MthStat 465, Spring 2005, Lecture Number 21
Overview of Testing Statistical Hypotheses, Part II

This a continuation of the previous lecture.

The decision rule splits the sample space into two complementary events. Outcomes in one of these events are judged to support the null hypothesis, and outcomes in the other event are judged not to support the null hypothesis. This latter set is called the **rejection region**, which we shall denote by D (for "don't support the null hypothesis"). D is an event. Our goal is that if $\Pr \in H_0$, and α is the level of the test, then $\Pr(D) \leq \alpha$. In other words, the probability of a Type I error should be less than or equal to the level of the test. On the other hand, if $\Pr \in H_1$, we want $\Pr(D^c)$ to be close to 0, or, more directly, $\Pr(D)$ should be close to 1. The function $\Pi : H_1 \to [0,1]$ defined by $\Pi(\Pr) = \Pr(D)$ is called the **power function** for the test.

Here is an illustration of these concepts in a coin tossing experiment. A coin is tossed 6 times, and sequence of length 6 is recorded of H's and T's corresponding to the outcomes of head and tail on each successive toss. The sample space consists of all such sequences (there are 64 of them) and the sigma algebra is all the subsets of this set. (There are 2^{64} sets in the sigma algebra.) We will only consider probability measures of the form $\Pr_p(\{s\}) := p^{h(s)}(1-p)^{6-h(s)}$ where $h(s)$ is the number of H's in the outcome s. Here $p \in [0,1]$.

First, consider $H_0 = \{\Pr_p : p = 1/2\}$ and $H_1 = \{\Pr_p :\neq 1/2\}$, and $D = \{s \in S : h(s) = 0 \text{ or } 6\}$. That is, we say that the data does not support the null hypothesis of a fair coin if we get all heads or all tails in 6 tosses. Since $\Pr_{1/2}(D) = 1/32$, the level of this test is any $\alpha \geq 1/32$. As for the power, $\Pi(\Pr_p) = \Pr_p(D) = p^6 + (1-p)^6$ for any $p \neq 1/2$). Notice that the closer p gets to 1/2 the closer the power gets to 1/32, which is not good (1/32 is not close to 1), but this is unavoidable. In fact, we only get the power to be greater that 0.95 if we have (approximately) $p < 0.00851$ or $p > 0.99149$). This is pretty lousy. We can only make substantial improvements by tossing the coin more. For example if we take 100 tosses, and make $D = \{s \in S : h(s) \leq 39 \text{ or } h(s) \geq 61\}$, then then $\Pr_{1/2}(D) \approx 0.0350$ and

$$\Pi(\Pr_p) = \sum_{k=0}^{39} \binom{100}{k} p^k (1-p)^{100-k} + \sum_{k=61}^{100} \binom{100}{k} p^k (1-p)^{100-k}.$$

Now we get the power to exceed 0.95 if (approximately) $p < 0.315$ or $p > 0.685$.

If we return to our experiment with 6 tosses, we could have $H_0 = \{\Pr_p : p \geq 1/2\}$ and $H_1 = \{\Pr_p : p < 1\}$, perhaps contrasting beliefs about whether or not the probability of heads was at least 1/2 or not. Now the probability of Type I error is not constant on H_0 unless our rejection region D is empty. Suppose that we try $D = \{s : h(s) leq 1$, that is, we say the data don't support the assertion that the probability of heads is at least 1/2 if we get 1 or fewer heads. We have

$$\Pr_p(D) = (1-p)^6 + 6p(1-p)^5.$$

We can check that for $0 \leq p \leq 1$ the expression $(1-p)^6 + 6p(1-p)^5$ decreases as p increases by computing the derivivative of the expression. Hence the level of this test is only $7/64 \approx 0.11$, and the power falls below 0.9 at about $p = 0.09$. Again, things can be improved by increasing the number of tosses, but, since the $\Pr_p(D)$ is continuous in p (it is a polynomial function) regardless of the number of tosses, we can never achieve the level we want and keep the power near 1 across the whole alternative hypothesis.