MA 578 — Bayesian Statistics

Homework 1

(Due: Tuesday, 09/17/19)

1. Hoff's book[1] problem 2.5.

2. Hoff's book problem 2.6.

3. (Simpson's paradox) An experiment was performed in the 80s that involves 800 subject; each subject received either treatment A or treatment B, and each subject was classified into one of the following four categories: older males, younger males, older females, and younger females. At the end of the experiment, it was determined for each subject whether the treatment that the subject had received was helpful or not. The results for each of the four categories of subjects are given in the table below.

	Helpful	**Not**
Older males		
Treatment I	120	120
Treatment II	20	10
Younger males		
Treatment I	60	20
Treatment II	40	10
Older females		
Treatment I	10	50
Treatment II	20	50
Younger females		
Treatment I	10	10
Treatment II	160	90

   (a) Show that treatment II is more helpful than treatment I within each of the four categories of subjects.

   (b) Show that if these four categories are aggregated into only the two categories, older subjects and younger subjects, then treatment I is more helpful than treatment II within each of these categories.

   (c) What are the (marginal) success rates of treatments I and II? How do they compare to the (conditional) success rates given the older subjects and younger subjects? How would you explain the discrepancies in light of the probabilities computed in (a) and (b)?

   (d)$*$ [2] 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_{1Young} and p_{1Old} the probabilities of a younger and an older subject, respectively,

[1] Hoff "A First Course in Bayesian Statistical Methods".

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

being assigned to treatment I. Plot a heatmap of the ratio ρ of success rates between treatments I and II as a function of p_{1Young} and p_{1Old}, and identify the region with $\rho < 1$, that is, where we see a paradox[3]. Explain any patterns you see in the region, e.g., when $p_{1Young} > p_{1Old}$.

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