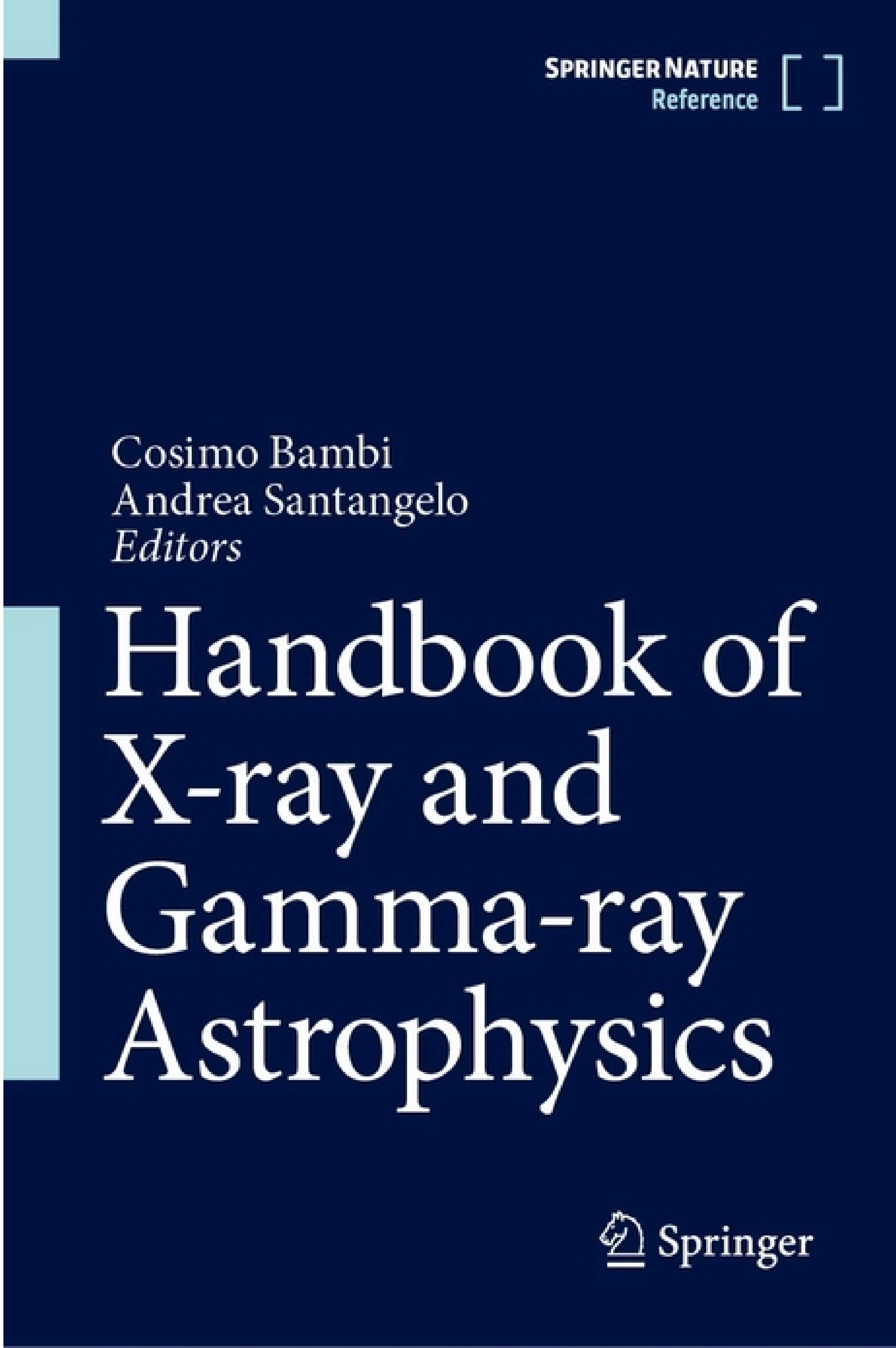
X-ray spectral analysis

eRO-SDAS workshop
21.11.2024
Johannes Buchner

Statistical Aspects of X-ray Spectral Analysis

Johannes Buchner & Peter Boorman

https://arxiv.org/abs/2309.05705



SPRINGER NATURE
Reference

Cosimo Bambi
Andrea Santangelo
Editors

Handbook of X-ray and Gamma-ray Astrophysics

Springer

X-ray Spectral Analysis

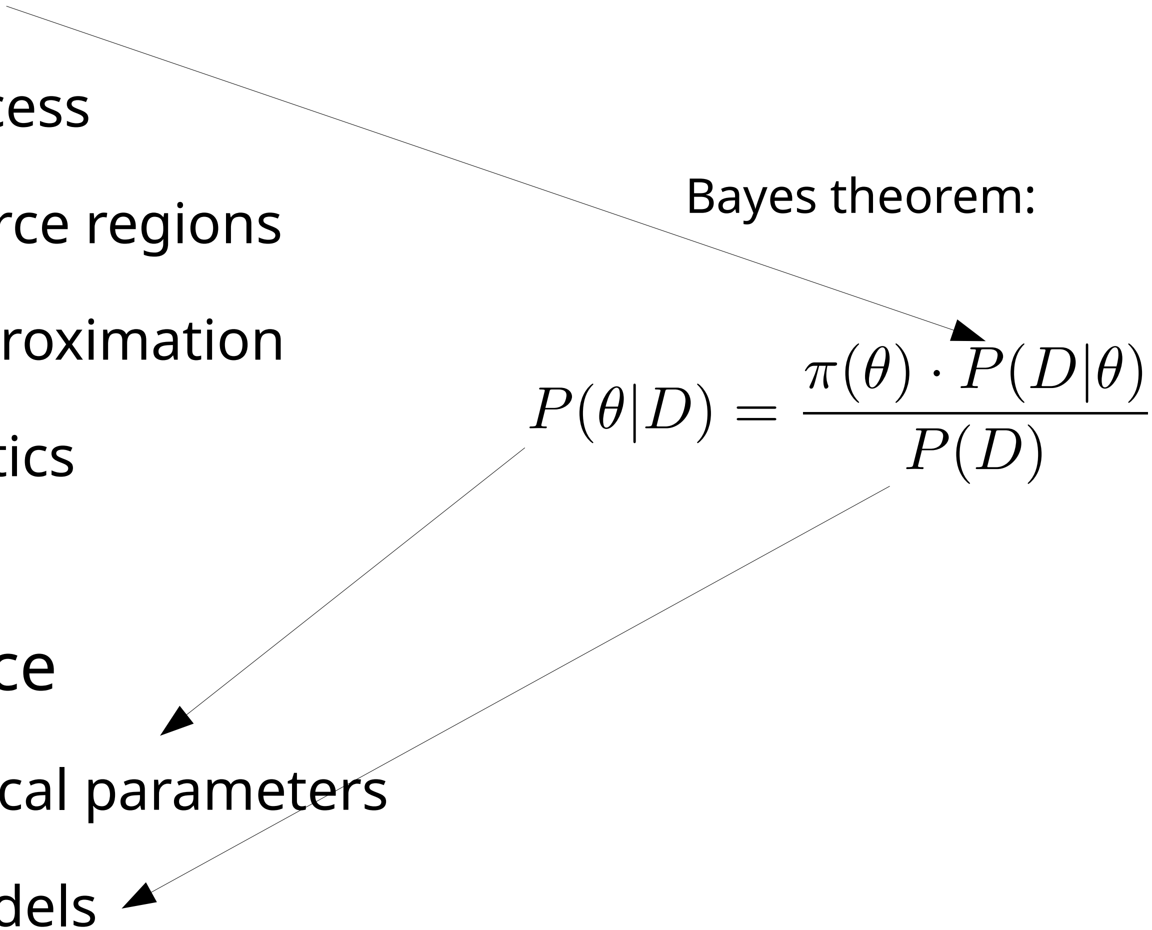
1) Likelihood
    - Measurement process
    - Background & source regions
    - Linear algebra approximation
    - Likelihood & statistics

Bayes theorem:

$$P(\theta|D) = \frac{\pi(\theta) \cdot P(D|\theta)}{P(D)}$$

2) Bayesian inference
    - Constraining physical parameters
    - Differentiating models
    - Treating backgrounds

X-ray Spectral Analysis



1) Likelihood
    - Measurement process
    - Background & source regions
    - Linear algebra approximation
    - Likelihood & statistics

Bayes theorem:

$$P(\theta|D) = \frac{\pi(\theta) \cdot P(D|\theta)}{P(D)}$$

2) Bayesian inference
    - Constraining physical parameters
    - Differentiating models
    - Treating backgrounds

Why X-rays?

- Hot plasmas
  - Stars, Galaxy clusters, Inter-galactic medium
  - Temperatures, turbulence, …

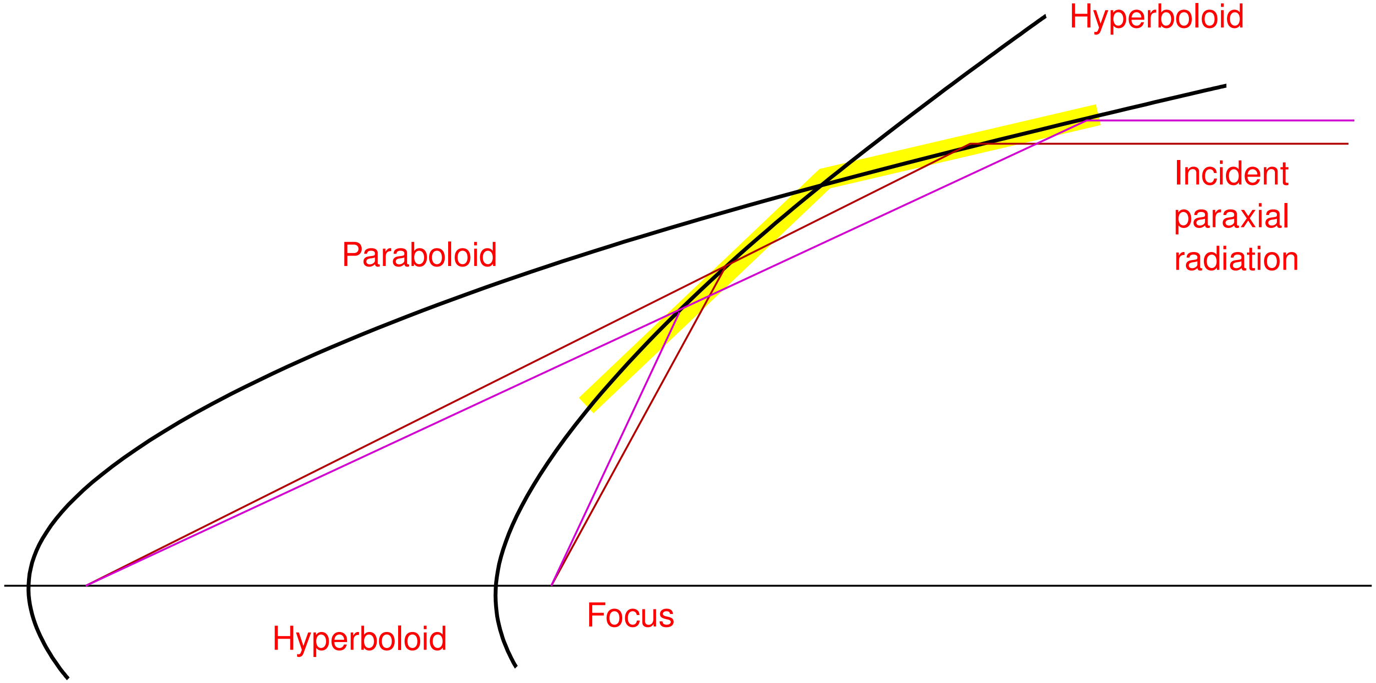
model examples:
blackbody, mekal, apec

- Compact objects
  - Black holes of all sizes, neutron stars, pulsars, …
  - Accretion rate, spin, …

