ВШБ Бизнес-информатика: ТВиМС 2025.
Лист задач для самостоятельного решения #11 (internal).
Точечные оценки. Метод моментов. Интервальные оценки.

Основные формулы

Свойства точечных оценок

- Несмещенность: $E[\hat{\theta}] = \theta$
- Смещение: $\mathrm{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$
- Среднеквадратичная ошибка: $\mathrm{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = \mathrm{Var}(\hat{\theta}) + [\mathrm{Bias}(\hat{\theta})]^2$

Доверительные интервалы

- Среднее μ, дисперсия известна:

$$\mu \in \left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

- Среднее μ, дисперсия неизвестна:

$$\mu \in \left(\bar{x} - t_{(n-1,\alpha/2)}\frac{s}{\sqrt{n}},\ \bar{x} + t_{(n-1,\alpha/2)}\frac{s}{\sqrt{n}}\right)$$

- Доля p:

$$p \in \left(\tilde{p} - z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}},\ \tilde{p} + z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}\right)$$

- Разность долей $p_1 - p_2$:

$$p_1 - p_2 \in \left(\tilde{p}_1 - \tilde{p}_2 - z_{\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{m}},\ \tilde{p}_1 - \tilde{p}_2 + z_{\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{m}}\right)$$

- Разность средних $\mu_X - \mu_Y$, дисперсии σ_X^2, σ_Y^2 известны:

$$\mu_X - \mu_Y \in \left(\bar{x} - \bar{y} - z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}},\ \bar{x} - \bar{y} + z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right)$$

- Разность средних $\mu_X - \mu_Y$, дисперсии неизвестны, но предполагаются равными:

$$\mu_X - \mu_Y \in \left(\bar{x} - \bar{y} - t_{(n+m-2,\alpha/2)} s_p\sqrt{\frac{1}{n} + \frac{1}{m}},\ \bar{x} - \bar{y} + t_{(n+m-2,\alpha/2)} s_p\sqrt{\frac{1}{n} + \frac{1}{m}}\right)$$

  где $s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}$ — объединённая выборочная дисперсия.

1. * Независимые случайные величины X_1, X_2 и X_3 имеют одинаковое матожидание μ но разные стандартные отклонения $\sigma, 2\sigma$ и 3σ соответственно. В качестве оценки матожидания мы рассматриваем три варианта: $\hat{\theta}_1 = \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3$, $\hat{\theta}_2 = \frac{1}{6}X_1 + \frac{1}{5}X_2 + \frac{1}{4}X_3$, $\hat{\theta}_3 = \frac{1}{6}X_1 + \frac{1}{3}X_2 + \frac{1}{2}X_3$. Какая из этих оценок лучше? Указание - проверить несмещенность, у несмещенных сравнить дисперсии.

2. * Пусть X_1, X_2 — случайная выборка из распределения с математическим ожиданием μ и дисперсией σ^2.

   Рассмотрим следующую оценку дисперсии σ^2:

   $$\hat{\theta} = c(X_1 - X_2)^2.$$

   Найти константу c такую, что $\hat{\theta}$ является несмещенной оценкой для σ^2.

3. Случайные величины X_1 и X_2 распределены по одному закону и независимы. Среди всех несмещенных оценок матожидания вида $c_1X_1 + c_2X_2$ найти оценку с наименьшей дисперсией.

4. * Количество покупок, совершаемых клиентами в интернет-магазине за день, подчиняется распределению Пуассона. У нас есть выборка данных по количеству покупок за несколько дней, результаты записаны в таблице. Необходимо определить параметр λ по этой выборке, используя метод моментов.

Количество покупок за день x_i	0	1	2	3	4	5
Количество дней с количеством покупок x_i	10	37	38	22	12	6

5. * Найти методом моментов по выборке $x_1, x_2, \ldots, x_n$ точечную оценку параметра p биномиального распределения.

6. При условии равномерного распределения случайной величины X произведена выборка:

3	5	7	9	11	13	15	17	19	21	21	16	15	26	22	14	21	22	18	25

   Найти оценку параметров a и b по методу моментов.

7. * Тренер и ученик стреляют в цель до первого попадания каждый. Известно, что тренер попадает в цель с вероятностью в два раза большей, чем ученик. Методом максимального правдоподобия оценить вероятность попадания учеником в цель при единичном выстреле, если известно, что тренер попал со второго раза, а ученик – с пятого. При решении задачи использовать логарифмическую функцию правдоподобия.

8. * В магазине работают три кассы. Время обслуживания покупателей каждым из кассиров распределено по показательному закону. При этом первый из кассиров самый опытный – среднее время обслуживания покупателя у него в два раза меньше чем у оставшихся двух. Первый кассир обслужил очередного покупателя за минуту, второй – за две минуты, третий – за полторы. Методом наибольшего правдоподобия оценить параметр λ – интенсивность для первого кассира.

