

Writing Bayesian Hierarchical Models

Models for Socio-Environmental Data

Chris Che-Castaldo, Mary B. Collins, N. Thompson Hobbs

December 18, 2020



Examples of

- ▶ Multi-level models (aka random effects)
- ▶ Sampling error in x's and y's
- ▶ Calibration error in y's

All of these will appear in the exercises.

Things to watch for today

Partitioning uncertainty

- ▶ Process variance
- ▶ Sampling variance
- ▶ Calibration variance (aka observation variance)
- ▶ Group level variance

Steps in writing Bayesian models

1. Write your deterministic model. Be careful about support.
2. Draw Bayesian network (DAG) describing relationships between observed and unobserved quantities.
3. Use the Bayesian network to write proportionality between posterior and joint distributions using bracket notation $[|]$.
 - 3.1 Posterior distribution: $[\text{unobserved quantities} | \text{data}]$
 - 3.2 Joint distribution
 - 3.2.1 All nodes in Bayesian network at the heads of arrows (children) must be on the left hand side of a conditioning symbol.
 - 3.2.2 All nodes in Bayesian network at the tails of arrows (parents) must be on the right hand side of a conditioning symbol $|$.
 - 3.2.3 All nodes at the end of an arrow with no arrow coming into them must be expressed unconditionally, i.e., they must have numeric arguments.
4. Assign specific PDF or PMF to each of the brackets.
5. Choose numeric values for parameters of prior distributions.
Do this sensibly! Do not default to vague priors. (Do as I say, not as I do.)

$n \rightarrow Y$

\uparrow

P

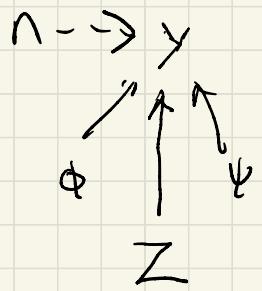
$Y = \text{number positive}$
 tests

$n = \text{number of}$
 tests

$$[P|Y] \text{ & } [Y|P] [P]$$

prevalence = P

$$Y \sim \text{binomial}(n, P)$$



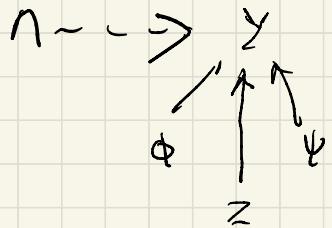
$\gamma = \text{number positive}$
 tests

$n = \text{number of}$
 tests

$Z = \text{true, unobserved}$
 number of infected
 individuals

$\phi = \text{sensitivity, probability that a}$
 test returns a positive conditional
 on an individual being truly infected.

$\psi = \text{specificity, probability that a}$
 test returns a negative result
 conditional on an individual being
 truly uninfected



γ = number positive tests

n = number of tests

z = true, unobserved number of infected individuals

ϕ = sensitivity, probability that a test returns a positive result on an individual being truly uninfected

ψ = specificity, probability that a test returns a negative result conditional on an individual being truly uninfected

$$[\phi, \psi, z | y] \propto [y] \left[\frac{\phi z + (1-\phi)(n-z)}{n}, n \right] [\phi] [\psi] [z]$$

