**MA 578 — Bayesian Statistics**

# Homework 1

(Due: Tuesday, 09/17/19)

1. BDA[1] problem 1.1.

2. BDA problem 1.3.

3. (Simpson's paradox) In a medical study[2] to determine the effectiveness of two different treatments, A and B, for two different types of kidney stones, "small" and "large", the researchers found the following success rates: for treatment A, 93% success on small stone cases and 73% success on large stone cases; for treatment B, 87% success on small stone cases and 69% success on large stone cases. It is assumed that 51% of kidney stone cases are due to small stones.

   (a) The experimental design of the study assigned 24% of small stone cases to treatment A and 77% of large stone cases to treatment A. If a patient is selected at random from treatment A, what is the probability that she has small stones? What is the probability of her having small stones given that she was now selected from treatment B?

   (b) What are the (marginal) success rates of treatments A and B? How do they compare to the (conditional) success rates given the conditions small and large stones? How would you explain the discrepancies in light of the probabilities computed in (a)?

   (c)$*$ [3] The discrepancy above is known as "Simpson's paradox", and in this case seems to stem from the poor experimental design. Let us investigate this further: call $p_{AS}$ and $p_{AL}$ the probabilities of a small stone case and a large stone case, respectively, being assigned to treatment A. Plot a heatmap of the ratio $\rho$ of success rates between treatments A and B as a function of $p_{AS}$ and $p_{AL}$, and identify the region with $\rho < 1$, that is, where we see a paradox[4]. Explain any patterns you see in the region, e.g., when $p_{AS} > p_{AL}$.

---

[1] Gelman *et al*'s "Bayesian Data Analysis".

[2] Charig, Webb, Payne, and Wickham, (1986) "Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy", Br Med J (Clin Res Ed) 292 (6524).

[3] The $*$ means "harder, so optional, but recommended".

[4] Can you identify the isolines for $\rho$? If possible, try to show that they are hyperbolae, and for a specific value of $\rho$ a parabola!