**The Instar Splitting Problem** or **Turning One Gaussian Distribution into Two**

**Background:**

**Instar analysis:**

* Snow crab size distributions, at least among smaller crab, a show a series of modes that corresponds to distinct cohorts or generations.
* It is possible to decompose the size distribution and estimate growth parameters and instar abundances.
* The method used to achieve this is a Gaussian mixture analysis.
* Each Gaussian component of this mixture models the mean size and variation of a particular instar.

**Maturation:**

* As snow crab grow, they go through different maturity phases and some effects are visible in the size distributions and the properties of the instars.
* For females, instars I through VII are all immature.
* Starting with instar VIII, some of females become adolescents.
* The gonads of adolescent females are larger and orange-colored, in contrast to the small white gonads of immature females.
* Thus there are two types of instar VIIIs : immature and adolescent. The average size of adolescent instar VIIIs (~ 41 mm) are larger than their immature counterparts (~ 37 mm).

**Growth:**

* Despite the size differences that is visible among instar VIIIs, we do not expect the growth to be very different between the two groups.
* Development of gonads requires a diversion of energy from somatic (body) growth into gonadic (reproductive) growth. Following this logic the size of adolescent instars VIII should be smaller that immatures VIIIs.
* What explains the difference is that maturation is a size-dependent process, in that larger instar VII have a higher probability of maturing to adolescent instar VIIIs.

**Analytical Approach:**

* A natural modelling choice would be to introduce a size-selective function, such as a logistic function to split the Gaussian component of immature instar VII.
* However, such a logistic partitioning of a Gaussian distribution yields two distributions which not Gaussian and have non-trivial distributions (also non-standard).
* Aiming to preserve the Gaussian mixture structure, we opt for the following approach.
* Consider an immature instar (e.g. VII) parametrize by mean and standard error .
* We want to partition the Gaussian into two components
* We want to split the Gaussian distribution into a two-component Gaussian mixture, with means and and standard errors and , mixed with proportions and .
* We impose the constraint that the mixture mean and standard error of the mixture matches the mean and error of the original Gaussian.

Formally we impose the following constraints:



where and are the means and variances of the original Gaussian.

* Of course, the Gaussian mixture is not itself Gaussian, but we will assume that is approximately so for reasonable parameter values.
* Given and and a specified value of , equations (1) and (2) impose some restrictions on the values that , , and can take.
* The variance equation (2) indicates that as the separation between and increase, so does the term . The largest value that can take can be calculated by setting

when approaches , then and approach zero, i.e. the components have zero variance and singular densities (Dirac deltas) at the points and .

From (1) we have that:



Thus,

(4)

Putting eqn. (4) in (3), we get an expression for the lower bound of :

(5)

We will now find the bound for . From equation (1) and isolating we have:

Equating this equation with eqn. (5), we have:

(6)

Thus the bounds and lie in the interval . Keep in mind that and are linear functions of each other through equation (1). Note that the interval is not symmetric about , but is a non-linear function of . Recall that at these bounds, and converge to zero.

By constraining and to lie in the interval , we have insured that :



Let’s define , we have from (2) that:



From this equation, we can calculate from and vice versa.

If we make the simplifying assumption , i.e. the error of the two components are equal, we have that



In summary, there are interval constraints on and , which then define and , along with and . To facilitate the parameterization of this model, we define a variable , which indexes on the interval :

From which we then derive:

We then calculate