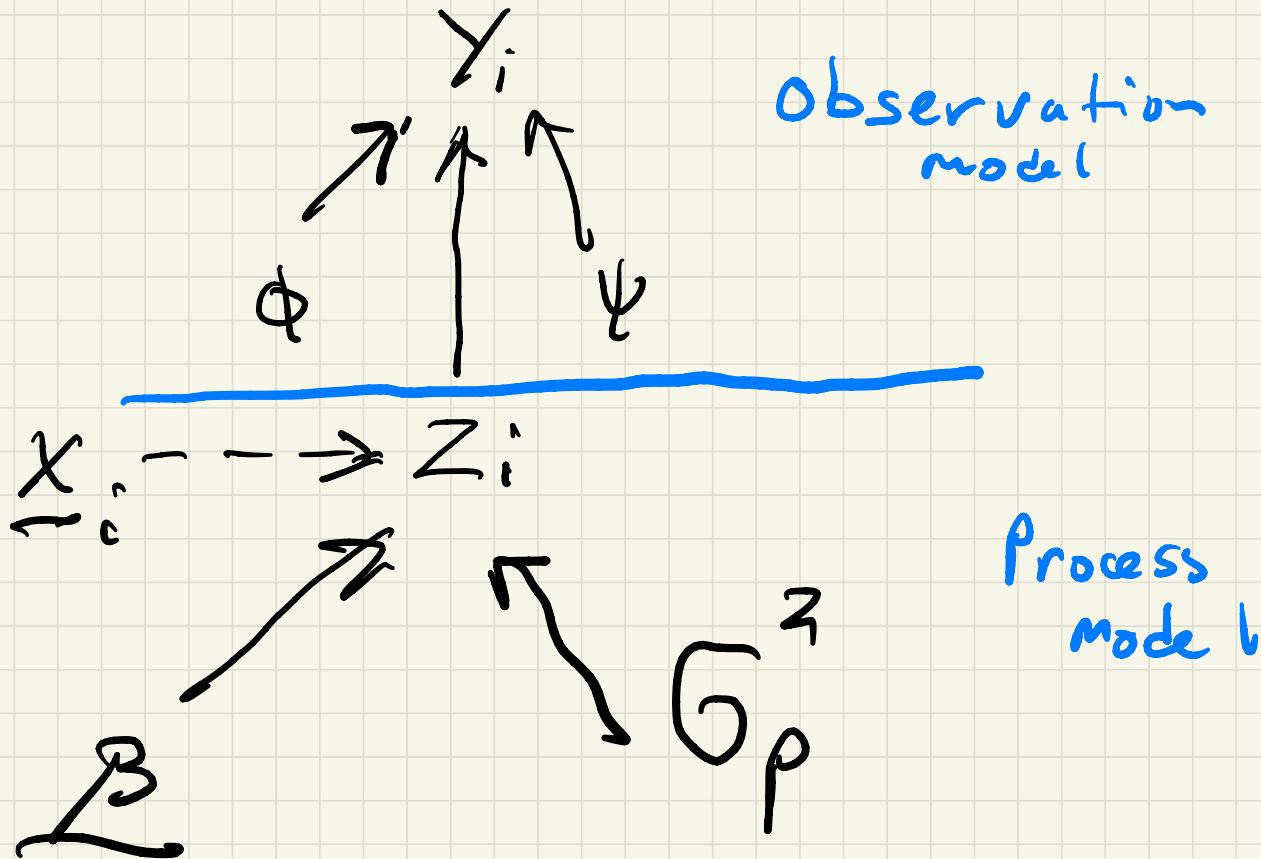
$$\text{prevalence} = \frac{z}{n}$$

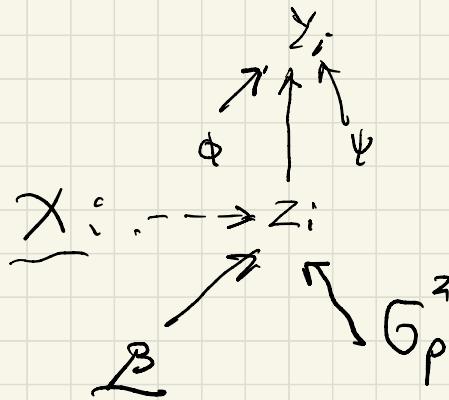
$$y \sim \text{binomial} \left(\frac{z\phi + (n-z)(1-\psi)}{n}, n \right)$$

$$\phi \sim \text{beta}(\alpha_1, \beta_1) \rightarrow \text{must be informed!}$$

$$\psi \sim \text{beta}(\alpha_2, \beta_2)$$

$$z \sim \text{Categorical}(v) \quad v = \left(\frac{1}{n}, \dots, \frac{1}{n} \right) \quad \text{length } n$$





ϕ = sensitivity, probability that a test returns a positive conditional on individual being truly infected

