Homework Set 5 , given Fri 3/7, due Friday 3/14

Problem 5.1. Consider n i.i.d. Bernoulli observations $X_1, \ldots, X_n$, $X_i \sim \text{Ber}(p)$ where $p \in (0,1)$ is unknown. Consider two sided hypotheses
$H_0 : p = p_0$
$H_a : p \neq p_0$.
The standard test takes the form

$$\phi(X_1, \ldots, X_n) = \mathbf{1}\left\{\left|\hat{Z}_n\right| > z^*_{\alpha/2}\right\},\ \hat{Z}_n = \frac{\sqrt{n}\,(\hat{p}_n - p_0)}{\sqrt{p_0(1-p_0)}}.$$

It was shown that this is an asymptotic α-test.
a) Show that the test is consistent, i.e the power functions tends to one for every parameter p in the alternative:

$$\Pr_p(\phi = 1) \to 1 \text{ for all } p \neq p_0.$$

b) In this situation, since consistency is a very basic property, one would like a more accurate evaluation of the power function. Instead of a fixed alternative p one may take *a sequence* p_n *of alternatives depending on* n, such that p_n moves closer to the null hypothesis. The philosophy behind this approach of **local alternatives** is that as sample size tends to infinity, i.e. information increases, one simultaneously makes the problem more difficult by considering alternatives closer and closer to H_0 (indeed these are harder to detect). Accordingly, set $p_n = p_0 + hn^{-1/2}$ where h is fixed; $\{p_n\}$ will be the sequence of alternatives. Show that along these alternatives,

$$\lim_{n\to\infty} \Pr_{p_n}(\phi = 1) = \gamma_0(h) \tag{1}$$

where γ_0 is the power function of the two sided Z-test based on one observation $Z \sim N(\mu, p_0(1-p_0))$ for $H_0 : \mu = \mu_0$. (For $Z \sim N(\mu, 1)$ this power function was given in (41), p. 49 handout). You may use that the CLT given in relation (48) handout, p. 58 holds also along $p = p_n$.

Problem 5.2. In the discussion of the Kolmogorov-Smirnov test statistic KS_n, and the Proposition A.2 which states that it is distribution free, it was argued that the empirical distribution function

$$\hat{F}_n(t) = n^{-1}\sum_{i=1}^{n} \mathbf{1}\{X_i \leq t\}$$

has jumps of height $1/n$ in the order statistics $X_{[i]}$. However as defined above, $\hat{F}_n$ may have jumps of height k/n if k of the observations $X_1, \ldots, X_n$ coincide. This is known as the case of ties. Show that ties occurs with probability 0 only under the current assumptions, more specifically: if X_1, X_2 are independent and at least one of them has a continuous cdf then $\Pr(X_1 = X_2) = 0$.

Problem 5.3. A historian examining British colonial records for the Gold Coast in Africa suspects that the death rate was higher among African miners than among European miners. In the year 1936, there were 223 deaths among 33,809 African miners and 7 deaths among 1541 European miners on the Gold Coast.
Consider this year as a sample from the pre-war era in Africa. Is there good evidence that the proportion of African miners who died was higher than the proportion of European miners who died? State hypotheses, calculate a test statistic, give a P-value, and state your conclusion. *(Theory for this was discussed in class, supplemental handout forthcoming).*