный момент
$E\widehat{\alpha}_k = \alpha_k$ – несмещенность
$\operatorname{Var}\widehat{\alpha}_k = \dfrac{1}{n}\operatorname{Var}(X_1^k) = \dfrac{1}{n}((EX_1^{2k}) - (EX_1^k)^2)$
По ЦПТ $\sqrt{n}\dfrac{\widehat{\alpha}_k - \alpha_k}{\sqrt{\alpha_{2k} - \alpha_k^2}} \approx N(0,1)$
$\sqrt{n}\dfrac{\widehat{\alpha}_k - \alpha_k}{\sqrt{\widehat{\alpha}_{2k} - \widehat{\alpha}_k^2}} = \sqrt{n}\dfrac{\widehat{\alpha}_k - \alpha_k}{\sqrt{\alpha_{2k} - \alpha_k^2}} \underbrace{\sqrt{\dfrac{\alpha_{2k} - \alpha_k^2}{\widehat{\alpha}_{2k} - \widehat{\alpha}_k^2}}}_{\to 1} \approx N(0,1)$

Пояснение
$\widehat{\alpha}_k \xrightarrow{P} \alpha_k$ (ЗБЧ)

$\widehat{\beta}_k = \overline{(X - \overline{X})^k} = \dfrac{1}{n}\sum\limits_{j=1}^{n}(X_j - X)^k$ – k-ый центральный выборочный момент
$\widehat{\beta}_2 = S_*^2$ – выборочная дисперсия
S_* – выборочное отклонение

Замечание
Выборочные моменты – моменты, посчитанные относительно эмпирического распределения
Тогда для них действуют утверждения, свойственные обычным моментам
$S_*^2 = \overline{X}^2 - (\overline{X})^2$
$\widehat{\beta}_k = Poly(\widehat{\alpha}_k, \ldots, \widehat{\alpha}_1)$
Т.к. $\widehat{\alpha}_k \xrightarrow{P} \alpha_k$, то $\widehat{\beta}_k \xrightarrow{P} \beta_k$

Отступление
Пусть ξ_n – последовательность случайных векторов и $\sqrt{n}(\xi_n - \mu) \xrightarrow{d} N(0, \Sigma)$

μ – какой-то вектор (необязательно матожидание)

1. $\xi_n \xrightarrow{P} \mu$
   Т.к. $(\xi_n - \mu)\dfrac{\sqrt{n}}{\sqrt{n}} \xrightarrow{d} 0$

2. Пусть $\phi : \mathbb{R}^m \to \mathbb{R}, \phi \in C^1(\mathbb{R})$
   $\phi(\xi_n) \approx \phi(\mu) + \nabla\phi(\mu)(\xi_n. - \mu)$
   $\phi(\xi_n) - \phi(\mu) \approx \nabla\phi(\mu)(\xi_n - \mu)$
   $\mathrm{Var}(\phi(\xi_n) - \phi(\mu)) \approx \mathrm{Var}(\nabla\phi(\mu)(\xi_n - \mu)) = \nabla\phi(\mu)\,\mathrm{Var}(\xi_n)(\nabla\phi(\mu))^T$
   Тогда $\sqrt{n}(\phi(\xi_n) - \phi(\mu)) \approx \sqrt{n}\nabla\phi(\mu)(\xi_n - \mu) \to N(0, \nabla\phi(\mu)\Sigma(\nabla\phi(\mu))^T)$

Теорема
Пусть $\xi_n = (\overline{X}, \ldots, \overline{X}^k)$
(Многоперная ЦПТ $\Rightarrow \sqrt{(}\xi_n - \alpha) \xrightarrow{d} N(0, \sigma), \alpha = (\alpha_1, \ldots, \alpha_k), \Sigma = \mathrm{Var}(X_1, \ldots, X_1^k))$
$\phi : \mathbb{R}^k \to R, \phi \in C^1(\mathbb{R})$
$\sigma^2 = \nabla\phi(\alpha)\Sigma(\nabla\phi(\alpha))^T > 0$
Тогда $\sqrt{n}\dfrac{\phi(\xi_n) - \phi(\alpha)}{\sigma} \xrightarrow{d} N(0, 1)$
Кроме того, $\sigma = \sigma(\alpha) \in C^1(\mathbb{R}) \Rightarrow \sqrt{n}\dfrac{\phi(\xi_n) - \phi(\alpha)}{\sigma(\xi_n)} \xrightarrow{d} N(0, 1)$
Упражнение
$\sqrt{n}\dfrac{S_*^2 - \sigma^2}{\sqrt{\widehat{\beta}_4 - S_*^4}} \approx N(0, 1)$
$ES_*^2 = \dfrac{n-1}{n}\sigma^2$
$S^2 := \dfrac{n}{n-1}S_*^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X - \overline{X})^2$ – исправленная (несмещенная) дисперсия
Коэффициент асимметрии
$\dfrac{E(X - EX)^3}{\sigma^3}$
Тогда $\dfrac{\widehat{\beta}_3}{S_*^3}$ – выборочный коэффициент асимметрии
Коэффициент эксцесса
$\dfrac{E(X - EX)^4}{\sigma^4} - 3 \mapsto \dfrac{\widehat{\beta}_4}{\sigma_*^4} - 3$
Ковариация