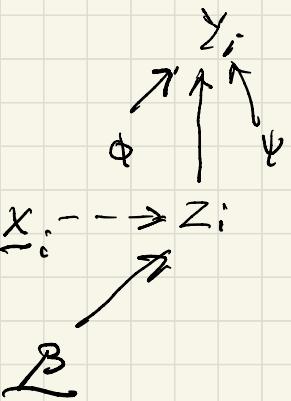
ψ = specificity, probability that a test returns a negative result conditional on an individual being truly uninfected.

$$Z_i = \begin{cases} 1 & \text{if sick} \\ 0 & \text{if healthy} \end{cases}$$

true, unobserved

$$Y_i = \begin{cases} 1 & \text{if positive} \\ 0 & \text{if negative} \end{cases}$$

X_i = vector of covariates



ϕ = sensitivity, probability that a test returns a positive conditional on individual being truly infected

ψ = specificity, probability that a test returns a negative result conditional on an individual being uninfected

$$[\beta, \phi, \psi, z_i | y_i] \sim \prod_{i=1}^n [z_i | x_i; \beta + (1-z_i)(1-\psi)] [z_i | p_i] [\beta] [\phi] [\psi] [\phi]$$

$$y_i \sim \text{Bernoulli}(z_i \phi + (1-z_i)(1-\psi))$$

$$z_i \sim \text{Bernoulli}(p_i)$$

$$\beta_i \sim \text{Normal}(0, 2.5)$$

$$\phi \sim \text{Beta}(-, -)$$

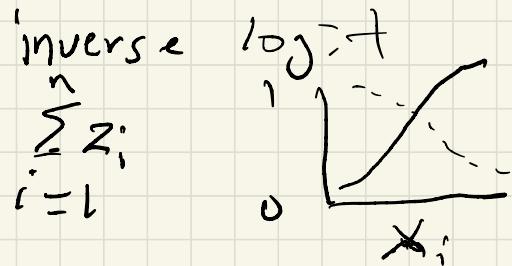
$$\psi \sim \text{Beta}(-, -)$$

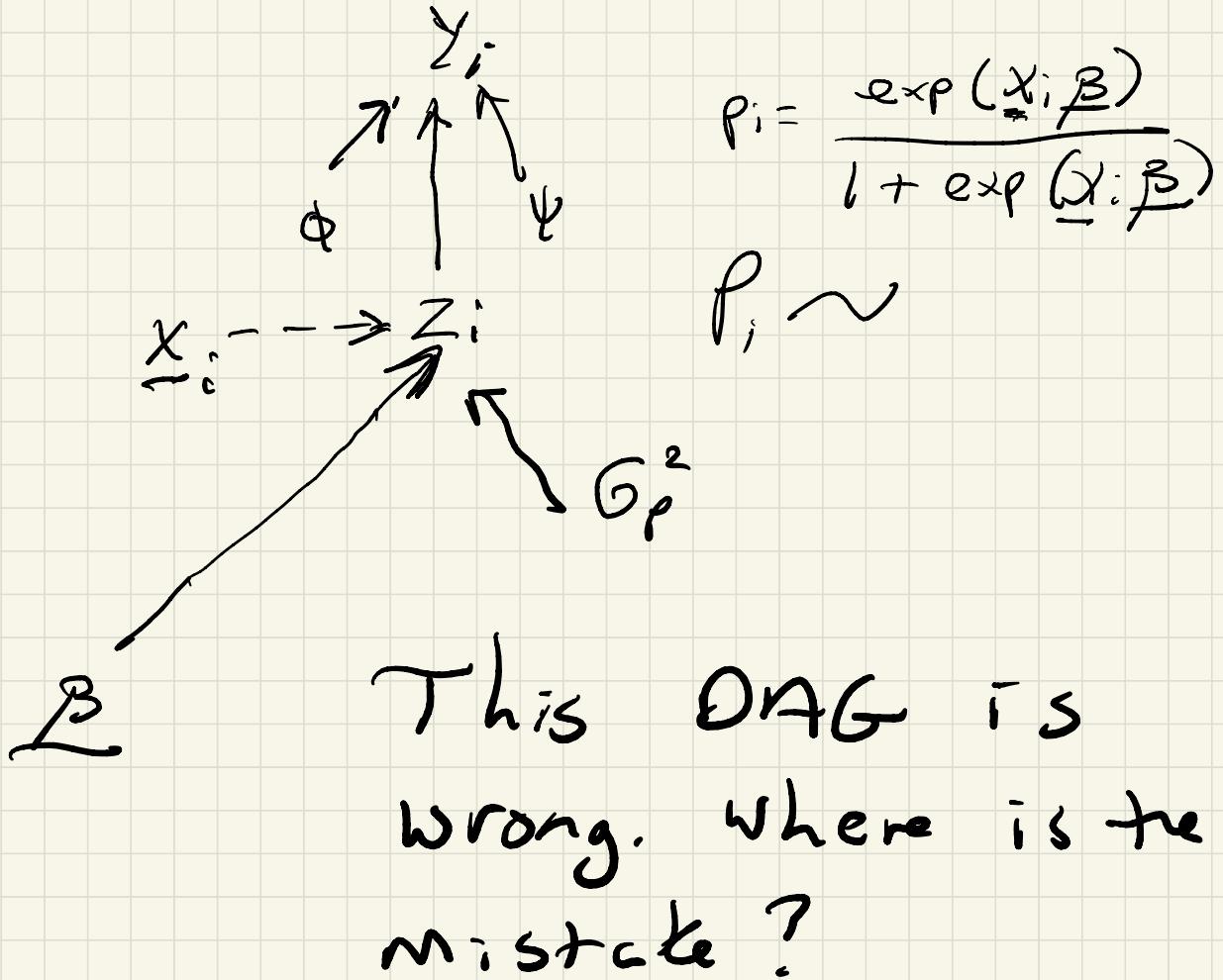
$$\text{prevalence} = \frac{1}{n} \sum_{i=1}^n z_i$$