model examples:
powerlaw, relxill, optxagn

Focussing x-rays

Wolter Telescopes



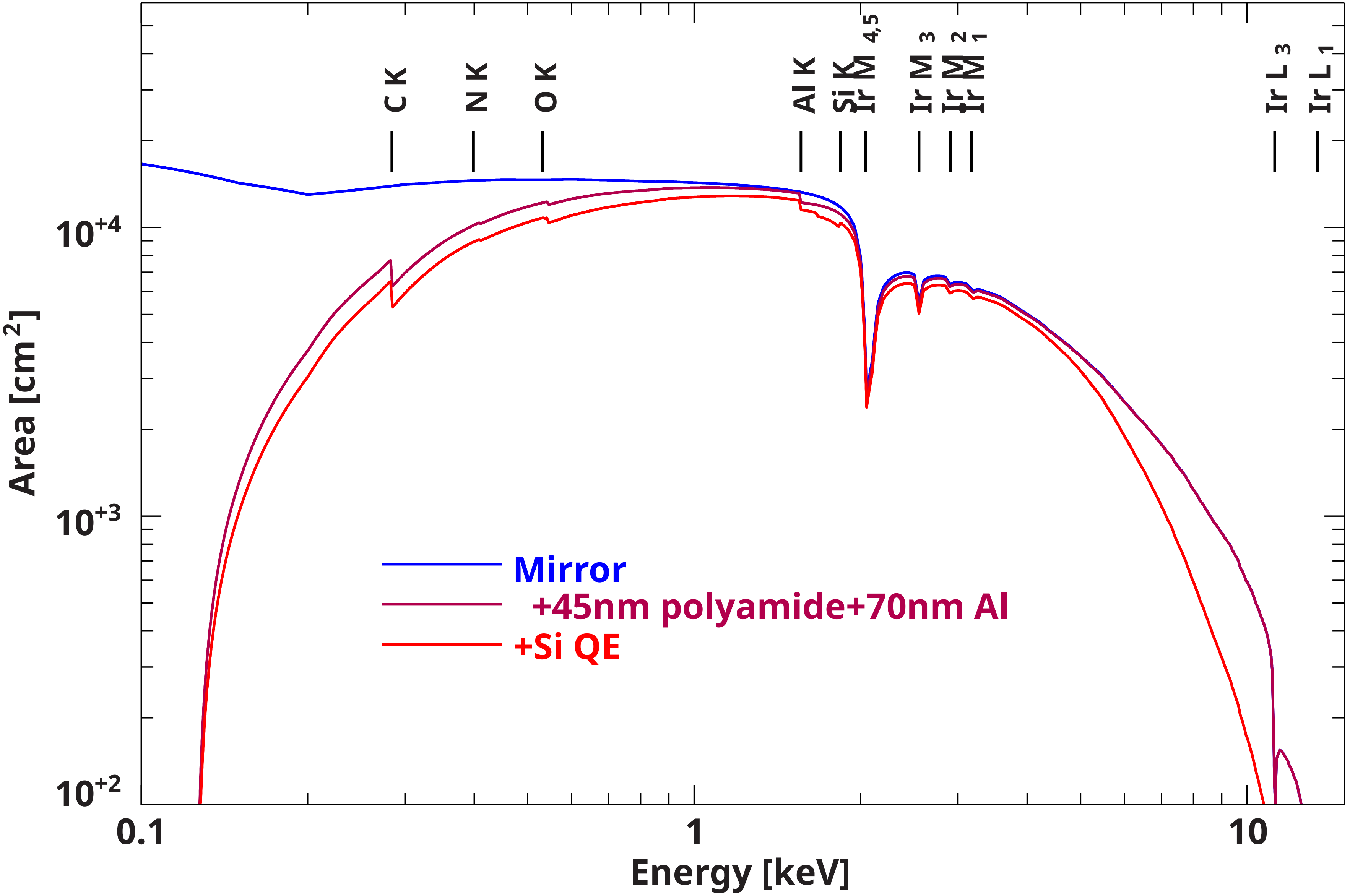
after ESA

To obtain manageable focal lengths (~10 m), use two reflections on a parabolic and a hyperboloidal mirror (Wolter) type

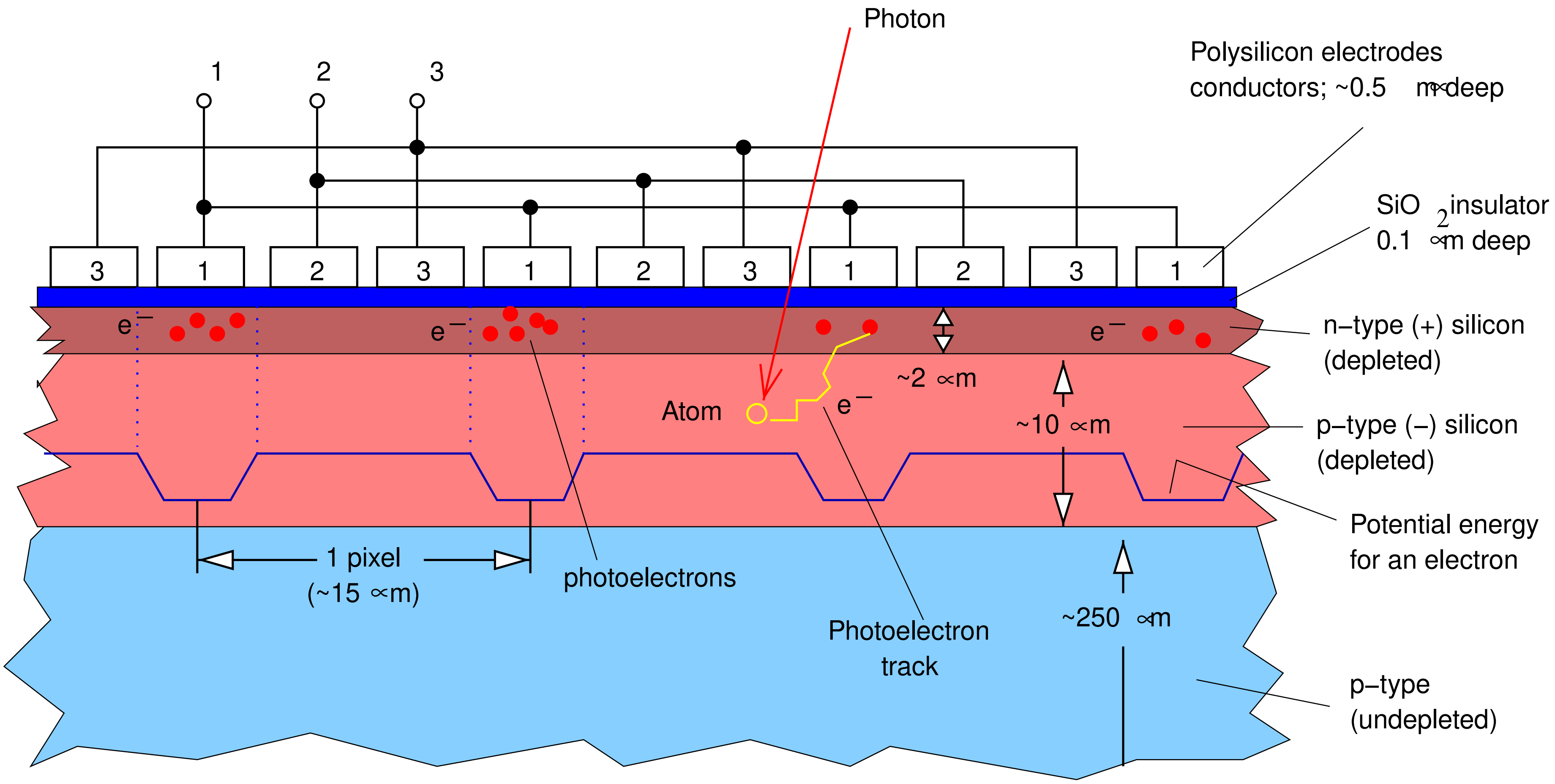
(Wolter 1952 for X-ray microscopes, Giacconi & Rossi 1960 for UV- and X-rays).

