WMS: 8.7

H. Designing an Experiment

1. Sample Size Determination

   a. Suppose you want to estimate parameter to within a certain accuracy e

      i. called *margin of error*.

   b. As measured by c.i. of level $1-\alpha$.

   c. Suppose you have pre-knowledge of the standard deviation.

   d. Then $\sigma z_{\alpha/2}/\sqrt{n} \leq e$

   e. Then $\sigma z_{\alpha/2}/e \leq \sqrt{n}$

   f. Then $n \geq \sigma^2 z_{\alpha/2}^2/e^2$

   g. Ex., to estimate binomial proportion (ex. poll result) to 2%,

      i. $\sigma^2 = \theta(1-\theta) \leq .25$

      ii. Can get by with $n = .25(1.96)^2/.02^2 \approx 2500$.

WMS: 8.8-8.9

I. Common Applications

1. We just did one-sample examples.

   a. One-sample normal means confidence intervals

      i. As above: $\bar{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$ for normal mean

- $\bar{X} \pm t_{\alpha/2;n-1} S/\sqrt{n}$ if σ must be estimated as S.

b. Two-sample binomial intervals:

i. $X_1 \sim$ Bin(n_1, π_1), $X_2 \sim$ Bin(n_2, π_2).

ii. Estimates are $\hat{\pi}_1 = X_1/n_1$, $\hat{\pi}_2 = X_2/n_2$.

iii. c.i. for $\pi_1 - \pi_2$ is $\hat{\pi}_1 - \hat{\pi}_2 \pm z_{\alpha/2} \sqrt{\hat{\pi}_1(1-\hat{\pi}_1)/n_1 + \hat{\pi}_2(1-\hat{\pi}_2)/n$

2. Two-sample mean difference

a. Assume

i. $X_1, \ldots, X_m$ same expectation and finite variance

ii. $Y_1, \ldots, Y_n$ same expectation and finite variance

iii. All independent

b. Estimate $\theta = \mathrm{E}[Y_i] - \mathrm{E}[X_i]$

i. Case with common variance:

- Pivotal quantity $S = (\bar{Y} - \bar{X} - \theta)/(\sigma\sqrt{1/m + 1/n})$ if common variance were known to be σ.
- Pivotal quantity $S = (\bar{Y} - \bar{X} - \theta)/(S_p\sqrt{1/m + 1/n})$
- $S_p = \sqrt{\frac{\sum_{i=1}^{m}(X_i - \bar{X}) + \sum_{i=1}^{n}(Y_i - \bar{Y})}{m+n-2}}$: pooled standard deviation. ***********
- Pivot has distribution approximately N$(0, 1)$
  - ▷ More closely, t_{m+n-2}.

ii. Case with variances not known to be common:

- Let $\sigma^2 = \mathrm{Var}[X_i]$, $\tau^2 = \mathrm{Var}[Y_i]$
- Pivotal quantity $S = (\bar{Y} - \bar{X} - \theta)/\sqrt{\sigma^2/m + \tau^2/n}$ if σ, τ known.
- If σ, τ unknown, estimate by $S_x = \sqrt{\sum_{i=1}^{m}(X_i - \bar{X})/(m-1)}$, $S_y = \sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})/(n-1)}$ respectively.
- Is $S = (\bar{Y} - \bar{X} - \theta)/\sqrt{S_x^2/m + S_y^2/n}$ pivotal? No.
  - ▷ If $\sigma = 0$, reduces to t_{n-1}.
  - ▷ If $\sigma = \tau$, $m = n$, t_{m+n-2}.
- Standard solution: approximate by t_d, where d is complicated formula of S_x, S_y, m, n.
- See Fig. 4.

