

Identifiability and reparameterisation methods for inverse problems

Oliver J. Maclaren (oliver.maclaren@auckland.ac.nz)

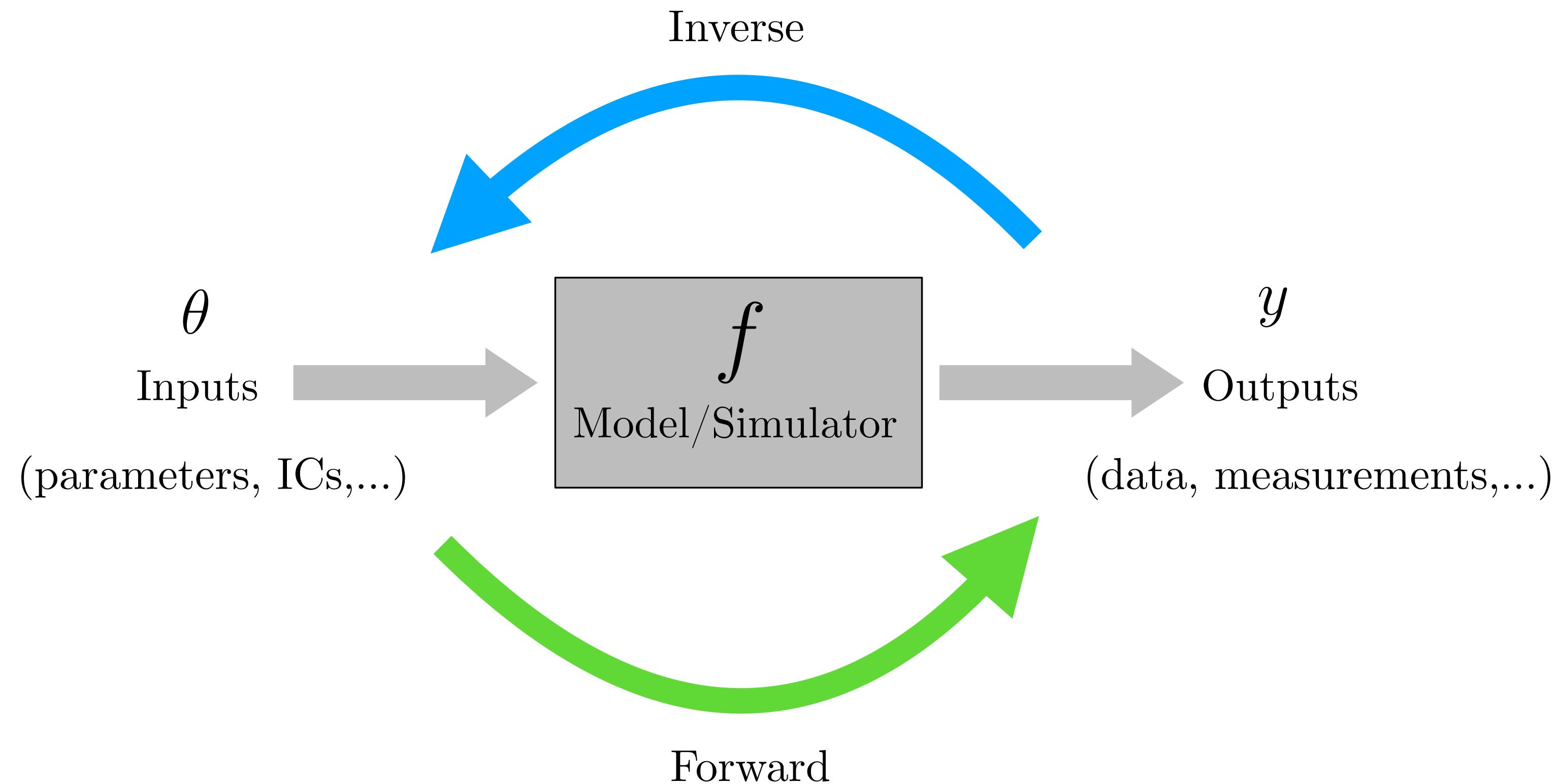
Ruanui Nicholson

Joel A. Trent

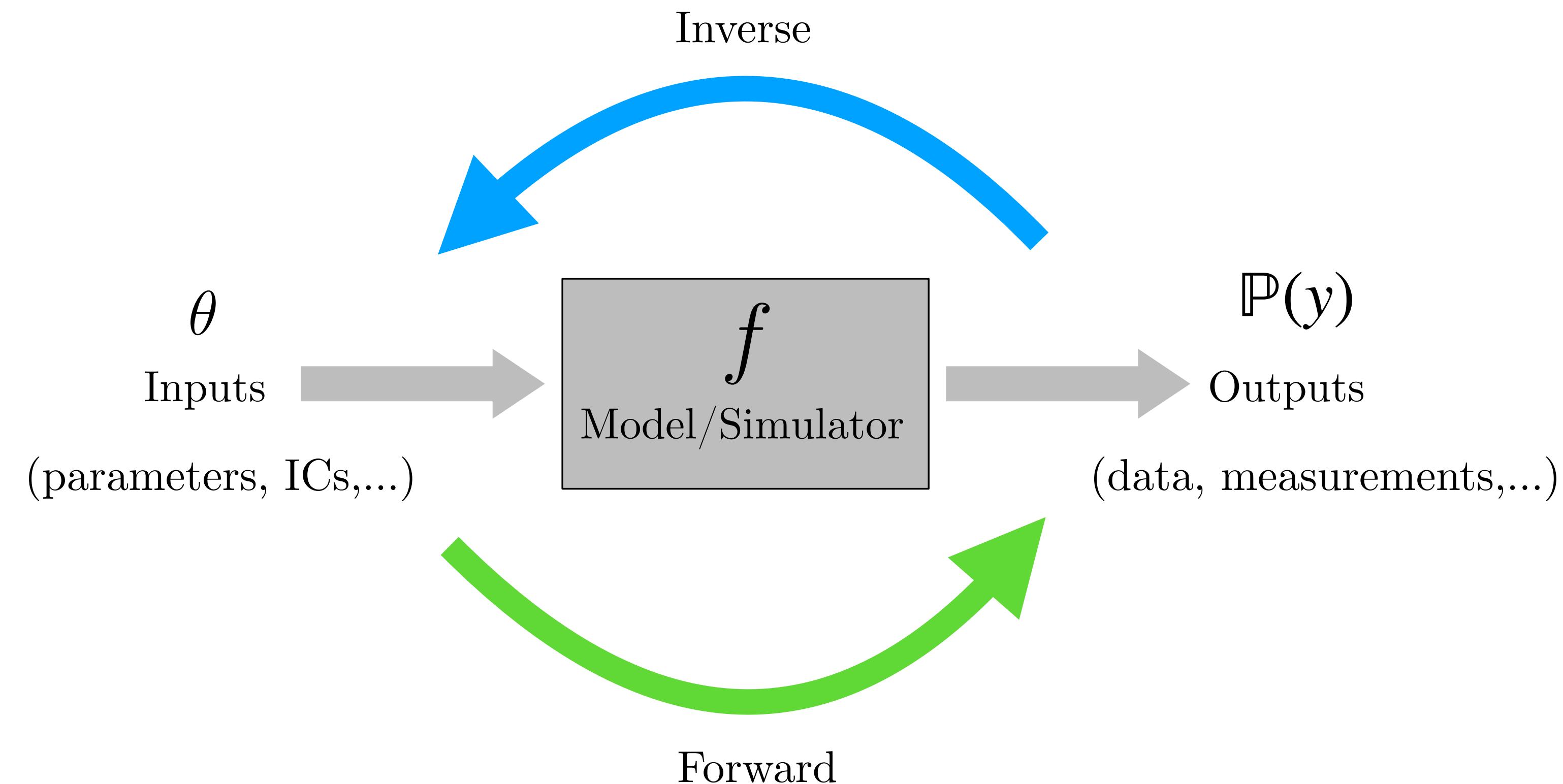
Joshua Rottenberry

Matthew J. Simpson

The basic framework



The basic framework



Hadamard's conditions revisited

Standard well-posedness conditions:

- **Existence:** 'true' data distribution can be supposed to lie in image of parameter-to-distribution mapping
- **Uniqueness:** identifiability of parameter-to-distribution mapping, i.e. injectivity
- **Stability:** given finite samples from the data distribution/given the empirical distribution, we can estimate the parameter sufficiently accurately, e.g. finite and useful confidence intervals

Hadamard's conditions revisited

Much of inverse problems, both classical and Bayesian, essentially focuses on **stability** and **regularisation**-type ideas (including introducing prior info)

Here we (re-)consider **starting from the uniqueness/identifiability** perspective keeping in mind **the need to consider stability**

