Statistics 225
Bayesian Statistical Analysis (Part 2)

Hal Stern

Department of Statistics
University of California, Irvine
sternh@uci.edu

March 28, 2019

Hierarchical models – motivation
James-Stein inference

- Suppose $X \sim N(\theta, 1)$
  - X is admissible (not dominated) for estimating θ with squared error loss
- Now $X_i \sim N(\theta_i, 1),\ \ i = 1, \ldots, r$
  - $X = (X_1, \ldots, X_r)$ is admissible if $r = 1, 2$ but not $r \geq 3$
  - for $r \geq 3$

$$\delta_i = (1 - \frac{r-2}{\sum_i X_i^2})X_i$$

    yields better estimates
  - known as James-Stein estimation

Hierarchical models – motivation

James-Stein inference (cont'd)

- The Bayes view: $X_i \sim N(\theta_i, 1)$ and $\theta_i \sim N(0, a)$
  - posterior distn: $\theta_i | X_i \sim N$
  - posterior mean is $(1 - \frac{1}{a+1})X_i$
  - need to estimate a; one natural approach yields James-Stein
- Summary
  - estimation results depend on loss function
  - squared-error loss do well on avg but maybe poor for one component
  - powerful lesson about combining related problems to get improved inferences

Hierarchical Models

Suppose we have data

$$Y_{ij} \quad \begin{array}{l} j = 1, \ldots, J \\ i = 1, \ldots, n_j \end{array}$$

such that $Y_{ij} \quad i = 1, \ldots, n_j$ are independent given θ_j with distribution $p(Y|\theta_j)$. e.g. $\underbrace{\textit{scores}}_{Y}$ for $\underbrace{\textit{students}}_{(i)}$ in $\underbrace{\textit{classrooms}}_{(j)}$ It might be reasonable to expect θ_j's to be "similar" (but not necessarily identical).

Therefore, we may perhaps try to estimate population distribution of θ_j's. This is achieved in a natural way if we use a prior distribution in which the θ_j's are viewed as a sample from a common *population distribution*.

Hierarchical Models

- **Key:** The observed data, y_{ij}, with units indexed by i within groups indexed by j, can be used to estimate aspects of the population distribution of the θ_j's even though the values of θ_j are not themselves observed.
- **How?** It is natural to model such a problem hierarchically
  - observable outcomes modeled conditionally on parameters θ
  - θ given a probabilistic specification in terms of other parameters, ϕ, known as *hyperparameters*.

Hierarchical Models

- Nonhierarchical models are usually inappropiate for hierarchical data. Why?
  - a single θ (i.e., $\theta_j \equiv \theta \ \forall j$) may be inadequate to fit a combined data set.
  - separate unrelated θ_j are likely to "overfit" data.
  - information about one θ_j can be obtained from others' data.
- Hierarchical model uses many parameters but population distribution induces enough structure to avoid overfitting.

Setting up hierarchical models

Exchangeability

Recall: A set of random variables $(\theta_1, \ldots, \theta_k)$ is **exchangeable** if the joint distribution is invariant to permutations of the indexes $(1, \ldots, k)$.
The indexes contain no information about the values of the random variables.

- hierarchical models often use exchangeable models for the prior distribution of model parameters
- iid random variables are one example
- seemingly non-exchangeable r.v.'s may become exchangeable if we condition on all available information (e.g., regression analysis)

Setting up hierarchical models
Exchangeable models

- Basic form of exchangeable model
  - $\theta = (\theta_1, \ldots, \theta_k)$ are independent conditional on additional parameters ϕ (known as hyperparameters)

$$p(\theta|\phi) = \prod_{j=1}^{k} p(\theta_j|\phi)$$

  - ϕ referred to as hyperparameter(s) with hyperprior distn $p(\phi)$
  - implies $p(\theta) = \int p(\theta|\phi)p(\phi)d\phi$
  - work with joint posterior distribution, $p(\theta, \phi|y)$
- One objection to exchangeable model is that we may have other information, say (X_j). In that case may take

$$p(\theta_1, \ldots, \theta_J|X_1, \ldots, X_J) = \prod_{i=1}^{J} p(\theta_i|\phi, X_i)$$

Setting up hierarchical models

- Model is usually specified in nested stages
  - sampling distribution of data $p(y|\theta)$
    (first level of hierarchy)
  - prior (or population) distribution for θ is $p(\theta|\phi)$
    (second level of hierarchy)
  - prior distribution for ϕ (hyperprior) is $p(\phi)$
  - Note: more levels are possible
  - hyperprior at highest level is often diffuse but improper priors must be checked carefully to avoid improper posterior distributions.