$\operatorname{Cov}(X,Y) = EXY - EXEY \mapsto S_{*XY} = \frac{1}{n}\sum_j (X_j - \overline{X})(Y_j - \widetilde{Y}) = \frac{1}{n}\sum_j X_i Y_i - \overline{XY}$

Корреляция

$\rho = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var} X \operatorname{Var} Y}} \mapsto \rho_n = \frac{S_{*XY}}{S_{*X}S_{*Y}}$

4 Порядковые статистики

Определение

Вариационный ряд – выборка $X_{(1)} \leq \ldots \leq X_{(m)}$

Определение

$X_{(k)}$ – k-ая порядковая статистика

Напоминание

Квантиль порядка α – $q_\alpha : P(X \geq q_\alpha) \geq 1 - \alpha, P(X \leq q_\alpha) \geq \alpha$

Едди F – строго монотонная, то $F(q_\alpha) = \alpha \Leftrightarrow q_\alpha = F^{-1}(\alpha)$

Определение

Выборочный квантиль порядка $0 = \min X_i$

Выборочный квантиль порядка $1 = \max X_i$

Выборочный квантиль порядка $\alpha \in (0,1)$:

$\exists\, 0 \leq k \leq n-1 : \frac{k}{n} \leq \alpha < \frac{k+1}{n}$

Тогда $X_{(k)}$ – искомый квантиль

$\alpha = \frac{1}{4}$ – первый (нижний) квартиль

$\alpha = \frac{1}{2}$ – второй квартиль, выборочная медиана

$\alpha = \frac{3}{4}$ – третий (верхний) квартиль

$\alpha = 1$ – четвертный квартиль / максимум

$n = 2m \Rightarrow med(X) = \frac{X_{(m)} + X_{(m+1)}}{2}$

$n = 2m+1 \Rightarrow \operatorname{med}(X) = X_{(m+1)}$

$P(X_{(k)} \leq t) = P(\mu_n(t) \geq k) = B(F(x), k, n-k+1)$

Пусть $p(t)$ – теоретическая плотность т.е. $p = F'$

$P(X_{(k)} \leq t)'_t = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} F^{k-1}(t)(1-F(t))^{n-k} p(t)$ – плотность k-ой

порядковой статистики
$g(x_1, x_2) = \dfrac{n!}{(k-1)!(r-k-1)!(n-r)!} F^{k-1}(x_1)(F(x_2) - F(x_1))^{r-k-1}(1 - F(x_2))^{n-r} p(x_1)p(x)2$ – совместная плотность вектора $(x_{(k)}, x_{(r)}), k < r$
$g(x_1, \dots, x_n) = n!p(x_1)\dots p(x_n)$ – плотность для вектора всех статистик $(X_{(1)}, \dots, X_{(n)})$
Определение
Средний член вариационного ряда – $X_{(k(n))}, \dfrac{k(n)}{n} \to \text{const} \in (0,1)$
Крайний член варианционного ряда – $X_{(r)}, r$ – ограничено по n или X_{n+1-s}, s – ограничено
Теорема (об асимптотике среднего члена вариационного ряда)
Пусть $0 < \alpha < 1, p$ – теоретическая плотность, q_α – теоретическая квантиль порядка α
$p \in C^1(U(q_\alpha))$
Тогда $\sqrt{n}p(q_\alpha)\dfrac{X_{(\lfloor n\alpha \rfloor)} - q_\alpha}{\sqrt{\alpha(1-\alpha)}} \xrightarrow{d} N(0,1)$
Доказательство
Комментарии к доказательству в лекции 3, 0:55

1. $k := \lfloor n\alpha \rfloor$. Выпишем плотность $X_{(k)}$
2. Напишем плотность преобразования над $X_{(k)}$

Теорема (об асимптотике крайнего члена вариационного ряда)
Пусть r, s – фиксированные, p – плотность
Тогда $nF(X_{(r)}) \xrightarrow{d} \Gamma(r, 1), nF(X_{(n+1-s)}) \xrightarrow{d} \Gamma(s, 1)$ – независимые

5 Точечное оценивание параметров

5.1 Постановка задачи точечного оценивания параметрова

Пусть $X_1, \dots, X_n \sim F_\theta, \theta \in \Theta \subset \mathbb{R}^d$ – параметр
В классической постановке θ – фиксированный неизвестный вектор
Цель: оценить θ в виде функции $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ от выборки
Замечание

1. Функции от выборки принято называть статистиками

2. Байесовская постановка: θ – случайная величина из известного априорного распределения

