Пояснения к статье из yahoo

1 Предложение 1 (стр. 27)

Приведём пояснение к предложению 1 статьи [1]. Допустим, у нас есть случайная величина θ. Мы проводим n независимых измерений, в результате i-го измерения мы получаем значение $\theta + \varepsilon_i$, где $\varepsilon_i \sim N(0, \sigma_i^2)$. Итак, в результате измерений мы получили $a_1, \ldots, a_n$:

$$\theta + \varepsilon_i = a_i.$$

Тогда мы находимся в условиях предложения 1 из [1]. Действительно, из i-го измерения мы получили $\theta \sim N(a_i, \sigma_i^2)$, и независимость измерений здесь как раз означает условную независимость a_i при условии θ.

Апостериорное распределение θ после наших измерений есть (prior неинформативный!)

$$Ap_1(a_1 - \theta)p_2(a_2 - \theta) \ldots p_n(a_n - \theta),$$

где p_i — плотность распределения ε_i, то есть

$$p_i(x) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{x^2}{2\sigma_i^2}},$$

A — нормировочная константа. Имеем

$$Ap_1(a_1 - \theta)p_2(a_2 - \theta) \ldots p_n(a_n - \theta) = A_1 e^{-\frac{\theta^2}{2\sigma_1^2} - \ldots - \frac{\theta^2}{2\sigma_n^2}} e^{\frac{a_1}{\sigma_1^2} + \ldots + \frac{a_n}{\sigma_n^2}\theta} = \\ = A_2 e^{-\frac{(\theta - m)^2}{2\sigma^2}}$$

(нормировочные константы A, A_1, A_2 зависят от a_i). Отсюда

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \ldots + \frac{1}{\sigma_n^2},$$

$$\frac{m}{\sigma^2} = \frac{a_1}{\sigma_1^2} + \ldots + \frac{a_n}{\sigma_n^2}.$$

Отсюда получаем формулы из предложения 1 из [1]

References

[1] Agarwal, Chen, Elango. Spatio-temporal models for estimating click-through rate.