But: small collecting area (A ~ πr^2 l/f where f: focal length)

ARF – sensitive area



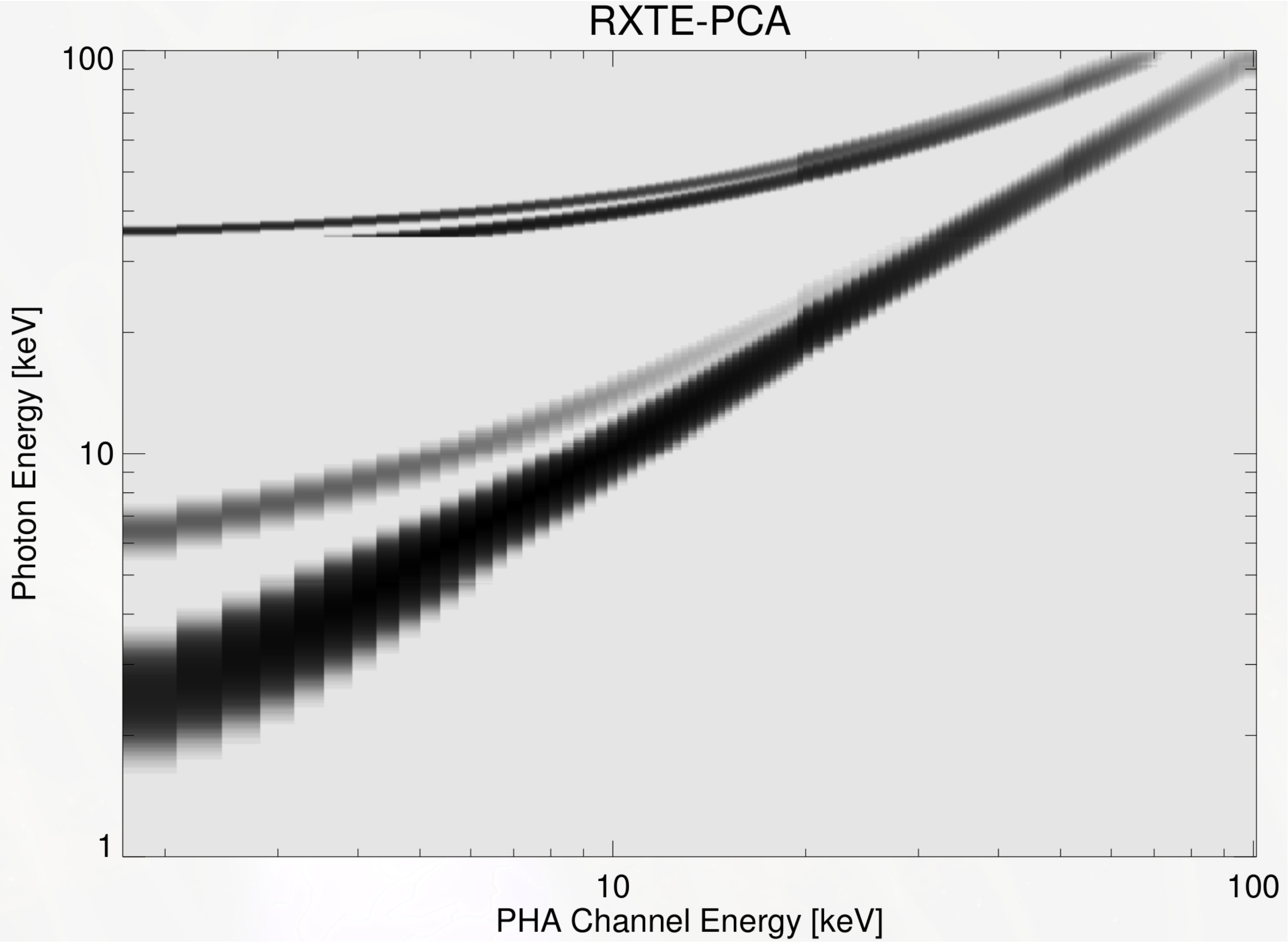
Detecting X-rays



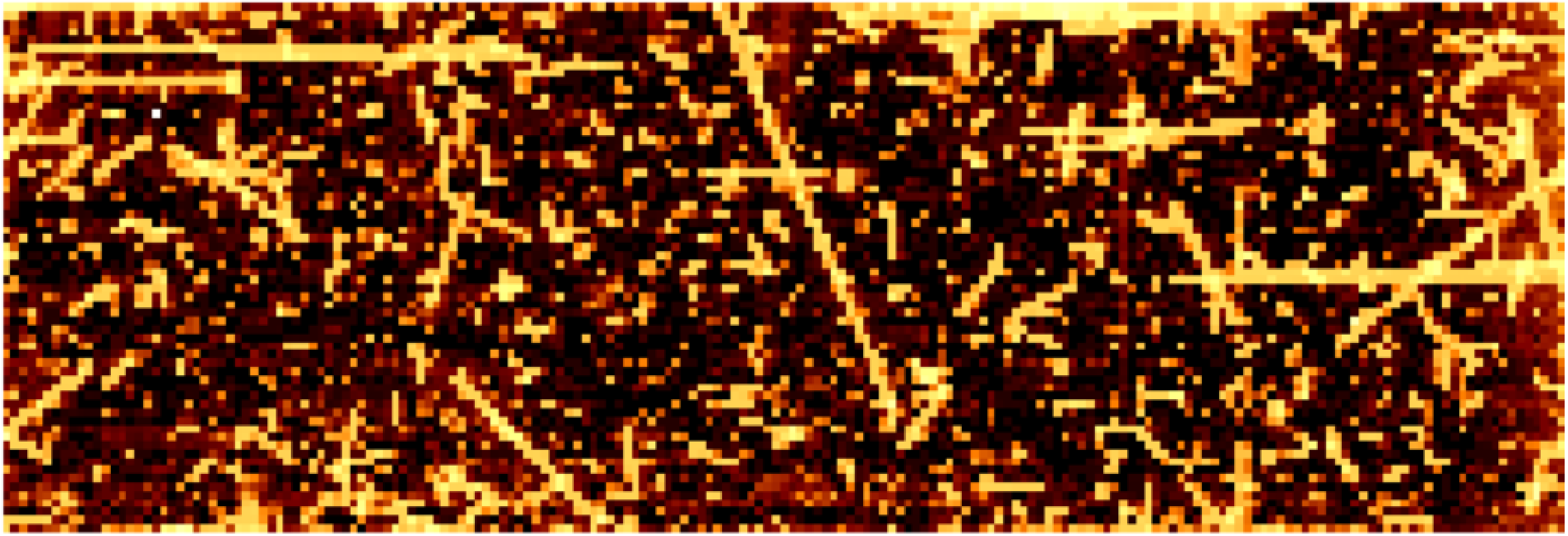
After Bradt

2d imaging with Charge Coupled Devices (CCDs)

RMF – detector response



Background



M. Wille

Cosmic rays & protons

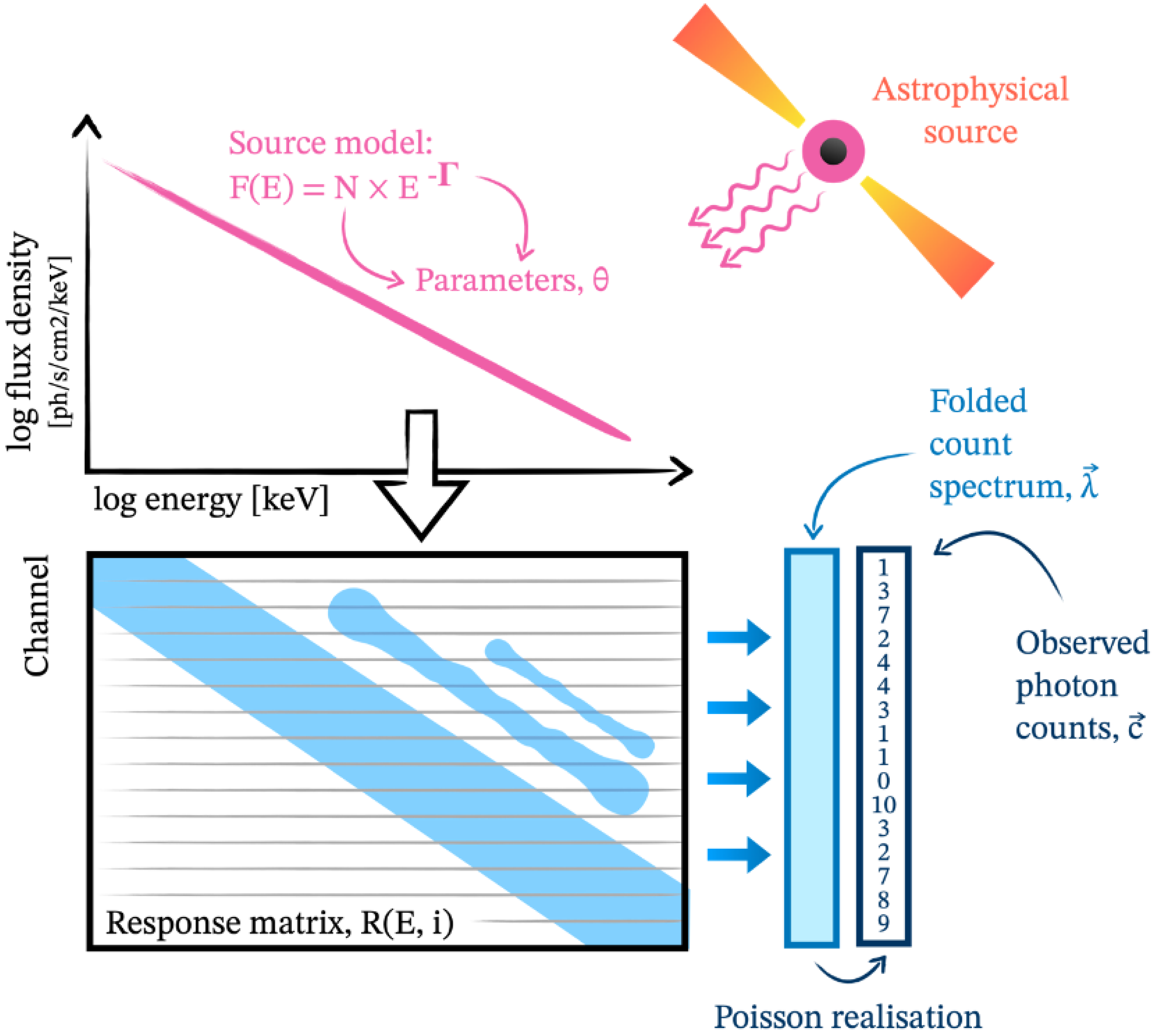
not going through the mirror

Files

- .pha or .pi – spectrum,
  counts in channels
- .rsp or .rmf – response matrix
- .arf – effective area
- bkg.pi – background spectrum

Data archive for other missions: heasarc.gsfc.nasa.gov

Search for previous observations with xamin interface


Astrophysical source
Source model:
F(E) = N × E -Γ
Parameters, θ
log flux density
[ph/s/cm2/keV]
log energy [keV]
Channel
Response matrix, R(E, i)
Folded count spectrum, λ⃗
Observed photon counts, c⃗
1
3
7
2
4
4
3
1
1
0
10
3
2
7
8
9
Poisson realisation

Formal data analysis

$$N(c) = \sum R(c, E) \times A(E) \times F(E) dE + b(c)$$

background

count rate in channel c (counts s^{-1})

detector response (∝ probability to detect photon of energy E in channel c).

effective area (cm^{2})

photon flux density (ph cm^{-2} s^{-1} keV^{-1}),

Formal data analysis

$$N(c) = \sum R(c,E) \times A(E) \times F(E)dE + b(c)$$

background

count rate in channel c (counts s^{-1})

detector response (∝ probability to detect photon of energy E in channel c).