Setting up hierarchical models

- Inference
  - Joint distn:

$$\begin{aligned} p(y, \theta, \phi) &= p(y|\theta, \phi)p(\theta|\phi)p(\phi) \\ &= p(y|\theta)p(\theta|\phi)p(\phi) \end{aligned}$$

  - Posterior distribution

$$\begin{aligned} p(\theta, \phi|y) &\propto p(\phi)p(\theta|\phi)p(y|\theta) \\ &= p(\theta|y, \phi)p(\phi|y) \end{aligned}$$

    - often $p(\theta|\phi)$ is conjugate for $p(y|\theta)$
    - if we know (or fix) ϕ: $p(\theta|y, \phi)$ follows from conjugacy
    - then need inference for ϕ: $p(\phi|y)$

Computational approaches for hierarchical models

- Marginal model

$$p(y|\phi) = \int p(y|\theta)p(\theta|\phi)d\theta$$

  do inference only for ϕ (e.g. marginal maximum likelihood)
  - this is the approach that is often used in traditional random effects models
  - no inference for θ

Computational approaches for hierarchical models

- Empirical Bayes

$$p(\theta|y, \hat{\phi}) \propto p(y|\theta)p(\theta|\hat{\phi})$$

  - estimate ϕ (often using marginal maximum likelihood)
  - inference for θ conditional on the estimated ϕ
  - underestimates the uncertainty about θ

Computational approaches for hierarchical models

- Hierarchical Bayes (a.k.a. full Bayes)

$$p(\theta, \phi | y) \propto p(y|\theta)p(\theta|\phi)p(\phi)$$

  inference for θ and ϕ

  - full posterior distribution of θ and ϕ is obtained
  - this is the approach we rely on

Hierarchical models and random effects

Animal breeding example

Consider the following mixed linear model commonly used in animal breeding studies

$$Y = X\beta + Zu + e$$

X = design matrix for fixed effects
Z = design matrix for random effects
β = fixed effects parameters
u = random effects parameters
e = individual variation $\sim N(0, \sigma_e^2 I)$

$$Y|\beta, u, \sigma_e^2 \;\sim\; N(X\beta + Zu, \sigma_e^2 I)$$

$$u|\sigma_a^2 \sim N(0, \sigma_a^2 A)$$

(can also think of β as random with $p(\beta) \propto 1$)

Hierarchical models and random effects

Animal breeding example

- Marginal model (after integrating out u)

$$Y|\beta, \sigma_a^2, \sigma_e^2 \sim N(X\beta, \sigma_a^2 ZAZ' + \sigma_e^2 I)$$

- Note: the separation of parameters into θ and ϕ is somewhat ambiguous here:
  - model specification suggests $\phi = \{\sigma_a^2\}$ and $\theta = \{\beta, u, \sigma^e\}$
  - marginal model suggests $\phi = \{\beta, \sigma_a^2, \sigma_e^2\}$ and $\theta = \{u\}$

Hierarchical models and random effects
Animal breeding example

- Empirical Bayes (known as REML/BLUP)

  We can estimate σ_a^2, σ_e^2 by marginal (restricted?) maximum likelihood ($\hat{\sigma}_a^2$, $\hat{\sigma}_e^2$).
  Then

  $$p(u, \beta | y, \hat{\sigma}_a^2, \hat{\sigma}_e^2) \propto p(y|\beta, u, \hat{\sigma}_e^2) p(u|\hat{\sigma}_a^2)$$

  (a joint normal distn)
- Hierarchical Bayes

$$p(\beta, \sigma_a^2, \sigma_e^2, \mu | y) \propto p(y|\beta, u, \sigma_e^2) P(u|\sigma_a^2) p(\beta, \sigma_a^2, \sigma_e^2)$$