Определение (Состоятельность)
$\hat{\theta}$ – состоятельная оценка $\theta \Leftrightarrow \hat{\theta} \xrightarrow{p} \theta \Leftrightarrow P(\|\hat{\theta}-\theta\| > \varepsilon) \xrightarrow[n\to\infty]{} 0$
Определение (несмещенность)
$bias(\hat{\theta}) = E\hat{\theta} - \theta$ – смещение
$bias(\hat{\theta}) = 0 \Leftrightarrow \hat{\theta}\ \forall n$ – несмещенная
$bias(\hat{\theta}) = 0 \Leftrightarrow \hat{\theta}$ при $n \to \infty$ – асимптотическая несмещенная
Определение (асимптотическая нормальность)
$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{n} N(0, \Sigma_\theta)$
Определение (эффективность/оптимальность)
$\hat{\theta}_1$ – эффективнее $\hat{\theta}_2 \Leftrightarrow MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$, где $MSE(\hat{\theta}) = E\|\hat{\theta}-\theta\|^2 = E(\hat{\theta}-\theta)^T(\hat{\theta}-\theta)$
$MSE(\hat{\theta}) = \mathrm{tr}(\mathrm{Var}\,\hat{\theta}) + \|bias(\hat{\theta})\|^2$
Доказательство
$E(\hat{\theta}-\theta)^T(\hat{\theta}-\theta) = E(\hat{\theta}-E\hat{\theta}+E\hat{\theta}-\theta)^T(\hat{\theta}-E\hat{\theta}+E\hat{\theta}-\theta) = E\underbrace{(\hat{\theta}-E\hat{\theta})^T(\hat{\theta}-E\hat{\theta})}_{\mathrm{Var}\,\hat{\theta}} + \|bias(\hat{\theta})\|^2$

Метод моментов
Рассмотрим $g_1, \ldots, g_d : \exists\, Eg_i(X_1) = m_j(\theta_1, \ldots, \theta_d) \xrightarrow{\text{выборочные аналоги}} \overline{g_i(X_1)} = m_i(\hat{\theta}_1, \ldots, \hat{\theta}_d)$ – получили уравнения от θ
Решая уравнения, получаем оценки
Часто берут $g_i(x) = x^i$ – отсюда метод моментов (но можно брать и другие функции)

1. Асимптотическая нормальность $\Rightarrow$ состоятельность
   Доказательство
   $\hat{\theta}-\theta = \dfrac{1}{\sqrt{n}}\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{p} 0$

2. Асимптотическая нормальность $\Rightarrow bias(\hat{\theta}) \to 0$
   Доказательство
   Пусть $d = 2$
   $P(|\theta-\hat{\theta}| > \varepsilon) = P(\dfrac{\sqrt{n}|\theta-E\hat{\theta}|}{\sigma} > \dfrac{\sqrt{n}\varepsilon}{\sigma}) = 1 - P(\dfrac{\sqrt{n}|\theta-E\hat{\theta}|}{\sigma} < \dfrac{\sqrt{n}\varepsilon}{\sigma}) \approx 2(1-\Phi(\dfrac{\sqrt{n}\varepsilon}{\sigma})) = 2(1-\Phi(\dfrac{\sqrt{n}\varepsilon}{\sigma})) \to 0$

3. Состоятельность $\Rightarrow bias(\hat{\theta}) \to 0$
   Доказательство
   Следует из УЗБЧ
   $\overline{X} \xrightarrow{a.s.} \mu \Rightarrow E\overline{X} \to \mu$

4. Пусть $d = 1, bias\hat{\theta} \to 0, \operatorname{Var}\widetilde{\theta} \to 0 \Rightarrow \hat{\theta}$ – состоятельная

Замечание

1. Если $(\overline{g_1(X)}, \dots, \overline{g_d(X)})$ – состоятельная оценка для $(\dots, \overline{Eg_i(X)}m_i, \dots)$, $\alpha_1, \dots, \alpha_d$ – непрерывные от $\overline{g_1(X)}, \dots, \overline{g_d(X)}$, то они состоятельные

2. Если $(\overline{g_1(X)}, \dots, \overline{g_d(X)})$ – асимптотически нормальные и $g_1, \dots, g_d$ – гладкие, то каждая оценка асимптотически нормальная

6 Метод максимального правдоподобия

$pmf : p(x, \theta) = p(x|\theta)$
$pdf : p(x, \theta) = p(x|\theta)$
Все это будем называть плотностью

$X_1, \dots, X_n \sim p(X|\theta)$
$L(X|\theta) = \prod\limits_i p(X_i|\theta)$ – функция правдоподобия
$\hat{\theta}_* = \operatorname{argmax} L(X|\theta_i)$

Предположим, что $\theta \in B$ – откр., $\theta_1 \neq \theta_2 \Rightarrow L(X, \theta_1) \neq L(X, \theta_2)$ Алгоритм

