Assignment 3 - CSC/DSC 262/462 - Fall 2017 - Due October 26

Q1: The binomial distribution $bin(n,p)$ can be approximated by a Poisson distribution with mean $\lambda = np$. To explore this, suppose we compute the probability $q_\lambda = P(X \le 8)$, where X is a Poisson random variable with mean $\lambda = 10$. Then we should have

$$q_{n,p} \approx q_\lambda$$

where $q_{n,p} = P(Y \le 8)$, for $Y \sim bin(n,p)$ with $p = \lambda/n$, provided n is large enough.

To get a sense of how large n should be, using `R`, construct a plot of $q_{n,p}$ against n, for $n = 10, 11, \ldots, 199, 200$, in each case setting $p = \lambda/n$. Superimpose on the plot a horizontal line at q_λ. Then find the smallest n for which $|q_{n,p} - q_\lambda| \le 0.01$.

Q2: Suppose X_1, X_2 are independent observations from a geometric distribution with mean $1/p$, $p \in (0,1)$.

(a) Derive the conditional distribution of X_1 conditional on $\{X_1 + X_2 = s\}$ for any $s \ge 2$, in the form of the probability mass function (PMF)

$$p_X(x) = P(X_1 = x \mid X_1 + X_2 = s).$$

(b) How does $p_X(x)$ depend on x and p?

Q3: The odds of an event A is denoted $Odds(A)$. Suppose the distribution of a random variable X depends on whether or not event A occurs. In particular, conditional on A, $X \sim bin(4, 0.5)$. Conditional on A^c, $X \sim bin(2, 0.9)$.

Determine the relationship between $Odds(A \mid X = x)$ and $Odds(A)$ for $x = 0, 1, 2, 3, 4$. For which values of x does evidence of the form $\{X = x\}$ increase the odds that A does not occur.

Q4 A test for a certain infection was evaluated experimentally. When administered to a test group of 285 individuals known to have the infection, the test was positive in 256 cases. The test was also administered to a control group of 220 subjects known to be free of the infection. The test was positive in 12 cases.

1. Estimate the sensitivity and specificity of the test directly from the data.
2. This test is intended to be used in clinical populations of varying infection prevalence. Use `R` to construct plots of PPV and NPV for values of prevalence ranging from 0 to 10%. Use the `type = 'l'` option of the `plot()` function.
3. Calculate prevalence, NPV and PPV directly from the data. How do these values compare to those shown in the plots of part (b)?

Q5: [For Graduate Students] Suppose X is a random variable. The *moment generating function* is a function of a real variable t, defined by

$$M_X(t) = E\left[e^{tX}\right]$$

for any fixed t.

(a) Assuming that $M_X(t)$ is finite in some open interval (a,b), where $a < 0$ and $b > 0$, show that the kth moment of X can be calculated by the kth derivative of $M_X(t)$ evaluated at $t = 0$, that is,

$$\left.\frac{d^k M_X(t)}{dt^k}\right|_{t=0} = E[X^k], \quad k = 1, 2, \ldots.$$

[HINT: Assume that you may exchange the order of differentiation and integration where needed.]

(b) Show that if X and Y are independent random variables, the moment generating function of $X + Y$ equals the product of the moment generating functions of X and Y, that is,

$$M_{X+Y}(t) = M_X(t) M_Y(t),$$

where they are finite.

(c) Derive the moment generating function for $X \sim bin(n,p)$.

(d) If $X \sim bin(n,p)$ and $Y \sim bin(m,q)$, show that $X + Y$ is a binomial random variable if and only if $p = q$.