Computation with hierarchical models

- Two cases
  - conjugate case ($p(\theta|\phi)$ conjugate prior for $p(y|\theta)$)
    - approach described below
  - non-conjugate case
    - requires more advanced computing
    - problem-specific implementations
- Computational strategy for conjugate case
  - write $p(\theta, \phi|y) = p(\phi|y)p(\theta|\phi, y)$
  - identify conditional posterior density of θ given ϕ, $p(\theta|\phi, y)$ (easy for conjugate models)
  - obtain marginal posterior distribution of ϕ, $p(\phi|y)$
  - simulate from $p(\phi|y)$ and then $p(\theta|\phi, y)$

Computation with hierarchical models

The marginal posterior distribution $p(\phi|y)$

- Approaches for obtaining $p(\phi|y)$
  - integration $p(\phi|y) = \int p(\theta, \phi|y)d\theta$
  - algebra - for a convenient value of θ

$$p(\phi|y) = \frac{p(\theta, \phi|y)}{p(\theta|\phi, y)}$$

- Sampling from $p(\phi|y)$
  - easy if known distribution
  - grid if ϕ is low-dimensional
  - more sophisticated methods (later)

Normal-normal hierarchical model

- Data model
  - $y_j|\theta_j \sim N(\theta_j, \sigma_j^2), j = 1, \ldots, J$ (indep)
  - σ_j^2's are assumed known for now
    (can release this assumption later)
  - motivation: y_j could be a summary statistic
    with (approx) normal distn from the j-th study
    (e.g., regression coefficient, sample mean)
- Prior distn
  - need a prior distn $p(\theta_1, \ldots, \theta_J)$
  - if exchangeable, then model θ's as iid given
    parameters ϕ

Normal-normal hierarchical model: motivation

- Can think of this data model as a one-way ANOVA model (especially if y_j is a sample mean of n_j obs in group j). Typical ANOVA analysis begins by testing:

$$\begin{aligned} H_0 &: \theta_1 = \ldots = \theta_J \\ H_a &: \text{not } H_0 \end{aligned}$$

  - If we don't reject H_0, we might prefer to estimate each θ_j by the pooled estimate,

$$\bar{y}_{..} = \frac{\sum_{j=1}^{J} \frac{1}{\sigma_j^2} y_j}{\sum_{j=1}^{J} \frac{1}{\sigma_j^2}}$$

  - If we reject H_0, we might use separate estimates, $\hat{\theta}_j = y_j$ for each j.
  - Alternative: compromise between complete pooling and none at all, e.g., a weighted combination,

$$\theta_j = \lambda_j y_j + (1 - \lambda)\bar{y}_{..} \text{ where } \lambda_j \in (0, 1)$$

Normal-normal hierarchical model

- **Constructing a prior distribution**
  (a) The pooled estimate $\hat{\theta} = \bar{y}_{..}$ is the posterior mean if the J values θ_j are restricted to be equal, with a uniform prior density on the common θ; i.e. $p(\theta) \propto 1$.
  (b) The unpooled estimate $\hat{\theta}_j = y_j$ is the posterior mean if the J values θ_j have independent uniform prior densities on $(-\infty, \infty)$; i.e. $p(\theta_1, \ldots, \theta_J) \propto 1$.
  (c) The weighted combination is the posterior mean if the J values θ_j are iid $N(\mu, \tau^2)$.

  Note: (a) corresponds to (c) with $\tau^2 = 0$
  (b) corresponds to (c) with $\tau^2 \to \infty$

Normal-normal hierarchical model

- Data model $p(y_j|\theta_j) \sim N(\theta_j, \sigma_j^2), j = 1, \ldots, J$

  σ_j^2's assumed known
- Prior model for θ_j's is normal (conjugate)

$$p(\theta_1, \ldots, \theta_J|\mu, \tau) = \prod_{j=1}^{J} N(\theta_j|\mu, \tau^2)$$

  i.e. θ_j's conditionally independent given (μ, τ)
- Hyperprior distribution $p(\mu, \tau)$
  - noninformative distribution for μ given τ, i.e., $p(\mu|\tau) \propto 1$ (this won't matter much because the combined data from all J experiments are highly informative about μ)
  - more on $p(\tau)$ later
  - $p(\mu, \tau) = p(\tau)p(\mu|\tau) \propto p(\tau)$

