18.650 Problem Set 5, due Wednesday, October 21, 2015

Here are some comments on "$p$-values," a common notion in statistics that Rice mentions, and which we've already seen for regression coefficients. If a hypothesis $H_0$ will be rejected for large values of a statistic $V$ such as $X^2$ or $W$, and a value $V_{\text{obs}}$ is observed, then the $p$-value is the probability that $V$ is greater or equal $V_{\text{obs}}$, assuming $H_0$.

If $V$ has approximately a chi-squared distribution, then you can determine whether the $p$-value is less than $\alpha$, for a few usual values of $\alpha$, from the chi-squared tables.

By the same computational methods that have been used to compute the numbers in the tables (which was done some decades before computers were invented, on very slow desk calculator machines) it is also possible to compute $p$-values for any $V_{\text{obs}}$, and Rice mentions some such $p$-values. But you aren't asked to do such computations, just to answer questions about $\alpha$ and $1-\alpha$ comparisons, and the tables are enough for that.

1. A statistic $X$ is measured in a medical test. Suppose that for a certain disease, $D$, the hypothesis $H_0$ is that the person being tested does not have $D$ and in that case the distribution of $X$ is $N(4,1)$. The alternative hypothesis $H_1$ is that the person does have $D$, although they may not have any symptoms yet. Under $H_1$, $X$ has distribution $N(7,1)$.

(a) Show that for any $c$, $X \geq c$ is a best critical region for testing $H_0$ vs. $H_1$, in the sense that it has largest power for given size (recall the Neyman-Pearson lemma).

(b) Suppose it costs \$50 each time $X$ is measured for one patient. Suppose this will be done for people in a "risk group" in which *a priori*, the probability of having $D$ is 0.04. If the test is positive ($H_0$ is rejected), a test based on a different statistic costing \$1000 will be done which will give a correct result in essentially all cases. If the original \$50 test is negative ($H_0$ is not rejected) then no further test or treatment will be done. In that case, if a patient has $D$, assume a loss of \$1,000,000. How should $c$ be chosen to minimize the expected loss?

(c) In a general population where the *a priori* probability of having $D$ is $10^{-5}$, is it cost-effective to do any such test procedure (for any $c$)? Hint: would it be, even if an initial \$50 test always gave the correct result?

2. Some random numbers were generated in R as follows. They are independent and normally distributed, all with mean 2.8, but with a variance chosen

at random: with probability 0.9 the variance is 1, and with probability 0.1 it is 25, so that the standard deviation is 5. Let's consider some numbers actually output by R and see if we can determine which variance they were generated from. For each of the following numbers, find the likelihood ratio that they came from $N(2.8, 1)$ relative to coming from $N(2.8, 25)$. Also find the posterior probability in each case. If answers are very large or small, give them in scientific notation $r \cdot 10^m$ where $m$ is an integer that may be positive, negative or 0 and $1 \leq r < 10$, giving $r$ to three significant digits.

(a) $X = 3.4477$ (b) $X = 0.1174$ (c) $X = 3.8779$.

3. In Buffon's needle problem, a needle of length $L$ is thrown at random onto ruled paper with lines at distance $D$ apart. If $L < D$, the probability that the needle hits a line is $2L/(\pi D)$. Lazzarini in 1901 reported on an experiment with $L/D = 5/6$, where the needle was thrown 3408 times and it hit the line 1808 times.

Usually $\chi^2$ tests are one-sided, and the hypothesis is rejected for large values of the statistic $X^2$. But, in this case, do a two-sided test for random sampling with the given hitting probability, where we'd reject the hypothesis if $X^2$ is too large (the hitting probability is wrong) or if $X^2$ is too small (the sampling may have been non-random), at the $\alpha = 0.05$ level. So we'll reject if $X^2$ is less than the .025 quantile or larger than the 0.975 quantile. of $\chi^2$.

4. Gregor Mendel, in one of his experiments on peas, found that of 556 total pea plants, 315 had round, yellow peas, 108 had round, green ones, 101 had wrinkled, yellow peas and 32 had wrinkled and green peas. A simple hypothesis is that "round" is dominant and has probability 3/4, with "wrinkled" having probability 1/4, and likewise "yellow" is dominant and has probability 3/4 with "green" having probability 1/4. Moreover, the simple hypothesis says that being round and being yellow are independent. Thus the four possible combinations of properties have probabilities 9/16 for round and yellow, and the others respectively 3/16, 3/16, and 1/16. Test the simple hypothesis by a $\chi^2$ test, rejecting it for large values of $X^2$, at level 0.05.

5. Consider the class of all normal distributions $N(\mu, 1)$, so that $H_1$ is the set of all $(\mu, 1)$ for any real $\mu$ and $\sigma = \sigma^2 = 1$. Let $H_0$ be the subset of $H_1$ where $\mu$ has a fixed value $\mu = \mu_0$. Show that in this case the Wilks statistic $W = -2 \log \Lambda$ (defined in the handout in general, as opposed to the multinomial case as in Rice) has exactly a $\chi^2(1)$ distribution for all $n$, not only asymptotically as $n \to \infty$.