1. Рассмотрим $\ln L(X, \theta)$

2. Приравняем производную к нулю

3. Найдем максимум

Пример
$Poly(1, p), p = (p_1, \dots, p_m)$
$\nu_1, \dots, \nu_m$ – количество наблюдений типа $1, \dots, m$
$L(X, p) = p_1^{\nu_1} \cdot \dots \cdot p_m^{\nu_m}$

$$\ln L(X,p) = \sum_{j=1}^{m-1} \nu_j \ln p_j + \nu_m \ln(1 - p_1 - \ldots - p_{m-1})$$

$$\frac{\partial \ln L}{\partial \ln p_j} = \frac{\nu_j}{p_j} - \frac{\nu_m}{1 - p_1 - \ldots - p_{m-1}}$$

Суммируем уравнения: $\nu_j(1 - \hat{p}_1 - \ldots - \hat{p}_{n-1}) = \hat{p}_j \nu_m$

$\hat{p}_m(n - \nu_m) = \nu_m(1 - \hat{p}_m)$

$\hat{p}_m = \dfrac{\nu_m}{n}$

Аналогично $\hat{p}_j = \dfrac{\nu_j}{n}$

Определение (информация Фишера)

Для $d = 1$

$L(X,\theta) = \prod p(X_j,\theta)$

$\ln L(X,\theta) = \sum \ln p(X_j,\theta)$

$V(X,\theta) = \dfrac{\partial \ln L}{\partial \theta} = \sum \dfrac{\partial \ln p}{\partial \theta}$ – вклад выборки

Пусть $\theta \in$ Н – открыто

$\theta_1 \neq \theta_2 \Rightarrow p(X,\theta_1) = p(X,\theta_2)$

Регулярность:

1. $\dfrac{\partial}{\partial \theta} \displaystyle\int_X T(X) L(X,\theta) \int_X \frac{\partial}{\partial \theta} L(X,\theta) T(X) \,\mathrm{d}\, x$
   Необходимое условие
   $\operatorname{supp} P_x$ нне зависит от θ
   $U[0,\theta] : \displaystyle\int_0^\theta \frac{1}{\theta} \,\mathrm{d}\, t = 1$
   $(\displaystyle\int_0^\theta \frac{1}{\theta} \,\mathrm{d}\, t)' = (\frac{1}{\theta} \int_0^\theta \mathrm{d}\, t)' = -\frac{1}{\theta^2} \int_0^\theta \mathrm{d}\, t + \frac{1}{\theta} \neq \int_0^\theta (\frac{1}{\theta})' \,\mathrm{d}\, t$

2. $EV^2(X,\theta) < \infty$
   $\displaystyle\int_X L(X,\theta) \,\mathrm{d}\, X = 1$
   $\displaystyle\int_X \frac{\partial L}{\partial \theta} \,\mathrm{d}\, X = \int_X \frac{\frac{L}{\theta}}{L} L \,\mathrm{d}\, X = \int_X V(X,\theta) L(X,\theta) \,\mathrm{d}\, X = EV(X,\theta) = 0$

$I(\theta) = VarV(X,\theta) = EV^2(X,\theta)$ – информация Фишера всей выборки

$$V(X,\theta) = \sum_j \frac{\partial \ln \hat{p}(X_j,\theta)}{\partial \theta} = \operatorname{Var} V(X,\theta) = n \qquad \underbrace{\operatorname{Var} \frac{\partial \ln p(X_j,\theta)}{\partial \theta}}_{i(\theta) \text{ – информация Фишера набора}}$$

$i(\theta) = E(\dfrac{\partial \ln p(X_j, \theta)}{\partial \theta})^2 = -E\dfrac{\partial\partial \ln p}{\partial \theta^2}$

$\dfrac{\partial}{\partial \theta}\int_{\mathbb{R}} \dfrac{\partial \ln p(x,\theta)}{\partial \theta} p(x,\theta)\,\mathrm{d}\,x = \int_{\mathbb{R}} \dfrac{\partial\partial \ln p(x,\theta)}{\partial \theta^2} p(x,\theta) + \int_{\mathbb{R}} \dfrac{\partial L}{\partial \theta}\dfrac{\partial p}{\partial \theta}\,\mathrm{d}\,x = E\dfrac{\partial\partial \ln p}{\partial \theta^2} +$

$\underbrace{E(\dfrac{\partial \ln p}{\partial \theta})^2}_{i(\theta)} = 0$

Для произвольного d

$i(\theta) = -(E\dfrac{\partial\partial \ln p(X_1,\theta)}{\partial \theta_i \partial \theta_j})_{1 \le i,j \le d}$ – информационная матрица для 1 набора

$I(\theta) = n \cdot i(\theta)$

Рассмотрим $N(\theta_1, \theta_2)$

$p(x, \theta_1, \theta_2) = \dfrac{1}{\sqrt{2\pi\theta_2}} \exp(-\dfrac{(x-\theta_1)^2}{2\theta_2})$

$\ln p(x, \theta_1, \theta_2) = -\dfrac{1}{2}\ln 2\pi - \dfrac{1}{2}\ln \theta_2 - \dfrac{(x-\theta_1)^2}{2\theta_2}$

