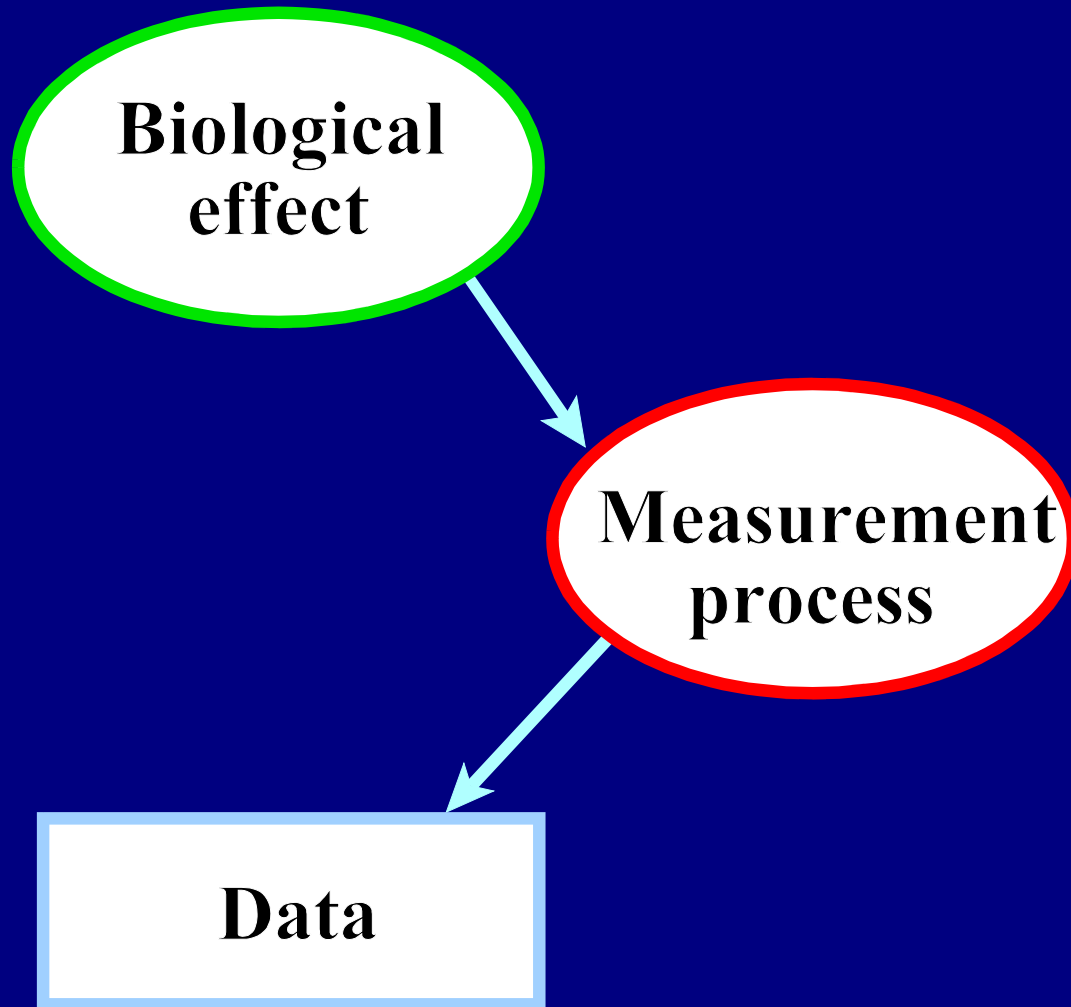


Analysis of Histogram Data and some Remarks on Roundoff Errors in WinBUGS

Krzysztof Muś
Bird Migration Research Station
University of Gdańsk, Poland

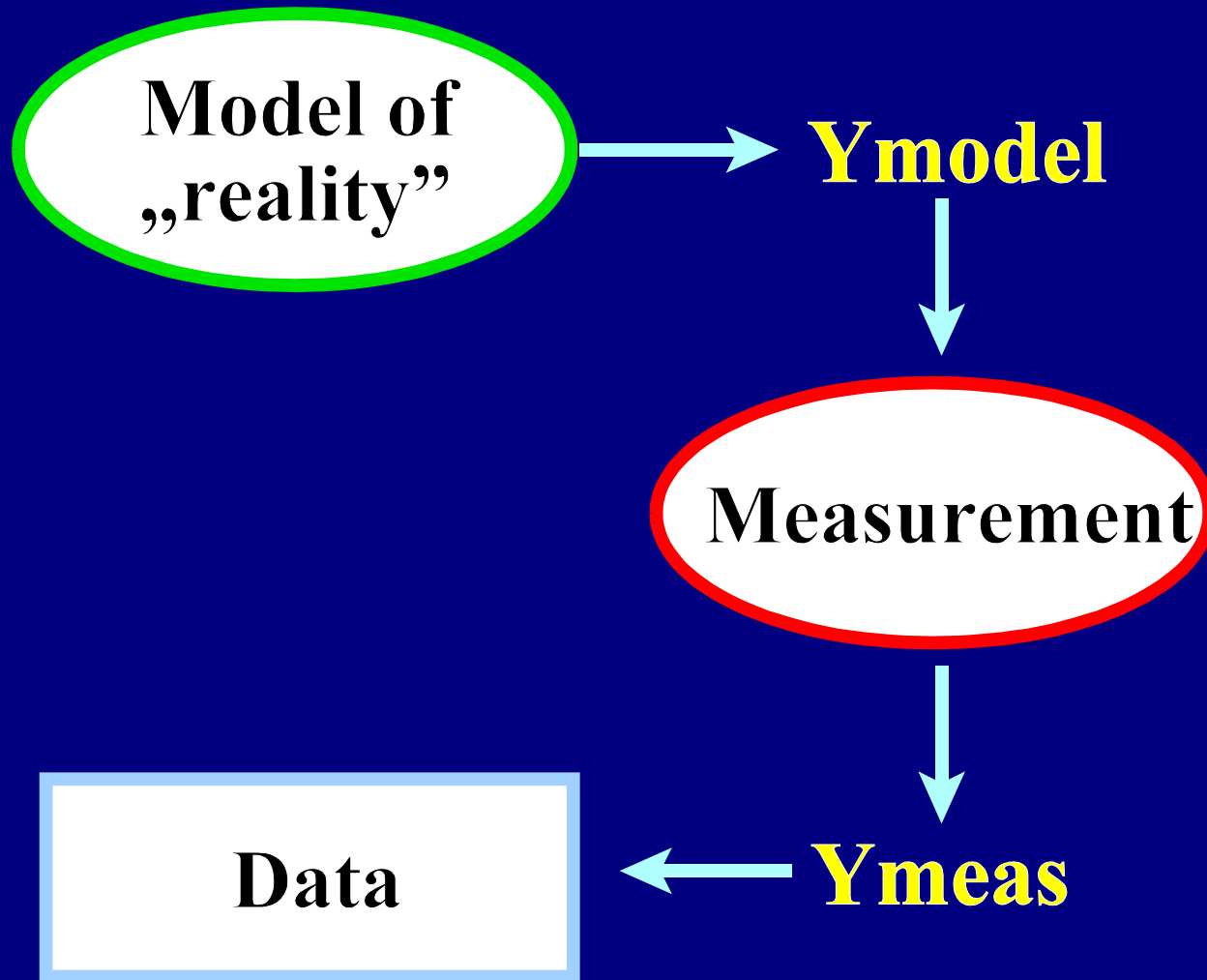
1 Bayesian modelling



2 Bayesian modelling

- „Deductive“ construction of model given model structure and parameters
- Proper stochastic and deterministic „links“ between model elements (nodes)
- Bayes theorem makes inductive (reverse) inference about model and parameters given data
- Model of measurement, i.e. model of transition from latent (unobservable, hypothetical) structure of reality to measured (observed) data should be integral part of the whole model

3 Measurement process



4 Measurement process

- Measurement procedure is an integral part of whole experiment
- In general, our knowledge about this process is much better than about „reality“, so we can construct this part of the whole model more precisely (calibration of devices etc.)
- In short: we want to construct proper mathematical „link“ between our theoretical expectations (Y_{model}) and measured data – it is simply a more or less precise, step by step reconstruction of measurement process