effective area (cm^{2})

photon flux density (ph cm^{-2} s^{-1} keV^{-1}),

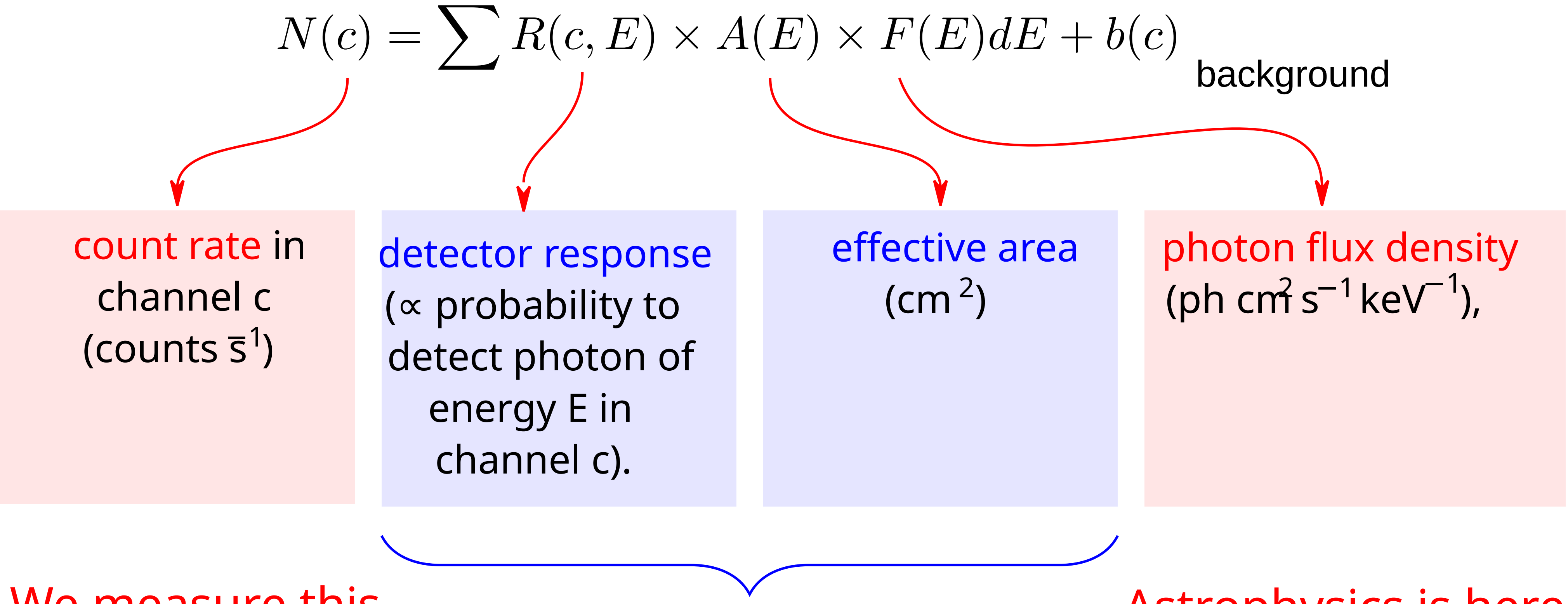
We measure this

Calibration

(“response” / “rsp”)

Astrophysics is here

Formal data analysis



$$N(c) = \sum R(c,E) \times A(E) \times F(E)dE + b(c)$$

background

count rate in channel c (counts s^{-1})

detector response (∝ probability to detect photon of energy E in channel c).

effective area (cm ^{2})

photon flux density (ph cm^{2} s^{-1} keV^{-1}),

We measure this

Calibration

(“response” / “rsp”)

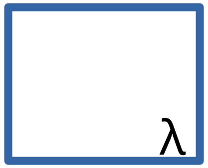
Astrophysics is here

1. Guess F(E) – astrophysical model + parameter values
2. Predict N(c)
3. Compare prediction to actual number – Poisson statistics
4. Modify guess

Comparing predicted and observed counts

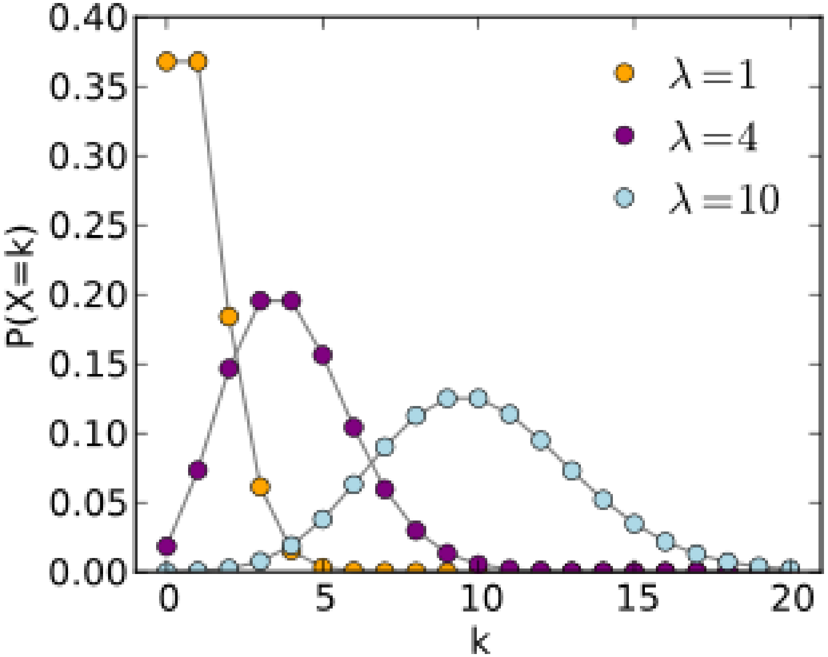
Single spectral bin

- Poisson



$$P(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

  - k: integer
  - λ: real (mean&variance)
  - Asymmetric
  - Integer
  - Positive



- Scaling

shape changes

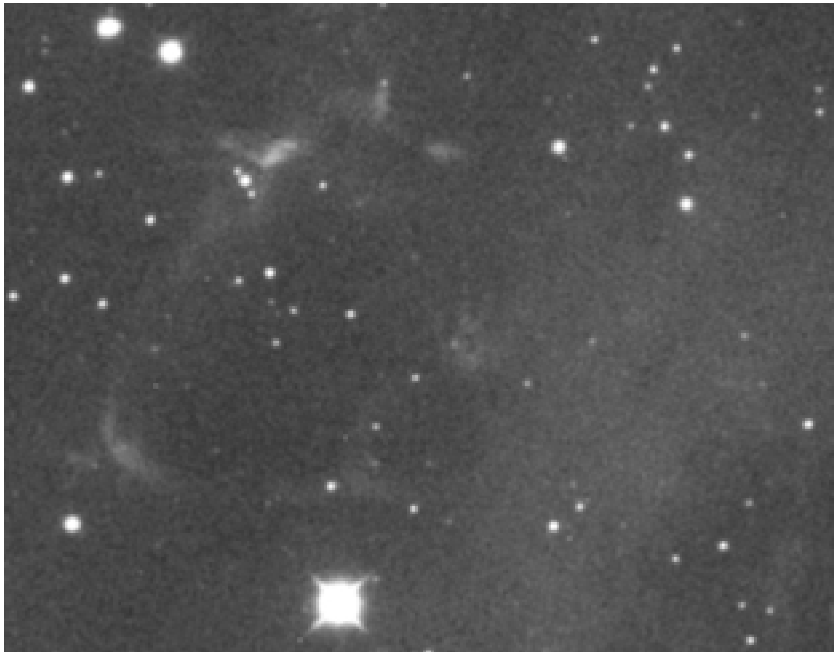
- Addition

(Poisson distribution)
Variability!

- Subtraction

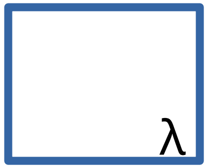
(Skellam distribution)

Samples
Electronics (shot noise)
Photon counting (Poisson nois



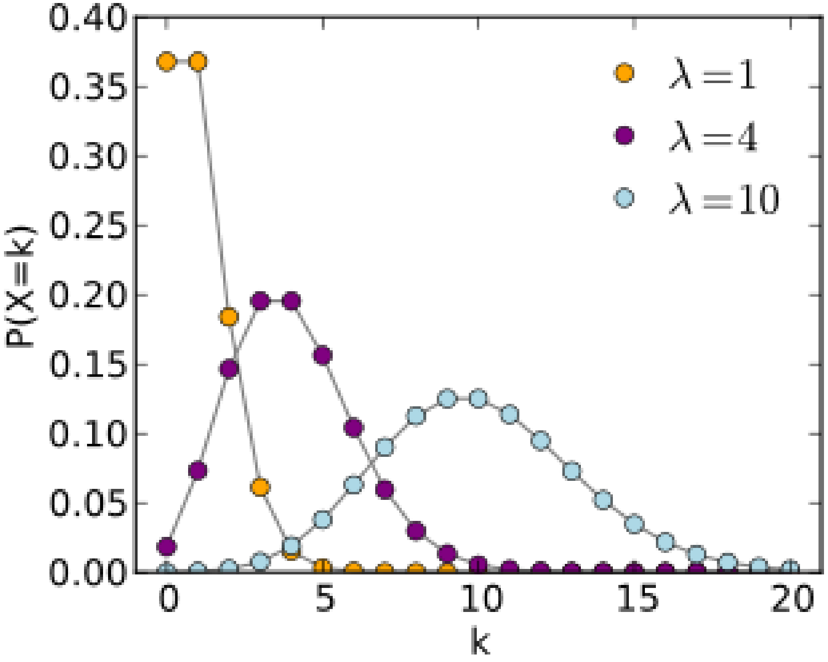
Single spectral bin

- Poisson



$$P(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

  - k: integer
  - λ: real (mean&variance)



- Gaussian
  - Mean (μ) & variance (σ^2) = λ
  - Mean (μ) & variance (σ^2) = k
  - real, can be negative

