МСПС | Введение

Схема проведения курса

1) Решение д/з учит. в итоговой оценке. Сдавать д/з – до след. семинара?
2) ~~Midterm~~ + экзамен в конце. Экзамен ~~будет~~ будет состоять из практ. и теор. части. письменной теорет.
3) ~~На Стэпике у Лусина есть много всего. Есть программа курса ≈. Чтобы получать информацию, послать мне e-mail. Д/з тоже есть на Стэпике~~ Д/з и программа на странице курса.
4) ~~Список хороших книг тоже есть на Стэпике.~~

Итак, основ. задача статистики: по данным/наблюдениям сделать выводы о распределении, которому подчиняются эти данные.

Будем изучать след. 3 парадигмы: параметрич., непараметрическая и байесовская.

1) Параметрическая: множество рассматриваемых распред. параметриз. одним или нек. параметрами:

$\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, где $\Theta \subset \mathbb{R}^k$, k – по сути, кол-во параметров.

Пример: $\mathcal{P} = \{N(a, \sigma^2)\}$, здесь $\theta = (a, \sigma^2)$; $\Theta = \mathbb{R} \times [0, +\infty)$.

Чем хороша: достаточно оценить параметр θ, чтобы решить основную задачу статистики – найти неизвест. распределение. Она так хороша, что ~~оценки можно находить с максимальной~~ так часто ∃ много методов нахождения оценок параметра θ, из которых можно выбрать наилучшую

Чем плоха: По виду гистограммы/других первичных методов анализа мы не всегда можем угадать нужную модель. Многие методы параметр. неустойчивы к выбросам

2) Непараметрическая: мы не можем параметризовать рассматриваемое семейство распределений, т.е.

$\mathcal{P} = \{P_\alpha, \alpha \in A\}$, где A – произвольное мн-во индексов.

Пример $\mathcal{P}_1 = \{$все распред., у которых нулевое математическое$\}$,

$\mathcal{P}_2 = \{$все симметричные распределения$\}$ → важный класс

Чем хороша: позволяет рассматривать широкие сем-ва распределений. Кроме того, методы непарамет. статистики, как правило, без каких-либо начальных предпол. о виде распределения, являются робастными, т.е. устойчивыми к далеко отстоящим наблюдениям, т.е. выбросам, ошибкам. (в отличие от большинства методов парам. стат.) Известно, что выбросы составляют от 0,1% до 1% членов выборки чаще всего, а есть методы типа t-критерия Стьюдента, которые очень чувствительны к ошибкам и перестают работать.

Чем плоха: Арсенал методов непарам. статистики гораздо беднее, чем у параметрической.

(3) Байесовская.

Наиболее развивающаяся за послед. 20 лет.

Здесь $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, но параметр θ полагается случайным. Т.е. θ имеет некое априорное распределение Q, которое мы выбираем, исходя из имеющихся на текущий момент данных о неизвестном распределении.

Пример $\mathcal{P} = \{N(a,1), a \in \mathbb{R}\}$, где a имеет нормальное распред. $N(0,1)$.

Чем хороша Суждения традиционной «частотной» статистики приблизительны - дается, как правило, доверит. интервал для параметра. Кроме того, мы не умеем отвечать на вопросы типа $P(0{,}3 < \theta < 0{,}9) = ?$ в завис. от выборки ответ может сильно отличаться? Байес. статистика дает точные ответы.

Чем плоха А откуда мы это априорное распред. берем? И всегда есть четкий алгоритм действий. Кажется, что «с потолка». Это основная претензия. И ответ зависит от prior. Далее см. Байес. статистика (7)

Семейства распределений

1) $\xi \sim N(a, \sigma^2)$. $E\xi = a$; $D\xi = \sigma^2$; $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}$.

Если $\xi \perp\!\!\!\perp \eta$, $\eta \sim N(\mu, \beta^2)$, то $\xi + \eta \sim N(a+\mu, \beta^2 + \sigma^2)$.

Верно правило 3 σ: $P(a - 3\sigma < \xi < a + 3\sigma) \geq 0{,}99$

2) $\xi \sim Exp(\alpha)$. Бывает со сдвигом.

$p(x) = \alpha e^{-\alpha x} \mathbb{I}(x > 0)$. $E\xi = \frac{1}{\alpha}$; $D\xi = \frac{1}{\alpha^2}$.

Если $\xi_1, \dots, \xi_n \sim Exp(\alpha)$ и неав., то $\sum_{i=1}^{n} \xi_i \sim \Gamma(\alpha, n)$.

Про многомер. норм. сказать.