$$z_i = \begin{cases} 1 & \text{if sick} \\ 0 & \text{if healthy} \end{cases}$$

$$y_i = \begin{cases} 1 & \text{if true, unobserved} \\ 0 & \text{if positive} \\ 1 & \text{if negative} \end{cases}$$

$$p_i = \frac{\exp(x_i \beta)}{1 + \exp(x_i \beta)}$$



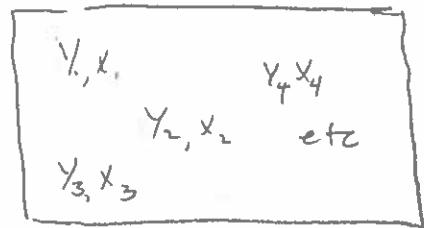


①

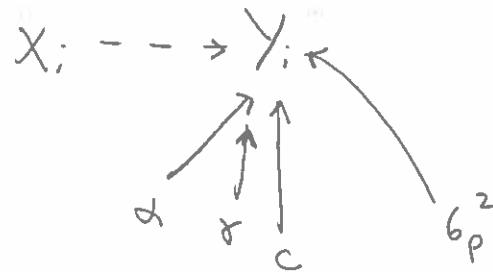
Simple Bayes

$$g(\alpha, \gamma, c, x_i) = \frac{\alpha(x_i - c)}{\alpha + (x_i - c)}$$

height increment



x_i 's
are measured
perfectly



Interpretation of σ^2 : process variance
because y_i are measured perfectly

$$[\alpha, \gamma, c, \sigma_p^2 | \underline{y}] \propto \prod_{i=1}^n [y_i | g(\alpha, \gamma, c, x_i; \sigma_p^2)] [\underbrace{[\alpha] [\gamma] [c] [\sigma_p^2]}_{\text{put these here}}]$$

$$y_i \sim \text{normal}(g(\alpha, \gamma, c, x_i; \sigma_p^2))$$

$$\alpha \sim \text{gamma}\left(\frac{35^2}{4.25^2}, \frac{35}{4.25^2}\right)$$

$$\gamma \sim \text{Uniform}(0, 10)$$

$$\sigma_p \sim \text{uniform}(0, 50) \quad \leftarrow \text{matrix inverse gamma}$$

talk about alternative notations for likelihood?