$\dfrac{\partial \ln p}{\partial \theta_1} = \dfrac{(x-\theta_1)}{\theta_2}$

$\dfrac{\partial \ln p}{\partial \theta_2} = -\dfrac{1}{2\theta_2} + \dfrac{(x-\theta_1)^2}{2\theta_2^2}$

$\dfrac{\partial^2 f}{\partial \theta_1^2} = -\dfrac{1}{2\theta_2}$

$\dfrac{\partial^2 f}{\partial \theta_2^2} = \dfrac{1}{2\theta_2^2} - \dfrac{(x-\theta_1)^2}{\theta_2^3}$

$\dfrac{\partial^2 f}{\partial \theta_1 \theta_2} = -\dfrac{(x-\theta_1)}{\theta_2^2}$

$i(\theta) = \begin{pmatrix} \dfrac{1}{\theta_2} & 0 \\ 0 & \dfrac{1}{2\theta_2^2} \end{pmatrix}$

По неравенству Рао-Крамера $\operatorname{Var}\hat{\theta}_1 \ge \dfrac{\theta_2}{n}$

Теорема (неравенство Рао-Крамера)

Пусть модель регулярна, $d = 1$

$\tau(\theta)$ – оцениваемая функция, $\tau \in C^1, \tau(\theta) \equiv \theta$

$\hat{\tau}(\theta)$ – оценка несмещенная, т.е. $E(\hat{\tau}(\theta)) = \tau(\theta)$

Тогда $\operatorname{Var}\hat{\tau}(\theta) \ge \dfrac{(\tau'(\theta))^2}{n \cdot i(\theta)}$

Доказательство
$\tau'(\theta) = \int \hat{\tau}(\theta)\frac{\partial L}{\partial \theta}\,\mathrm{d}\,x = \frac{\int \tau(\theta)V(X,\theta)L(X,\theta)\,\mathrm{d}\,x}{E\hat{\tau}(\theta)V(x,\theta)} - EV(X,\theta)E\hat{\tau}(\theta) = \operatorname{Cov}(V(X,\theta),\hat{\tau}(\theta))$
$\operatorname{Cov}^2(V(X,\theta),\hat{\tau}(\theta)) \le \operatorname{Var} V(X,\theta)\operatorname{Var}\hat{\tau}(\theta)$
Замечание

1. $E\hat{\tau}(\theta) - \tau(\theta) = bias(\theta) \neq 0$
   $E\hat{\tau}(\theta) = \tau(\theta) + bias(\theta)$
   $\operatorname{Var}\hat{\tau}(\theta) \ge \dfrac{(\tau'(\theta) + bias'(\theta))^2}{ni(\theta)}$
   $MSE(\tau(\theta)) \ge \dfrac{(\tau'(\theta) + bias'(\theta))^2}{ni(\theta)} + bias^2(\theta)$
2. $\operatorname{Cov}^2(V(X,\theta),\hat{\tau}(\theta)) = \operatorname{Var} V(X,\theta)\operatorname{Var}\hat{\tau}(\theta)$, если $\hat{\tau}(\theta) = \alpha(\theta)V(X,\theta) + \tau(\theta)$

Теорема (неравенство Рао-Крамера)
Пусть модель регулярна, $d > 1$
$\tau(\theta) : \mathbb{R}^d \to \mathbb{R}, \tau \in C^1, \tau(\theta) \equiv \theta$
$\hat{\tau}(\theta)$ – оценка несмещенная, т.е. $E(\hat{\tau}(\theta)) = \tau(\theta)$
Тогда $\operatorname{Var}\hat{\tau}(\theta) \ge \dfrac{\nabla\tau(\theta)i^{-1}(\theta)\nabla^T\tau(\theta)}{n}$
Свойства оценки максимального правдоподобия Если существует несмещенная оптимальная оценка в регулярном случае, то она совпадает с оценкой максимального правдоподобия

6.1 Состоятельность оценки максимального правдоподобия

Пусть θ_0 – реальный параметр, $\theta \neq \theta_0$
Тогда $P_{\theta_0}(L(X,\theta_0) > L(X,\theta)) \to 1$
Доказательство
$\dfrac{L(X,\theta)}{L(X,\theta_0)} < 1$
$\dfrac{1}{n}\sum \ln \dfrac{p(X_j,\theta)}{p(X_j,\theta_0)} < 0$
По ЗБЧ $E_{\theta_0} \ln \dfrac{p(X_j,\theta)}{p(X_j,\theta_0)} \le E_{\theta_0}(\dfrac{p(X_j,\theta)}{p(X_j,\theta_0)} - 1) = \int_X p(X,\theta)\,\mathrm{d}\,X - \int p(X,\theta_0)\,\mathrm{d}\,X =$