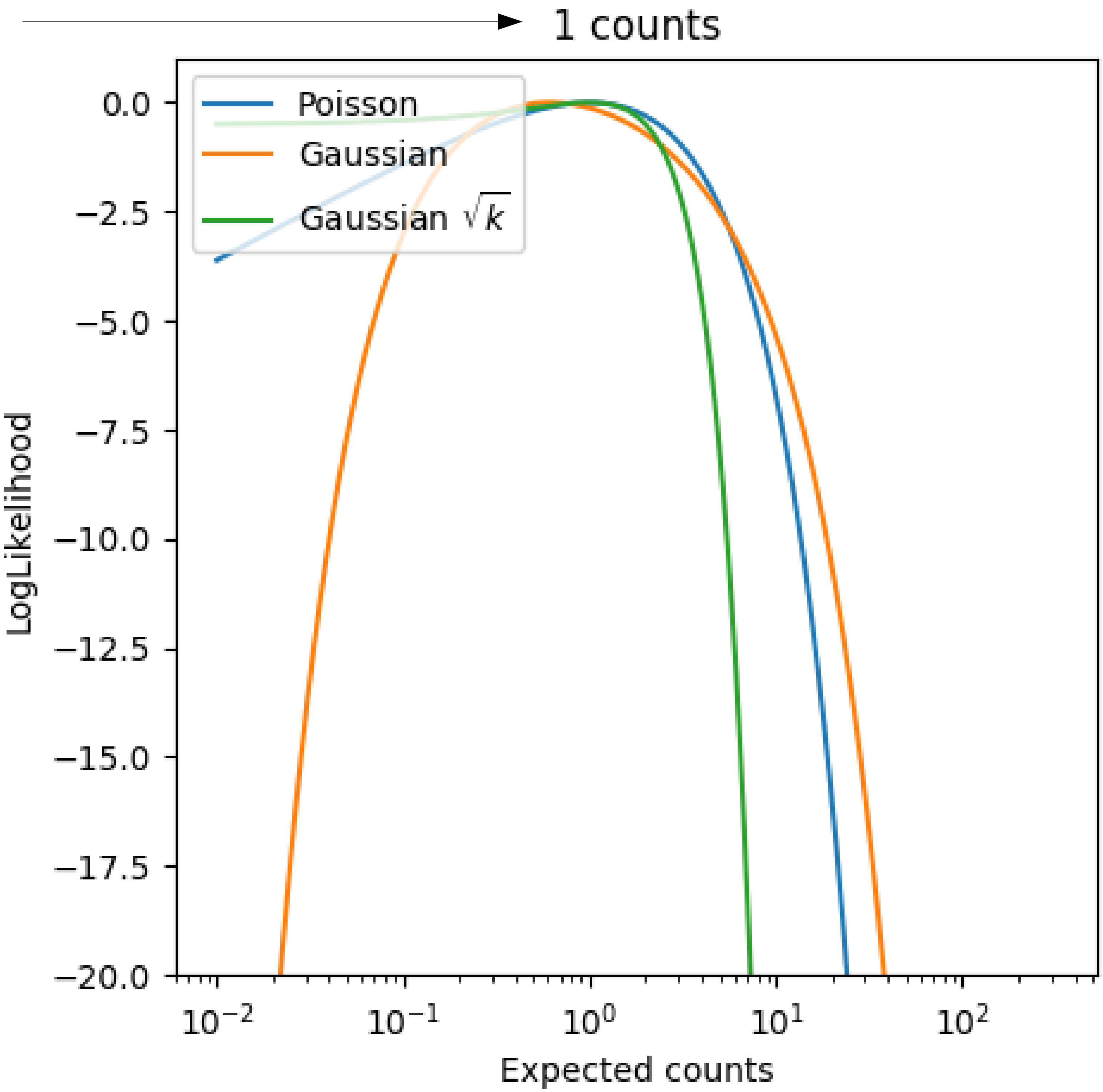
Known data

$$P(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Unknown rate

Likelihood

Probability (frequency) to produce exactly this data



Known data

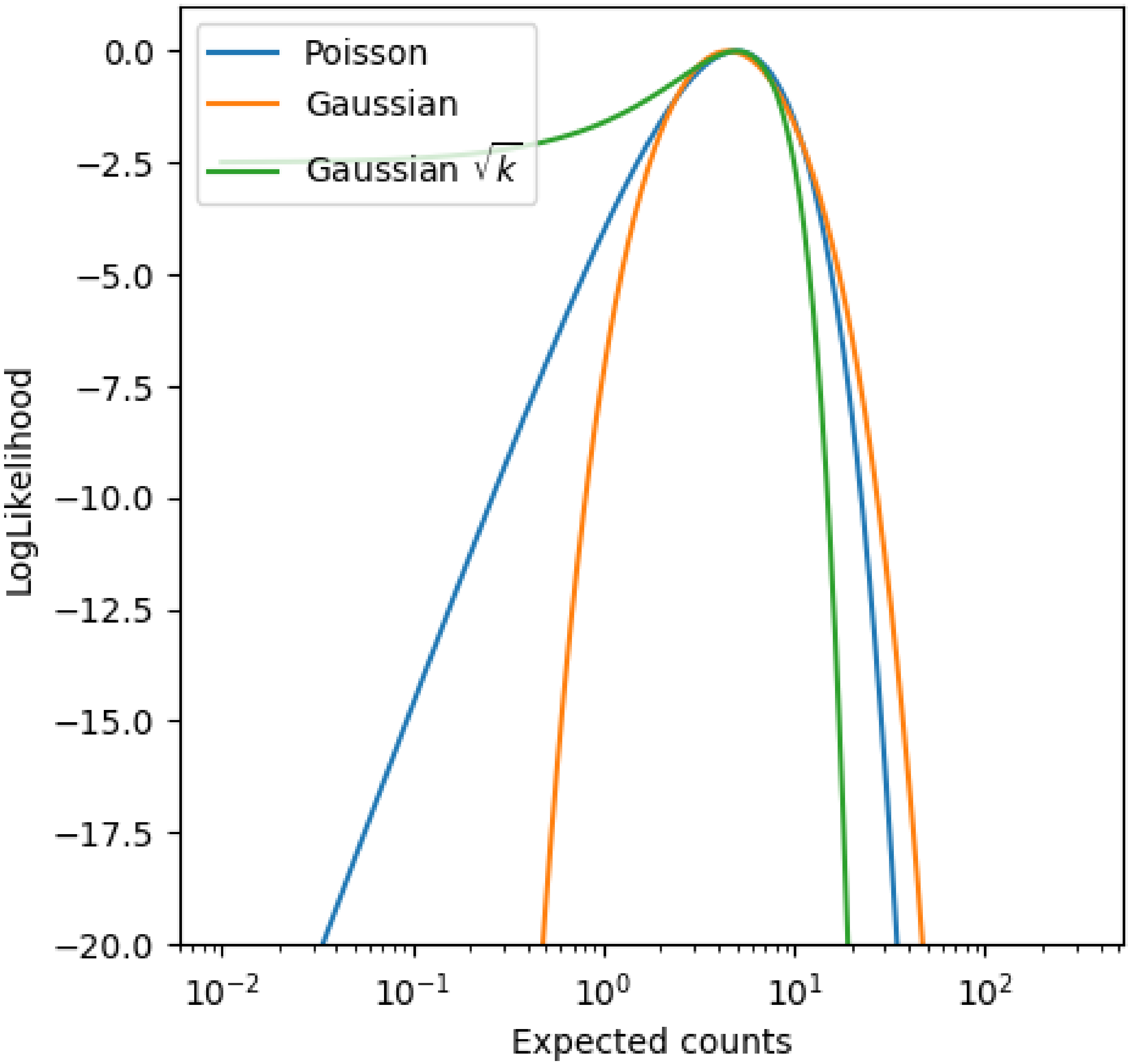
$$P(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Unknown rate