3) $\xi \sim U[a,b]$ (или $R[a,b]$): $p(x) = \frac{1}{b-a} \cdot I(x \in [a,b])$.
$E\xi = \frac{b+a}{2}$; $D\xi = \frac{(b-a)^2}{12}$.

4) $\xi \sim Cauchy(\theta)$. Бывает со сдвигом. $E\xi$ не $\exists$.

Если $\xi_1 \sim Cauchy(\theta_1)$; $\xi_2 \sim Cauchy(\theta_2)$, $\xi_1 \perp\!\!\!\perp \xi_2$ (и независ.), то $\xi_1 + \xi_2 \sim Cauchy(\theta_1 + \theta_2)$

5) $\xi \sim \Gamma(\alpha, \gamma)$: $p(x) = \frac{x^{\gamma-1}\alpha^\gamma}{\Gamma(\gamma)} e^{-\alpha x} \cdot I(x>0)$. $E\xi = \frac{\gamma}{\alpha}$; $D\xi = \frac{\gamma}{\alpha^2}$.

$\xi_1 \perp\!\!\!\perp \xi_2$, $\xi_1 \sim \Gamma(\alpha, \gamma_1)$; $\xi_2 \sim \Gamma(\alpha, \gamma_2)$, то $\xi_1 + \xi_2 \sim \Gamma(\alpha, \gamma_1 + \gamma_2)$.

6) $\xi \sim B(\alpha, \beta)$, $p(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \cdot I(x \in [0,1])$, бывает сдвиг. и растяж.

- $\xi \sim Pareto(\alpha)$: $p(x) = \frac{\alpha}{x^{\alpha+1}} I(x \geq 1)$.
- $\xi \sim Weibull(\alpha, \beta)$ $F(x) = 1 - e^{-(\alpha x)^\beta} \cdot I(x>0)$.

Дискретное

- $\xi \sim Pois(\lambda)$, $\lambda > 0$. $E\xi = \lambda$; $D\xi = \lambda$. $P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$
  $\xi \perp\!\!\!\perp \eta$, $\xi \sim Pois(\lambda_1)$; $\eta \sim Pois(\lambda_2) \Rightarrow \xi + \eta \sim Pois(\lambda_1 + \lambda_2)$.
- $\xi \sim Geom(p)$, $p \in (0,1)$: $P(\xi = k) = p(1-p)^{k-1}$. $E\xi = \frac{1}{p}$; $D\xi = \frac{1-p}{p^2}$.
- $\xi \sim Bin(n, \theta)$, $\theta \in (0,1)$: $P(\xi = k) = C_n^k \theta^k (1-\theta)^{n-k}$
  $E\xi = n\theta$; $D\xi = n\theta(1-\theta)$
- $\xi \sim Bern(\theta)$ – бином. с пар. $Bin(1, \theta)$.

Скалана - виды оценок

Методы построения оценок

Опр. Статистика – (измеримая) ф-ция от выборки.

Опр. Оценка – статистика, которая имеет значение в мн-ве Θ.
(т.е. оценивает θ или ф-ции от него)

1) Метод моментов. $\theta = (\theta_1 \dots \theta_k)$

a) Пусть у нас k параметров $\Rightarrow$ выбираем k проб. ф-ций (желательно – довольно простых) $g_1(x), \dots, g_k(x)$. Пусть $m_i(\theta_1 \dots \theta_k) = E_\theta g_i(X_1)$

б) Составляем систему:

$$\begin{cases} m_1(\theta_1 \dots \theta_k) = E_\theta g_1(X_1) = \overline{g_1(X)} \\ \dots \\ m_k(\theta_1 \dots \theta_k) = E_\theta g_k(X_1) = \overline{g_k(X)} \end{cases}, \text{ где } \overline{g_i(X)} = \frac{1}{n}\sum_{j=1}^{n} g_i(X_j).$$

в) Находим решение этой системы $\hat{\theta}_1, \dots \hat{\theta}_k$.

Пример $N(a, \sigma^2)$. Есть $X_1 \dots X_n \sim N(a, \sigma^2)$ - выборка

Выберем $g_1(x) = x$; $g_2(x) = x^2$. Тогда $Eg_1(X_1) = a$; $Eg_2(X_1) = \sigma^2 + a^2 \Rightarrow$

$\Rightarrow \begin{cases} a = \overline{X} \\ \sigma^2 + a^2 = \overline{X^2} \end{cases} \Rightarrow$ решая эту сист. относительно a и σ^2, получаем

$\hat{a} = \overline{X}$ и $\hat{\sigma}^2 = \overline{X^2} - (\overline{X})^2$.

Оценка мет. моментов состоятельна, если m^{-1} (где $m = (m_1 \dots m_k)$) непрерывна.