Normal-normal model: computation

- Joint posterior distribution:

$$
\begin{aligned}
& p(\theta, \mu, \tau | y) \\
& \quad \propto \; p(\mu, \tau) p(\theta | \mu, \tau) p(y | \theta) \\
& \quad \propto \; p(\tau) \prod_{j=1}^{J} N(\theta_j | \mu, \tau^2) \prod_{j=1}^{J} N(y_j | \theta_j, \sigma_j^2) \\
& \quad \propto \; p(\tau) \frac{1}{\tau^J} \exp\left[-\frac{1}{2} \sum_j \frac{1}{\tau^2} (\theta_j - \mu)^2 \right] \exp\left[-\frac{1}{2} \sum_j \frac{1}{\sigma_j^2} (y_j - \theta_j)^2 \right]
\end{aligned}
$$

- Factors that depend only on y and $\{\sigma_j\}$ are treated as constants because they are known
- Posterior distn is a distn on $J + 2$ parameters
- Can compute using MCMC (later) or
- Hierarchical computation:
  1. $p(\theta_1, \ldots, \theta_J | \mu, \tau, y)$
  2. $p(\mu | \tau, y)$
  3. $p(\tau | y)$

Normal-normal model: computation

Conditional posterior distn of θ given μ, τ, y

- Treat (μ, τ) as fixed in previous expression
- Given (μ, τ), the J separate parameters θ_j are independent in their posterior distribution
- $\theta_j | y, \mu, \tau \sim N(\hat{\theta}_j, V_j)$ with

$$\hat{\theta}_j = \frac{\frac{1}{\sigma_j^2}\ y_j + \frac{1}{\tau^2}\ \mu}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}} \quad \text{and} \quad V_j = \frac{1}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}}$$

- Result from simple normal-normal conjugate analysis
- $\hat{\theta}_j$ is weighted average of hyperprior mean and data

Normal-normal model: computation

Marginal posterior distribution of μ, τ given y

- We can analytically integrate the full posterior distn $p(\theta, \mu, \tau|y)$ over θ

$$p(\mu, \tau|y) = \int p(\theta, \mu, \tau|y)\, d\theta$$

- An alternative is to use the marginal model $p(\mu, \tau|y) \propto p(y|\mu, \tau)p(\mu, \tau)$
- Marginal model

$$p(y|\mu, \tau) = \prod_{j=1}^{J} \int \underbrace{N(\theta_j|\mu, \tau)N(\bar{y}_{.j}|\theta_j, \sigma_j^2)}_{\text{quadratic in } y_j}\, d\theta_j$$

$$\Rightarrow \quad y_j|\mu, \tau \sim \text{Normal}$$

$$\begin{aligned} E(y_j|\mu, \tau) &= E(E(y_j|\theta_j, \mu, \tau)) = E(\theta_j) = \mu \\ Var(y_j|\mu, \tau) &= E(Var(y_j|\mu, \tau, \theta_j)) + Var(E(y_j|\mu, \tau, \theta_j)) \\ &= E(\sigma_j^2) + Var(\theta_j) = \sigma_j^2 + \tau^2 \end{aligned}$$

Normal-normal model: computation

Marginal posterior distribution of μ, τ given y

- End result is

$$
\begin{aligned}
p(\mu, \tau | y) &\propto p(\tau) \prod_{j=1}^{J} N(y_j | \mu, \sigma_j^2 + \tau^2) \\
&\propto p(\tau) \prod_{j=1}^{J} (\sigma_j^2 + \tau^2)^{-1/2} \exp\left(- \frac{(y_j - \mu)^2}{2(\sigma_j^2 + \tau^2)} \right)
\end{aligned}
$$

- Note: in non-normal models, it is not generally possible to integrate over θ and rely on the marginal model, so that more elaborate computational methods are needed