Likelihood

Probability (frequency) to produce exactly this data

5 counts



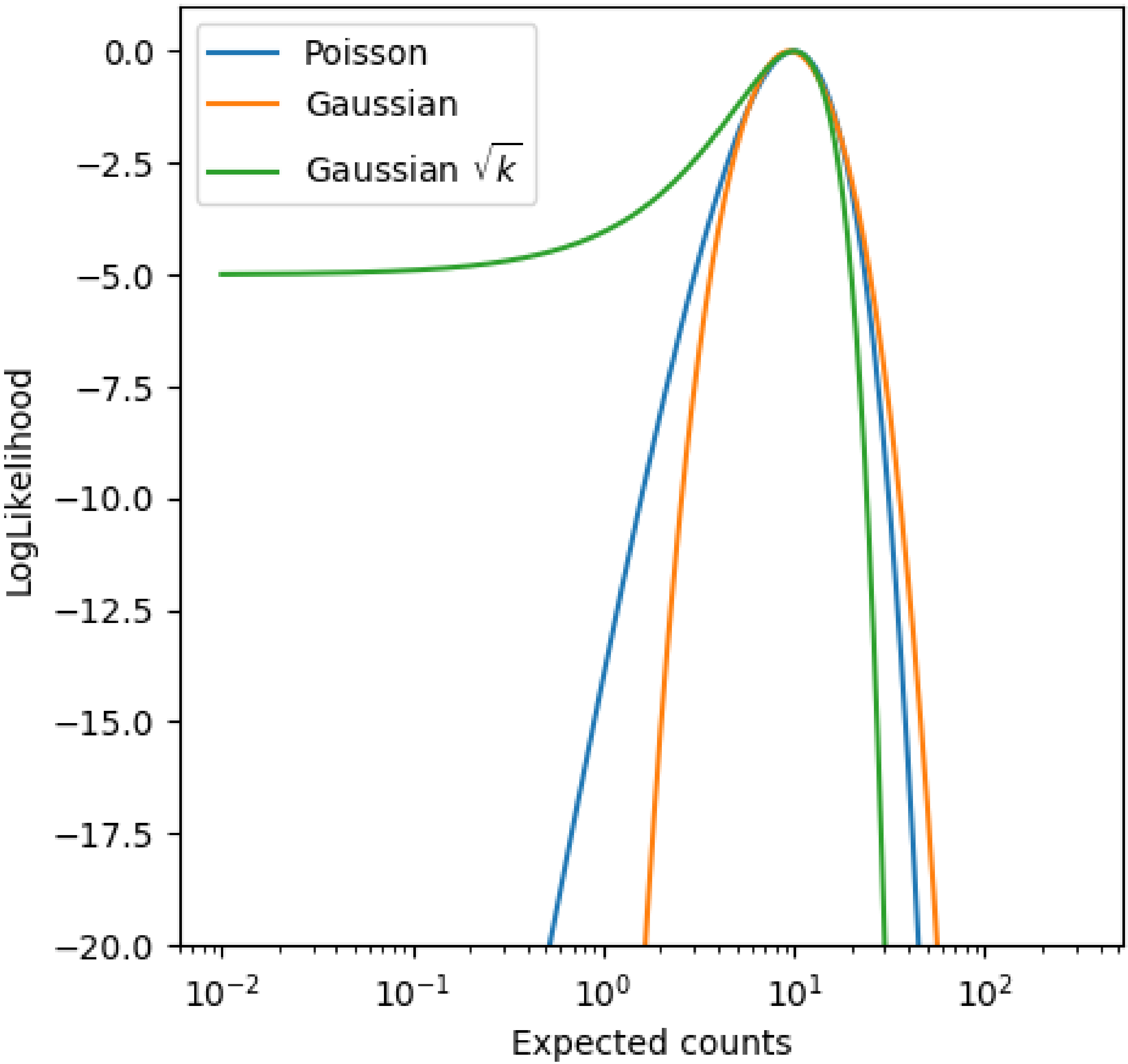
Known data

$$P(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Unknown rate

Likelihood

Probability (frequency) to produce exactly this data



10 counts

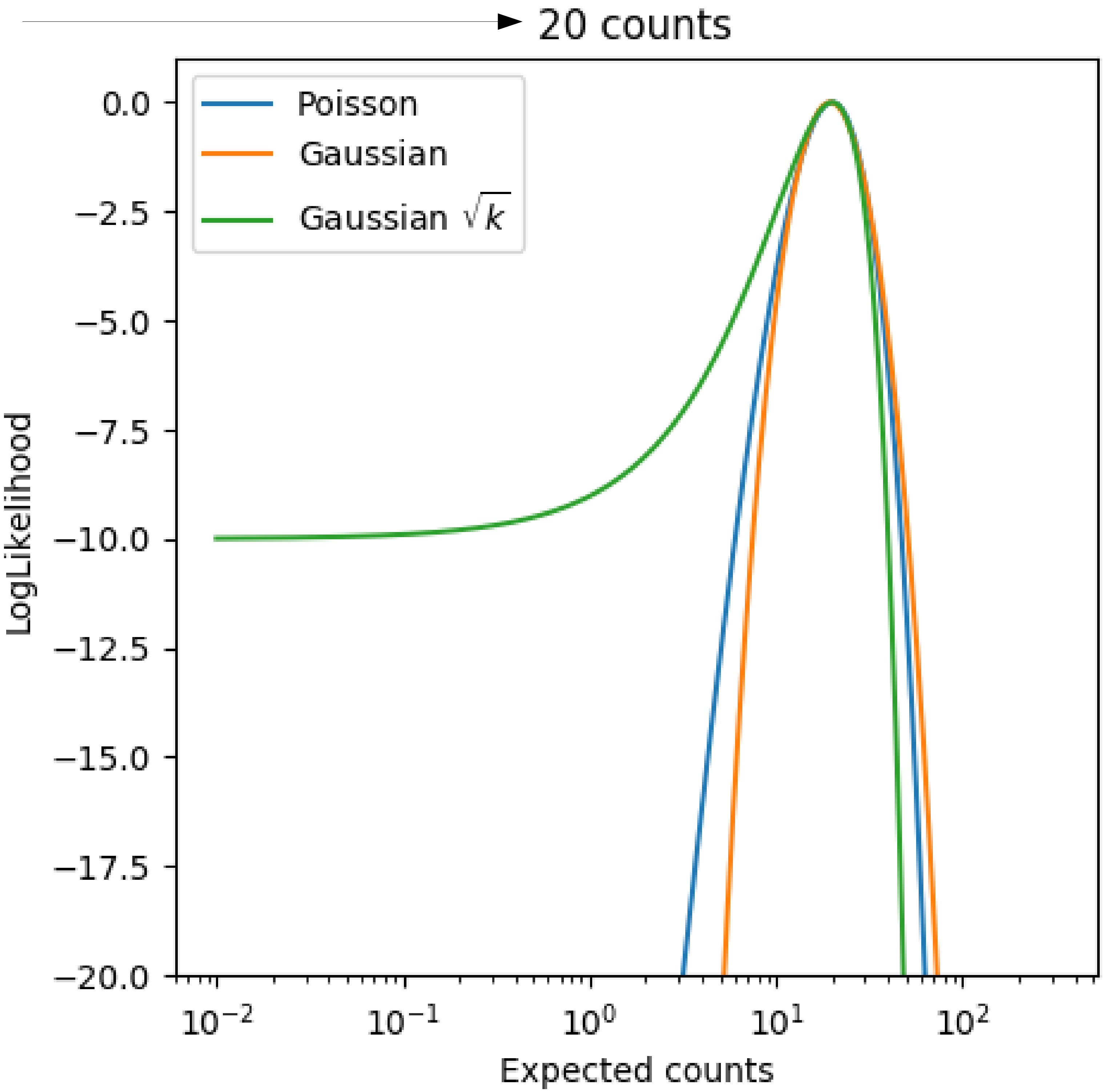
Known data

$$P(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Unknown rate

Likelihood

Probability (frequency) to produce exactly this data



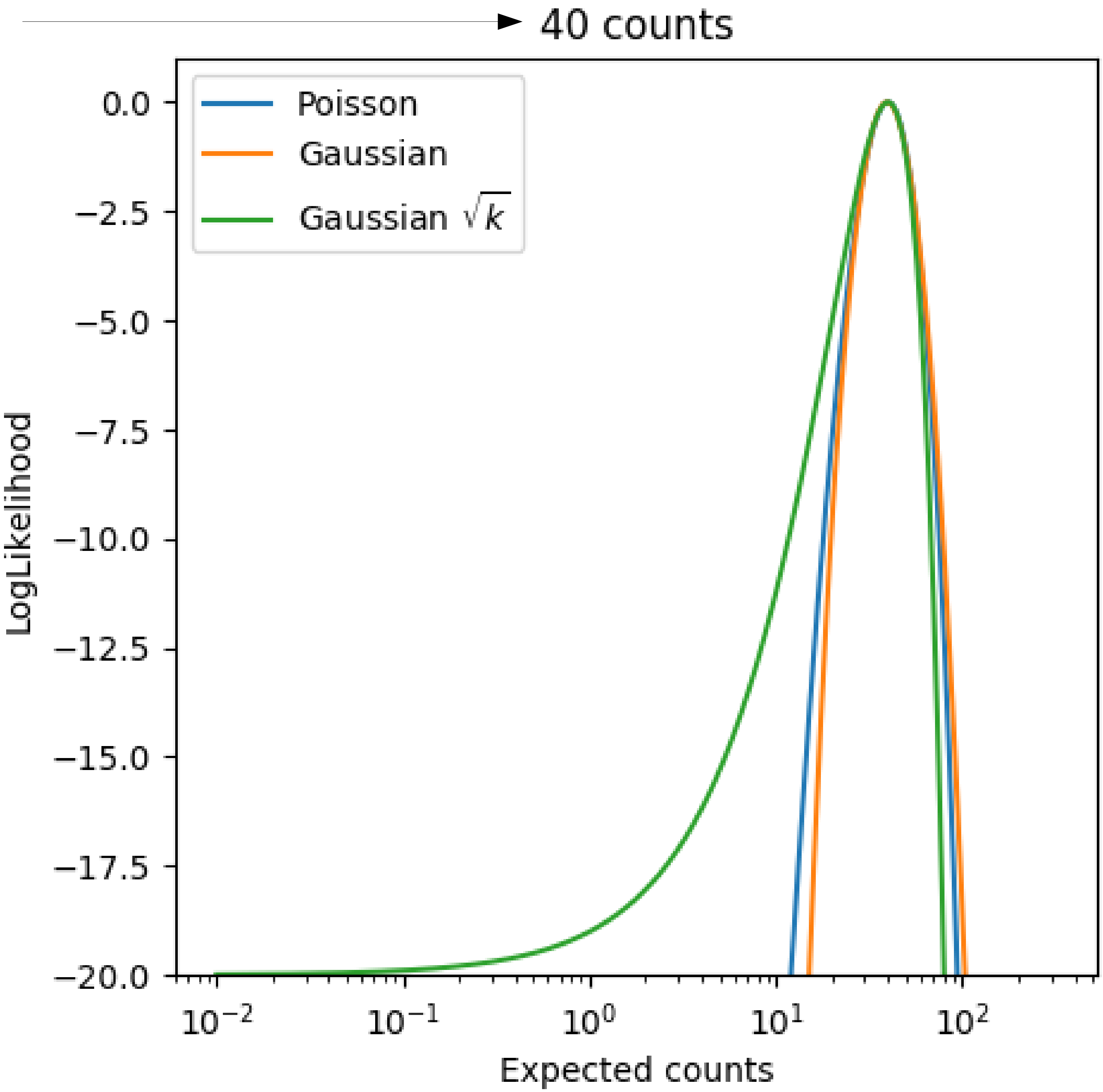
Known data

$$P(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Unknown rate

Likelihood

Probability (frequency) to produce exactly this data



Known data

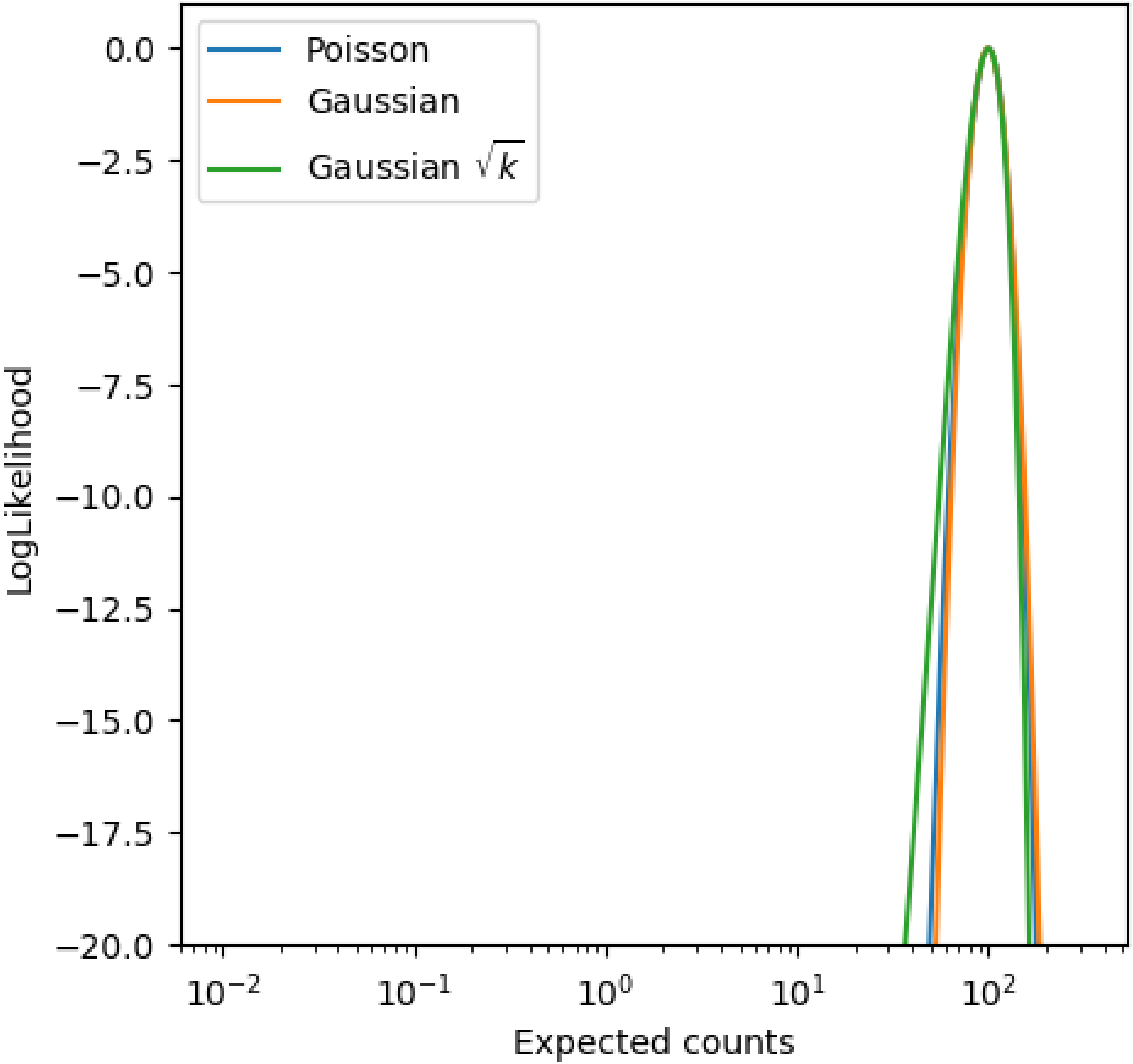
$$P(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Unknown rate