5 Examples

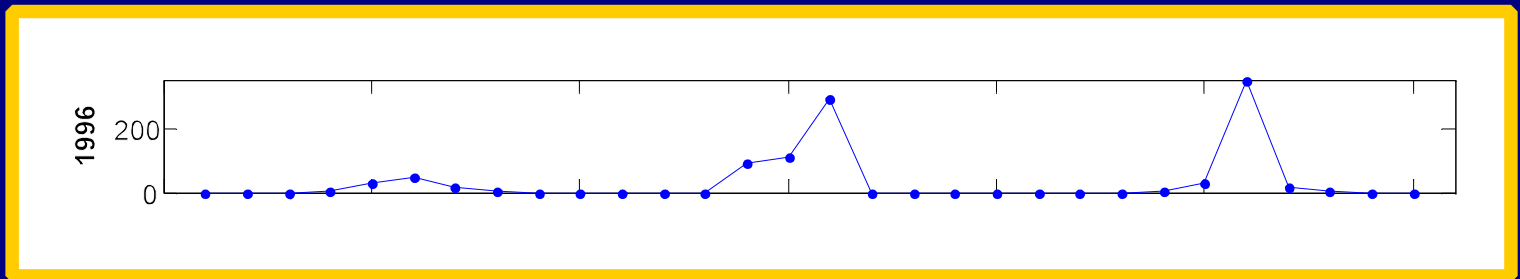
- Most common example is a so-called „random effect“ between „true“ and measured value (here $Y_{\text{meas}} \equiv \text{Data}$):

$$\begin{aligned} Y_{\text{model}} &\leftarrow \text{fun}(\text{params}) \\ \text{Data} &\sim \text{dnorm}(Y_{\text{model}}, \text{tau}) \end{aligned}$$

- In following I would like briefly analyse two types of measurements of frequent occurrence in biology:
 - ▶ *data are coming as counts of some events*
 - ▶ *data are coming as values rounded to some accuracy*