Identifiability analysis



Matrix workshop on Parameter Identifiability in Mathematical
Biology

Basic setting

Standard setting definitions

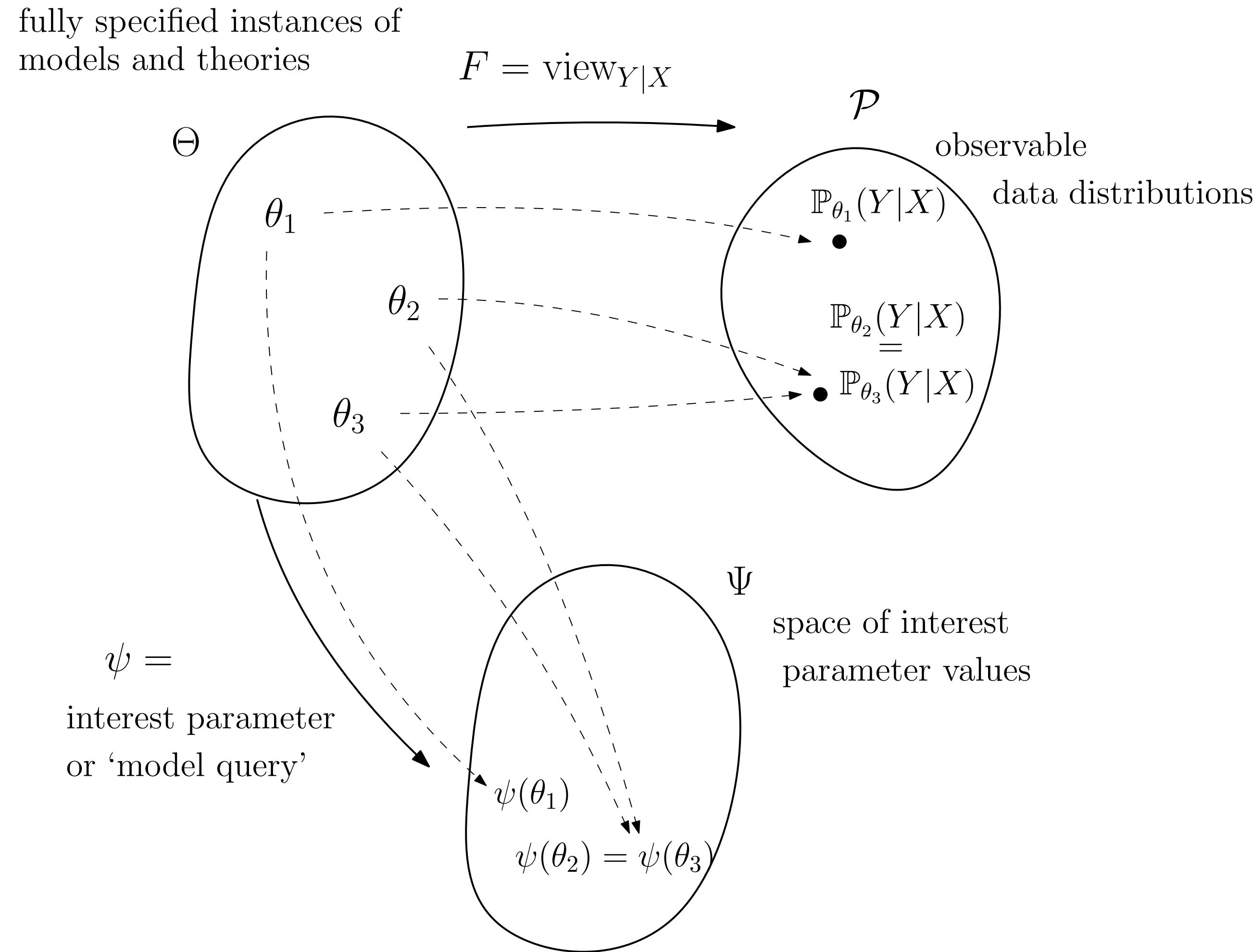
Θ : Parameter space

$\theta_1, \theta_2 \in \Theta$: Specific parameters

$\psi(\theta)$: Interest parameter

$F : \Theta \rightarrow \mathcal{P}, \quad \theta \mapsto F(\theta)$: Forward mapping/statistical experiment

Basic setting



Adapted from Maclaren & Nicholson (2019), 'What can be estimated?'

Definitions

Standard (structural) identifiability definition:

An interest parameter $\psi : \Theta \rightarrow \Psi$ is identifiable with respect to F if and only if for all $\theta_1 \neq \theta_2$,

$$F(\theta_1) = F(\theta_2) \implies \psi(\theta_1) = \psi(\theta_2),$$

or, equivalently, if for all $\theta_1 \neq \theta_2$,

$$\psi(\theta_1) \neq \psi(\theta_2) \implies F(\theta_1) \neq F(\theta_2).$$

Practical identifiability definition: essentially, replace equalities by approximate equalities! I.e., *stability*!