Normal-normal model: computation

Posterior distribution of μ given τ, y

- Instead of sampling (μ, τ) on a grid, factor the distribution: $p(\mu, \tau|y) = p(\tau|y)p(\mu|\tau, y)$
- $p(\mu|\tau, y)$ is obtained by looking at $p(\mu, \tau|y)$ and thinking of τ as known:

$$\Rightarrow \quad p(\mu|\tau, y) \propto \prod_{j=1}^{J} N(y_j|\mu, \sigma_j^2 + \tau^2)$$

- This is the posterior distn corresponding to a normal sampling distribution with a noninformative prior density on μ
- Result: $\mu|\tau, y \sim N(\hat{\mu}, V_\mu)$ with

$$\hat{\mu} = \frac{\sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau^2} y_j}{\sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau^2}} \quad \text{and} \quad V_\mu = \frac{1}{\sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau^2}}$$

Normal-normal model: computation

Posterior distribution of τ given y

- $p(\tau|y)$ can be found in two equivalent ways
  - integrate $p(\mu, \tau|y)$ over μ
  - use algebraic form $p(\tau|y) = p(\mu, \tau|y)/p(\mu|\tau, y)$, which must hold for any μ
- Choose the second option, and evaluate at $\mu = \hat{\mu}$ (for simplicity):

$$
\begin{aligned}
p(\tau|y) &\propto \frac{\prod_{j=1}^{J} N(y_j|\hat{\mu}, \sigma_j^2 + \tau^2)}{N(\hat{\mu}|\hat{\mu}, V_\mu)} \\
&\propto V_\mu^{1/2} \prod_{j=1}^{J} (\sigma_j^2 + \tau^2)^{-1/2} \exp\left(-\frac{(y_j - \hat{\mu})^2}{2(\sigma_j^2 + \tau^2)}\right)
\end{aligned}
$$

- Note that V_μ and $\hat{\mu}$ are both functions of τ
- Compute $p(\tau|y)$ on a grid of values of τ

Normal-normal model: computation

Summary

- To simulate from joint posterior distribution $p(\theta, \mu, \tau|y)$:
  1. draw τ from $p(\tau|y)$ (grid approximation)
  2. draw μ from $p(\mu|\tau, y)$ (normal distribution)
  3. draw $\theta = (\theta_1, \ldots, \theta_J)$ from $p(\theta|\tau, y)$ (independent normal distribution for each θ_j)
- Choice of $p(\tau)$
  - $p(\tau) \propto 1$ - proper posterior distribution
  - $p(\log \tau) \propto 1$ - improper posterior distribution (equivalent to $p(\tau^2) \propto 1/\tau^2$ but this common noninformative prior for variances doesn't work in this case
  - discuss further on the next slide
- Then illustrate with SAT coaching example (add to slides or do separately)

Normal-normal model: computation
Hyperprior distribution

- Non-informative or weakly informative prior distributions for τ
  - $p(\tau) \propto 1$ - yields a proper posterior distribution ($J > 2$); can be thought of as limit of $U(0, A)$; sometimes useful to use $U(0, A)$ with A determined by context of problem
  - $p(\log \tau) \propto 1$ - yields an improper posterior distribution; why??
    - this is a common noninformative prior for variances
    - here $1/\tau^2$ assigns infinite mass near $\tau = 0$ and the data can never rule out $\tau = 0$ because the θ_j's are not observable
    - can contrast with σ^2 in usual normal model where data (assuming all y's are not equal) rules out $\sigma^2 = 0$
  - $p(\tau) =$ inverse-gamma(ϵ, ϵ) - proper prior distribution; but does not yield a proper posterior in the limit as $\epsilon \to 0$ so choice of ϵ matters
  - $p(\tau) \propto (1 + \tau^2/A^2\nu)^{-(\nu+1)/2}$ - known as half-t; distn of absolute value of a mean zero t distribution with scale parameter A and degrees of freedom ν (see Gelman 2006)

