# 1 Finite Sample Inference Beyond Normality – 14.382. Spring 2006

The basic idea could be illustrated using the following example.

**Example 1.** Suppose $Y = X\beta + \epsilon$, $E[\epsilon|X] = 0$, $E[\epsilon\epsilon'|X] = \sigma^2 I$, $\epsilon = (\epsilon_1, ...\epsilon_n)' = \sigma U$, where

$$U = (U_1, ..., U_n) \text{ are i.i.d. with law } F_U \tag{1}$$

where $F_U$ is known. For instance, taking $F_U = t(3)$ will better match the features of many financial return datasets than $F_U = N(0, 1)$.

Consider testing $H_0 : \beta_j = \beta_j^0$ vs. $H_A : \beta_j > \beta_j^0$. Under $H_0$

$$t_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{s^2 (X'X)^{-1}_{jj}}} \overset{\text{by } H_0}{=} \frac{((X'X)^{-1}X'\epsilon)_j}{\sqrt{\frac{\epsilon' M_X \epsilon}{n-K}(X'X)^{-1}_{jj}}} = \frac{((X'X)^{-1}X'U)_j}{\sqrt{\frac{U' M_X U}{n-K}(X'X)^{-1}_{jj}}} \tag{2}$$

P-value for this test can be computed via Monte-Carlo. Simulate many draws of the t-statistics under $H_0$:

$$\{t^*_{j,d}, d = 1, ..., B\}, \tag{3}$$

where $d$ enumerates the draws and $B$ is the total number of draws, which needs to be large. To generate each draw $t^*_{j,d}$, generate a draw of $U$ according to (1) and plug it in the right hand side of (2). Then p-value can be estimated as

$$\widehat{pval} = \frac{1}{B}\sum_{d=1}^{B} 1\{t^*_{j,d} \geq t_j\}, \tag{4}$$

where $t_j$ is the empirical value of the t-statistic. The p-value for testing $H_0 : \beta_j = \beta_j^0$ vs. $H_A : \beta_j \neq \beta_j^0$ can be estimated as

$$\widehat{pval} = \frac{1}{B}\sum_{d=1}^{B} 1\{|t^*_{j,d}| \geq |t_j|\}. \tag{5}$$

Critical values for confidence regions and tests based on the t-statistic can be obtained by taking appropriate quantiles of the sample (3).

**Example 2.** Next generalize the previous example by allowing $F_U$ to depend on the unknown nuisance parameter $\gamma$, which true value $\gamma_0$ is known to belong to region $\Gamma$. Denote the dependence as $F_U(\gamma)$.

For instance, suppose $F_U(\gamma)$ is t-distribution with the "degrees of freedom" parameter $\gamma \in \Gamma = [3, 30]$, which allows to nest distributions that have a wide range of tail behavior, from very heavy tails to light tails. The normal case is also approximately nested by setting $\gamma = 30$.

Then, obtain P-value for each $\gamma \in \Gamma$ and denote it as $pval(\gamma)$. Then use

$$\sup_{\gamma \in \Gamma} pval(\gamma)$$

for purposes of testing. Since $\gamma_0 \in \Gamma$, this is a valid upper bound on the true P-value $pval(\gamma_0)$. Likewise, one can obtain critical values for each $\gamma \in \Gamma$ and use the *least favorable* critical value. The resulting confidence regions could be quite conservative if $\Gamma$ is large.

The question that comes up naturally is: why not use an estimate $\hat{\gamma}$ of the true parameter $\gamma_0$ and obtain $pval(\hat{\gamma})$ and critical values using MC where we set $\gamma = \hat{\gamma}$? This method is known as the (parametric) *bootstrap*. Bootstrap is simply a MC method for obtaining p-values and critical values using the estimated data generating process. Bootstrap provides asymptotically valid inference, but bootstrap does not necessarily provide valid finite sample inference. However, bootstrap often provides a more accurate inference in finite samples than the asymptotic approach does.

**Example 3. (HW)** Consider Temin's (2005) paper that models the effect of distance from Rome on wheat prices in the Roman Empire. There are only 6 observations. Calculate the P-values for testing the null that the effect is zero versus the alternative that the effect is negative. Consider first the case with normal disturbances (no need to do simulations for this case) , then anlayze the second case where disturbances follow a t-distribution with 8 degrees of freedom (calculating this analytically will require that you be a mighty mountain of mathematical genius, and it may be impossible anyway. how could mere mathematics mortals calculate p-values?). Also, provide an economic justification for the form of the null and the alternative hypotheses. What in the paper justifies the use of a one-sided alternative hypothesis? Can you think of reasons why a two-sided alternative might be a better choice? If the alternative were two-sided, what would the effect on the outcome of the hypothesis test and the p-value be (for both the normal and t(8) distribution cases)?