Stat 611: Homework 1.

Moulinath Banerjee

University of Michigan

January 19, 2003

1 **Variance Stabilizing Transformations.** Let $X_1, X_2, \ldots, X_n$ be i.i.d random variables with common probability mass function/probability density function $f(x, \theta)$ with $\theta \in \Theta$. Consider the following parametric models:

$$(a) f(x, \theta) = \theta^x (1 - \theta)^{1-x} \quad x \in \{0, 1\}, \quad 0 < \theta < 1 .$$

Thus, the X_i's are i.i.d. Bernoulli(θ).

$$(b) f(x, \theta) = \frac{e^{-\theta}\,\theta^x}{x\,!} \quad x = 0, 1, 2, \ldots, \quad 0 < \theta < \infty .$$

Thus, the X_i's are i.i.d. Poisson(θ).

$$(c) f(x, \theta) = (1 - \theta)^{x-1}\,\theta, \quad x = 1, 2, \ldots, \quad 0 < \theta < 1 .$$

Thus, the X_i's are i.i.d. Geometric(θ).

$$(d) f(x, \theta) = \theta\, e^{-\theta\, x}, \quad x > 0, \quad 0 < \theta < \infty .$$

Thus, the X_i's are i.i.d. Exponential(θ).

In each of these examples, use the asymptotic distribution of the MLE along with the method of variance stabilizing transformations to construct a level $1 - \alpha$ confidence interval for θ.

2 **Information inequality.**(a) Let $X_1, X_2, \ldots, X_n$ be i.i.d Bernoulli(θ). Consider unbiased estimation of θ^2.

(i) Find the MVUE (minimum variance unbiased estimator) of $q(\theta) = \theta^2$. Does this attain the information bound?

(ii) Find the asymptotic distribution of $q(\hat{\theta}_n)$. Does this attain the information bound asymptotically?

(iii) Show that there does not exist any unbiased estimator of the odds ratio, $q(\theta) = \theta/(1-\theta)$.

(b) Let $\mathcal{F}$ denote the class of densities with mean $\theta^{-1} \equiv q(\theta)$ and variance θ^{-2} $(\theta > 0)$ that satisfy the conditions of the information inequality. Construct a family of densities $\{f(x,\theta) : \theta > 0\}$ that minimizes the Fisher information $I(\theta)$ over $\mathcal{F}$. (Hint: What happens if you apply the information inequality to $T(X) = X$?)

3 **Delta method and influence functions:** (i) Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d vectors with values in $\mathbb{R}^k$ with $E\, X_1 = \mu$ and $E(X_1^T X_1) < \infty$ so that $\Sigma = E\,(X_1 - \mu)\,(X_1 - \mu)^T$ is well-defined. Let g be a real valued function defined on $\mathbb{R}^k$ and suppose that g is continuously differentiable in a neighborhood of μ; thus, $\bigtriangledown g(\cdot)$ exists and is continuous in a neighborhood of μ. Show then that $g(X_n)$ is *asymptotically linear* at μ, i.e.

$$\sqrt{n}\,(g(\overline{X}_n) - g(\mu)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i) + o_p(1)\,,$$

for some function $\psi(x)$ (which you need to identify).

(ii) Show that in a regular paramteric model, with θ_0 denoting the true underlying value of the parameter and $\hat{\theta}_n$ denoting the MLE and q being a continuously differentiable map, $\hat{q}(theta_n)$ is asymptotically linear, i.e.

$$\sqrt{n}\,(q(\hat{\theta}_n) - q(\theta_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\theta_0}(X_i) + o_p(1)\,,$$

for some function $\psi_{\theta_0}(\cdot)$ that you need to identify. Do you recognize this function?

4 **Nonregular parametric models:** A classic example of a nonregular parametric model is the set of uniform distributions on $(0,\theta)$ where $\theta > 0$. Consider i.i.d. observations $X_1, X_2, \ldots, X_n$ from $U(0,\theta)$. It is well known (derive for yourself if not comfortable with this fact) that the MLE of θ is $X_{(n)}$. However, unlike regular parametric models, $X_{(n)}$ is not asymptotically normal. The following exercises reveal certain facts about $X_{(n)}$.

(a) Show that there exist sequences of constants $\{a_n\}$ and $\{b_n\}$, possibly depending upon θ such that $(X_{(n)} - a_n)/b_n$ converges to a limiting distribution. Identify the limit. Is $X_{(n)}$ consistent for θ?

(b) Find the UMVUE of θ. Call this T_n. Compute the mean squared errors of T_n, $X_{(n)}$ and $R_n \equiv 2\,\overline{X}_n$ as estimates of θ and comment on their relative behavior for fixed and increasing n.

