**FISH 559: Numerical Computing for the Natural Resources**
**Homework 2 : Estimation of proportions**

One of the inputs to a stock assessment is the "mean" sex-ratio over the period of years. The objective of this assignment is to evaluate methods for estimating this sex-ratio and its standard error when the data available are numbers of males and females by year. The tasks to be undertaken are:

1. Show that the conventional estimator $\sum_y M_y / \sum_y (M_y + F_y)$ where $M_y$ is the number of males in year *y*, and $F_y$ is the number of females in year *y* is the maximum likelihood estimate when it is assumed that the sex-ratio is independent of time and each year's data is the result of a binomial sample, i.e.

$$M_y \sim Bi(\bar{p}, M_y + F_y) \tag{1}$$

where $\bar{p}$ is the mean sex-ratio

2. Use TMB to fit model (1) to the data set in HOME2.DAT and report the estimate of $\bar{p}$ and its standard error. Only use data for years for which the sample size for both males and females is at least 1.
3. The assumption that *p* is constant over time is unrealistic. Use simulation to examine whether the coverage probability of this estimator equals the nominal 95% level when the annual sample size is 100 and $p \sim Beta(2,1)$, i.e. how often the estimated 95% confidence interval contains the true value. Base your simulations on 1,000 replicates and 25 years of data.
4. One way to account for the overdispersion caused by *p* not being constant over time is to allow for "process error", using the following negative log-likelihood:

$$-\ell nL = \sum_y \left( \ell \mathrm{n}\sigma_y + \frac{1}{2(\sigma_y)^2} \left[ p_y - \bar{p} \right]^2 \right) \tag{2}$$

where $p_y$ is the observed sex-ratio for year *y*, and $\sigma_y$ is standard error of the sex-ratio for year *y*, accounting for over-dispersion, i.e.:

$$\sigma_y = \sqrt{\tau^2 + p_y(1-p_y)/n_y} \tag{3}$$

where $n_y$ is the number of animals sexed during year *y* ($= M_y + F_y$) and $\tau$ is the standard deviation of the process error.

   4.a Apply this estimator to the data in HOME2.DAT and report the estimates of $\bar{p}$ and $\tau$ and their standard errors.

   4.b. Apply this equation to the simulated data (task 3) to evaluate how the coverage probability is improved.

Equation 2 is still subject to criticism. Comment on whether you think the assumption of a binomial variance for an individual year is valid and whether the assumption that the *p*s are normal is valid. If not, suggest alternatives which you believe are more appropriate.