

ВОПРОСЫ НА ПОНИМАНИЕ

Упражнение 1 (10 баллов). Ответьте на следующие вопросы:

1. Что такое доверительный интервал?
2. Чем асимптотический доверительный интервал отличается от неасимптотического?
3. Что больше: квантиль стандартного нормального распределения уровня 0.05 или уровня 0.1? Почему?
4. Пусть $X_1, X_2, X_3 \sim \mathcal{N}(0, 1)$. Какое распределение будет у $Z = \sqrt{2}X_1/\sqrt{X_2^2 + X_3^2}$?
5. Какие плюсы и недостатки есть у бутстрэпа? Как думаете, почему этот метод так популярен?

ЗАДАЧИ

Упражнение 2 (10 баллов). В этом упражнении мы обсудим правила двух и трех сигм для произвольной случайной величины. Пусть X имеет некоторое распределение с математическим ожиданием $\mathbb{E}X = a$ и конечной дисперсией $\text{Var}(X) = \sigma^2 < \infty$. С помощью неравенства Чебышёва покажите, с какой вероятностью эта случайная величина лежит в отрезках $[a - 2\sigma, a + 2\sigma]$ и $[a - 3\sigma, a + 3\sigma]$. Сравните полученные вероятности с соответствующими вероятностями для стандартного нормального распределения $\mathcal{N}(0, 1)$. Какой вывод можно сделать?

Упражнение 3 (15 баллов). Пусть имеется реализация выборки x_1, \dots, x_n из равномерного распределения на $[0, \theta]$ с неизвестным параметром $\theta > 0$.

- (1) Постройте точный доверительный интервал для параметра θ уровня доверия $1 - \alpha$ с помощью статистики $\hat{\theta}_1 = x_{(n)} = \max\{x_1, \dots, x_n\}$. Используйте тот факт, что распределение максимума из n независимых равномерно распределённых на отрезке $[0, 1]$ случайных величин имеет функцию распределения

$$F(u) = \begin{cases} 0, & \text{если } u < 0, \\ u^n, & \text{если } u \in [0, 1], \\ 1, & \text{если } u > 1. \end{cases}$$

- (2) Постройте асимптотически точный доверительный интервал для неизвестного параметра θ уровня доверия $1 - \alpha$, используя статистику $\hat{\theta}_2 = 2\bar{x} = 2(x_1 + \dots + x_n)/n$.

Упражнение 4 (15 баллов). Пусть имеется реализация выборки x_1, \dots, x_n из равномерного распределения на $[0, \theta]$. Допустим, мы оценили θ с помощью $2\bar{x} = 2(x_1 + \dots + x_n)/n$. Затем мы генерируем новую выборку из равномерного распределения на $[0, 2\bar{x}]$ и оцениваем с ее помощью какую-то величину, которая нам интересна (то есть мы используем параметрический бутстрэп для оценки этой величины). Какую дисперсию будут иметь случайные величины из равномерного распределения на $[0, 2\bar{x}]$? Сравните ее с дисперсией равномерного распределения на $[0, \theta]$. Какой вывод можно сделать?

Упражнение 5 (25 баллов). Рассмотрим нормальное распределение $\mathcal{N}(\theta, \sigma^2)$ с неизвестными параметрами $\theta \in \mathbb{R}$ и $\sigma^2 > 0$. В этой задаче мы численно сравним следующие доверительные интервалы для параметра θ : 1) теоретический, 2) на основе параметрического бутстрэпа, 3) на основе непараметрического бутстрэпа. Для этого:

- (1) сгенерируйте неизвестный параметр θ из равномерного распределения на $[10, 20]$ и σ — из равномерного распределения на $[1, 2]$;
- (2) сгенерируйте выборку из нормального распределения $\mathcal{N}(\theta, \sigma^2)$ размера $n = 50$;
- (3) постройте доверительные интервалы 1), 2), 3) уровня доверия $1 - \alpha$;
- (4) сравните длины полученных доверительных интервалов;
- (5) проведите этот эксперимент на 10 000 выборках; с какой частотой θ попадает в полученные доверительные интервалы?

Упражнение 6 (25 баллов). В файле `Banner_small.npy` даны две выборки из распределения Бернулли. Они соответствуют кликам по двум рекламным баннерам в Интернете. Постройте доверительные интервалы для частоты кликов (то есть параметра «успеха»). Какой из изученных вариантов доверительных интервалов лучше применять в этой задаче? Как думаете, можно ли сказать, что один из баннеров лучше, чем другой? Если да, то с какой вероятностью? Проверьте свой вывод на выборке большего размера, которая записана в файл `Banner_large.npy`.