Another toy example

You tag 1000 territorial animals with transmitting tags.

Every month you survey the area. Assume that you can detect every tag attached to a living organism.

You know (from other studies) that the probability of a tag falling off are: 0.10 in the first month, 0.15 in the second month, 0.2 in the third month, and 0.25 for every month after that.

Can you estimate the per-month probability of death?

You have studied bill width in a population of finches for many years.

You have a standardization technique that converts the widths to standardized widths that follow a Normal distribution with $\mu = 0$ and $\sigma = 1$. The standardization relies on your historical studies of the finches.

There is a drought, and you want to know if the mean width has changed.

Indiv.	standardized bill width
1	0.01009121
2	3.63415088
3	-1.40851589
4	3.70573177
5	-0.94145782

Numerical optimization – minimizing a function by evaluating it at many trial points.

Main points:

1. optimizers can fail to find the global optimum:
   (a) multiple modes are a problem.
   (b) result is often starting point dependent.
2. limited precision in computers $\rightarrow$ rounding error, which complicates termination criteria.

Numerical optimization – practical recommendations.

1. Try multiple starting points.
2. Try multiple optimization algorithms.
3. Reparameterization can help
4. Using derivatives from finite differences can be surprisingly effective – consider BFGS even if you can't calculate the gradient.

Summary of LRT example:

- A test based on the likelihood ratio test statistic is the most powerful hypothesis test.

- If we do not know the value of a parameter that occurs in the likelihood equation, we can estimate it.

- Even if we don't care about the parameter (e.g. σ in our original question); its value *can* affect our hypothesis tests.

- When the likelihood around the MLE looks "normalish" (not at a boundary and not a weird likelihood function), then the the χ^2_k distribution does a nice job of describing the null distribution of the LRT statistic for nested models.

Alternative forms of model selection

The following methods do not assume that models are nested:

minimizing the Akaike Information Criterion:

$$AIC(M|X) = 2k - 2\ln L(\hat{\theta}|X, M)$$

Bayes Factors: B_{01} is the Bayes factor in favor of model 0 over model 1:

$$B_{01} = \frac{\mathbb{P}(X|M_0)}{\mathbb{P}(X|M_1)}$$

This is just a likelihood ratio, but it is not the likelihood evaluated at it maximum, rather it is:

$$\mathbb{P}(X|M_0) = \int \mathbb{P}(X|\theta_0)\mathbb{P}(\theta_0)d\theta_0) \tag{1}$$

where θ_0 is the set of parameters in model 0.

Bayes factors can be approximated using differences in:

$$BIC(M|X) = 2k\ln(n) - 2\ln L(\hat{\theta}|X, M)$$

Better approximations of the Bayes factor are available, but they are usually much more expensive.

Parametric bootstrapping

1. Estimate $\hat{\theta}$
2. Estimate $\hat{\theta}_0$ – the value of θ that agrees with the null, which has the highest likelihood.
3. Calculate $\Lambda = \log L(\hat{\theta}) - \log L(\hat{\theta}_0)$, the observed LR test statistic.
4. Simulate M data sets from $\hat{\theta}_0$. For each replicate i:S
   (a) Estimate $\hat{\theta}_i$
   (b) Estimate $\hat{\theta}_{0i}$
   (c) Calculate $\Lambda_i = \log L(\hat{\theta}_i) - \log L(\hat{\theta}_{0i})$, the observed LR test statistic.

The proportion of the M simulations for which $\Lambda_i \geq \Lambda$ approximates the P-value.

You suspect that a population of big horn sheep are made up of two classes of males based on their fighting ability: Strong and Weak. The proportion of strong individuals is unknown.

Experiment:

- You randomly select 10 pairs of males from a large population.
- For each pair you randomly assign one of them the ID 0 and the other the ID 1.
- You record the # of winner from 2 contests.

Model:

- If two individuals within the same class fight, you expect either outcome to be equally likely.
- If a Strong is paired against a Weak then you expect that the probability that the stronger one wins with some probability, w.
- w is assumed to be the same for every pairing of Strong versus Weak and the same for every fight within such a pairing.

