

ВОПРОСЫ НА ПОНИМАНИЕ

Перед тем как приступить к решению домашних задач, попробуйте ответить на следующие вопросы. Это простые вопросы на понимание. Ответы на них включать в домашнюю работу не нужно.

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

ЗАДАЧИ

Упражнение 1 (15 баллов). В этом упражнении мы обсудим правила двух и трех сигм для произвольной случайной величины. Пусть X имеет некоторое распределение с математическим ожиданием $\mathbb{E}X = \mu$ и конечной дисперсией $\text{Var}(X) = \sigma^2 < \infty$. С помощью неравенства Чебышёва покажите, с какой вероятностью эта случайная величина лежит в отрезках $[\mu - 2\sigma, \mu + 2\sigma]$ и $[\mu - 3\sigma, \mu + 3\sigma]$. Сравните полученные вероятности с соответствующими вероятностями для стандартного нормального распределения $\mathcal{N}(0, 1)$. Какой вывод можно сделать?

Упражнение 2 (15 баллов). Пусть имеется реализация выборки x_1, \dots, x_n из равномерного распределения на $[0, \theta]$ с неизвестным параметром $\theta > 0$. Постройте точный доверительный интервал для параметра θ уровня доверия $1 - \alpha$ с помощью статистики $\hat{\theta} = \max\{x_1, \dots, x_n\}$. Используйте тот факт, что распределение максимума из n независимых равномерно распределённых на отрезке $[0, 1]$ случайных величин имеет функцию распределения

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

Упражнение 3 (30 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 $1 - \alpha$ (положите $\alpha = 0.05$);
- (4) проведите этот эксперимент на 10 000 выборках. Какая средняя длина доверительных интервалов? С какой частотой θ попадает в полученные доверительные интервалы?