Behrens-Fisher (Extra Credit)

Problem Setup

Recall the two sample testing scenario

$$\begin{aligned} x_i &\sim N(\theta_x, \sigma_x^2), \qquad i = 1, \ldots, N_x, \\ y_i &\sim N(\theta_y, \sigma_y^2), \qquad i = 1, \ldots, N_y. \end{aligned}$$

There are *three* important cases to distinguish regarding the variances σ_x^2 and σ_y^2: case 1) both σ_x^2 and σ_y^2 known, case 2) $\sigma_x^2 = \sigma_y^2$ but unknown, and case 3) $\sigma_x^2 \neq \sigma_y^2$ with both unknown..

Cases 1 and 2 do not typically involve approximations (since the sampling distributions are known to be Normally distributed).
For case 1, we know: $\bar{X} - \bar{Y} \sim N(\theta_x - \theta_y, \sigma_x^2 + \sigma_y^2)$.
For case 2, we know: $\bar{X} - \bar{Y} \sim T_\nu(\theta_x - \theta_y, S_p^2)$, where $S_p^2 = \frac{(N_x-1)S_x^2 + (N_y-1)S_y^2}{N_X + N_Y - 2}$, with $S_x^2 = \frac{\sum_{i-1}^{N_x}(x_i - \bar{X})^2}{N_x - 1}$, $S_y^2 = \frac{\sum_{i-1}^{N_y}(y_i - \bar{Y})^2}{N_y - 1}$, and $\nu = N_x + N_y - 2$.

One-sided Testing

Consider the one sided hypothesis test:

$$\begin{aligned} H_0 : \theta_x - \theta_y &= 0 \\ H_A : \theta_x - \theta_y &< 0. \end{aligned}$$

For case 1, with data generated under $H_0 : (\theta_x = \theta_y)$, we incur a "Type 1 Error" if, $\frac{\bar{X} - \bar{Y}}{(\sigma_x^2/N_x + \sigma_y^2/N_y)^{1/2}} > R$. For this particular test, if $R = 1.6449$ (very approximately the 95^{th} quantile), the theoretical Type 1 Error is 0.05.

Similarly, for case 2, with data generated under $H_0 : (\theta_x = \theta_y)$, we incur a Type 1 Error if, $\frac{\bar{X} - \bar{Y}}{(S_P^2(1/N_x + 1/N_y))^{1/2}} > R$. For the Type 1 Error to be 0.05, R would be specified as the 95^{th} quantile from a *standard* T-distribution (shift=0, scale=1), with degrees of freedom $N_x + N_Y - 2$.

For cases where $\sigma_x^2 \neq \sigma_y^2$, The "Welch" statistic follows as:

$$T = \frac{\bar{X} - \bar{Y}}{(s_x^2/N_x + s_y^2/N_y)^{1/2}}. \tag{1}$$

It is a fact that with $\sigma_x^2 \neq \sigma_y^2$, T does not follow a T-distribution. However, in the "*old -days*", people were inclined to suggest that T was very close to a T-distribution, with degrees of freedom:

$$\nu_{ws} = \frac{(S_x^2/N_x + S_y^2/N_y)^2}{S_x^4/(N_x^2(N_x - 1)) + S_y^4/(N_y^2(N_y - 1))},$$

which is often referred to as the Welch-Satterthwaite correction. In class, we showed that (note that we used normally distributed data) if we let $T > R$ define a Type 1 Error, where R was specified as the 95^{th} quantile from a *standard* T-distribution (shift=0, scale=1), with degrees of freedom ν_{ws}, the Type 1 Error was *approximately 0.05.*

10% Midterm Credit Points

Verification

1. Simulate data according to Cases 1, 2, and 3, and check the actual Type 1 Error rates. It is your job to decide how many variations of data simulations are necessary.

Bayesian Testing

1. Under marginal Jefferys priors, code up a Gibbs Sampler for obtaining posterior draws: $\delta_{x,y}^{(i)} = \theta_x^{(i)} - \theta_y^{(i)}$. After "burning-in", and collecting *enough* of these samples, show how to compute $Pr(H_0|\{x_1, \ldots, x_{N_x}\}, \{y_1, \ldots, y_{N_y}\})$.

2. Specify a reasonable prior for testing H_0 VS. H_A, and provide some justification for why it is reasonable.

3. Let the rule: $Pr(H_0|\{x_1, \ldots, x_{N_x}\}, \{y_1, \ldots, y_{N_y}\}) < 0.05$, be the rule you use for rejecting H_0. From the data that you simulated in the preceding section, report your Type 1 Error rates. Just to be clear, you're simulating this.

4. How would you modify your prior after observing your results?

5% Additional Midterm Credit Points

Repeat the exercises above, but this time consider the *sharp* test:

$$
\begin{aligned}
H_0 : \theta_x - \theta_y &= 0 \\
H_A : \theta_x - \theta_y &\neq 0.
\end{aligned}
$$

5% Additional Midterm Credit Points

Show how the exercises above change under varied specifications of the sampling distributions. That is, let

$$
\begin{aligned}
x_i &\sim f, \qquad i = 1, \ldots, N_x \\
y_i &\sim g, \qquad i = 1, \ldots, N_y,
\end{aligned}
$$

where f and g are arbitrary distributions. Of course you won't be able to consider all cases concerning arbitrary f and g, so as a step, perhaps check how the results change under '*heavily skewed*' distributions.