(2) Метод макс. правдоподобия

Опр. Ф-ция правдоподобия выборки $X_1 \dots X_n$ - $f(X_1 \dots X_n, \theta) = \prod_{i=1}^{n} p_\theta(X_i)$, где $p_\theta(x)$ - обобщ. плотность X_1.

Опр. Оценка макс. правдоподобия $\hat{\theta} = \arg\max_\theta f(X, \theta)$.

Пример Найдём ОМП для выборки из $Exp(\alpha)$.

Ф-ция пр-я: $f(X, \alpha) = \prod_{i=1}^{n} \alpha e^{-\alpha X_i} = \alpha^n \cdot e^{-\alpha \sum_{i=1}^{n} X_i}$

Как ищется max? Возьмём ln, а потом продифференцируем и приравняем к 0, чтобы найти экстремум.

$L(X, \alpha) = \ln f(X, \alpha) = n \ln \alpha - \alpha \sum_{i=1}^{n} X_i$.

$\frac{\partial}{\partial \alpha} L(X, \alpha) = \frac{n}{\alpha} - \sum_{i=1}^{n} X_i = 0 \Rightarrow \hat{\alpha} = \overline{X}$ ✓

Теорема Если корень ур-ния пр-я единственный, то ОМП асимпт. нормальна и ас. эффективна при нек. условиях на распределение.

(3) Другие методы.

Опр. z_α - квантиль ф.р. $F(x)$, если $z_\alpha = \inf\{x : F(x) \geq \alpha\}$.

Опр. Вариац. ряд выборки $X_{(1)} \leq \dots \leq X_{(n)}$ - знач. выборки поставим по порядку (на всех ω).

Опр Выборочная квантиль ур. α: $\hat{z}_\alpha = X_{([\alpha n])}$.

Теорема (о выборочной квантили) Пусть плотность дифференцируема, тогда $\sqrt{n}(\hat{z}_\alpha - z_\alpha) \xrightarrow{d} N\left(0, \frac{\alpha(1-\alpha)}{f^2(z_\alpha)}\right)$.

Можно ещё рассматривать ф-ли от $F(x)$ (типа E_ξ, D_ξ) и подставлять туда $F_n(x)$ - эмпирич. ф-ция распред. (т.н. метод подстановки)

Пример $p(x) = \frac{1}{1 + (x-\theta)^2} \cdot \frac{1}{\pi}$. Тогда оц. для θ служит выборочная медиана, т.е. $X_{([\frac{n}{2}])}$.

Бутстрепное исправление смещенности

Виды оценок

Опр. $\hat\Theta$ – несмещ. для $\tau(\theta)$, если $\forall \theta \in \Theta \quad E_\theta \hat\Theta = \tau(\theta)$.

Пример $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ всегда несмещ. оценивает $E_\theta X_1$ (если оно есть)

Опр. Если $\forall \theta \in \Theta \quad \hat\Theta_n \xrightarrow{P_\theta} \theta$, то $\hat\Theta_n$ наз. состоят. оц. θ. Борьба за вероятность стремление!

Пример Выборочная медиана $\hat\mu = X_{(\lfloor \frac{n}{2} \rfloor)} \xrightarrow{P_\theta} z_{1/2}$, где $F(z_{1/2}) = \frac{1}{2}$ (для непр. распред) (или, точнее, $z_{1/2} = \inf(x : F(x) \geq \frac{1}{2})$)

Опр. Если $\forall \theta \in \Theta \quad \sqrt{n}(\hat\Theta_n - \theta) \xrightarrow{d_\theta} N(a(\theta), \sigma^2(\theta))$, то $\hat\Theta_n$ наз. асимпт. нормальной. $\sigma^2(\theta)$ наз. асимпт. дисперсией, если эффективностью $\hat\Theta_n$, $a(\theta)$ (как правило, $=0$) – асимпт. смещением. Пример: $\overline{X}$, по ЦПТ следует

Довериt. интервалы

Опр. Пара статистик $(T_1(X), T_2(X))$ (выборка) наз. (неасимпт.) доверит. интервалом для пар-ра θ, если $\forall \theta \in \Theta \quad P_\theta(T_1(X) < \theta < T_2(X)) \geq 1-\alpha$. Уровня доверия $1-\alpha$

Опр. — // — асимпт. дов. интервалом — // —, если $\forall \theta \in \Theta \quad P_\theta(T_{1n}(X) < \theta < T_{2n}(X)) \to 1-\alpha, \ n \to \infty$

Пример Построим ас. дов. интервал для пар-ра α в модели $Exp(\alpha)$.

Строятся они по общей схеме – применение ЦПТ, если есть второй момент, и теоремы о выборочных квантилях – иначе.