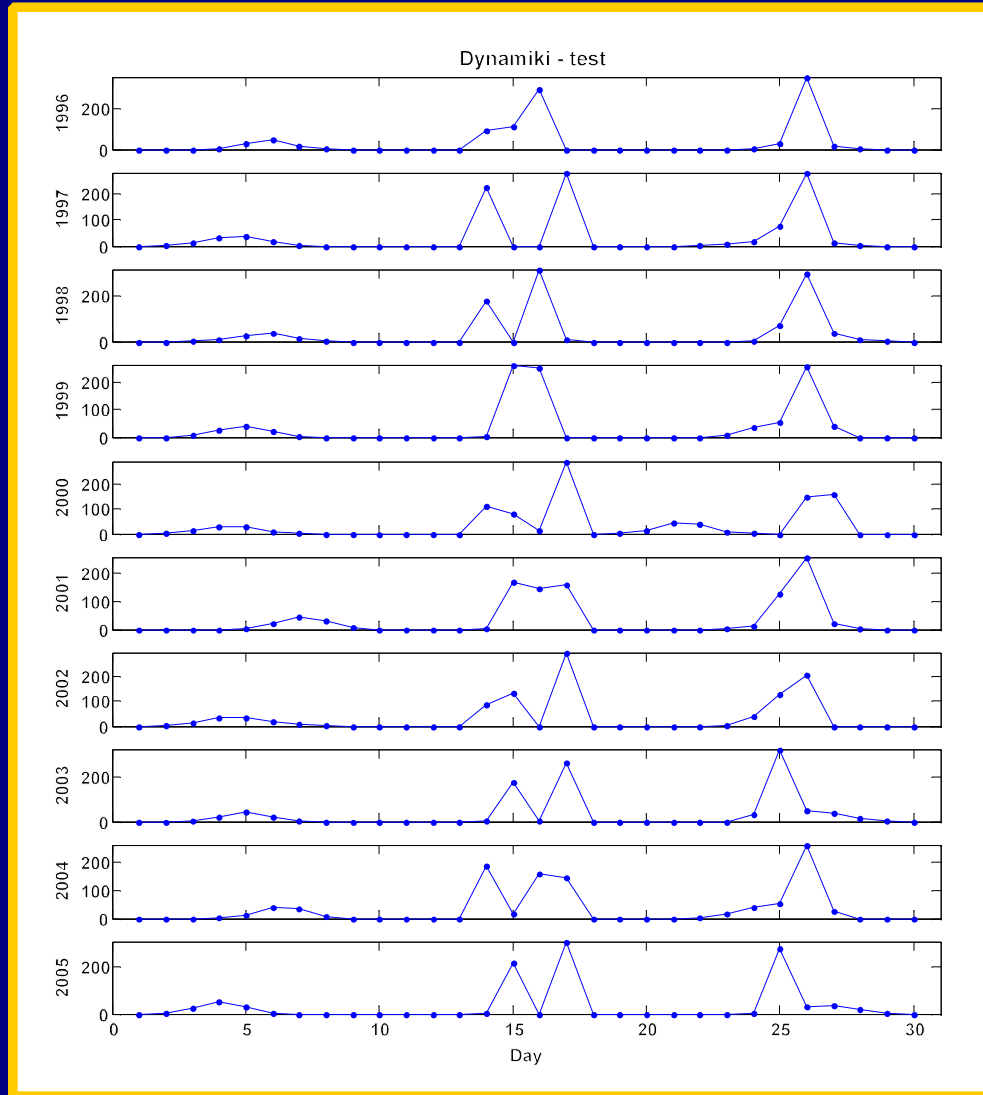
6 Counting of birds

- Our data are simply daily counts of birds of some species caught in ornitological station
- We can easily see that there are groups or „waves“ of birds flying together



- Our goal is to determine repeated structure of these waves over years

7 Repeated structure of „waves”



8 Model of „waves“

- There is sequence of K waves in each year
- $m[1] < m[2] < \dots < m[K]$ are the multi-year means of the centers of waves
- $s[1:K]$ are the typical „durations“ of each wave

9 Model of „waves“ – cont.

- In the year yr ($yr=1,\dots,YR$) wave k has its individual position:

$$mm[yr,k] \sim dnorm(m[k], tau_m)$$

- Also „dispersions“ have individual freedom:

$$ss[yr,k] \sim dnorm(s[k], tau_s)$$

- Relative weights (or probabilities) of waves are totally free:

$$ww[yr,1:K] \sim ddirch(alpha[1:K])$$

10 Model of „waves“ – cont.

- We model $wave[yr,k]$ as gaussian with the mean $mm[yr,k]$, standard deviation $ss[yr,k]$ and relative weight (probability) $ww[yr,k]$
- So probability of catching a bird from the wave k in the day d in the year yr we can write as:

$$ppp[yr,d,k] \leftarrow ww[yr,k] * \left(\text{phi} \left(\frac{d + 1 - mm[yr,k]}{ss[yr,k]} \right) - \text{phi} \left(\frac{d - mm[yr,k]}{ss[yr,k]} \right) \right)$$

- Now we can determine probability of catching any bird (of selected species) in a day d :

$$pp[yr,d] \leftarrow \text{sum}(ppp[yr,d,1:K])$$

11 Model of „waves“ – likelihood

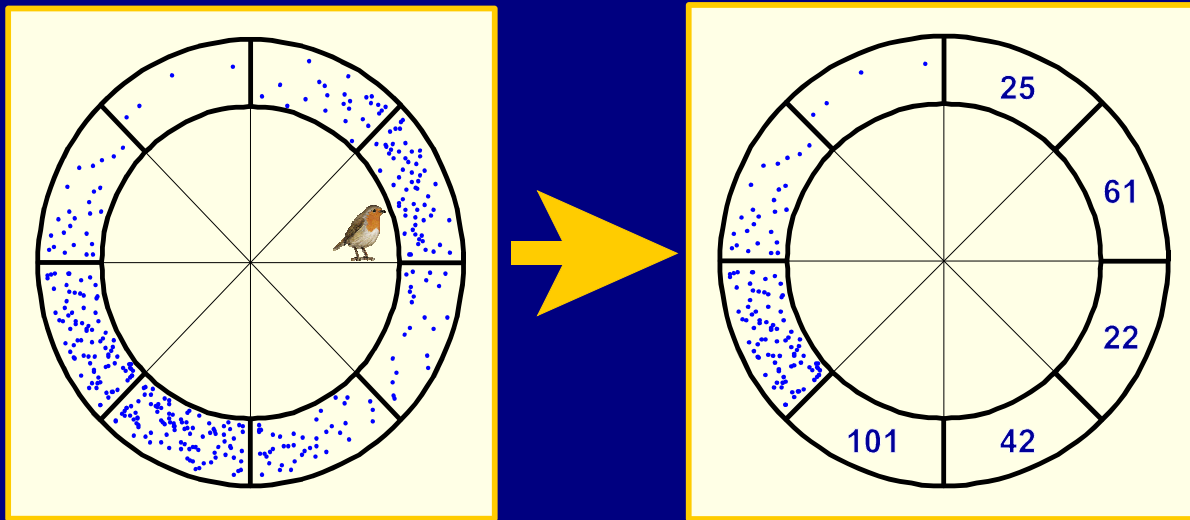
- Given $pp[yr,1:D]$ we can model observed counts as:

$$nn[yr,1:D] \sim \text{dmulti}(pp[yr,1:D], NN[yr])$$

- Of course this one-dimensional model is too simple and can lead to misidentification of similar waves
- It is natural way to extend this simplified picture to multivariate distribution taking also into account biometrical data