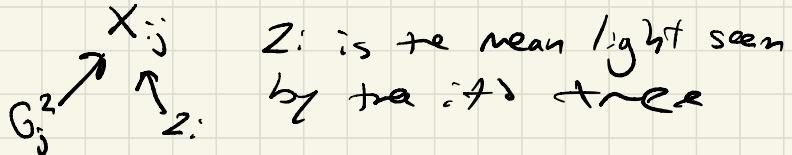
$$[y_i | \mu_i, \sigma^2] \quad \mu_i = \alpha$$

$$\boxed{[y_i | \theta]} \quad \theta = (\gamma, \alpha, c, \sigma^2)$$

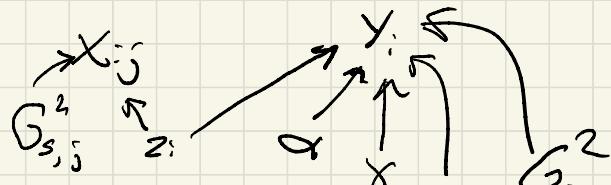
$$[y_i | \alpha, \gamma, c, \sigma^2]$$

Errors in the Covariate: we have $j=1, \dots, 8$ replicate measures of light intensity at random points in the canopy

How would we estimate the mean light seen by the tree?



We now attach this to our model:

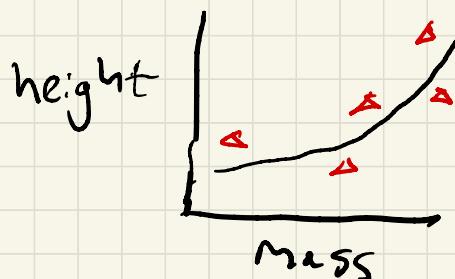


$$\left[G_p^2 \mid G_s^2, z_i, \alpha, \gamma, c \right] \propto \prod_{i=1}^n \prod_{j=1}^8 \left[x_{ij} \mid g(\alpha, \gamma, c, z_i), G_p^2 \right] \left[z_i \mid G_s^2 \right]$$

$$\times \left[\alpha \right] \left[\gamma \right] \left[c \right] \left[G_p^2 \right]$$

× priors

Errors in the y 's: incorporating a calibration model.



$$\text{mass} = M$$

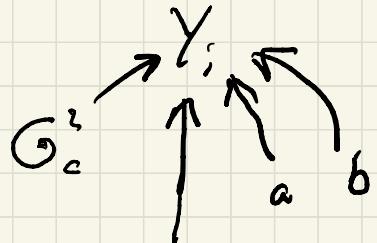
$$\text{height} = h$$

$$M = ah^b$$

$$h = \left(\frac{M}{a}\right)^{\frac{1}{b}}$$

$$[a, b, G_c^2) y] \propto \prod_{i=1}^n \text{gamma} \left(y_i, \sqrt{\frac{(m_i/a)^b}{G_c^2}}, \frac{(m_i/a)^{1/b}}{G_c^2} \right)$$

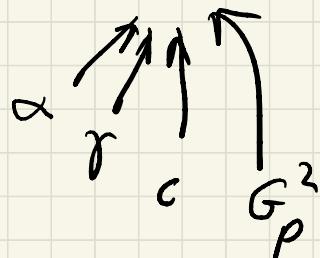
$$\times \text{gamma}(al. 001, .001) \text{gamma}(bl. 001, .001) \text{gamma}(cl. 001, .001)$$



Y_i = observed height

$X_i \rightarrow M_i$ M_i = true, unobserved

mass



$$[m, a, b, G_c^2, \alpha, \beta, \gamma, G_p^2 | Y] \sim \prod_{i=1}^n [Y_i | \left(\frac{m_i}{a}\right)^{\frac{1}{b}}, G_c^2]$$

$$\times [m_i | g(\alpha, \beta, c, x_i), G_p^2]$$

$$\times [\alpha] [\beta] [G_c^2] [\alpha] [\beta] [c] [G_p^2]$$

$$Y_i \sim \text{gamma} \left(\frac{(m_i/a)^{\frac{2}{b}}}{G_c^2}, \frac{(m_i/a)^{\frac{1}{b}}}{G_c^2} \right)$$