9. * Андрей и Борис независимо друг от друга играют в покер в интернете. Выигрыш каждого из них за день – это случайная величина, распределенная по нормальному закону, причем известно, что у них одинаковое матожидание выигрыша m (тысяч рублей), но разные стандартные отклонения – у Андрея 1 а у Бориса 2 тысячи рублей. За последний день они выиграли по 2 и 3 тысячи рублей соответственно. Методом наибольшего правдоподобия оценить значение параметра m.

10. Для определения среднего возраста своих клиентов крупный производитель одежды провёл случайную выборку из 50 клиентов и получил $\bar{x} = 36$. Известно, что стандартное отклонение генеральной совокупности $\sigma = 12$:

    (a) Постройте 98% доверительный интервал для среднего возраста μ всех клиентов.

    (b) Предположим, что требуется, чтобы 92% доверительный интервал был строго равен $[\bar{X} - 2, \bar{X} + 2]$. Какой размер выборки для этого потребуется?

11. * Тут во втором вопросе как раз можно напомнить о том, что когда мы берем разные выборки, мы по факту получаем разные реализации случайных величин, в том числе и статистик тоже. Поэтому и интервалы будут отличаться.

    Проведена случайная выборка из 200 студентов. 30 из них говорят, что им "очень нравится"статистика.

    (a) Вычислите долю студентов в этой выборке, которые говорят, что им "очень нравится"статистика, и затем постройте 95% доверительный интервал для истинной доли.

(b) Теперь предположим, что вы решили спросить снова, но уже других студентов того же возраста. На этот раз в выборке 40 студентов, и 16 из них говорят, что им "очень нравится"статистика. Постройте 95% доверительный интервал для истинной доли в этом случае. Подумайте, почему два одинаковых по уровню доверия интервала для "поимки"одного и того же параметра могут отличаться.

12. * Рассмотрим случайную выборку размера 20 из распределения $\mathcal{N}(\mu, \sigma^2)$. Наблюдаемые значения выборочного среднего и выборочной дисперсии равны $\bar{x} = 81.2$ и $s^2 = 26.5$. Найдите соответственно 90%, 95% и 99% доверительные интервалы для среднего генеральной совокупности μ. Отметьте и прокомментируйте, как увеличивается ширина доверительных интервалов при увеличении уровня доверия.

13. Есть опасения по поводу скорости автомобилей, проезжающих по определённому участку шоссе. Для случайной выборки из 7 автомобилей радар зафиксировал следующие скорости (в милях в час):

$$79 \quad 73 \quad 68 \quad 77 \quad 86 \quad 71 \quad 69$$

(a) Найдите выборочное среднее и выборочную дисперсию.

(b) Указав все необходимые предположения, постройте 90% доверительный интервал для средней скорости всех автомобилей, проезжающих по этому участку шоссе.

14. Рассмотрим оригинальное клиническое исследование вакцины Солка от полиомиелита, проведённое в 1954 году. Случайным образом одна группа детей получила вакцину (группа лечения), а другая группа получила плацебо (контрольная группа). Пусть p_c и p_T обозначают истинные доли заболевших полиомиелитом в контрольной группе и группе лечения соответственно. Результаты исследования представлены в таблице:

Группа	Количество детей	Количество случаев полиомиелита
Лечение	$200,745$	57
Контроль	$201,229$	199

Постройте 95% доверительный интервал для разности $(p_c - p_T)$. (Не округляйте слишком сильно, оставьте хотя бы $4-5$ знаков в дробной части). Что можно сказать об эффективности вакцины на основе построенного доверительного интервала?

15. Пусть $\bar{x}$ и $\bar{y}$ — выборочные средние двух независимых случайных выборок $X_1, \ldots, X_8$ и $Y_1, \ldots, Y_{10}$ из распределений $\mathcal{N}(\mu_X, \sigma^2)$ и $\mathcal{N}(\mu_Y, \sigma^2)$ соответственно, где общая дисперсия неизвестна. Известны собранные данные: $\bar{x} = 5$, $\bar{y} = 3$, $\sum_{i=1}^{8} x_i^2 = 215.75$, $\sum_{i=1}^{10} y_i^2 = 107.64$.

(a) Вычислите 95% доверительные интервалы для μ_X и μ_Y.

(b) Вычислите 90% доверительный интервал для $\mu_X - \mu_Y$.

(c) Получите теоретическую формулу, а потом на имеющихся данных вычислите 95% доверительный интервал для $\theta = \frac{1}{3}\mu_X + \frac{2}{3}\mu_Y$.