Counting Coffee Cups and Photons

James Ritchie

A mathematician is a device for turning coffee into theorems

— Alfréd Rényi

Does something similar hold for Informatics?

What's the distribution of coffee cup usage in the forum?

1. What's the distribution of coffee cup capacity?
2. How often are they used?

We can put reasonably informative priors on both of these.

Experimental approaches?

1. Visit everyone's office and measure their coffee cups?
   — Difficult to co-ordinate
2. Stand by the machine and measure mugs when people get coffee?
   — Time consuming
3. Check the machines at regular intervals and measure the total volume of coffee dispensed?
   — Simple, but complicates the statistics.

An example

We check the coffee machine after an hour, and find 1 litre of coffee was dispensed.

— 1 1L cup?

— 20 50ml cups?

— 5 200ml cups?

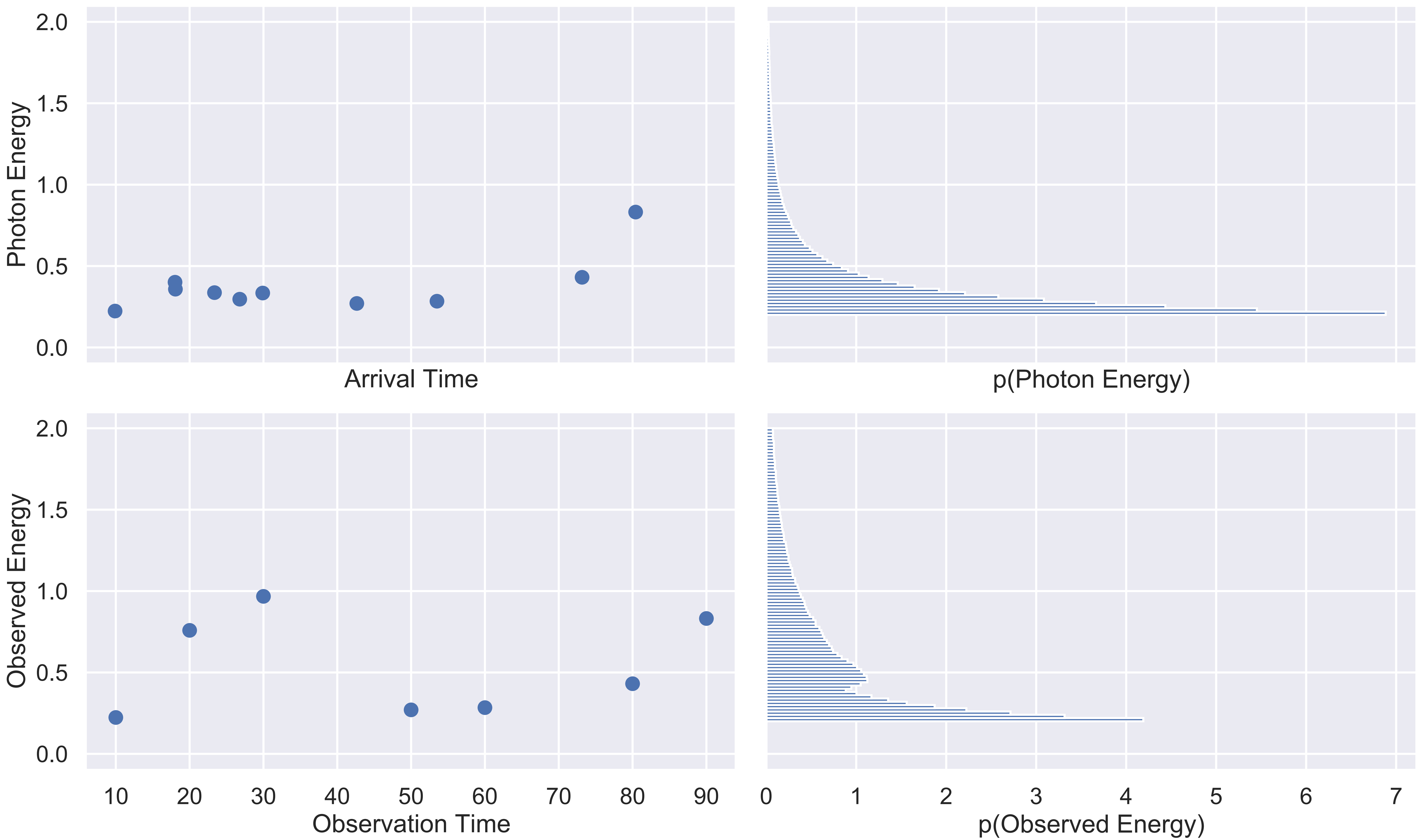
— 4 250ml cups?

— 3 50ml cups, 3 200ml cups and 1 250 ml cup?

Maybe the problem is unidentifiable?

CCD Pile-up

- Photons generated by a astronomical object arrive at a telescope
- Pixels in a CCD sensor measure the energy of the photons
- Each pixel is checked at a given interval
- The pixel reports the total amount of photon energy absorbed



Photon Energy
Arrival Time
p(Photon Energy)
Observed Energy
Observation Time
p(Observed Energy)

Generative Model

For each observation timebin t:

1. Draw a photon count N_t from a Poisson distribution $P(N_t \mid \lambda)$
2. Draw N_t photon energies e_{it} from a power-law distribution $p(e_{it} \mid \alpha)$
3. Sum the energies up and add some Gaussian noise drawn from $\mathcal{N}(\epsilon \mid 0, \sigma)$ to get observed energy E_t.

Inference

The posterior we want to work with:

$$p(\alpha, \lambda, \sigma \mid \{E_t\}_{t=1}^{T}) \propto p(\{E_t\}_{t=1}^{T}) \mid \alpha, \lambda, \sigma) p(\alpha) p(\lambda) p(\sigma)$$

The posterior we have to work with:

$$p\left(\alpha, \lambda, \sigma, \left\{N_t, \{e_{it}\}_{i=1}^{N_t}\right\}_{t=1}^{T} \mid \{E_t\}_{t=1}^{T}\right) \propto \ldots$$

Use MCMC to draw samples from it.

Problem 1

The posterior has a high dimensionality, typically $D > 100$

Solution: Use Adaptive Hamiltonian Monte Carlo (HMC)

Massively oversimplified description: Use the gradients of the posterior w.r.t to the parameters to scale up MCMC.

Problems 2 and 3

- The dimensionality of our posterior is variable
  - The number of parameters e_{it} depends on N_t
- The posterior contains discrete parameters.
  - HMC only works with continous parameters

Solution: Marginalise out the count variables N_t

$$\sum_{N_t=1}^{\infty} p(\alpha, \lambda, \sigma, \left\{N_t, \{e_{it}\}_{i=1}^{N_t}\right\}_{t=1}^{T} \mid \{E_t\}_{t=1}^{T})$$

Problem 4

If the noise is small, the e_{it}'s must add up to something close to E_t, i.e. they are constrained to lie on a manifold.

Solution: Reparameterise e_{it} in terms of a latent noise-free total energy L_t and fractions f_{it},

$$e_{it} = L_t f_{it}.$$

f_{it} can then be parameterised using unconstrained variables via a stick-breaking transformation.

Summary

1. CCD Pile-up makes inference on simple astromical models difficult.
2. We can deal with it by making it part of our observation model.
3. Make it easier with standard MCMC tricks:
   - Use Adaptive HMC
   - Marginalise out discrete variables
   - Reparameterise to remove constraints

Experimental proof-of-concept work ongoing