

University of Waterloo
STAT 845

Diagnostic Questions

(Wednesday, September 6, 2023)

1. (*Probability*) What is the *law of total variance*? For example, if $S_n = X_1 + X_2 + \cdots + X_n$ is the sum of n independent and identically distributed random quantities, what is the difference between $\text{Var}(S_{100})$ and $\text{Var}(S_N)$, where N is also random but with $\mathbb{E}(N) = 100$?

2. (*Mathematical Statistics*) What does *efficiency* mean in statistics? For example, given two independent observations X and Y from the same distribution with mean θ , which one,

$$\frac{2X + 8Y}{10} \quad \text{or} \quad \frac{3X + 7Y}{10},$$

is a more *efficient* way of estimating the unknown parameter θ ?

3. (*Computational Statistics*) How is the *EM algorithm* similar to and different from *coordinate ascent*?

4. (*Linear Models*) You are trying to predict Y with a certain linear function of X_1, X_2, \dots, X_d . Why is *ridge regression* useful when some of X_1, X_2, \dots, X_d are highly correlated, or when d is relatively large?

5. (*Sampling and Design*) What is the main reason why a *stratified sample* may be better than a *simple random sample* when conducting a survey, and why a *randomized block design* may be better than a *completely randomized design* when conducting an empirical experiment?