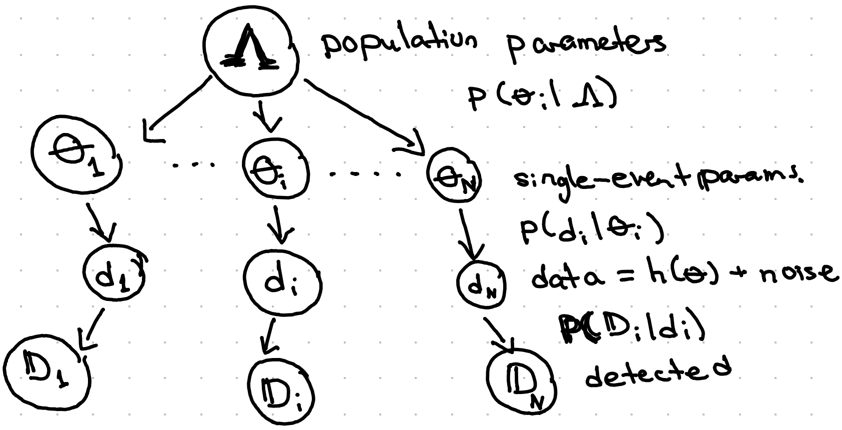
Lecture 9 - Hierarchical Bayesian Inference



Last time: $p(\theta \mid d, H) = \dfrac{p(d \mid \theta, H)\, p(\theta \mid H)}{p(d \mid H)}$

where $p(d \mid H) = \int p(d \mid \theta, H)\, p(\theta \mid H)\, d\theta$

from PE to pop:

Idea: find the "best" prior $p(\theta \mid H)$

pop model: family of priors described by some params Λ

$$p(\theta \mid H) \rightarrow p(\theta \mid \Lambda)$$

Goal: infer Λ, i.e. evaluate posterior

$$p(\Lambda \mid d) \propto p(d \mid \Lambda)\, p(\Lambda)$$

What is likelihood $p(d \mid \Lambda)$?

In the absense of selection effects, this is just what we called the evidence in single-event PE.

For a single event,

$$p(d_i \mid \Lambda) = \int p(d_i \mid \theta)\, p(\theta \mid \Lambda)\, d\theta$$

Assuming N independent observations:

$$p\underbrace{(d_1, \dots, d_i, \dots d_N}_{\text{notation: } \{d_i\}} \mid \Lambda) = \prod_{i=1}^{N} \int p(d_i \mid \theta)\, p(\theta \mid \Lambda)\, d\theta$$

How do we evaluate this pop. likelihood?

We usually already have some PE samples $\{\theta_i^s\}$ for each event i,

$$\theta_i^s \sim \underbrace{p(\theta_i|d_i)}_{\text{PE posterior}} \propto \underbrace{p(d_i|\theta_i)}_{\text{PE likelihood}} \underbrace{\pi_{PE}(\theta_i)}_{\text{interim PE prior}}$$

We can evaluate each integral over $p(d_i|\theta)\,d\theta$ as a Monte Carlo integral over these samples!

$$\int p(d_i|\theta)\, p(\theta|\Lambda)\, d\theta \propto \int p(\theta|d_i) \frac{p(\theta|\Lambda)}{\pi_{PE}(\theta)}\, d\theta$$

$$\approx \left\langle \frac{p(\theta|\Lambda)}{\pi_{PE}(\theta)} \right\rangle_{\theta_i^s \sim p(\theta_i|d_i)}$$

$$= \frac{1}{N_{samps}} \sum_{s=1}^{N_{samps}} \frac{p(\theta_i^s|\Lambda)}{\pi_{PE}(\theta_i^s)}$$

Inhomogeneous Poisson Process

Often we are not only interested in the (normalized) probability distribution function (pdf)