0

Пусть $S_n = \{X : \ln L(X, \theta_0) > \ln L(X, \theta_0 - \alpha)\} \cap \{X : \ln L(X, \theta_0) > \ln L(X, \theta_0 + \alpha)\}$
$P_{\theta_0}(S_n) \to 1$
$A_n = \{X : |\hat{\theta} - \theta_0| < \alpha\}$
$B_n = \{X : \frac{\partial \ln L(X, \theta)}{\partial \theta}\Big|_{\theta = \hat{\theta}} = 0\}$
$S_n \subset A_n B_n \subset A_n$
Отсюда $P(A_n) \to 1$ – т.о. оценка состоятельная

6.2 Принцип инвариантности правдоподия

$\theta \in \text{Ⓗ} \overset{\text{биекция}}{\underset{\phi}{\leftrightarrow}} \gamma \in \Gamma$
$\gamma = \Phi(\theta)$
$\theta = \phi^{-1}$
$\sup\limits_{\theta} L(X, \phi(\theta)) = \sup\limits_{\gamma} L(X, \theta)$
Пример
Дано $Exp(\lambda)$
Тогда $\lambda e^{-\lambda x} \mapsto \frac{1}{\overline{X}}$
$\frac{1}{\lambda} e^{-\frac{x}{\lambda}} \mapsto \overline{X}$
Теорема (асимптотическая нормальность оценки максимального правдоподобия)
Пусь модель регулярна
$|\frac{\partial^3 \ln F(X, \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}| \le M(X), EM(X) < \infty$
θ_* – оценка максимального правдоподия
$\nabla \ln F(X, \theta) = 0$ – имеет единственное решение
Тогда

1. $\sqrt{n}(\theta_* - \theta) \to N(0, i^{-1}(\theta))$
2. Если $\tau(\theta)$ – оцениваемая и $\in C^1$, то
   $\sqrt{n}(\tau(\theta_*) - \tau(\theta)) \to N(0, \theta^2)$
   $\sigma^2 = \nabla \tau(\theta) i^{-1}(\theta) \nabla \tau(\theta)$

3. Если σ^2 непрерывная от σ, то $\sqrt{n}\dfrac{\tau(\sigma_*) - \tau(\theta)}{\sigma(\theta_*)} \to N(0,1)$

Доказательство 1

$V(X,\theta) = \dfrac{\partial \ln L(X,\theta)}{\partial\theta}$

θ_0 – реальный параметр

$V(X,\theta) = V(X,\theta_0) + V'_\theta(X,\theta_0)(\theta-\theta_0) + V''_\theta(X,\widetilde{\theta})\dfrac{(\theta-\theta_0)^2}{2}, \widetilde{\theta} \in (\theta,\theta_0)$

Выполним подстановку $\theta = \theta_*$

$0 = V(X,\theta_0) + V'_\theta(X,\theta_0)(\theta_*-\theta_0) + V''_\theta(X,\widetilde{\theta})\dfrac{(\theta_*-\theta_0)^2}{2}$

$V'_\theta(X,\theta_0)(\theta_*-\theta_0) = -V(X,\theta_0) - V''_\theta(X,\widetilde{\theta})\dfrac{(\theta_*-\theta_0)^2}{2}$

$\sqrt{n}V'_\theta(X,\theta_0)(\theta_*-\theta_0) = -\sqrt{n}V(X,\theta_0) - \sqrt{n}V''_\theta(X,\widetilde{\theta})\dfrac{(\theta_*-\theta_0)^2}{2}$

Заметим, что $V(X,\theta_0)$ на самом деле представляет сумму независимых одинаково распределенных величин с матожиданием 0 и дисперсией, равной информации Фишера

Тогда по ЦПТ $-\sqrt{n}V(X,\theta_0) = A_n = N(0, i(\theta))$

$$\sqrt{n}V''_\theta(X,\widetilde{\theta})\frac{(\theta_*-\theta_0)^2}{2} = n^{\frac{3}{2}} \quad \underbrace{\frac{V''_\theta(X,\widetilde{\theta})}{n}}_{\text{ограниченно по ЗБЧ}} \quad \frac{(\theta_*-\theta_0)^2}{2}$$

$V'_\theta(X,\theta_0) = n\dfrac{V'_\theta(X,\theta_0)}{n} \to -ni(\theta)$ по ЗБЧ

$\sqrt{n}V'_\theta(X,\theta_0)(\theta_*-\theta_0) \to N(0,i(\theta))$

TO BE CONTINUED

Какая-то лажа, смотри `https://t.me/c/2069367863/1/726`

Определение

Вспомним неравенство Рао-Крамера

Показатель эффективности – $\dfrac{(\tau'(\theta))^2}{n \cdot i(\theta) \operatorname{Var}\widehat{\tau}(\theta)} \in [0,1]$ для регулярных моделей

Если ПЭ $= 1$, то оценка эффективная

Определение

Пусть $\sqrt{n}(\widehat{\theta} - \theta_0) \to N(0, \underbrace{\frac{\sigma^2}{n}}_{\text{какое-то число}})$