(c) Use both the exact and the limit distributions of $X_{(n)}$ to construct level $1-\alpha$ confidence sets (exact and asymptotic respectively) for θ.

5 **Exchangeability and conditional independence for Bernoulli random variables.** This is the most non–standard problem in the Homework and therefore the most interesting (and not very difficult). Michael Jordan's talk on Friday afternoon where he discussed the use of exchangeability to develop internet search engines triggered the urge to set this problem. Its history dates back to the 1700's and originates according to Stigler in the work of Bayes (see Chapter 3 on Inverse Probability by Stigler – The History of Statistics, pages 122 – 131 for a detailed discussion).

Let $X_1, X_2, \ldots,$ be (a possibly infinite sequence of) random variables defined on a common probability space. Call this sequence exchangeable if for all n and for all permutations Π of $\{1, 2, \ldots, n\}$, the joint distribution of $(X_1, X_2, \ldots, X_n)$ is the same as the joint distribution of $(X_{\Pi(1)}, X_{\Pi(2)} \ldots, X_{\Pi(n)})$. Thus exchangeability amounts to the invariance of the joint distribution under permutations of the data.

In addition to the above, suppose that each X_i is a Bernoulli random variable with parameter p_i.

(a) Show that the p_i's are all equal; in other words the random variables are identically distributed.

(b) Let Θ be a random variable on $(0, 1)$ with distribution denoted by G. Conditional on Θ, generate $X_1, X_2, \ldots,$ as i.i.d. Bernoulli(θ). Show that the joint distribution of the X_i's is exchangeable. What is the common p_i for all these X_i's?

(c) Here is a much stronger fact (the converse of (b)). Suppose we have a sequence of exchangeable Bernoulli random variables $X_1, X_2, \ldots,$. Then there exists a random variable Θ assuming values between 0 and 1, such that given Θ, $X_1, X_2, \ldots$ are all i.i.d. Bernoulli(θ). Thus, infinite exchangeability amounts to conditional independence. This is essentially a version of the *De Finetti representation theorem.* This is not officially part of the homework but those of you taking the 620's/that have seen Advanced Probability should give it a shot.

(d) Consider an infinite sequence of exchangeable Bernoulli random variables, $X_1, X_2, \ldots,$. By (c), without loss of generality you can assume that these are conditionally independent given some Θ, assuming values in (0,1). Suppose that you know that the chance that each of the first n trials ends in a success is $1/n + 1$ for any n, i.e.

$$P(X_1 = 1, X_2 = 1, \ldots, X_n = 1) = \frac{1}{n+1}\,, \quad n = 1, 2, 3, \ldots\,.$$

Show that this implies that for each $1 \leq n < \infty$,

$$P(k\ out\ of\ the\ first\ n\ trials\ are\ successes) \equiv P(S_n = k) = \frac{1}{n+1}\,, \quad k = 0, 1, \ldots, n\,.$$

In the above display, $S_n = X_1 + X_2 + \ldots + X_n$.

Hint: Can you say anything about the distribution of Θ in this problem?

(e) Now consider a finite sequence of exchangeable Bernoulli random variables $X_1, X_2, \ldots, X_N$ and assume that assume that,

$$P(X_1 = 1, X_2 = 1, \ldots, X_n = 1) = \frac{1}{n+1}\,, \quad n = 1, 2, \ldots, N\,.$$

Then $S_n \equiv X_1 + X_2 + \ldots + X_n$, the number of successes in the first n trials has a discrete uniform distribution for each $n \leq N$; i.e.

$$P(k\ out\ of\ the\ first\ n\ trials\ are\ successes) \equiv P(S_n = k) = \frac{1}{n+1}\,, \quad k = 0, 1, \ldots, n\,.$$

The proof I have uses induction, but other proofs would be welcome. There is a subtle difference between scenarios (d) and (e). In (e) we only have finitely many random variables and therefore we cannot conclude that they are conditionally i.i.d. given some fixed Θ. The De Finetti representation can only be invoked for an infinite sequence. The assertion in (d) can be established using the distribution of Θ; in (e) the exchangeability hypothesis needs to be used more fundamentally. Indeed, as (e) shows, the fact that the distribution of the total number of successes S_n has a discrete uniform distribution is fundamentally a consequence of exchangeability, rather than the De Finetti representation. While the result in (d) seems to be well-known, the result in (e) does not seem to be. Thomas Richardson and myself stumbled on this fact a few years ago and the proof I'll provide in the homework solutions is mine. The fact itself is fairly basic, so we must have rediscovered the wheel, so to speak.