

# Scale Uncertainty in ALDEx2

Michelle Nixon

May 13, 2024

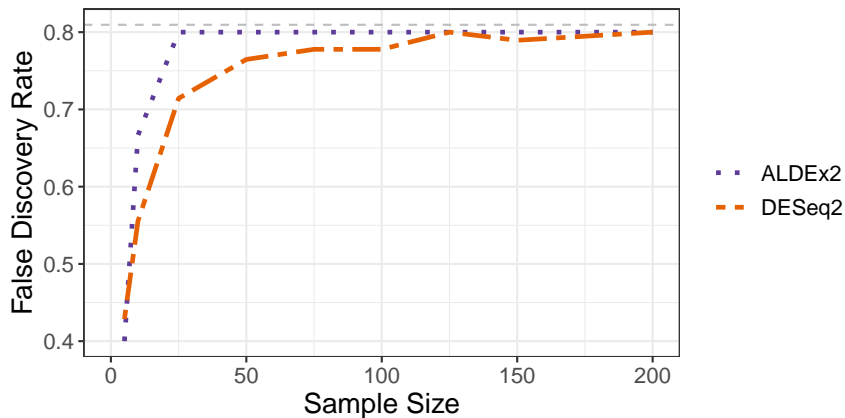
## Recap: Sequencing depth can confound conclusions.

Observed data (Y)	Sample 1	Sample 2	Sample 3	
Condition	Health	Health	Disease	Conclusion
Entity 1	5	10	100	Increase
Entity 2	10	25	3	Decrease
Entity 3	0	1	8	Increase
Entity 4	0	0	19	Increase
Sampling Depth	15	36	130	

This can mislead analyses.

System data (W)	Sample 1	Sample 2	Sample 3	
Condition	Health	Health	Disease	Conclusion
Entity 1	227	351	154	Decrease
Entity 2	684	891	3	Decrease
Entity 3	48	32	15	Decrease
Entity 4	43	39	27	Decrease
Scale ( $W^{\perp}$ )	1,002	1,313	200	

... and lead to unacknowledged bias.



## Problem Set-Up

## Observed Data as a Sample from the System

# Notation

- ▶  $Y$  is a measurement of the underlying system  $W$ .
- ▶  $W$  depends on both the composition ( $W_{dn}^{\parallel}$ ) and system scale ( $W_n^{\perp}$ ):

$$W_{dn} = W_{dn}^{\parallel} W_n^{\perp}$$

$$W_n^{\perp} = \sum_{d=1}^D W_{dn}$$

- ▶  $\theta$  is what we want to estimate.

# Differential Abundance/Expression Analysis

- ▶ Question: How do entities (e.g., taxa or genes) change between conditions?
- ▶ In this case,  $\theta$  is the log-fold change (LFC):

$$\theta_d = \text{mean}_{\text{case}}(\log W_{dn}) - \text{mean}_{\text{control}}(\log W_{dn})$$



# The Original ALDEx2 Model

## Step 1: Model Sampling Uncertainty

$$Y_{.n} \sim \text{Multinomial}(W_{.n}^{\parallel})$$

$$W_{.n}^{\parallel} \sim \text{Dirichlet}(\alpha)$$

## Step 2: Centered Log-Ratio Transformation

$$\log W_{.n} = \left[ \log W_{1n}^{\parallel} - \text{mean}(\log W_{.n}^{\parallel}), \dots, \log W_{Dn}^{\parallel} - \text{mean}(\log W_{.n}^{\parallel}) \right]$$

## Step 3: Calculate LFCs and Test if Different from Zero.

$$\theta_d = \text{mean}_{\text{case}}(\log W_{dn}) - \text{mean}_{\text{control}}(\log W_{dn})$$

# Implied Assumptions about Scale

## Step 1: Model Sampling Uncertainty

$$Y_{\cdot n} \sim \text{Multinomial}(W_{\cdot n}^{\parallel})$$

$$W_{\cdot n}^{\parallel} \sim \text{Dirichlet}(\alpha)$$

## Step 2: Centered Log-Ratio Transformation

$$\log W_{\cdot n} = \left[ \log W_{1n}^{\parallel} - \text{mean}(\log W_{\cdot n}^{\parallel}), \dots, \log W_{Dn}^{\parallel} - \text{mean}(\log W_{\cdot n}^{\parallel}) \right]$$

## Step 3: Calculate LFCs and Test if Different from Zero.

$$\theta_d = \text{mean}_{\text{case}}(\log W_{dn}) - \text{mean}_{\text{control}}(\log W_{dn})$$

## Implied Assumptions about Scale, cont.

Adding Uncertainty in Scale can Help.

## Scale Reliant Inference (Informal)

# Scale Reliant Inference: The Basics

# Scale Reliant Inference: The Basics

- ▶ What happens if  $\theta$  depends on  $W^\perp$ ?
- ▶ Consider LFCs: how are taxa changing between two conditions?

$$\begin{aligned}\theta_d &= \text{mean}_{\text{case}}(\log W_{dn}) - \text{mean}_{\text{control}}(\log W_{dn}) \\ &= \dots \\ &= (\text{mean}_{\text{case}}(\log W_{dn}^\parallel) - \text{mean}_{\text{control}}(\log W_{dn}^\parallel)) \\ &\quad - (\text{mean}_{\text{case}}(\log W_n^\perp) - \text{mean}_{\text{control}}(\log W_n^\perp)) \\ &= \theta^\parallel + \theta^\perp\end{aligned}$$

Don't we need  $\theta^\perp$ ?

# Scale Reliant Inference: Theory Intro

Recall for LFCs:

$$\begin{aligned}\theta_d &= \text{mean}_{\text{case}}(\log W_{dn}) - \text{mean}_{\text{control}}(\log W_{dn}) \\ &= \theta^{\parallel} + \theta^{\perp}\end{aligned}$$

- ▶ What can we say about  $\theta$  from  $\theta^{\parallel}$  alone?
- ▶ E.g. If  $\theta^{\parallel} = 20$ , what does that say about  $\theta$ ? If there are no restrictions, nothing!
- ▶ Statistical perspective:  $\theta$  is not identifiable without  $\theta^{\perp}$ .
- ▶ Practical issues: unbiased estimators, calibrated confidence sets, and type-I error control *NOT* possible!



# Scale Simulation Random Variables

**Goal:** Estimate  $\theta = f(W^{\parallel}, W^{\perp})$ .

1. Draw samples of  $W^{\parallel}$  from a measurement model (can depend on  $Y$ ).
2. Draw samples of  $W^{\perp}$  from a scale model (can depend on  $W^{\parallel}$ ).
3. Estimate samples of  $\theta = f(W^{\parallel}, W^{\perp})$ .

## The Updated ALDEx2 Software

## ALDEx2 as an SSRV

# Benefits of Moving Past Normalizations to Scale

## Coding Changes to ALDEx2

Including scale

## Option 1: Default Scale Model

## Option 2: More Complex Scale Models



# Sensitivity Analyses

## Real Data Examples

## Real Example: SELEX

## Real Example: Vandputte