Likelihood

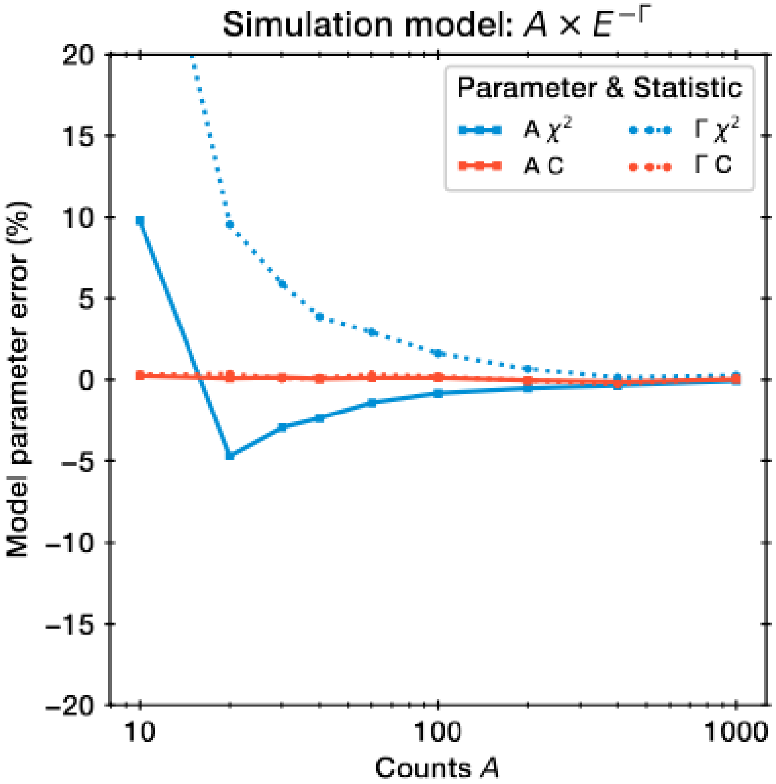
Probability (frequency) to produce exactly this data

100 counts



Approximation quality

- Tails have different slopes
  - Gauss high-end more permissive
  - Poisson low-end more permissive
- Right way: Poisson
- Historically: Gauss faster to evaluate



Best fits are biased (in %) if assuming chi² statistics

Humphrey et al. (2009), Mighell (1999)

Wheaton et al. (1995); Nousek & Shue (1989); Mighell (1999; van Dyk et al. (2001)

“Statistics”

- Poisson
  - Likelihood $\mathcal{L}(k|\lambda) = e^{-\lambda}\lambda^k/k!$

    -2*log → $-2\log\mathcal{L}(k|\lambda) = 2\lambda - 2k\log\lambda + C$

- Gaussian
  - Likelihood $\mathcal{L}(k|\mu,\sigma) = \exp[-((x-\mu)/\sigma)^2/2]/\sqrt{2\pi\sigma^2}$

    -2*log → $-2\log\mathcal{L}(x|\mu,\sigma) = ((x-\mu)/\sigma)^2 + C$

“Statistics”

- Poisson
  - Likelihood $\mathcal{L}(k|\lambda) = e^{-\lambda}\lambda^k/k!$

    -2*log → $-2\log\mathcal{L}(k|\lambda) = \underbrace{2\lambda - 2k\log\lambda + C}_{\text{CStat, Cash}}$

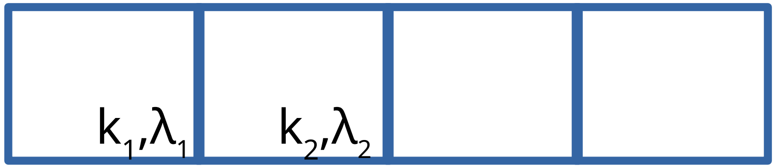
Cash (1979)

- Gaussian
  - Likelihood $\mathcal{L}(k|\mu,\sigma) = \exp[-((x-\mu)/\sigma)^2/2]/\sqrt{2\pi\sigma^2}$

    -2*log → $-2\log\mathcal{L}(x|\mu,\sigma) = \underbrace{((x-\mu)/\sigma)^2}_{\text{Chi}^2} + C$

Does not mean they follow a chi² distribution!

Multiple bins



- Poisson

$$\mathcal{L}(k_1,k_2|\lambda_1,\lambda_2) = e^{-\lambda_1}\lambda^{k}_{1}{}_{1}e^{-\lambda_2}\lambda^{k}_{2}{}_{2}$$

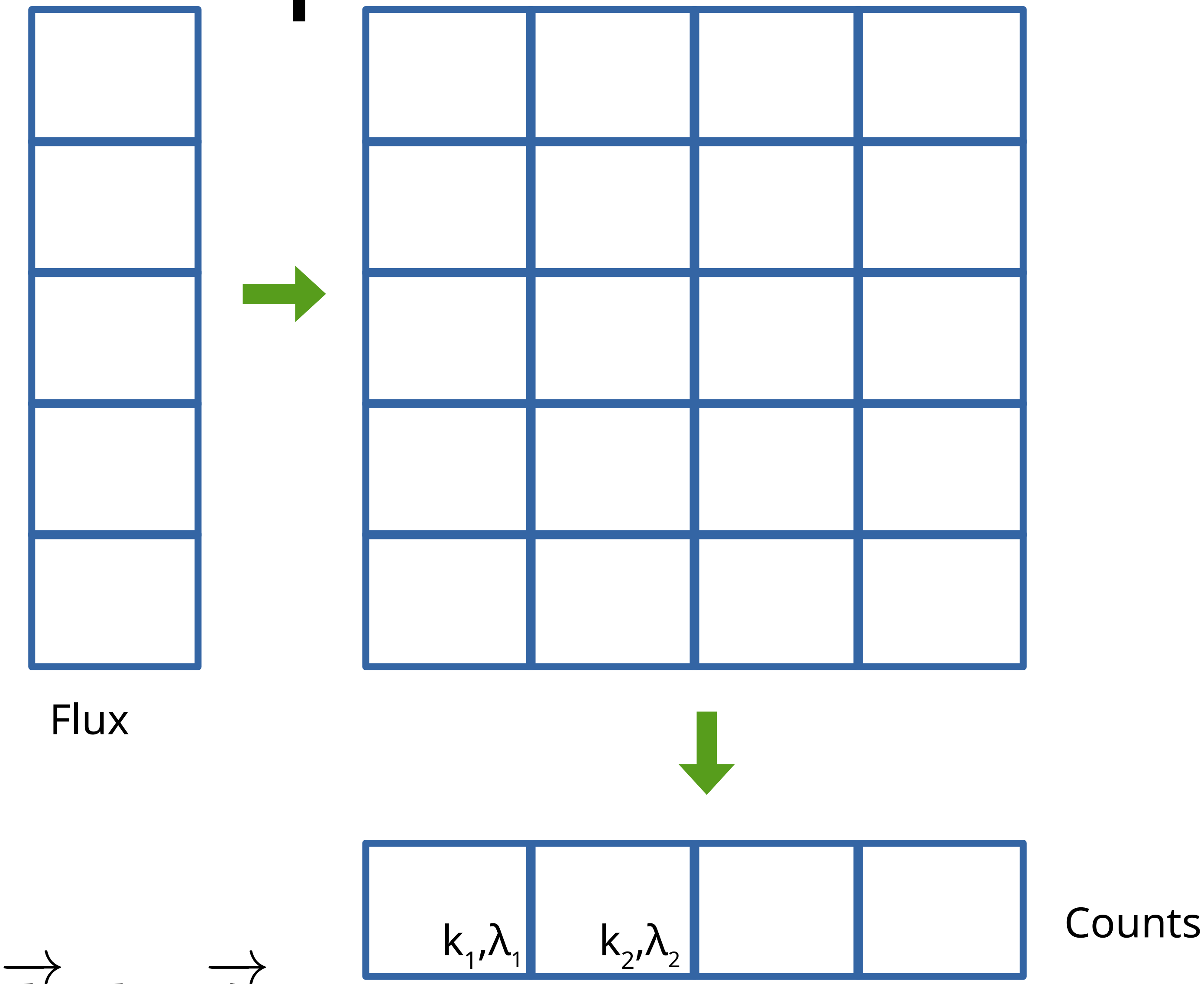
$$2\lambda_1 - 2k_1\log\lambda_1 + 2\lambda_2 - 2k_2\log\lambda_2 + C$$

- Gaussian

$$\mathcal{L}(x_1,x_2|\mu_1,\sigma_1,\mu_2,\sigma_2) = \exp[-((x_1-\mu_1)/\sigma_1)^2]\exp[-((x_2-\mu_2)/\sigma_2)^2]$$

$$((x_1-\mu_1)/\sigma_1)^2 + ((x_2-\mu_2)/\sigma_2)^2$$

Multiple bins



$$\vec{\lambda} = \vec{F} \cdot \underline{R}$$

$$C = 2\,\vec{\lambda} \cdot \vec{\lambda} - 2\,\vec{k} \cdot \log \vec{\lambda}$$

Remember:
λ=number / cm^2 / s / keV * dE * dt * dA
k=number

Inference with likelihoods

-0.5 Cstat, -0.5 chi²

$$\mathcal{L}(\overrightarrow{k}|\theta_1, \theta_2, ..., \theta_d, M, R, B, ...)$$

Higher L: model under these parameters often makes this data
Lower L: less frequently

$$P(D|\theta)$$

→ Frequency of data

Likelihood function at D, at parameter values   (not a density)

Inference desiderata

- Parameter ranges allowed or probable (L, T, …, physical parameters)