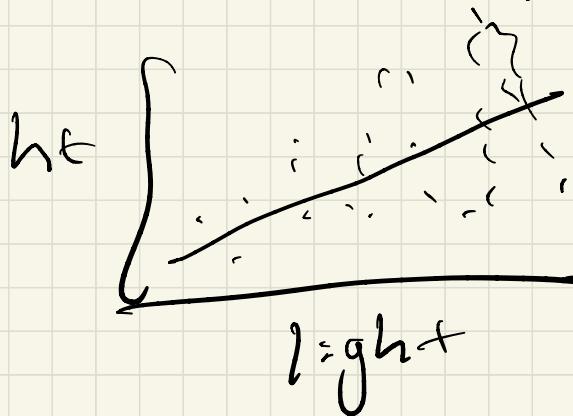
Be Careful with choice of likelihood and moment match.

These will not work:

$$\log(y_i) \sim \text{normal} \left(\log \left(\frac{m}{a} \right)^{\frac{1}{5}}, G_c^2 \right)$$

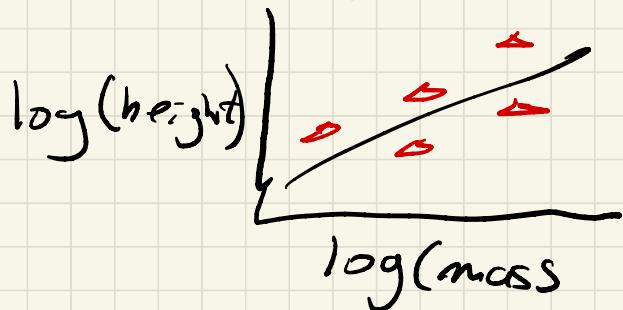
$$y_i \sim \text{lognormal} \left(\log \left(\frac{m}{a} \right)^{\frac{1}{5}}, G_c^2 \right)$$

Why?



If you want to use the lognormal:

- 1) Do Calibration regression on log scale.

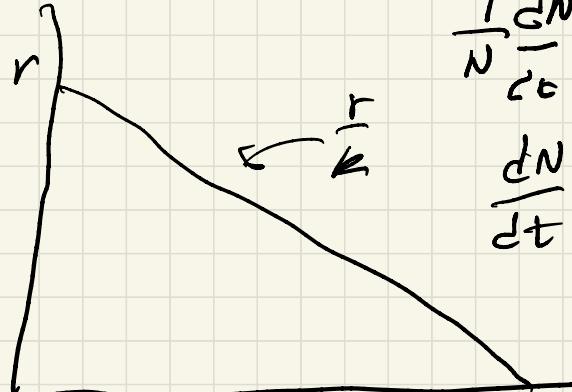


and use previous likelihoods.

- 2) Do calibration as before and moment match for mean and variance, e.g.

$$M_i = \left(\frac{M_i}{\sigma}\right)^{\frac{1}{2}}$$
$$Y_i \sim \text{lognormal} \left(\frac{1}{2} \log \left(\frac{G_c^2 + M_i^2}{\mu^2} \right), \sqrt{\log \left(\left(G_c^2 + M_i^2 \right) / \mu^2 \right)} \right)$$

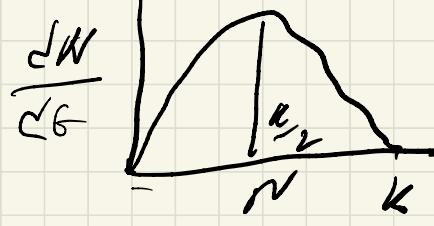
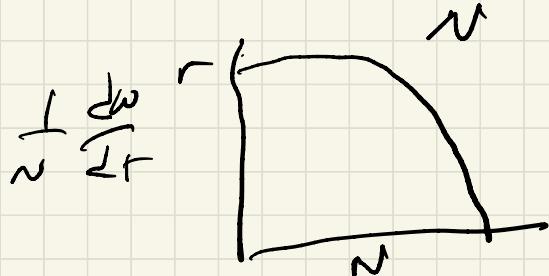
$$\frac{1}{N} \frac{dN}{dt}$$



