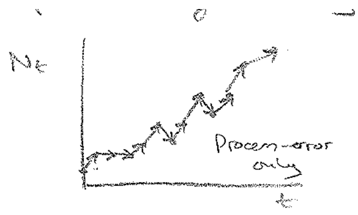
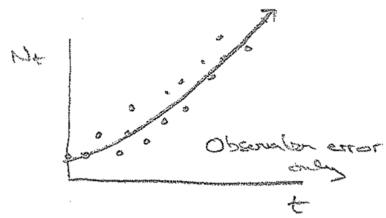
Lecture 6 – Model-fitting - observation & process error

Announcements:
Finish up MSY first ($\sim$ 50 min.)
Next time: Maximum Likelihood, AIC, Paper discussion & start PS3
Concepts:
Process vs. Observation error

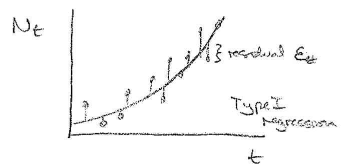
Types of Error
Process error
Environmental stochasticity
Demographic stochasticity
...both include biotic (intra & interspecific intxns.) an abiotic effects on vital rates.
Observation error (aka Measurement error)
Variation due to observers (e.g., bias: experienced vs. novice observer)
Sample estimation (only subset of total observed)

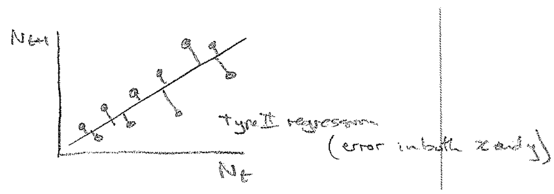
Implications of attributing error:





Estimation of Observation error
"Trajectory matching": Calculate standard deviations of "expected" trajectory given hypothesized model and expected distribution of residuals.





Estimate "best" model parameters using:
- Least-squares regression (Type I or Type II) if $\epsilon \sim \mathcal{N}(\mu, \sigma)$
- Maximum likelihood *will explain later* (in non-Normal residuals)
  (equivalent to least-squares if $\epsilon \sim \mathcal{N}(\mu, \sigma)$)

E.g., Type I regression:

$$N_t = N_0 e^{rt}$$

$$log(N_t) = log(N_0) + rt \qquad \text{"Deterministic/Process model"} \qquad \Rightarrow y = a + bx$$

$$log(N_t) = log(N_0) + rt + \epsilon \qquad \text{"Stochastic/Statistical model"} \qquad \Rightarrow y = a + bx + \epsilon$$

$$\epsilon \sim iid\mathcal{N}(\mu, \sigma)$$

Residual at point t: $\quad \epsilon_t = N_{obs,t} - N_{pred,t}$

"Best" least-squares parameter estimates (i.e. *"most likely given the data."*)
are at $\mu = 0$ and $\sum \epsilon^2 = \min \sum \epsilon^2$ (i.e. *least* squares).

MLE's for slope and intercept (using $\hat{x}$ to indicate estimate):

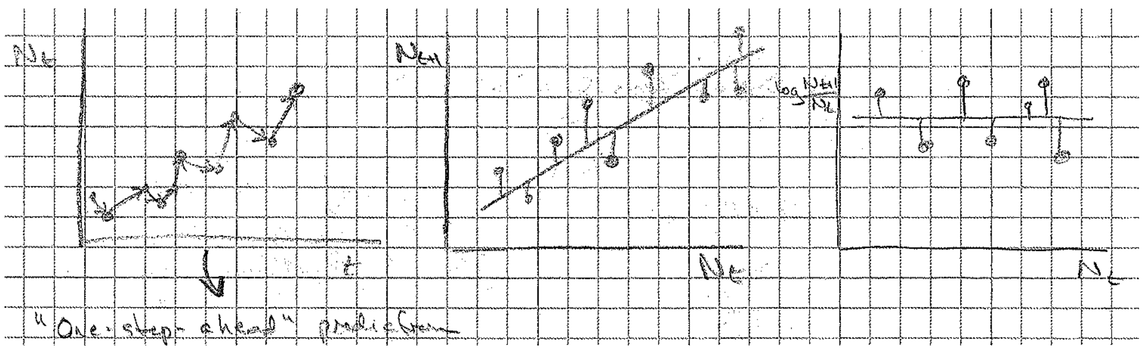
$$\hat{b} = \rho_{xy}\frac{\sigma_y^2}{\sigma_x^2} \quad \text{where } \rho_{xy} \text{ is the Pearson correlation coefficient.} \qquad \Rightarrow \hat{r} = \rho_{N_t,t}\frac{\sigma_{N_t}^2}{\sigma_t^2}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \qquad \Rightarrow \widehat{log(N_0)} = \bar{N}_t - \hat{r}\bar{t}$$

Note, for series of n observations of N:

$$\text{Observer error } = \hat{\sigma}_{obs}^2 = \underbrace{\frac{1}{n-1}}_{\text{small sample correction}} \sum_t^n \left(log(N_{t,obs}) - log(N_{t,pred})\right)^2$$

Estimating Process Error



"One step-ahead" prediction

Estimating error of process at each time-step.