Показатель асимптотической эффективности – $\dfrac{1}{i(\theta)\sigma^2}$

6.3 Экспоненциальное семейство распределений

Пусть модель регулярна
Если $p(x,\theta) = \exp(A(\theta)B(x) + C(\theta) + D(x))$, то распределение относится к регулярному распределению
Примеры: $N, \Gamma, Pois, Bin, NB$
Свойства
$\ln p(x,\theta) = A(\theta)B(x) + C(\theta) + D(x)$
$\dfrac{\partial \ln p(x,\theta)}{\partial\theta} = A'(\theta)B(x) + C'(x)$
$V(X,\theta) = A'(\theta)\sum B(X_i) + nC'(\theta)$
$V(X,\theta) = n(A'(\theta)\overline{B(X)} + C'(\theta))$
$\dfrac{V(X,\theta)}{n} - C'(\theta) = A'(\theta)\overline{B(X)}$
$\overline{B(X)} = \dfrac{V(X,\theta)}{nA'(\theta)} - \dfrac{C'(\theta)}{A'(\theta)}$
Тогда $\overline{B}(X)$ – оптимальная оценка для $-\dfrac{C'(\theta)}{A'(\theta)}$ по критерию оптимальности

6.4 Байесовская постановка

$X_1, \ldots, X_n \in F_\theta$
$\theta = ?..\theta \sim \Pi(\theta)$ – априорное распределение
$l(\widehat{\theta}, \theta)$ – функция потерь
$R(\widehat{\theta}, \theta) = E_{F_\theta} l(\widehat{\theta}, \theta)$ – риск
$r(\widehat{\theta}) = E_{\pi(\theta)} R(\widehat{\theta}, \theta)$ – байесовский риск
$\widehat{\theta}_B = \operatorname{argmin}_{\widehat{\theta}} r(\widehat{\theta})$
$r(\widehat{}) = E_{(\pi(\theta), F_\theta)} l(\widehat{\theta}, \theta)$
Теорема Байеса для плотностей
$p(\theta|X) = \dfrac{L(X|\theta)\pi(\theta)}{\int L(X|\theta)\pi(\theta)\,\mathrm{d}\,\theta}$
Утверждение
$\widehat{\theta}_B = \operatorname{argmin}_{\widehat{\theta}} E[l(\widehat{\theta}, \theta)|X]$
Доказательство
Пусть $\theta_* = \operatorname{argmin} \ldots$
$r(\theta_*) = EE[l(\theta_*, \theta)|X] \leq EE[l(\widehat{\theta}, \theta)|X] = r(\widehat{\theta})$
Замечание

$l(\widehat{\theta}, \theta) = (\theta - \widehat{\theta})^2 \Rightarrow \widehat{\theta_B} = E[\theta|X]$

6.5 Минимаксная оценка

$m(\hat{\theta}) = \sup\limits_{\theta} R(\hat{\theta}, \theta)$
$\hat{\theta}_{wc} = \operatorname{argmin} m(\hat{\theta})$ – минимаксная оценка
Утверждение
$r(\hat{\theta}) \leq \mu(\hat{\theta})$ – по определению
Утверждение
Если $\exists\, \pi(\theta)$ – априорное распределение : $R(\hat{\theta}_B, \theta) \equiv \text{const}$
Тогда $\hat{\theta}_{wc} = \hat{\theta}_B$
Доказательство
Пусть $\exists\, \hat{\theta} : m(\hat{\theta}) < m(\theta)$
Тогда $r(\hat{\theta}) \leq m(\hat{\theta}) < m(\hat{\theta}_B) = r(\hat{\theta}_B)$ – противоречие с определением $\hat{\theta}_B$

6.6 Интервальное оценивание

Определение (доверительный интервал)
$X_1, \ldots, X_n \sim F_\theta, \theta \in \Theta \subset \mathbb{R}$
$1 - \alpha = \gamma \in (0, 1)$ – уровень доверия
Рассмотрим $(T_l(X), T_r(X))$ – доверительный интервал уровня $\gamma = 1 - \alpha$, если $P(\theta \in (T_l(X), T_r(X))) \geq \gamma$
Классическая схема построения доверительных интервалов
Пусть $T(X, \theta) \sim G$ – не зависит от θ
Рассмотрим $P(q_1 < T(X, \theta) < q_2) = 1 - \alpha$ – доверительный интервал
Потребуем, чтобы $P(T(X, \theta) \leq q_1) = P(T(X, \theta) \geq q_2) = \dfrac{\theta}{2}$
Тогда $q_1 = q_{\frac{\alpha}{2}}, q_2 = q_{1-\frac{\alpha}{2}}, q_\bullet$ – квантили

6.7 Доверительные интервалы параметров нормального закона

Лемма о независимости линейной и квадратической статистик
$X_1, \ldots, X_n \sim N(\theta, \sigma^2) - -i.i.d.$