12 Orientation experiment

- Another example of histogram data is an orientation experiment on individual bird in Emlen's or Busse's orientation cage
- Now we have counts of points and scratches in each of 8 sectors of the cage



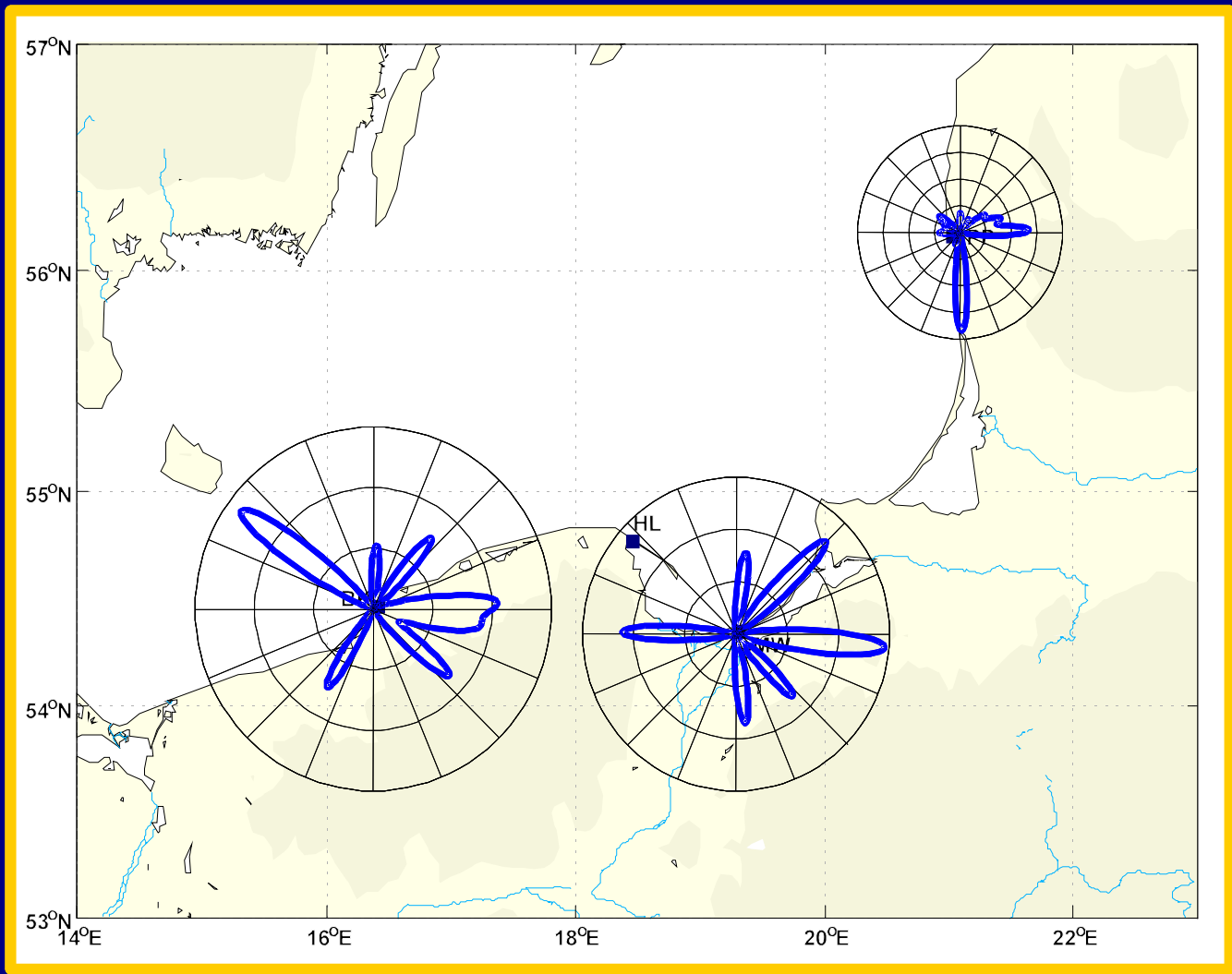
13 Orientation experiment – likelihood

- We model the counts as previously:

$$nn[bird,1:8] \sim dmulti(pp[bird,1:8], Ntot[bird])$$

- $pp[bird,sector]$ is calculated as a weighted sum of K wrapped normal distributions using $\text{phi}((\text{theta}-\mu) / \text{sig})$ with additional sum over angle from -720° to 540°
- Some preliminary results come from the Busse's cage experiments: autumn migration of Robin *Erithacus rubecula*, 1995-2004