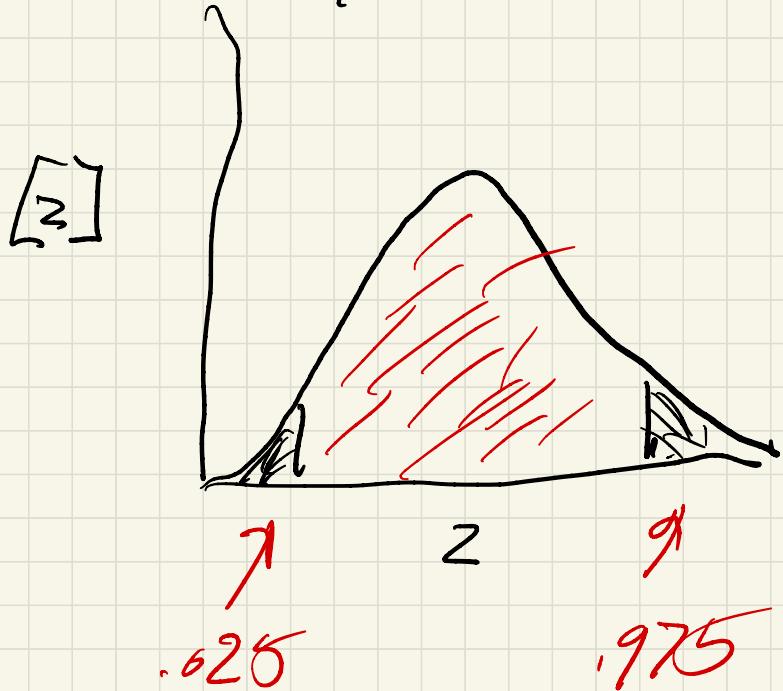
$$\frac{1}{N} \frac{dN}{dt} = r - \frac{r}{k} N$$

$$\frac{dN}{dt} = N \left(r - \frac{r}{k} N \right)$$

$$rN \left(1 - \frac{N}{k} \right)^a$$



equal-tailed



.625

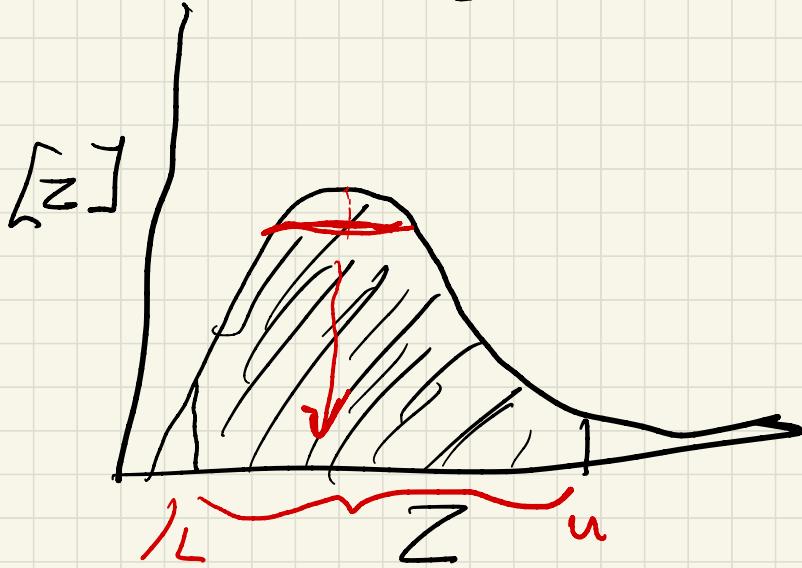
1

2

3

.975

HPOI



The median of the parameter Θ was 42.6 (the highest posterior density in real, BT, HPOI =)