Simple examples

- Poisson limit of normal approximation to binomial distribution

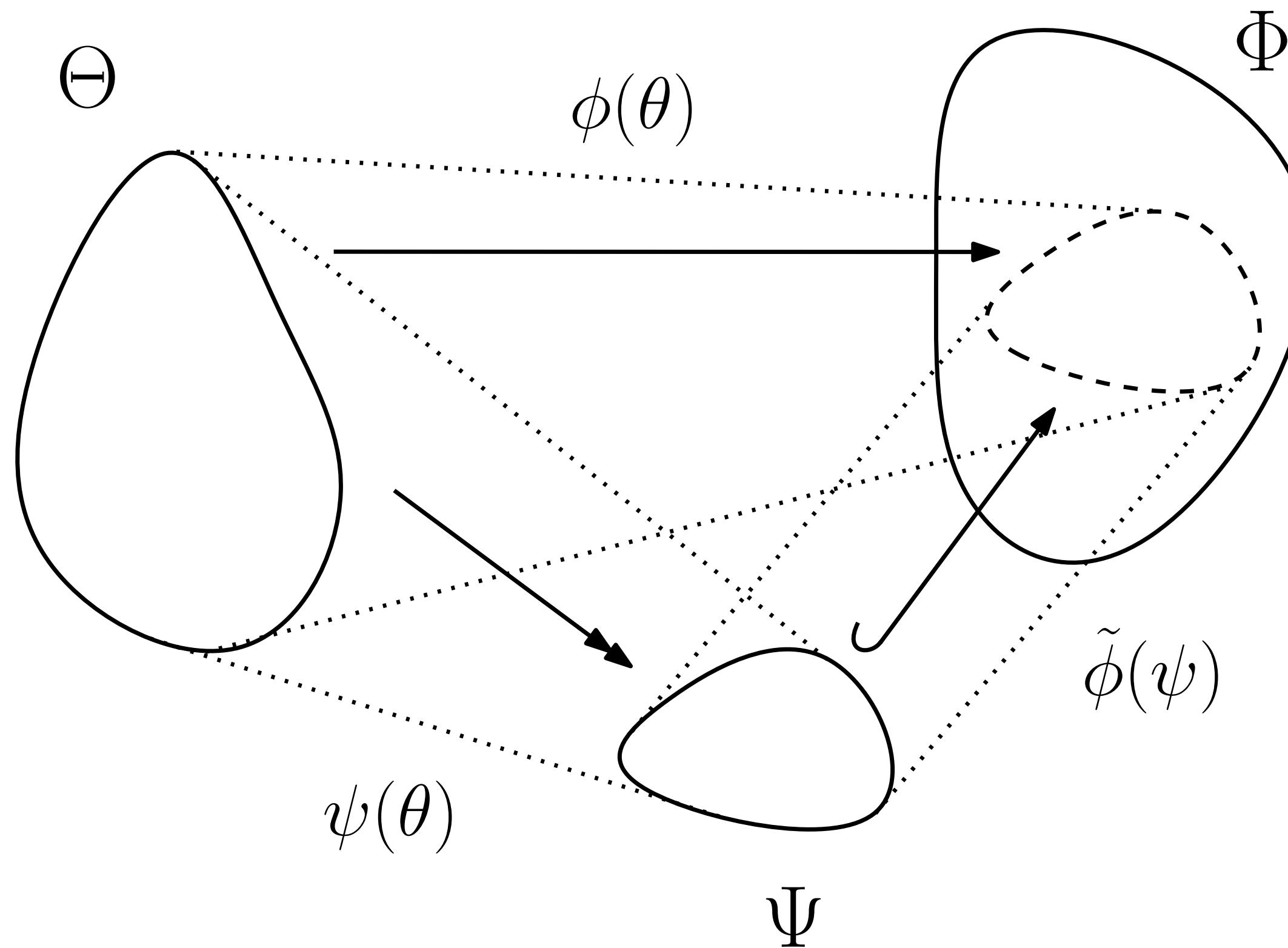
$$\begin{bmatrix} n \\ p \end{bmatrix} \mapsto \mathcal{N}(np, np) \quad \text{Identified: } np \quad (\text{can't get components separately})$$

- Fisher-Kolmogorov PDE: wave speed for certain initial conditions

$$\text{Identified: } c = 2\sqrt{D\lambda} \quad (\text{can't get diffusion or proliferation rate separately})$$

- Etc!

The exhaustive summary + reparameterisation approach



'Epi-mono' factorisation:

$$\phi(\theta) = \tilde{\phi}(\psi(\theta))$$

Where:

θ

Original parameters

Exhaustive **s**ummary/

data distribution parameters

On**t**o mapping giving
identified parameters

$\phi(\theta)$

$\psi(\theta)$

$\tilde{\phi}(\psi)$

One-to-one function of
identified parameters

Different implementations of common idea

Different communities approach identifiability in different ways

(See e.g. ‘Parameter Redundancy and Identifiability’ by Cole, 2020)

- **Structural identifiability** (existence): Symbolic differentiation, Taylor series, generating series, similarity transforms, differential algebra, etc...‘symbolic’ or algebraic approaches
- **Practical identifiability** (stability/estimation): Profile likelihood, Fisher information, sloppiness, nonconvergent MCMC for Bayesians...
- The structural methods are all (in my opinion...) **essentially closely related ways** to implement the same basic ‘exhaustive summary reparameterisation’ idea.
- Here we propose another! Why? Respects factorisation idea of **structural identifiability analysis**, while allowing for **stability/practical identifiability analysis** and **numerical rather than symbolic** implementation

The symbolic/algebraic approach and derivative factorisation

If we further assume that our epi-mono factorisation is preserved under **differentiation** and hence the **chain rule**, i.e. that we have a **submersion** followed by an **immersion**, we get:

$$D_\theta \phi(\theta) = D_\psi \tilde{\phi}(\psi(\theta)) D_\theta \psi(\theta) \quad (\text{Chain rule})$$

Which Implies: $\ker D_\theta \psi(\theta) = \ker D_\theta \phi(\theta)$

Hence we solve: $D_\theta \psi(\theta) \alpha(\theta) = 0$

For $\psi(\theta)$ where $\alpha(\theta)$ are the (symbolic) null-space vectors of $D_\theta \phi(\theta)$

(Slightly different perspective but equivalent to that reviewed by Cole, 2020)

Simple example

Poisson limit of normal approximation to binomial

$$\begin{bmatrix} n \\ p \end{bmatrix} \mapsto \mathcal{N}(np, np) \quad \text{with ex. summary} \quad \phi : \begin{bmatrix} n \\ p \end{bmatrix} \mapsto \begin{bmatrix} np \\ np \end{bmatrix}$$

Jacobian: (symbolic) nullspace vectors:

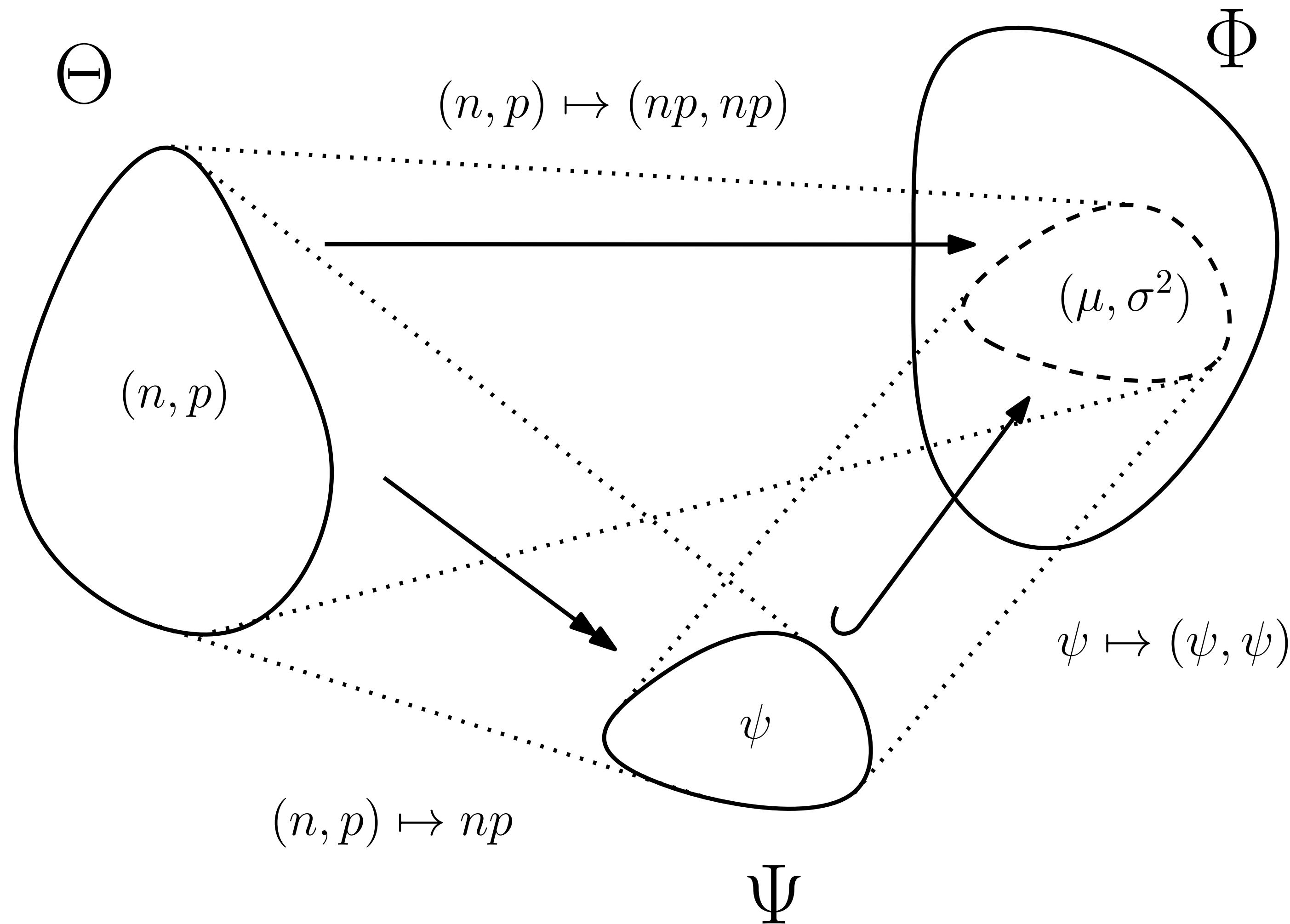
$$D\phi = \begin{bmatrix} p & n \\ p & n \end{bmatrix} \qquad \alpha(n, p) = \begin{bmatrix} n \\ -p \end{bmatrix}$$

Re-parameteration PDE to solve

$$D\psi(n, p)\alpha(n, p) = 0 \quad \text{i.e.} \quad \frac{\partial\psi}{\partial n}n + \frac{\partial\psi}{\partial p}(-p) = 0 \quad \text{Gives:}$$

$$\psi(n, p) = np$$

Simple example: picture



Unifying different approaches?

- What if we have a **numerical** forward model (possibly differentiable though)?
- Symbolic/algebraic approach likely not feasible. However, **algorithmic/automatic differentiation** often is (returns numerical evaluation of derivative)
- Can also fall back on e.g. **finite differences**
- Can easily formulate equivalent **likelihood-based** version of factorisation and consider finite, noisy data scenario, i.e. **practical identifiability** not just structural identifiability

Log transformations and parameter groups

Common parameter groups (Buckingham Pi):

$$\pi_i = \theta_1^{r_{i1}} \theta_2^{r_{i2}} \cdots \theta_p^{r_{ip}}$$

i.e.

