M4056 Hypothesis Testing II. **October 29, 2010**

I am going to jump ahead to error analysis, because this reinforces the conceptual distinctions you need to be able to make to understand the structure of the testing situation. A second reason to hurry to methods of evaluating tests is that it seems a bit peculiar to describe how to make tests with no idea about how good they are. The techniques used to evaluate tests provide framing for the discussion in 8.2 of methods of finding tests; I'll return to topics in 8.2 after treating these important ideas.

Type I and Type II Errors. Regardless of whether a hypotheses is true or false, the test may result in acceptance or rejection. Thus, there four possible outcomes when a test is completed: accept a true hypothesis, reject a true hypothesis, accept a false hypothesis, reject a false hypothesis.

- Rejecting H_0 when it is true is called a Type I Error.
- Accepting H_0 when it is false is called a Type II Error.

In practice, there are costs and benefits associated with each of the possible outcomes, and obviously one would seek a test that is optimal when the costs and benefits are taken into account.

Quick Exercise 1. Suppose a machine generates random numbers 0 and 1, and that the proportion of 1s generated is θ, an unknown parameter. We wish to test the hypothesis $\theta < \theta_0$. Let M be the average of the first n numbers generated by the machine. We pick a number q and resolve to accept the hypothesis if $M < q$. Find the probability of each kind of error.

Solution. A Type I error is made if $\theta < \theta_0$ and $q \le M$. Let $u(q)$ be the smallest integer such that $q \le u(q)/n$. The probability that $q \le M$ is $\sum_{i=u(q)}^{n} \binom{n}{i}\theta^i(1-\theta)^{n-i}$. The larger q is, the *smaller* the probability of a Type I error. A Type II error is made if $\theta_0 \le \theta$ and $M < q$. The probability that $M < q$ is $\sum_{i=0}^{u(q)-1} \binom{n}{i}\theta^i(1-\theta)^{n-i}$. The larger q is, the *larger* the probability of a Type II error.

As the exercise emphasizes, the probability of each kind of error is dependent upon the true value of the parameter. Yet if we knew the parameter value there would be no reason to perform a test. So how can error analysis be meaningful? The answer to this puzzle is that in many cases it is enough to find an upper bound on the probability of an error, and our ignorance of the actual value of θ need not stand in the way of finding such bounds.

Quick exercise 2. (a) Is the following true? *The probability of an error (of either kind) is between the probabilities of the two error types. That is, if the test were performed numerous times with possibly different values of θ on different occasions, then no matter how the values of θ arise, the error probability will be between these bounds.* (b) Is the following true? *If the probability of a Type I error is less that p, regardless of θ, and the probability of a Type II error is less than q, regardless of θ, then the probability of an error is less than the maximum of p and q.*

Solution. (a) This statement cannot be interpreted without clarifying the meaning of "probability of error". In the classical framework, both the type of the error that may be made and the probability of making it depend on the parameter θ. In this framework, it is difficult to understand what might be meant by **between** the probability of a Type I error and a Type II error, since after θ is set at a specific value, only one type of error is possible. However, in the Bayesian framework, θ is viewed as a random variable. In this case, we can make sense of the probabilities of the two different kinds of errors within a single probabilistic model that encompasses random variation in θ as well as random variation in the sample $\vec{X}$. In this setting, an error of either type could be regarded as an event. Since the two types are mutually exclusive, the overall probability of an error is the sum of the probabilities of the errors of each type. (b) In the classical setting, this is true. Once θ has a number value, only one kind of error can be made and the probability of that is either p or q.