Beta-binomial example

- Series of toxicology studies
- Study j: n_j exchangeable individuals
  y_j develop tumors
- Model specification:
  - $y_j|\theta_j \sim \text{Bin}(n_j, \theta_j), j = 1, \ldots, J$ (indep)
  - $\theta_j, j = 1, \ldots, J \mid \alpha, \beta \sim \text{Beta}(\alpha, \beta)$ (iid)
  - $p(\alpha, \beta)$ – to be specified later, hopefully "non" or "weakly" informative
- Marginal model:
  - can integrate out $\theta_j, j = 1, \ldots, J$ in this case

$$
\begin{aligned}
p(y|\alpha, \beta) &= \int \cdot \int \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1} \binom{n_j}{y_j} \theta_j^{y_j}(1-\theta_j)^{n_j-y_j} \, d\theta_1 \cdot d\theta_J \\
&= \prod_{j=1}^{J} \binom{n_j}{y_j} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}
\end{aligned}
$$

  - $y_j, j = 1, \ldots, J$ are ind
  - distn of y_j is known as beta-binomial distn

Beta-binomial example

- Conditional distn of θ's given α, β, y
  - $p(\theta|\alpha, \beta, y) = \prod_j \text{Beta}(\alpha + y_j, \beta + n_j - y_j)$
  - independent conjugate analyses
  - find this by algebra or by inspection of $p(\theta, \alpha, \beta|y)$
  - analysis is thus reduced to finding (and simulating from) $p(\alpha, \beta|y)$
- Marginal posterior distn of α, β

$$p(\alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + y_j)\Gamma(\beta + n_j - y_j)}{\Gamma(\alpha + \beta + n_j)}$$

  - could derive from marginal distn on previous slide
  - could also derive from joint posterior distn
  - not a known distn (on α, β) but easy to evaluate

Beta-binomial example

- Hyperprior distn $p(\alpha, \beta)$
  - First try: $p(\alpha, \beta) \propto 1$ (flat, noninformative?)
  - equivalent to $p(\alpha/(\alpha+\beta), \alpha+\beta) \propto (\alpha+\beta)$ (relevant because $\alpha/(\alpha+\beta)$ is the mean and $1/(\alpha+\beta)$ is roughly proportional to variance)
  - equivalent to $p(\log(\alpha/\beta), \log(\alpha+\beta)) \propto \alpha\beta$
  - check to see if posterior is proper
    - consider diff't cases (e.g., $\alpha \to 0, \beta$ fixed)
    - if $\alpha, \beta \to \infty$ with $\alpha/(\alpha+\beta) = c$, then $p(\alpha, \beta | y) \propto$ constant (not integrable)
    - this is an improper distn
    - contour plot would also show this (lots of probability extending out towards infinity)

Beta-binomial example

- Hyperprior distn $p(\alpha, \beta)$
  - Second try: $p(\alpha/(\alpha+\beta), \alpha+\beta) \propto 1$
    (flat on prior mean and precision)
    - more intuitive, these two params are plausibly independent
    - equivalent to $p(\alpha, \beta) \propto 1/(\alpha+\beta)$
    - still leads to improper posterior distn
  - Third try: $p(\log(\alpha/\beta), \log(\alpha+\beta)) \propto 1$
    (flat on natural transformation of prior mean and variance)
    - equivalent to $p(\alpha, \beta) \propto 1/(\alpha\beta)$
    - still leads to improper posterior distn
  - Fourth try: $p(\alpha/(\alpha+\beta), (\alpha+\beta)^{-1/2}) \propto 1$
    (flat on prior mean and prior s.d.)
    - equivalent to $p(\alpha, \beta) \propto (\alpha+\beta)^{-5/2}$
    - "final answer" - proper posterior distn
    - equivalent to $p(\log(\alpha/\beta), \log(\alpha+\beta)) \propto \alpha\beta(\alpha+\beta)^{-5/2}$ (this will come up later)

Beta-binomial example

- Computing
  - later consider more sophisticated approaches
  - for now, use grid approach
    - simulate α, β from grid approx to posterior distn
    - then simulate θ's using conjugate beta posterior distn
  - convenient to use $(\log(\alpha/\beta), \log(\alpha+\beta))$ scale because contours "look better" and we can get away with smaller grid
- Illustrate with rat tumor data (add slides or do separately?)