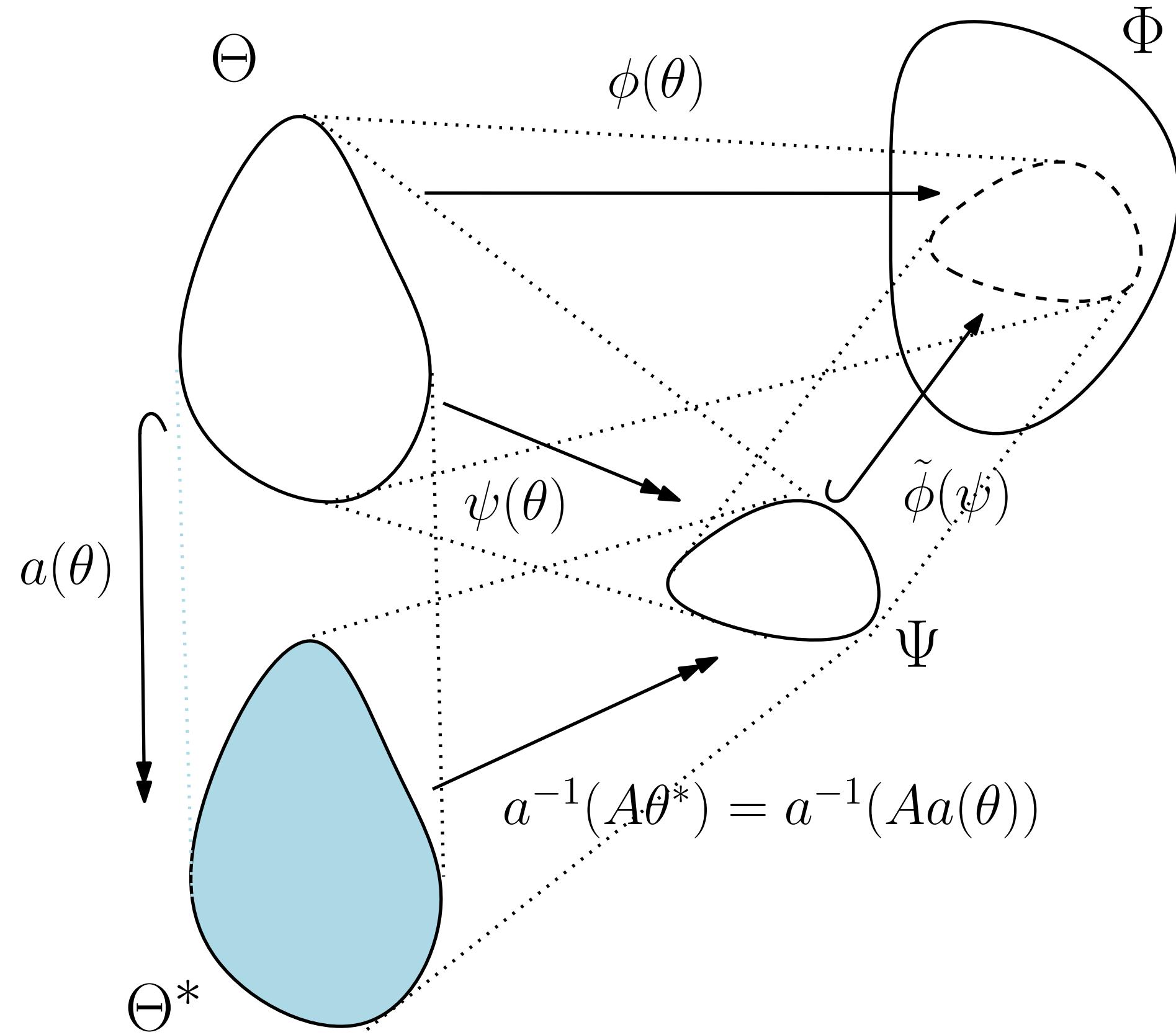
$$\log(\pi_i) = r_{i1} \log(\theta_1) + r_{i2} \log(\theta_2) + \cdots + r_{ip} \log(\theta_p)$$

Hence define:

$$\theta^* = \log(\theta)$$

and consider...

Log transformations and parameter groups



$$\phi(\theta) = \tilde{\phi}(\exp(A\theta^*))$$

where in $\exp(A\theta^*)$

- The exponential is applied componentwise (can also consider general componentwise immersion a)
- The A matrix is unknown to be determined, giving linear combination of log parameters
- Result is vector of monomial combinations of original parameters

Derivative factorisation revisited

Transformation leads to:

$$D_{\theta^*} \phi_*(\theta^*) = D_{\psi_*} \tilde{\phi}(\psi_*(\theta^*)) \text{diag}(\exp(A\theta^*)) A$$

Can be shown to imply:

$$\ker A = \ker D_{\theta^*} \phi_*(\theta^*)$$

i.e. unknown matrix A defining identifiable parameters can be found via **null space of Jacobian derivative wrt log variables**

Importantly, this is **constant**, independent of the parameter!

