STA 6448 - Take home (due May 2 (5 pm ET) by email at debdeep@stat.fsu.edu)

1. Let $\{X_n\}$ be an i.i.d sequence with common continuous distribution function. Call $\{X_k\}$ a *record value* for the sequence if $X_k > X_r$ for all $1 \leq r < k$. Let I_k be the indicator function for the event that X_k is a record value.
   (a) Show that I_k and I_l are independent for $k < l$.
   (b) Show that $\mathbb{E}[I_k] = 1/k$.
   (c) Let $S_n = \sum_{k=1}^{n} \left[I_k - 1/k\right]$. Show that $\{S_n\}$ defines a martingale sequence.
2. A hypothesis may be strongly rejected by a frequentist test of significance and yet be awarded high odds by a Bayesian analysis. This is known as *Lindley's paradox*. To explain this, we conduct the following experiment. Suppose $X_i, i = 1, \ldots, n$ are drawn independently and identically distributed as $\mathrm{N}(\mu, \sigma^2)$ and μ is believed to be μ_0 with probability 1/2 and $\mu \sim \mathrm{N}(\mu_0, \sigma^2)$ with probability 1/2. Suppose the observed data $(x_1, \ldots, x_n)$ satisfy $(1/n) \sum_{i=1}^{n} x_i = \mu_0 + 1.96\sigma/\sqrt{n}$. Compare the posterior probabilities of the event $\mu = \mu_0$ for $n = 5$ and $n = 50$. Comment on the findings.
3. Assume $X_i, i = 1, \ldots, n$ are drawn from a Weibull distribution with probability density function

$$g(x; \alpha, \phi) = \left(\frac{\alpha}{\phi}\right)\left(\frac{x}{\phi}\right)^{\alpha-1} \exp\{-(x/\phi)^{\alpha}\}, \quad \alpha, \phi > 0, \quad x > 0.$$

   A hapless researcher tries the exponential distribution instead

$$f(x; \theta) = \theta^{-1} \exp\{-(x/\theta)\}, \quad \theta > 0, \quad x > 0.$$

   Find the asymptotic distribution of the maximum likelihood estimate $\hat{\theta}_n$. You may use that for $Y \sim g$, $\mathbb{E}(Y) = \phi\Gamma(1 + \alpha^{-1})$ and $\mathbb{E}(Y^2) = \phi^2\Gamma(1 + 2\alpha^{-1})$, where $\Gamma(\cdot)$ is the gamma function.
4. Consider testing equality of two continuous cumulative distribution functions F and G based on independent and identically distributed samples $X_i, i = 1, \ldots, n$ and $Y_i, i = 1, \ldots, m$ from F and G respectively. Show that

$$P(X_1 < Y_1, X_2 < Y_1) + P(Y_1 < X_1, Y_2 < X_1) = \frac{2}{3} + \frac{1}{2}\int [F(x) - G(x)]^2 d[F(x) + G(x)] \tag{1}$$

   and construct an appropriate U-statistic based on (1). (*No need to derive asymptotic variance of the statistic.*)
5. Let $X_1, \ldots, X_n$ be independent $0 - 1$ random variables with $\mathbb{E}[X_i] = p_i$ (not necessarily equal). Let $X = \sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}[X] = \sum_{i=1}^{n} p_i$ and $p = \mu/n$. Prove that
   (a) $\mathbb{P}[X > \mu + \lambda] \leq \exp\{-nH_p(p + \lambda/n)\}$ for $0 < \lambda < n - \mu$;
   (b) $\mathbb{P}[X \leq \mu - \lambda] \leq \exp\{-nH_{1-p}(1 - p + \lambda/n)\}$ for $0 < \lambda < \mu$,

   where $H_p(x) = x\log(x/p) + (1-x)\log\{(1-x)/(1-p)\}$ popularly called the *Bernoulli entropy function* of x with respect to p.

   Hint: Exponentiate through e^t, apply Markov, use independence, plug in the m.g.f of each X_i, use concavity as a function of p and minimize the final bound as a function of t, using basic calculus.