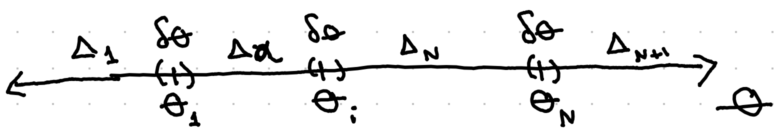
$$\theta_i \sim p(\theta | \Lambda)$$

But in the astrophysical <u>rate</u> density

Notation: $$\frac{dR}{d\theta}(\Lambda) = R\, p(\theta | \Lambda) \quad \text{s.t.}$$

$$\int_{\text{all } \theta} \frac{dR}{d\theta}(\Lambda)\, d\theta = \int_{\text{all } \theta} R\, p(\theta | \Lambda)\, d\theta = R$$

R is the total number of astrophysical sources
(R is a pop. param. in addition to Λ)



For now, pretend all sources are detectable and θ_i are known perfectly.
In an empty interval $\Delta\alpha$, we expect

$$\mu_\alpha = \int_{\Delta\alpha} \frac{dR}{d\theta}\, d\theta \quad \text{sources}$$

Poisson prob. of observing 0: $e^{-\mu_\alpha}$

In an interval $\delta\theta$ around source θ_i, we expect

$$\mu_i = \delta\theta \frac{dR}{d\theta}(\theta_i) \quad \text{sources}$$

Poisson prob. of observing 1: $\mu_i e^{-\mu_i}$

Multiply together all intervals:

$$p(\{\theta_i\}, N | \Lambda, R) \propto \prod_{i=1}^{N} R\, p(\theta_i | \Lambda)\, e^{-R} \qquad R = \int_{\text{all } \theta} \frac{dR}{d\theta}\, d\theta$$

$$= R^N \prod_{i=1}^{N} p(\theta_i | \Lambda)\, e^{-R}$$

What about when we have selection effects and measurement uncertainty?

In each empty interval $\Delta\alpha$, we expect

$$\mu_\alpha = \int_{\Delta\alpha} d\theta \frac{dR}{d\theta} \underbrace{\int dd\, p(d|\theta) P(\mathbb{D}|d)}_{\text{probability of detecting event with params } \theta\text{, i.e. prob that } \theta \text{ gives rise to } d \text{ that is detected}} \quad \text{sources}$$

In each $\delta\theta$ interval, we expect

$$\mu_i = \delta\theta \frac{dR}{d\theta}(\theta_i)\, p(d_i|\theta_i)\, P(\mathbb{D}_i|d_i)$$

so the joint probability

$$\boxed{p(\{\theta_i\}, \{d_i\}, \{\mathbb{D}_i\}, N | \Lambda, R) \propto R^N \prod_{i=1}^{N} p(\theta_i|\Lambda)\, p(d_i|\theta_i) P(\mathbb{D}_i|d_i) \times \exp\left[-R P(\mathbb{D}|\Lambda)\right]} \quad (1)$$

where $P(\mathbb{D}|\Lambda) \equiv \int_{\text{all } \theta} d\theta\, p(\theta|\Lambda) \int_{\text{all } d} dd\, p(d|\theta) P(\mathbb{D}|d)$

"fraction of detectable sources in population described by Λ"

Now define $K = R P(\mathbb{D}|\Lambda)$, the expected number of detections

Rewrite (1) as:

$$\boxed{p(\{\theta_i\}, \{d_i\}, \{\mathbb{D}_i\}, N | \Lambda, K) \propto \frac{K^N e^{-K}}{P(\mathbb{D}|\Lambda)^N} \prod_{i=1}^{N} p(\theta_i|\Lambda) p(d_i|\theta_i) P(\mathbb{D}_i|d_i)} \quad (2)$$

Ultimately, we are interested the joint posterior over $\{\theta_i\}$, Λ and K or R

Move Λ, K over to the left of $|$:

$$p(\{\theta_i\}, \{d_i\}, \{\mathbb{D}_i\}, \Lambda, K, N) \propto \overbrace{p(\Lambda, K)}^{\text{joint prior on } \Lambda, K} \times (2)$$

Move $\{d_i\}$ and $\{\mathbb{D}_i\}$ to the right:

$$p(\{\theta_i\}, \Lambda, K, N | \{d_i\}, \{\mathbb{D}_i\}) \propto p(\Lambda, K) \times (2) \div \underbrace{p(\{d_i\}, \{\mathbb{D}_i\})}_{\prod_{i=1}^{N} P(\mathbb{D}_i|d_i)\, p(d_i)}$$