● Polish Baltic coast



15 Some remarks on roundoff errors

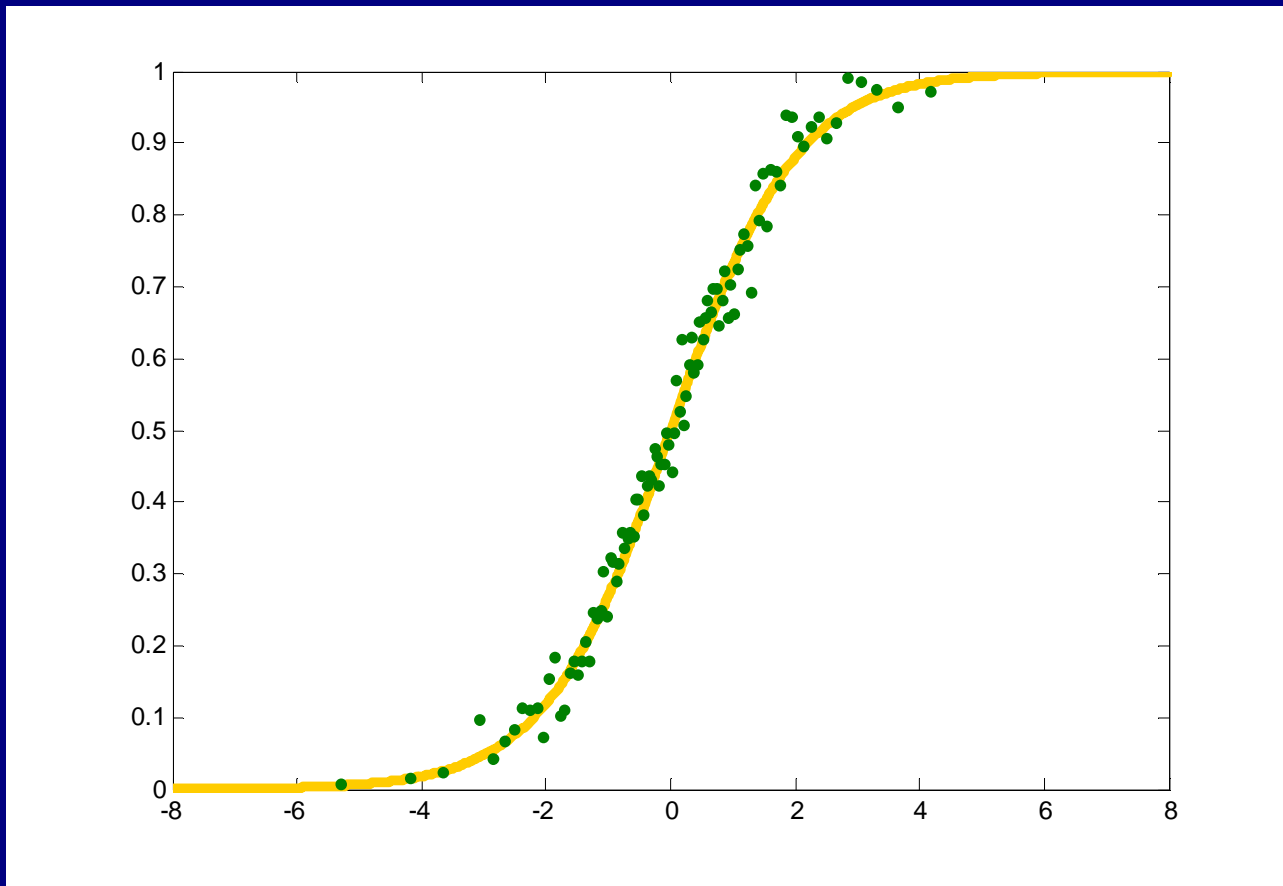
- In regression statistics we usually model errors as $\sim \text{dnorm}(0, \tau)$
- But if we have rounded data we should model errors as uniform in the rounding area:

$$Y_{meas} \sim \text{dunif}(Data - dy/2, Data + dy/2)$$

- But reverse operation is strictly deterministic: having Y_{meas} and dy we can determine $Data$ in unique way
- It is not a trivial problem even for a small roundoff errors

