Sampling and Sample Size Issues I

Estimating Relative Stock Composition w/GSI

Issues in Designing Sampling Programs for GSI of Salmon Populations

- Objective(s) of sampling
- Criteria for meeting objectives
- Definition of the sampling frame
- Simultaneous estimation for several stocks
- Absolute or relative precision
- Stratification
- Finite populations
- Misclassification
- Additional measurement error
- Sampling procedures
- Oversampling/subsampling

Objectives for Sampling

- To test hypotheses
- To detect presence or absence of a stock
- **To estimate stock composition from a mixture**
  - **Estimate a series of proportions (relative stock composition) from a mixture of stocks (usually a landed catch)**
  - **Multiply estimated catch by relative stock composition to estimate stock composition of the mixture**

Criteria for Meeting the Objective

- Criteria should promote levels of certainty in estimates of relative stock composition sufficient for
  - Determining terminal run size
  - Exploitation rate analysis
  - Forecasting
  - Stock-recruit analysis
- Specific levels of certainty should be explicitly stated in criteria as
  - CV
  - Size of Confidence (or Credibility) Intervals

Criteria for Meeting the Objective

Example of Sources for Determining Criteria:

- USCTCChinook. 1997. A review of stock assessment data procedures for U.S. Chinook stocks. Pacific Salmon Commission. Report (97)-1.
- TCChinook. 1999. Maximum sustained yield or biologically based escapement goals for selected Chinook salmon stocks used by the Pacific Salmon Commission's Chinook Technical Committee for escapement assessment. (99)-3. Pacific Salmon Commission
- Chapter 3, ¶1(b)(v) of the 1999 Annex to the PST, the Chinook Technical Committee is to "recommend standards for the minimum assessment program to effectively implement this chapter …."
- Ad hoc standards required by co-managers for terminal fisheries

Definition of the Sampling Frame

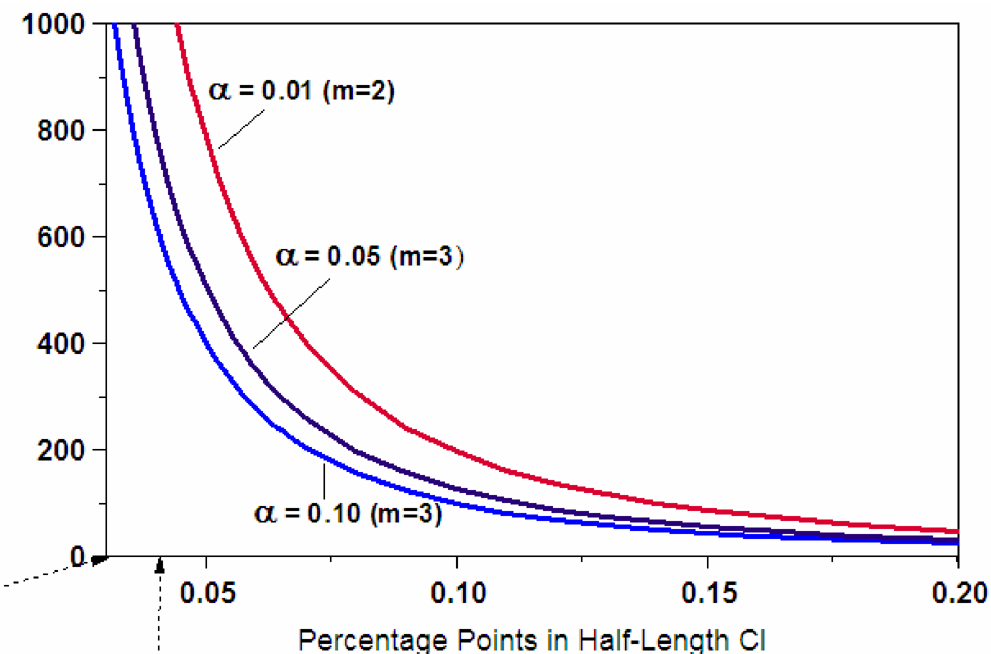
- Salmon are the basic sampling units, usually as landed catch
- Catches are grouped into strata according to time and area of fishing
- Needs of management determines focus within the sampling frame
  - Shaping fisheries in-season is focused on hypothesis testing or detecting presence/absence of a stock in <u>one spatial-temporal stratum at a time</u>
  - Mapping distribution of small stocks is focused on <u>one stratum at a time with data aggregated over several years</u>
  - **Forecasting and stock-recruit analysis is focused on stock composition in <u>all strata collectively in the same year</u>**