Здесь $EX = \frac{1}{\alpha}$; $DX = \frac{1}{\alpha^2}$.

Тогда по ЦПТ $\frac{\sum X_i - n\cdot\frac{1}{\alpha}}{\sqrt{n\frac{1}{\alpha^2}}} \xrightarrow{d} N(0,1)$, т.е. асимпт. дов. интервал такой:

$$\left(z_\gamma < \frac{\sum X_i - n\cdot\frac{1}{\alpha}}{\sqrt{n\cdot\frac{1}{\alpha^2}}} < z_{1-\gamma}\right) = \left(z_\gamma < \sqrt{n}\cdot\alpha\cdot\overline{X} - \sqrt{n} < z_{1-\gamma}\right) = \left(\frac{z_\gamma}{\sqrt{n}} + 1 < \alpha\overline{X} < \frac{z_{1-\gamma}}{\sqrt{n}} + 1\right)$$

$$= \left(\frac{z_\gamma + \sqrt{n}}{\sqrt{n}\,\overline{X}} < \alpha < \frac{z_{1-\gamma} + \sqrt{n}}{\sqrt{n}\,\overline{X}}\right), \text{ где } z_\gamma \text{ и } z_{1-\gamma} \text{ – квантили уровня } \gamma \text{ и } 1-\gamma \text{ у } N(0,1)$$

Часто мы дисперсию заменяем на её состоятельную оценку, так можно делать по т. Слуцкого.

Отличие байесов. статистики от традиционной: в традиционн. мы оцениваем параметр, в байесовской мы находим его распределение.

Т.е. результатом действий трад. статистики не всегда является число, байесовская - вероятностная мера.

Плюсы
1) Заранее понятно, что надо делать (есть алгоритм)
2) Для сложных (сильно параметризованных) моделей гораздо лучше работает
3) Интуитив. ясна: позволяет отвечать на вопросы типа $P(\theta>0|y)$, где y - наши данные. При проверке гипотез $P(H_0|y) = 1 - P(H_1|y)$, что неверно для традиц. статистики.

Минусы
1) Использование априор. распред. для пар-ра θ. Непонятно, с какого потолка оно берется.
2) Разные люди могут прийти к разным выводам на одних и тех же данных.
3) Большие вычислительные сложности

В общем, можно (и нужно) исп. и те, и другие методы для получения ясной картины о данных.

Итак, байесовский анализ состоит из 3 шагов:

1) Установл. совместного распред. $p(y, \theta) = p(y|\theta) \cdot p(\theta)$

$p(y|\theta)$ - распред. выборки, т.е. правдоподобие; $p(\theta)$ - априорное распред., которое мы выбираем

2) Поиск услов. распред. (наш результат!)

$$p(\theta|y) = \frac{p(y,\theta)}{p(y)} = \frac{p(y|\theta)\cdot p(\theta)}{\int p(y|\theta)\cdot p(\theta)\, d\theta}$$

($p(y)$ - вообще говоря, нормировочная константа :)

Байес. оценка $\hat{\theta} = \mathbb{E}_{(4)} \int \theta\, p(\theta|y)\, d\theta$

"Байесовская теорема" $p(\theta|y)$ содержит всю инф. из выборки + априор. распред

3) Проверка модели (студенты чистой байес. статистики обходят этот шаг)

Есть 3 величины, кот. нам очень интересны:

1) $p(y) = \int p(y, \theta)\, d\theta$ "prior predictive" (априор. предсказание? прочность?)

2) Пусть $\theta = (\theta_1 \dots \theta_m)$.
Нам будет интересна $p(\theta_i|y) = \int p(\theta_i|\hat{\theta}_i, y)\cdot p(\hat{\theta}_i|y)\, d\hat{\theta}_i \left(= \int p(\theta_i, \hat{\theta}_i|y)\, d\hat{\theta}_i\right)$
где $\hat{\theta}_i = \theta \setminus \theta_i$ - вектор $(\theta_1, \theta_2 \dots \theta_{i-1}, \theta_{i+1} \dots \theta_m)$.

3) Апостериорный прогноз
Пусть $\tilde{y}$ - будущие данные, связ с нашим экспериментом (т.е. будущие наблюдения)

Тогда $p(\tilde{y}|y) = \int p(\tilde{y}|\theta, y)\cdot p(\theta|y)\, d\theta = \int p(\tilde{y}|\theta)\cdot p(\theta|y)\, d\theta$,

поскольку при выбранном θ $\tilde{y}$ и y независимы (также понятно, что $p(\tilde{y}, y)$ зависима, потому что и y и $\tilde{y}$ одной моделью сделаны)