Only need **single numerical evaluation at any point**

Singular value decomposition

Can conveniently implement using SVD, either for exhaustive summary or associated Fisher information at any (non-unique) ‘best estimate’ (e.g. MLE)

$$D_{\theta^*} \phi_*(\hat{\theta}^*) = U \Sigma V^T$$

Gives: $\theta \mapsto \begin{bmatrix} \psi(\theta) \\ \lambda(\theta) \end{bmatrix} = \exp(V^T \log(\theta))$

Identifiable: $\psi(\theta)$
Nonidentifiable: $\lambda(\theta)$

Can use these as target parameters for finite data, likelihood-based analysis, e.g. Simpson and Maclaren (2023) *Profile-wise Analysis (PWA)* approach

Simple example revisited

Poisson limit of normal approximation to binomial

$$\begin{bmatrix} n \\ p \end{bmatrix} \mapsto \mathcal{N}(np, np) \quad \text{with ex. summary } \phi : \begin{bmatrix} n \\ p \end{bmatrix} \mapsto \begin{bmatrix} np \\ np \end{bmatrix}$$

Using $\theta^* : \begin{bmatrix} \log n \\ \log p \end{bmatrix}$ and $\phi_* : \begin{bmatrix} \log n \\ \log p \end{bmatrix} \mapsto \exp \begin{bmatrix} \log n + \log p \\ \log n + \log p \end{bmatrix}$

Get: $D_{\theta^*} \phi_*(\theta^*) = \begin{bmatrix} np & np \\ np & np \end{bmatrix} = \underbrace{\begin{pmatrix} 1 \\ 1 \end{pmatrix}}_{D_{\psi_*} \tilde{\phi}(\psi_*(\theta^*))} \underbrace{\begin{pmatrix} np \\ np \end{pmatrix}}_{\text{diag}(\exp(A\theta^*))} \underbrace{\begin{pmatrix} 1 & 1 \end{pmatrix}}_A$

(Same result, but null space is independent of parameter)

Another simple example

Normal approximation to binomial

$$\begin{bmatrix} n \\ p \end{bmatrix} \mapsto \mathcal{N}(np, np(1-p))$$

with ex. summary

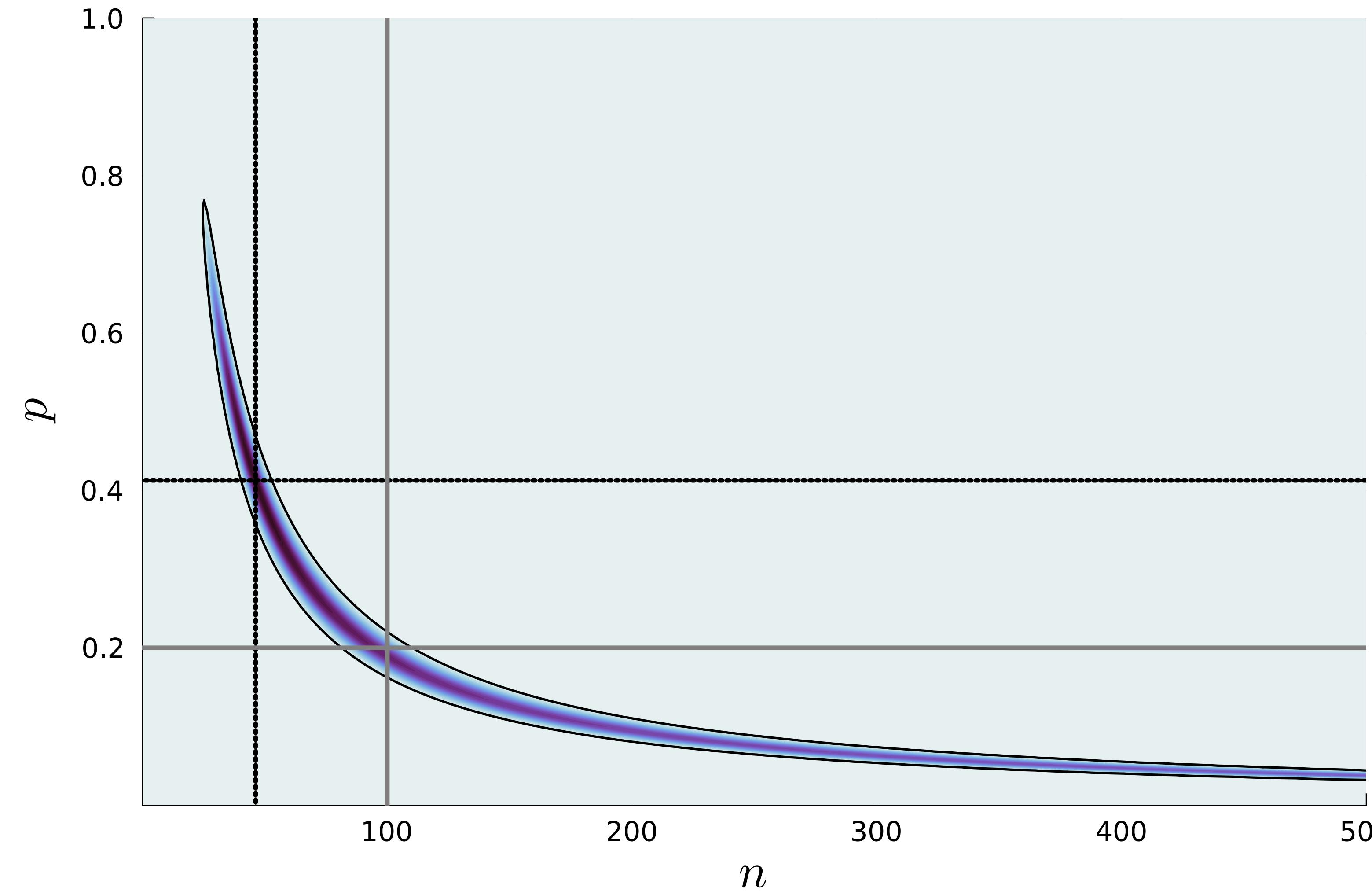
$$\phi : \begin{bmatrix} n \\ p \end{bmatrix} \mapsto \begin{bmatrix} np \\ np(1-p) \end{bmatrix}$$

