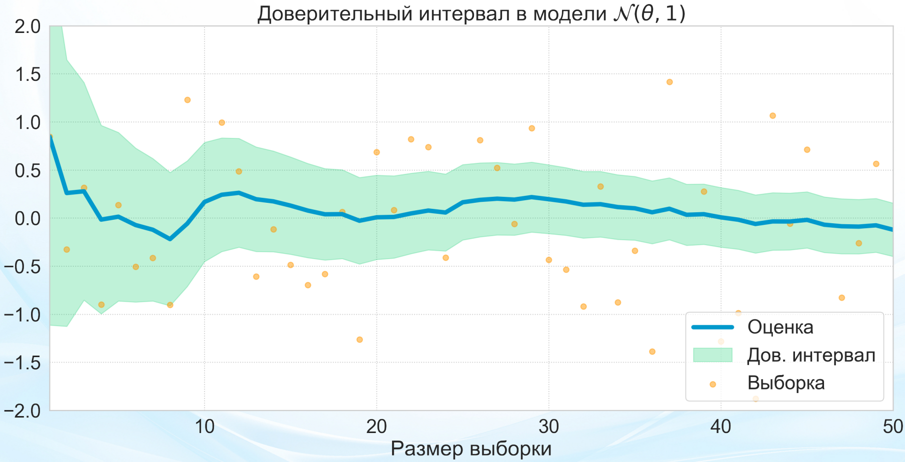
Статистика

прикладной поток

Лекция 7



Доверительный интервал в модели $\mathcal{N}(\theta, 1)$
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
10
20
30
40
50
Размер выборки
Оценка
Дов. интервал
Выборка

Распределение хи-квадрат

Обозначение: χ^2_k — хи-квадрат с k степенями свободы

- Параметр k — кол-во степеней свободы;
- Плотность

$$p(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$

- Если $\xi_1, ..., \xi_k$ — независимые $\mathcal{N}(0, 1)$, то $\xi_1^2 + ... + \xi_k^2 \sim \chi^2_k$
- Если $\eta \sim \chi^2_k$, то $\mathsf{E}\eta = k, \mathsf{D}\eta = 2k$
- $\chi^2_{k,p}$ — p-квантиль распределения χ^2_k
- `scipy.stats.chi2(df=`k`)`

Распределение Стьюдента

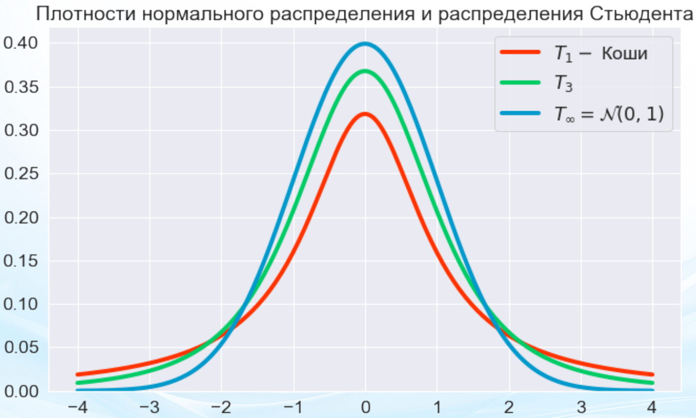
Обозначение: T_k — распределение Стьюдента с k степенями свободы

- Параметр k — кол-во степеней свободы;
- T_1 — распределение Коши
- $T_\infty = \mathcal{N}(0,1)$
- Плотность

$$p(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{\pi k}\,\Gamma(k/2)} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}$$

- Если $\xi \sim \mathcal{N}(0,1)$ и $\eta \sim \chi^2_k$ независимы, то $\zeta = \frac{\xi}{\sqrt{\eta/k}} \sim T_k$
- Если $\zeta \sim T_k$, то $\mathsf{E}\zeta = 0$ при $k > 1$
- Если $\zeta \sim T_k$, то $\mathsf{D}\zeta = \frac{k}{k-2}$ при $k > 2$
- $T_{k,p}$ — p-квантиль распределения T_k
- `scipy.stats.t(df=`k`)`

Сравнение распределений



Уильям Сили Госсет

Работал на пивоваренном заводе

Гиннеса в Дублине.

Чтобы предотвратить дальнейшее раскрытие

конфиденциальной информации, Гиннесс

запретил своим работникам публикацию

любых материалов, независимо

от содержавшейся в них информации.

Госсет выбрал себе псевдоним Student.