Pair #	winner fight 1	fight 2
1	1	1
2	1	0
3	0	1
4	1	1
5	0	0
6	0	1
7	1	1
8	0	0
9	1	0
10	1	1

What can we say about w?

Pair #	winner fight 1	fight 2
1	1	1
2	1	0
3	0	1
4	1	1
5	0	0
6	0	1
7	1	1
8	0	0
9	1	0
10	1	1

X

$x_1 = 1 \quad x_{11} = 1$
$x_2 = 1 \quad x_{12} = 0$
$x_3 = 0 \quad x_{13} = 1$
$x_4 = 1 \quad x_{14} = 1$
$x_5 = 0 \quad x_{15} = 0$
$x_6 = 0 \quad x_{16} = 1$
$x_7 = 1 \quad x_{17} = 1$
$x_8 = 0 \quad x_{18} = 0$
$x_9 = 1 \quad x_{19} = 0$
$x_{10} = 1 \quad x_{20} = 1$

$$L(w) = \prod_{i=1}^{20} \mathbb{P}(x_i|w)$$

Pair #	winner fight 1	fight 2
1	1	1
2	1	0
3	0	1
4	1	1
5	0	0
6	0	1
7	1	1
8	0	0
9	1	0
10	1	1

$$X$$

$$\begin{array}{ll} x_1 = 1 & x_{11} = 1 \\ x_2 = 1 & x_{12} = 0 \\ x_3 = 0 & x_{13} = 1 \\ x_4 = 1 & x_{14} = 1 \\ x_5 = 0 & x_{15} = 0 \\ x_6 = 0 & x_{16} = 1 \\ x_7 = 1 & x_{17} = 1 \\ x_8 = 0 & x_{18} = 0 \\ x_9 = 1 & x_{19} = 0 \\ x_{10} = 1 & x_{20} = 1 \end{array}$$

$$\mathbb{P}(x_{11} = 1 | x_1 = 1, w) \neq \mathbb{P}(x_{11} = 1 \mid x_1 = 0, w)$$

$$L(w) = \prod_{i=1}^{10} \mathbb{P}(x_i | w) \mathbb{P}(x_{10+i} | x_i, w)$$

	winner	
Pair #	fight 1	fight 2
1	1	1
2	1	0
3	0	1
4	1	1
5	0	0
6	0	1
7	1	1
8	0	0
9	1	0
10	1	1

Y

$y_1 = (1, 1)$
$y_2 = (1, 0)$
$y_3 = (0, 1)$
$y_4 = (1, 1)$
$y_5 = (0, 0)$
$y_6 = (0, 1)$
$y_7 = (1, 1)$
$y_8 = (0, 0)$
$y_9 = (1, 0)$
$y_{10} = (1, 1)$

$$L(w) = \prod_{i=1}^{10} \mathbb{P}(y_i | w)$$

0 = same ram wins both bouts
1 = different rams win

Pair #	winner fight 1	winner fight 2	Z
1	1	1	$z_1 = 0$
2	1	0	$z_2 = 1$
3	0	1	$z_3 = 1$
4	1	1	$z_4 = 0$
5	0	0	$z_5 = 0$
6	0	1	$z_6 = 1$
7	1	1	$z_7 = 0$
8	0	0	$z_8 = 0$
9	1	0	$z_9 = 1$
10	1	1	$z_{10} = 0$

$$L(w) = \prod_{i=1}^{10} \mathbb{P}(z_i|w)$$

Pair #	winner fight 1	fight 2	Z
1	1	1	$z_1 = 0$
2	1	0	$z_2 = 1$
3	0	1	$z_3 = 1$
4	1	1	$z_4 = 0$
5	0	0	$z_5 = 0$
6	0	1	$z_6 = 1$
7	1	1	$z_7 = 0$
8	0	0	$z_8 = 0$
9	1	0	$z_9 = 1$
10	1	1	$z_{10} = 0$

$$L(w) = \prod_{i=1}^{10} \mathbb{P}(z_i|w)$$

$$A = \sum_{i=1}^{10} z_i = 4$$

$$L(w) = \mathbb{P}(Z=0|w)^{(n-A)}\mathbb{P}(Z=1|w)^{A}$$