...Jacobian full rank (just), analysis leads to:

$$\psi(n, p) = \begin{bmatrix} np \\ n/p \end{bmatrix}$$

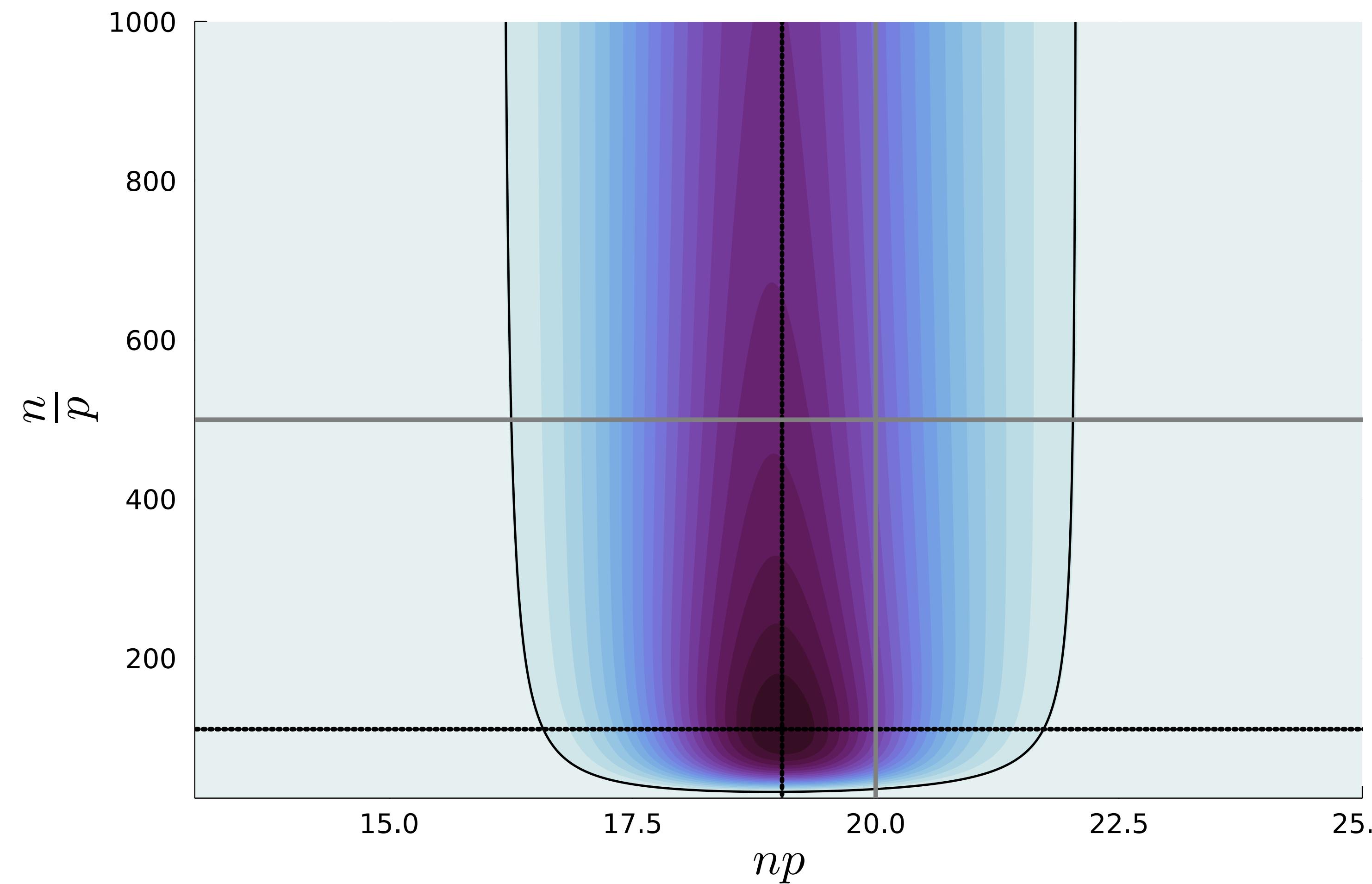
Normal approximation to binomial: finite data

Joint likelihood, 'banana' shaped: poorly identified/sloppy etc



Normal approximation to binomial: finite data

Joint likelihood, reparameterised: separate well- and poorly-identified parameter combinations:



More examples?

Preprint coming soon with:

- Monod/Michaelis Menten kinetics
- Non-homogeneous diffusion groundwater model
- Embryonic development in morphogen gradients model
- Etc