$T = AX; X = (X_1, \ldots, X_n)^T, A \in M_{m\times n}(\mathbb{R})$ – линейная статистика
$Q = X^T BX, B \in M_n(\mathbb{R}), B = B^T$ – квадратичная статистика
$AB = 0$
Тогда T, Q – независимые
Доказательство
$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m, 0, \ldots, 0), \lambda_i \neq 0$ – потенциально нули в конце
$\Lambda - U^T BU$ – в силу симметричности
$U = (u_1, \ldots, u_n)$ – собственные вектора, образуют ортонормированный базис
$B = U\Lambda U^T = \sum\limits_{j=1}^{m} \lambda_j u_j u_j^T$
$Q = \sum\limits_{j=1}^{m} \lambda_j (X^T u_j)(u_j^T X) = \sum\limits_{j} \lambda_j (u_j^T X)^2$
$A(\sum\limits_{j=1}^{m} \lambda_j u_j u_j^T) = 0$
$\sum\limits_{j=1}^{m} \lambda_j A u_j u_j^T = 0$
Возьмем $k \in [1, m]$
Домножим справа на u_k
$\sum\limits_{j=1}^{m} \lambda_j A u_j \underbrace{u_j^T u_k}_{\mathbb{1}(j=k)} = 0$
$\lambda_k A u_k = 0$
Отсюда $Au_k = 0$
Рассмотрим вектор $\begin{pmatrix} u^T \\ A \end{pmatrix} X$ – гауссовский вектор
Проверим, что $A_i X$ и $u_k^T X$ – независимые $\forall\, i, k$
$\mathrm{Cov}(A_i X, u_k^T X) = A_i \underbrace{\mathrm{Cov}(X, X^T)}_{\neq 0} u_k = A_i \underbrace{\mathrm{Var}\, X}_{\sigma^2 \cdot E} u_k = \sigma^2 A_i u_k = 0$
Т.е. статистики независимые
Лемма о независимости двух независимых статистик
$Q_1 = X^T B_1 X$
$Q_2 = X^T B_2 X$
$B_1 B_2 = B_2 B_1 = 0$
Тогда Q_1, Q_2 – независимые
Доказательство

Аналогично

Определение

$X_1, \ldots, X_n \sim N(0, 1)$

Тогда $\sum\limits_{k=1}^{n} X_k^2 \sim \chi^2(n)$ – распределение хи-квадрат с n степенями свободы

Лемма о распределении квадратичной статистики

$X_1, \ldots, X_n \sim N(0, 1)$

$Q = X^T B X, B = B^2$

$r := \operatorname{rg}(B)$

Тогда $Q \sim \chi^2(r), r = \operatorname{tr}(B)$

Доказательство

Заметим, что собтсвенные числа либо 0, либо 1: $\lambda \underbrace{u}_{\text{с.в.}} = Bu = B^2 u = B\lambda u = \lambda^2 u$

$B = \sum\limits_{k=1}^{r} u_k u_k^T$

$Q = \sum\limits_{k=1}^{n} (u_k^T X)^2$

$u_k^T X \sim N(\underbrace{u_k^T EX}_{0}, \underbrace{u_k^T E u_k}_{1})$

$\operatorname{Cov}(u_k^T X, u_j^T X) = 0$

Тогда $Q \sim \chi(r)$

$B = U \Lambda U^T$

$\operatorname{rg} B = \operatorname{rg} \Lambda = \operatorname{tr} \Lambda$

$B_{jj} = \lambda_j u_j u_j^T = \lambda_j$

Тогда $\operatorname{tr} B = \operatorname{tr} \Lambda = \operatorname{rg} B$

Теорема Фишера

Пусть $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$

Тогда

1. $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$

   Доказательство очевидно

2. $\frac{nS_*^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

   Доказательство

   $Y_j = \frac{X_j - \mu}{\sigma}$

$$\overline{Y} = \frac{1}{\sigma}(\overline{X} - \mu)$$
$$S_*^2(Y) = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \overline{Y})^2 = \frac{S_*^2(X)}{\sigma^2}$$
$$\overline{Y} = \frac{\sum Y_j}{n} = \underbrace{\begin{pmatrix}\frac{1}{n} & \dots & \frac{1}{n}\end{pmatrix}}_{b}\begin{pmatrix}Y_1 \\ \vdots \\ Y_n\end{pmatrix} = bY$$
$$nS_*^2(Y) = (Y - bY)^T(Y - bY) = Y^T(E - B)^T(E - B)Y, B = \begin{pmatrix}b \\ \vdots \\ b\end{pmatrix}$$
$$(I - B)^T(I - B) = I - B$$
По предыдущей лемме $Y^T(E - B)^T(E - B)Y \sim \chi^2(\mathrm{tr}(I - B))$
$$\mathrm{tr}(I - B) = \sum_{k=1}^{n}(1 - \frac{1}{n}) = n - 1$$

3. $S^2, \overline{X}$ – независимые $S_*^2, \overline{X}$ – независимые
   Доказательство
   $b(I - B) = b - b = 0$
   Тогда по лемме 1