Move N to the right: $\div\, p(N)$ (normalization constant)

So we get

$$\boxed{p(\{\theta_i\}, \Lambda, K | N, \{d_i\}, \{\mathbb{D}_i\}) \propto p(\Lambda, K) \frac{K^N e^{-K}}{P(\mathbb{D}|\Lambda)^N} \prod_{i=1}^{N} p(\theta_i|\Lambda) p(d_i|\theta_i)} \quad (3)$$

What if we don't care about K, only the shape of the population?

→ Marginalize over K

If we assume $p(\Lambda, K) = p(\Lambda) p(K)$ [separable prior],

then

$$p(\{\theta_i\}, \Lambda | N, \{d_i\}, \{\mathbb{D}_i\}) \propto$$

$$\underbrace{\left[\int p(K) K^N e^{-K} dK\right]}_{\text{some constant that doesn't depend on } \{\theta_i\} \text{ or } \Lambda \text{, so ignore}} p(\Lambda)\, P(\mathbb{D}|\Lambda)^{-N} \prod_{i=1}^{N} p(\theta_i|\Lambda) p(d_i|\theta_i)$$

What if we only care about Λ, not single-event params θ_i?

Marginalize over $\{\theta_i\}$, to get

$$\boxed{p(\Lambda \mid \{d_i\}, \{\mathbb{D}_i\}, N) \propto p(\Lambda)\, P(\mathbb{D}\mid\Lambda)^{-N} \prod_{i=1}^{N} \int d\theta\, p(d_i \mid \theta)\, p(\theta \mid \Lambda)} \quad (4)$$

What we had before, but with extra $P(\mathbb{D}\mid\Lambda)^{-N}$ term

In practice, we can sample Λ from (4), then sample $K \sim \underbrace{p(N\mid K)}_{K^N e^{-K}} p(K)$ and $R = \frac{K}{P(\mathbb{D}\mid\Lambda)}$

We mentioned that we compute $\int d\theta\, p(d_i\mid\theta)\, p(\theta\mid\Lambda)$ as a Monte-Carlo integral over PE samples

how do we compute $P(\mathbb{D}\mid\Lambda)$?

Also a Monte Carlo integral (over <u>injections</u>)

<u>Injections</u>:

Draw N_{draw} simulated signals from some astro. distribution,

$\theta_{inj} \sim P_{draw}(\theta)$

Generate mock data $d_{inj} = \theta_{inj} + n$ $\quad p(d_{inj}\mid\theta_{inj})$

Run search and record $\mathbb{D}_{inj}$, $\quad P(\mathbb{D}_{inj}\mid d_{inj})$

$$P(\mathbb{D}\mid\Lambda) = \iint p(\theta, d, \mathbb{D}\mid\Lambda)\, d\theta\, dd$$
$$= \iint p(\theta\mid\Lambda)\, p(d\mid\theta)\, p(\mathbb{D}\mid d)\, d\theta\, dd$$
$$= \iint \frac{p(\theta\mid\Lambda)}{P_{draw}(\theta)} P_{draw}(\theta)\, p(d\mid\theta)\, p(\mathbb{D}\mid d)\, d\theta\, dd$$
$$\approx \left\langle \frac{p(\theta\mid\Lambda)}{P_{draw}(\theta)} P(\mathbb{D}\mid d) \right\rangle_{\{\theta_{inj}, d_{inj}\}}$$

Typically detection threshold is <u>deterministic</u>, so $P(\mathbb{D}\mid d)$ is either 0 or 1

Then we can keep "found" injections for which $P(\mathbb{D}\mid d_{inj}) = 1$

We drew N_{draw} total injections

$$\boxed{P(\mathbb{D}\mid\Lambda) \approx \frac{1}{N_{draw}} \sum_{\text{found inj}} \frac{p(\theta_{inj}\mid\Lambda)}{P_{draw}(\theta_{inj})}}$$