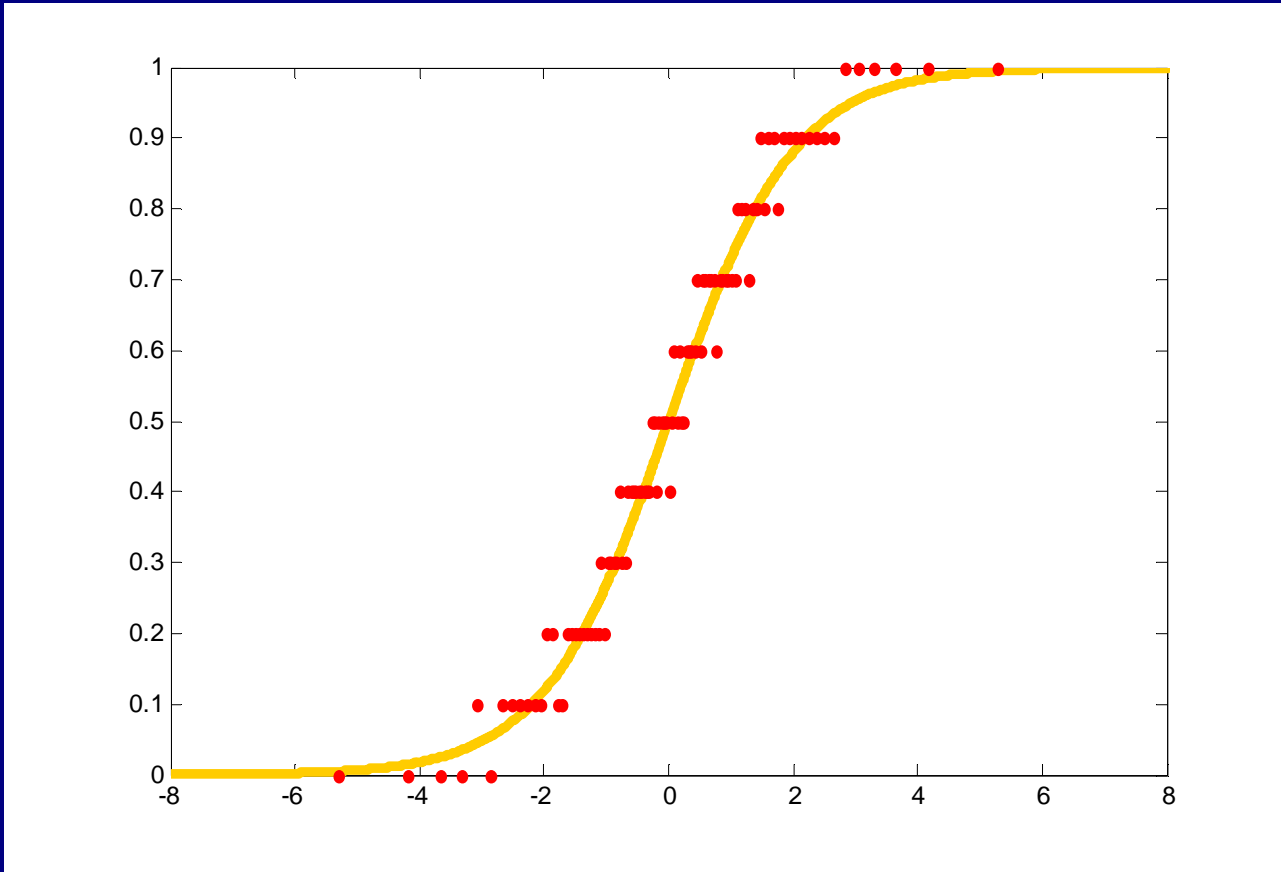
16 Illustration of problem

- Points are normally distributed around the logistic curve

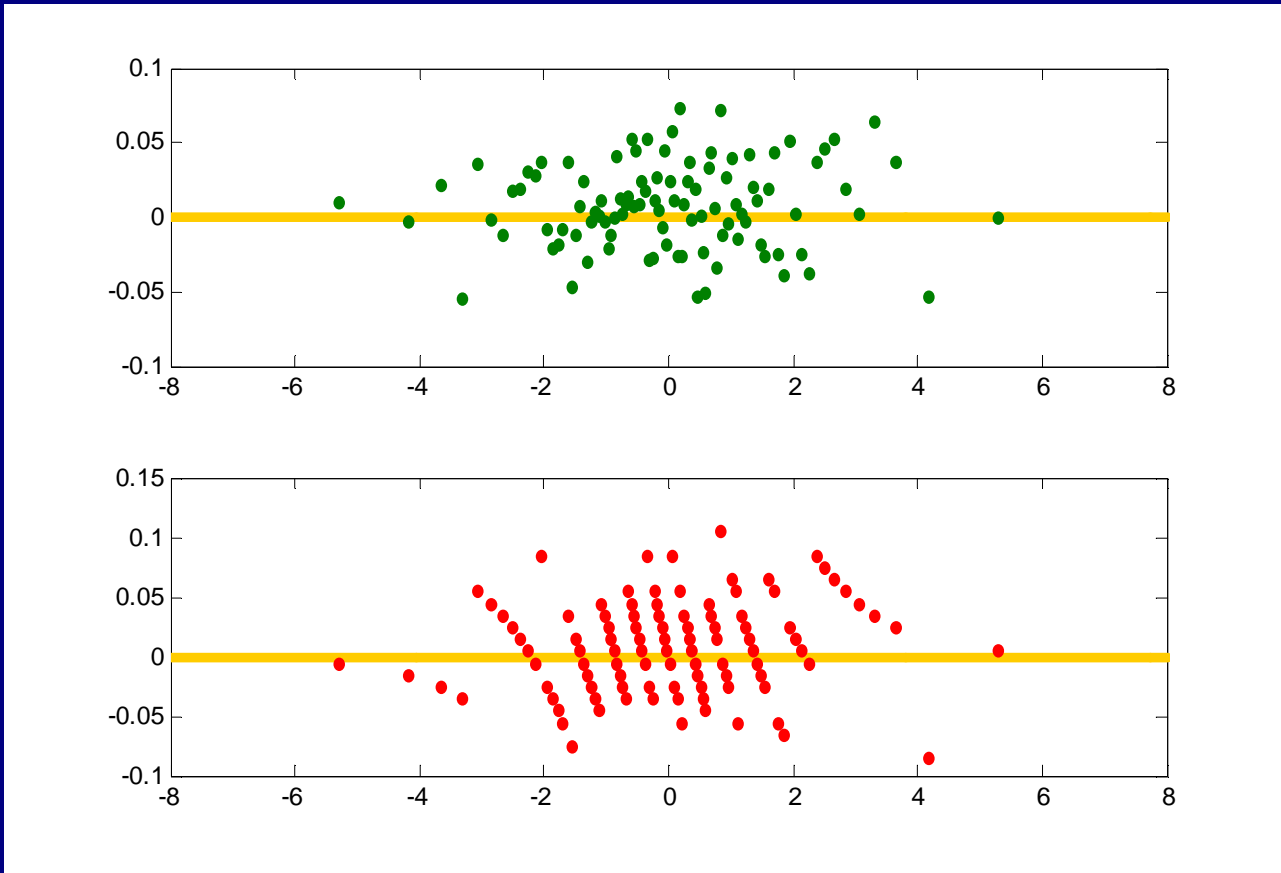


17 Illustration – cont.

- Same points rounded to 0.1



18 Residuals – are they gaussian?



19 Model of rounding

- $Y_{model} \leftarrow \text{fun}(\text{params}, \dots)$
- $Y_{meas} \sim \text{dnorm}(Y_{model}, \text{tau}_{meas})$ (important!)
- $\text{Data} \leftarrow \text{round}(Y_{meas}, dy)$
- Last operation is not possible in WinBUGS. It would be perfectly to implement such construction:

$$\begin{aligned} \text{round}(Y_{meas}, dy) &\leftarrow \text{Data} \\ Y_{meas} &\sim \text{dnorm}(Y_{model}, \text{tau}_{meas}) \\ Y_{model} &\leftarrow \text{fun}(\text{params}, \dots) \end{aligned}$$

- but left-handed functions are not permissible in that tricky construction (also `round()` is not left-handed)

20 Model of rounding in WinBUGS

- In my calculations I found the following solution:

$$Data \sim dnorm(Yround, tau_round)$$
$$Yround \leftarrow dy \star round(Ymeas / dy)$$
$$Ymeas \sim dnorm(Ymodel, tau_meas)$$
$$Ymodel \leftarrow fun(params, \dots)$$

- Model parameters were determined slightly more precisely, but model itself was much more evident than without including of roundoff step ($\Delta DIC \sim 300$ at $dy \approx 0.5\%$)