"Process" = growth rate

Step 1: Calculate log-transformed growth rates from data:

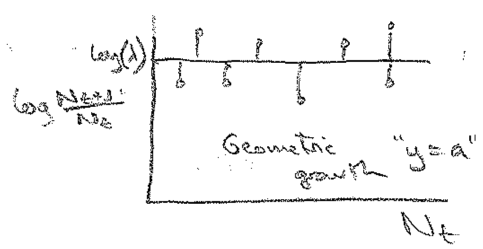
$$\frac{\ln\left(\frac{N_{t+\Delta t}}{N_t}\right)}{\Delta t} \qquad \text{or} \qquad \ln\left(\frac{N_{t+1}}{N_t}\right)$$

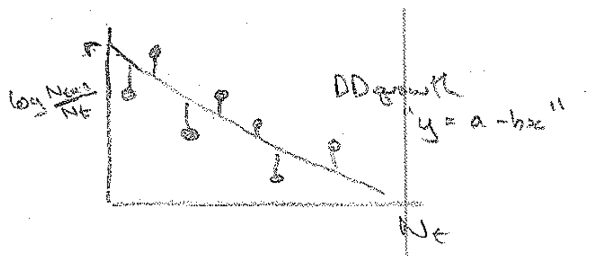
Step 2: Specify hypothesized process model, and rearrange to solve for $\ln\left(\frac{N_{t+1}}{N_t}\right)$

e.g., for geometric growth model:

$$\ln\left(\frac{N_{t+1}}{N_t}\right) = \ln\left(\frac{F(N_t)}{N_t}\right) = \ln\left(\frac{\lambda N_t}{N_t}\right) = \ln\lambda = \ln(e^r) = r$$

Step 3: Specify statistical model and fit to data (LS or MaxLik, etc.):





Estimated process error:

$$\hat{\sigma}_{est}^2 = \frac{1}{n-1}\sum\left(\ln\left(\frac{N_{t+1}}{N_t}\right)_{obs} - \ln\left(\frac{N_{t+1}}{N_t}\right)_{pred}\right)^2$$

R exercise - Load *Class-6.R* and data (Grizzlies, SAFseal1, Redsword)

Walk through code

- observation error
- process error
- prediction w/ process error

Observation error:

$$N_t = N_0 e^{rt} \quad \Rightarrow \quad \ln N_t = \ln N_0 + rt + \epsilon_t$$

in R: $lm(y \sim x)$

Process error:

$$\ln\left(\frac{N_{t+1}}{N_t}\right) = r + \epsilon_t \qquad = \beta_0 + \cancelto{0}{\beta_1} N_t + \epsilon_t \quad \text{for density-independence}$$

in R: $lm(y \sim 1)$ or just $mean()$!

In words, the best estimate for $\hat{r}$ is the mean of all $\ln\left(\frac{N_{t+1}}{N_t}\right)_t$

What if both types!?!

Ignoring observation error will inflate process error.

Standard methods can't disentangle them.

(Rule of thumb is that you're okay if observation error is $< 10\%$ of values.)

Most common solutions:

1 Use independent estimates (replicate surveys of same population)

Let $\hat{N}_t = \frac{1}{n}\sum_i^n N_{t,i}$ for n independent surveys at time t.

Then $\hat{N}$ is nearly unbiased estimator of true N.

For process error:

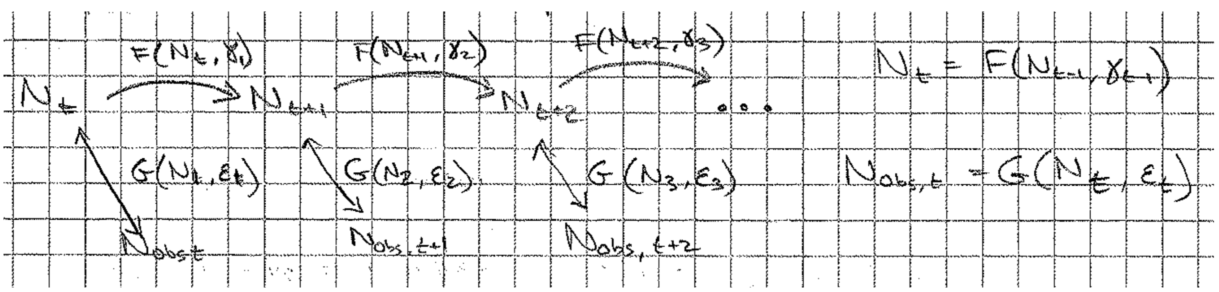
Use $\hat{N}$ to calculate $\ln\left(\frac{\hat{N}_{t+1}}{\hat{N}_t}\right)$ and proceed to estimate σ^2_{proc} (see Morris & Doak eqn. 5.5).

For observation error:

$$\sigma^2_{obs} = \frac{1}{n-1}\sum_i^n \left(N_{t,i} - \hat{N}_t\right)^2$$

2 State-space models

Combine process- w/ observation model:



Methods include Maximum likelihood & Bayesian, using Kalman filters, etc.

...disentangle γ from ϵ assuming specified distributions for these.
