Третье задание по курсу БМСО

Бурнаев Е., Зайцев А., Янович Ю.

1. (2 балла) Для пары вероятностных распределений P, Q, заданных на сигма-алгебре $\mathbb{B}$, *общей дисперсией (total variation)* называется

$$\|P - Q\|_{TV} = 2 \cdot \sup_{B \in \mathbb{B}} |P(B) - P(Q)|. \tag{1}$$

Для непрерывных распределений P и Q с плотностями p и q соответственно, обозначим

$$\|P - Q\|_d = \int |p - q| d\mu.$$

Доказать, что для непрерывных распределений определения $\| \star \|_{TV}$ и $\| \star \|_d$ эквивалентны.

2. (1 балл) Вычислить расстояние общей дисперсии (total variation) (1) между двумя одномерными нормальными распределениями $\mathcal{N}(\mu_1, \sigma_1^2)$ и $\mathcal{N}(\mu_2, \sigma_2^2)$.

3. (3 балла) Показать, что если

(У1) множество параметров $\theta \in \Theta \subset \mathbb{R}$ — ограниченное открытое множество.

(У2) существует такое $\sigma > 0$, что для $\theta \in \Theta$ информация Фишера $I(\theta)$:

$$I(\theta) > \sigma > 0,$$

то для экспоненциального семейства распределений выполнено "предположение (5)" из теоремы Берштейна-фон-Мизеса.

То есть, пусть плотность распределения имеет вид:

$$f(x|\theta) = \exp(-c(\theta) + \theta t(x)),$$

и выполнены условия (У1) и (У2). Тогда для любого $\delta > 0$ существует $\varepsilon > 0$, такое что

$$P_{\theta_0}\left(\sup_{|\theta - \theta_0| \geq \delta} \frac{1}{n}(L_n(\theta) - L_n(\theta_0)) \leq -\varepsilon\right) \to 1$$

для $n \to \infty$. Здесь $L_n(\theta)$ — логарифм правдоподобия для выборки независимых одинаково распределенных случайных величин $x_1, \ldots, x_n$ размера n, θ_0 — истинное значение параметра.

4. (2 балла) Получить оценку максимального правдоподобия $\hat{\theta}_n$ и Байесовскую оценку $\theta_n^* = \int \theta p(\theta | x_1, \ldots, x_n) d\theta$ для среднего μ нормального распределения $\mathcal{N}(\mu, \sigma^2)$ и априорного распределения $\theta \sim \mathcal{N}(\mu_\Pi, \sigma_\Pi^2)$. Величины σ^2, μ_Π и σ_Π^2 известны и фиксированы.

5. (3 балла) Сравнить сходимость к нормальному распределению распределения величин $\sqrt{n}(\theta_0 - \theta_n^*)$ и $\sqrt{n}(\theta_0 - \hat{\theta}_n)$:

- Нарисовать распределения $\sqrt{n}(\theta_0 - \theta_n^*)$ и $\sqrt{n}(\theta_0 - \hat{\theta}_n)$, показать как меняются эти распределения с ростром размера выборки.
- Оценить насколько отличаются полученные распределения от нормальных любым удобным способом.