****

WMS: 9.1

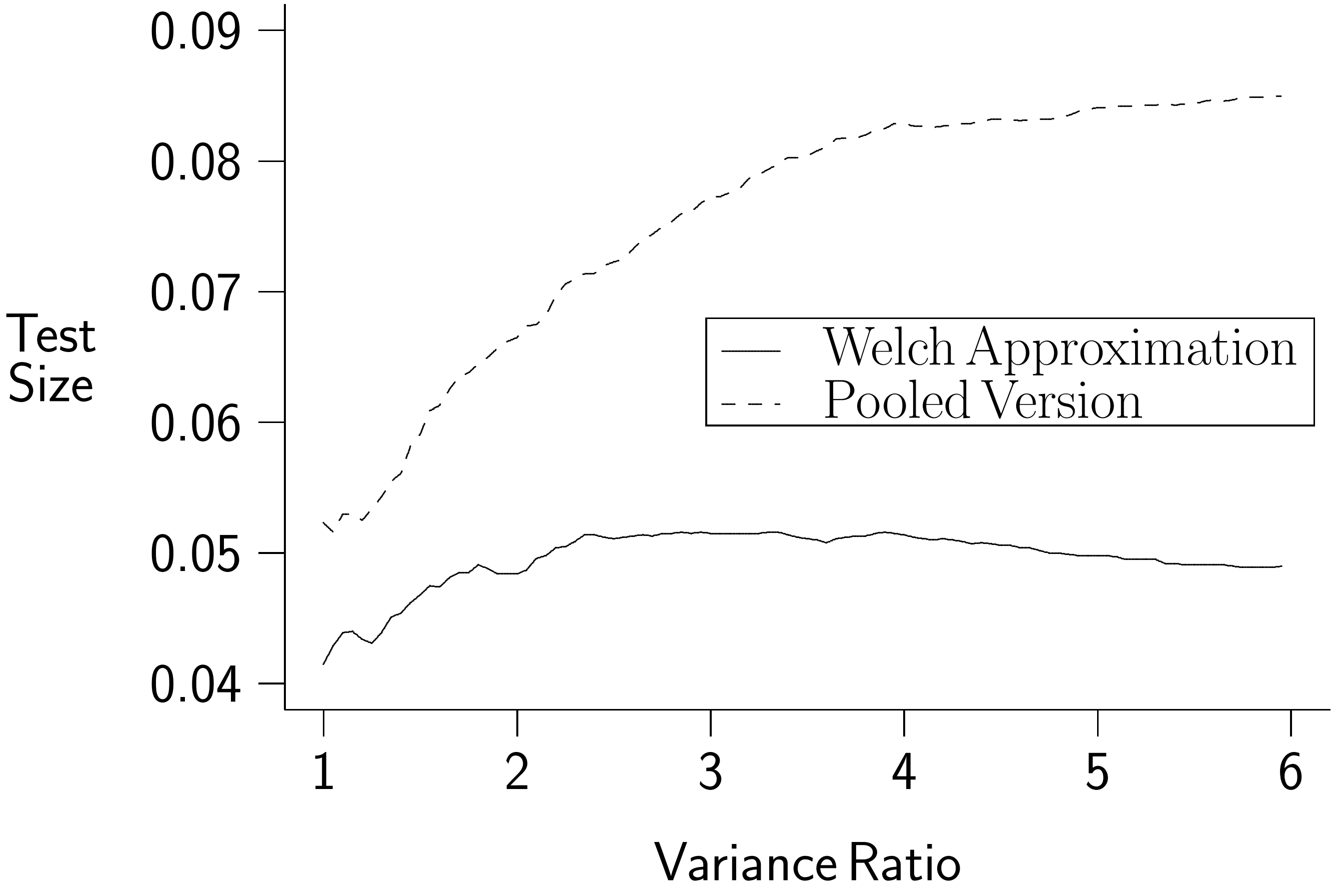
3. Variance Bounds for estimators:

a. How small a variance can one get for an unbiased estimate?

b. Precision of estimator comes from derivatives of log likelihood.

i. 1st derivative tells how fast density changes with θ.

Fig. 4: Dependence of the Two Sample Test on Variance Ratio



ii. 2nd derivative tells how fast density curves with θ .

c. Idea:

i. information about θ depends on how quickly on average $f_X(X;\theta)$ as a function of θ drops away from its peak

ii. This is measured by the inverse of the curvature.

iii. For this course always interpret log as natural logs.

WMS: 9.2

J. Relative Efficiency

1. Efficiency measures variances of estimates.

   a. Definition: The ratio $\mathrm{Var}\left[\hat{\theta}_1\right] / \mathrm{Var}\left[\hat{\theta}_2\right]$ is the *relative efficiency* of $\hat{\theta}_2$ re $\hat{\theta}_1$.