**Introduction:**

* The observations do not arise from a single, usually simple, probability distribution, but rather a specified number of distribution, each associated with a proportion.
* For each component of the mixture, in this case snow crab instars, we can estimate its statistical characteristics, such as the means and variances, the proportion that each component instar represents, as well as the component membership probability, e.g. the probability of a crab of a given size belonging to a particular instar.
* Mixtures can be thought of as an ANOVA whereby the groups are unknown a priori, so that membership is provided by a simultaneous clustering of the observations.
* So instar analysis has the potential to provide information instar statistics, growth, and moult transition, all from a set of size observations, which are often among the easier data to obtain.

**Problems with mixture modelling:**

* However, there are issues with mixture analyses.
* One is known as the identifiability problem: a mixture fit to the data is invariant (i.e. does not change) when we exchange the component labels. So some ad hoc constraint is usually applied, such as forcing the component means to be increasing, needs to be applied to properly constrain the parameter space.
* A related problem is the existence of multiple local solutions, in that different solutions can give similar fits to the data.
* Some solutions are degenerate, e.g. shrinking the variance around a particular observations yields arbitrarily high likelihood.
* Analytical solutions are not available, iterative methods must be used.
* Often the number of components is not known a priori.

**Problems specific to instar analysis:**

* There is no way to independently determine the instar stage. i.e. there is no structure that we can analyse and count the number of moults a crab has undergone.
* The range of size values increases with instar size. Mature or maturing crab has additional variation.
* Thus, overlap in larger instars, generally above instar VIII, means that unstructured mixture models do work properly, i.e. converge to solutions which are at odds with biological knowledge or assumptions.
* However, while larger, maturing instars are harder to separate, they are often more clearly resolved in specific areas in specific years, i.e. global samples are aggregated of growth variations which have occurred over space and time, but specific samples are subject to less growth variation. Although local sample sizes are generally not sufficient to perform an independent analysis.

**Approach:**

* Build a structured mixture model tailored for analyzing snow crab on a single sample, relatively fast.
* Extend the structure to account for hierarchical samples, applicable for a generating space and time-specific (i.e. inferences).

**Tables:**

**Table 1** : Female snow crab instar mean sizes and standard errors (in parentheses) by maturity stage. Measurement units are carapace widths in millimeters. Data are pooled from the snow crab surveys 1998 to 2022.

|  |  |  |  |
| --- | --- | --- | --- |
| **Instar** | **Immature** | **Pubescent** | **Primiparous** |
| **IV** | 9.8 (0.9) | - | - |
| **V** | 14.7 (1.4) | - | - |
| **VI** | 21.0 (2.0) | - | - |
| **VII** | 28.7 (2.7) | - | - |
| **VIII** | 37.9 (3.5) | 42.2 (3.7) | - |
| **IX** | 48.3 (4.5) | 50.8 (4.5) | 49.4 (4.3) |
| **X** | - | 59.8 (5.3) | 60.0 (5.3) |
| **XI** | - | - | 71.1 (6.3) |

Chart, histogram

Description automatically generated

**Figure 1** : Instar mixture model fits for immature (top panel), pubescent (middle panel) and primiparous (bottom panel) female snow crab. Data are pooled from the snow crab surveys 1998 to 2022.