random variable Y has pdf or pmf $f_\theta(y)$ where θ is a possibly vector-valued parameter – e.g., Y has a normal distribution with mean μ, variance $\sigma^2 \rightarrow \theta = \mu, \sigma^2$

(jumping ahead a bit, the general idea is that the data are assumed to have been generated from a parametric model that is specified up to the value of θ and you want to use the data for inference about θ)

a collection of random variables $Y_1, Y_2, \dots, Y_n$ (which in this class will typically constitute a collection of observations) has a joint pdf or pmf $f_\theta(y_1, y_2, \dots, y_n)$ again depending on a possibly vector-valued parameter θ

if these random variables are independent then their joint pdf factors into the product of their (univariate) marginal pdfs:

$$f_\theta(y_1, y_2, \dots, y_n) = \prod_{j=1}^{n} f_{\theta j}(y_j)$$

if, in addition, the random variables are identically distributed then these marginal pdfs are the same $f_{\theta j}(y_j) = f_\theta(y_j)$ – in the salp case, the observations are independent if they are taken far enough away in space but not identically distributed because their means depend on the temperature history which is different for different observation times

two ways to look at $f_\theta(y_1, y_2, \dots, y_n)$:

fix θ, vary $y_1, y_2, \dots, y_n \rightarrow$ tells you how likely different outcomes are for a fixed distribution

fix $y_1, y_2, \dots, y_n$, vary $\theta \rightarrow$ tells you how likely fixed outcome is for different values of θ

the likelihood takes the second viewpoint where $y_1, y_2, \dots, y_n$ are the observed values of the data:

$L(\theta) = f_\theta(y_1, y_2, \dots, y_n)$ it is a function of θ

if observations are independent:

$$L(\theta) = \prod_{j=1}^{n} f_{\theta j}(y_j)$$

the maximum likelihood (ML) estimate $\hat{\theta}$ is the value of θ that maximizes $L(\theta)$

comment: almost always work with the log likelihood $\log L(\theta)$ which, in the independent case, is:

$$\log L(\theta) = \sum_{j=1}^{n} \log f_{\theta j}(y_j)$$

comment: invariance property – if $\hat{\theta}$ is MLE of θ then $g(\hat{\theta})$ is MLE of $g(\theta)$ – this is useful because it is sometimes easier to find the MLE of a function of the parameter of interest (an example will come up shortly) – other estimates like the least squares estimate do not have this property

comment: key point – in non-Bayesian (or 'frequentist') statistics (of which ML estimation and related methods are a part), θ is fixed but uncertain – it is not a random variable so probability statements about it (e.g., 'the probability that $\theta < 10$ is 0.18') are meaningless – the situation is different in Bayesian statistics where θ is treated as a random variable and such statements can be made (we'll talk a bit about Bayesian methods in this class) – returning to the non-Bayesian perspective, while θ is not a random variable, $\hat{\theta}$ is – it is because it's a function of the observations which are random variables – they are random variables in the sense that under putative repetitions of the process that generated them, you will get different values and these will give different values of $\hat{\theta}$ – so $\hat{\theta}$ has a distribution (usually called its sampling distribution) and, among other things, it has a mean and variance →

comment: ML estimation has (under regularity conditions) optimality properties:

consistency: $E(\hat{\theta}) \to \theta$ as $n \to \infty$ asymptotically unbiased (not always unbiased)

efficiency: $Var\,\hat{\theta}$ is smallest among all consistent estimates

so, provided n is not too small, the average value of $\hat{\theta}$ over repetitions of the experiment is guaranteed to be close to the true value and the variance of these values of $\hat{\theta}$ is as small as it can be for a consistent estimator

comment: as we will see, ML estimation comes with a nice asymptotic theory that allows the construction of confidence intervals and hypothesis tests – this often works very well in moderate or even small samples – if not, bootstrap and related methods usually do the trick

<u>salps</u>

simplify notation: $\mu_o = \mu, Y_j = Y(t_j), \mu_j = \mu(t_j), c_j = \int_0^{t_j} x(u)du$

model is $Y_j \sim Poisson(\mu_j), \mu_j = \mu_o \exp(\beta\, c_j)$

seek MLE of $\theta = \mu_o, \beta$

from Poisson pmf: $f_{\theta j}(y_j) = \frac{\mu_j{}^{y_j} \exp(-\mu_j)}{y_j!}$

so after a little arithmetic:

$$\log L(\mu_o, \beta) = \sum_{j=1}^{n} (y_j \log \mu_j - \mu_j - \log y_j!)$$

a as a general proposition, you can omit additive terms in the log likelihood (or multiplicative terms in the likelihood) that do not depend on θ – they won't affect the maximization and they also won't affect confidence intervals or hypothesis tests – you don't have to omit them but it is usually convenient to (one less thing to calculate) – in this case, you can omit the terms $\log y_j!$ so that:

$$\log L(\mu_o, \beta) = \sum_{j=1}^{n} (y_j \log \mu_j - \mu_j)$$

to find the MLEs $\hat{\mu}_o, \hat{\beta}$, you maximize this numerically

<u>Homework</u>

Here are some data. Fit the model. Plot the 'fitted values' $\hat{\mu}_j = \hat{\mu}_o \exp(\hat{\beta}\, c_j)$ along with the data against t. Also plot $\hat{\mu}_j$ against y_j. Comment (nothing profound).

t_j	c_j	y_j
1	0.46	6
2	1.42	21
3	1.95	44
4	1.65	32
5	0.72	8
6	0.04	7
7	0.25	10
8	1.15	12
9	1.91	35
10	1.84	34
11	1.00	11
12	0.16	3