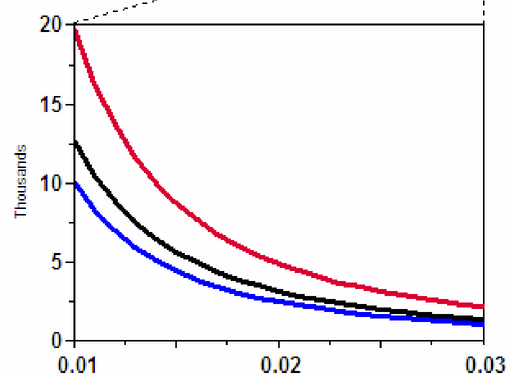
Simultaneous Estimation of Proportions for Several Stocks

Sample size n_o for a <u>simple</u> random sample

$n_o \cong$ **500** when d = 0.05
α = 0.05

$n_o \cong$ **5000** when d = 0.015
α = 0.05





Absolute Precision (d)

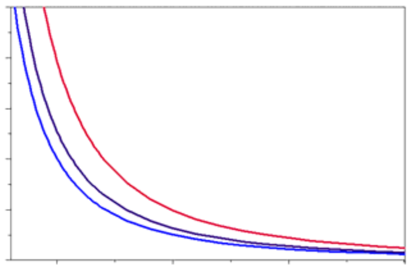
Developed from Thompson, S.K. 1992. Sampling. John Wiley, New York.

Absolute vs. Relative Precision

- Absolute precision is usually expressed as
  - the half length of the desired CI
  - the desired standard error
- Relative precision is usually expressed as
  - half length of the desired CI divided by the worst-case proportion
  - the desired CV of the worst-case proportion

Stratification

When estimating relative stock composition across several strata, sample size n_o needed to attain desired precision in a simple random sample is spread across all strata according several rules:



w/ optimal allocation across strata -

$$n_h = n_o \frac{C_h \sqrt{Var(p_h)}}{\sum_{h'} C_{h'} \sqrt{Var(p_{h'})}}$$

w/ proportional allocation across strata -

$$n_h = n_o \frac{C_h}{\sum_{h'} C_{h'}}$$

With 10 strata of equal size , prop allocation is

$n_h \cong$ **50** when d = 0.05
α = 0.05

$n_h \cong$ **500** when d = 0.015
α = 0.05

Sampling from a Finite Population

- Sample size n_h from a stratum needed to achieve desired precision for estimating relative stock composition will always be less than the size C_h of that stratum
- Sample size is less than expected when sampling from a finite population

$$n_{h(corrected)} = \frac{1}{1/n_h + 1/C_h}$$

Misclassification in GSI

Sample size can be increased to counter misclassification, but only to a point because sample size $n_h \leq C_h$

If misclassification errors are random, have mean 0, and variance σ^2, variance from sampling is inflated by σ^2/n

Option 1:
As stocks are aggregated, classification improves, $\sigma^2 \rightarrow 0$, and no increase in sample size may be needed

Option 2:
Get an *a priori* estimate of misclassification error variance σ^2, and use it to augment initial calculation of n_o , such as for independent errors:

$$n_o \leftarrow n_o + \frac{Z_{\alpha/2}^2 \hat{\sigma}^2}{d^2}$$

Additional Measurement Error

Reality of additional measurement error in planning GSI programs:

- Stock composition is the product of catch and relative stock composition
- Additional measurement error arises when catch is estimated with uncertainty (usually catch in a recreational fishery)
- Precision of estimated stock composition <u>can not be better</u> than precision in estimated catch regardless of the extent of GSI sampling

Sampling Procedures

Sampling protocols are important because usually a small fraction of the catch can be sampled and <u>sampling the catch is NEVER random</u>, but systematic at best and opportunistic at worst

- Sampling should be proportional to catch within a stratum
  - Across opportunities to access catch
  - Across time
- Secondary sampling units (boats?)
  - Should be systematically selected for sampling
  - Should have entire catch subsampled systematically

Oversampling/Subsampling

Number samples taken > number analyzed

- Advantages:
  - Samples to be analyzed can be selected from those taken to correct for non-random sampling and therefore improve accuracy
  - More amenable to having a universal number of samples taken
- Disadvantages:
  - Some delay in analysis
  - Needs a clearinghouse to direct inventory and analysis of samples
  - Taking more samples than are needed is marginally more expensive

Recommendations

- Management goals should drive objectives for GSI programs , objectives which in turn should be used to determine the number of samples that should be <u>analyzed</u>
- Establishing sampling protocols to increase the probability of obtaining a representative sample is extremely important because sampling will not be random and sample size will be a small fraction of the catch
- A policy of oversampling/subsampling should be employed to improve accuracy of estimates from GSI
- Largest degree of stock aggregation that is appropriate to management objectives should be employed to reduce misclassification error
- Sample sizes should be based on simultaneous estimation of stock aggregates
- A stock can be too small to be effectively sampled in a GSI program given sampling and misclassification error