$P(\theta|D)d\theta$ Probability <u>density</u>

In infinitely small region: zero probability

$$\underbrace{\underset{\text{Density}}{\underset{\uparrow}{P(\theta|D)}}\,\underset{\text{Volume}}{\underset{\uparrow}{d\theta}}}_{\text{Probability mass}}$$

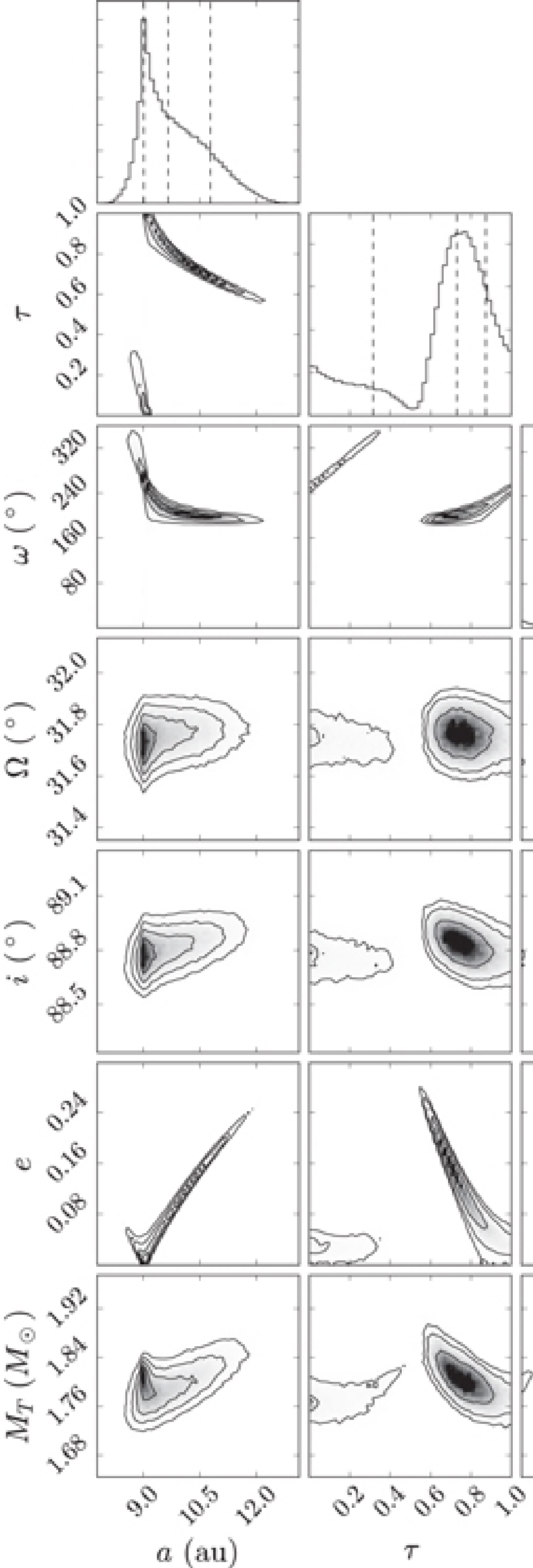
⇒ Find regions with high probability mass

$P(D|\theta)$

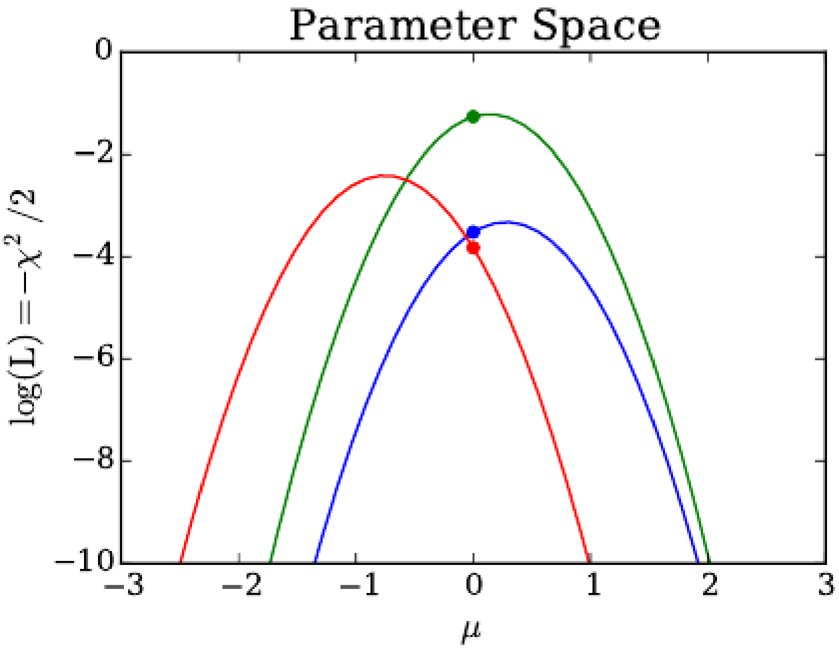
Parameter space exploration

Parameter space exploration

- Local optimization
  - LM, simplex, … (many)
  - Monte carlo optimization
- Local sampling: MCMC
  - Tempering
  - Limitations
- Global optimization
  - Genetic algorithms (DE)
- Global sampling
  - Nested sampling



Best fit parameters



- If away from boundary
- If model is linear
- If ndata → high (symmetric, single gauss)
- If θ is true parameter

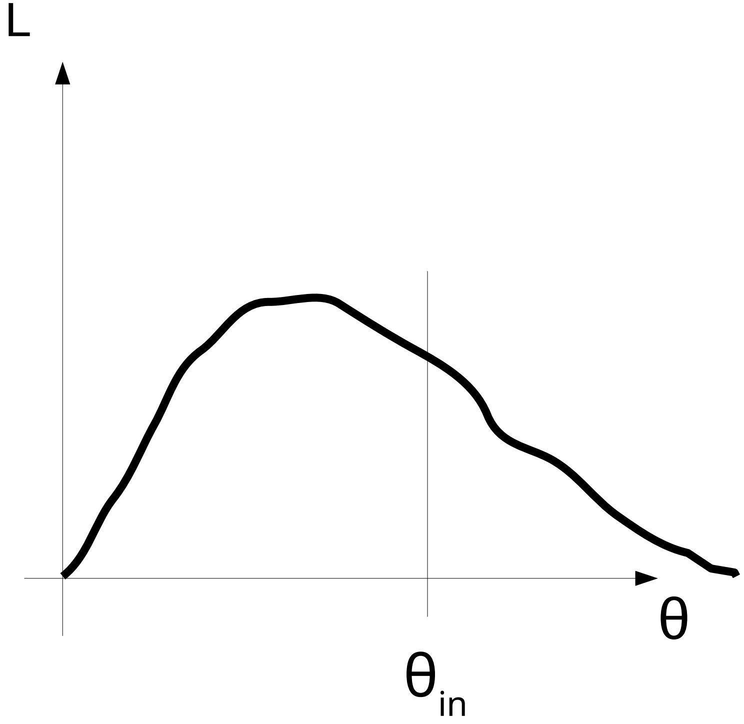
→ then

If many data are created under $\hat{\mu_D}$
logL interval -1/2 below best fit (Wilks' theorem)
Contains true value 68% of realisations

$P(\hat{\mu}|\hat{\mu_D})$ Confidence interval

What was the question again? $P(\mu|D)$
Are conditions fulfilled?
What do unequal "errors" mean?
2d?

Best fit parameters



- If away from boundary
- If model is linear
- If ndata → high

(symmetric, single gauss)

- If θ is true parameter

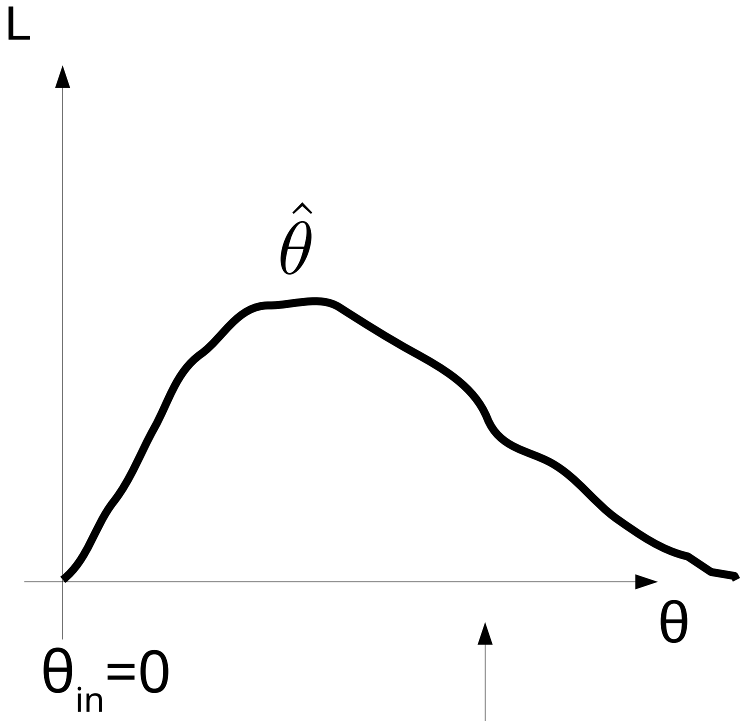
(always)

If conditions are not met

→ Monte Carlos simulations (parametric bootstrap)

Calibrate a Confidence interval

Detection

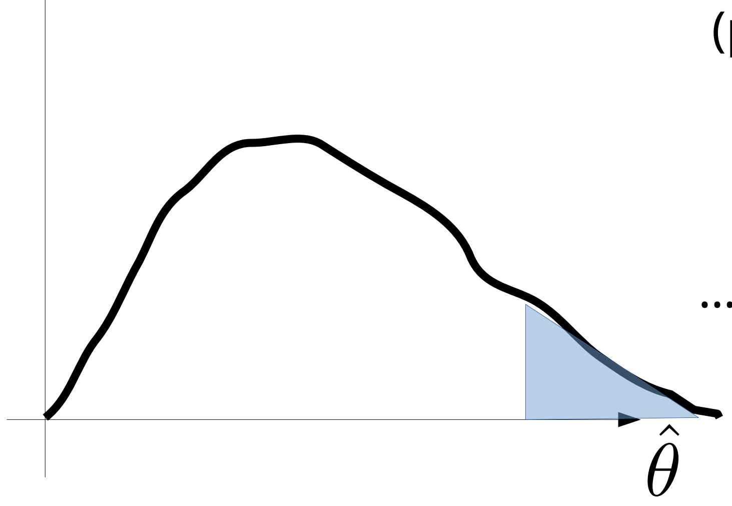


If conditions are not met

(always)

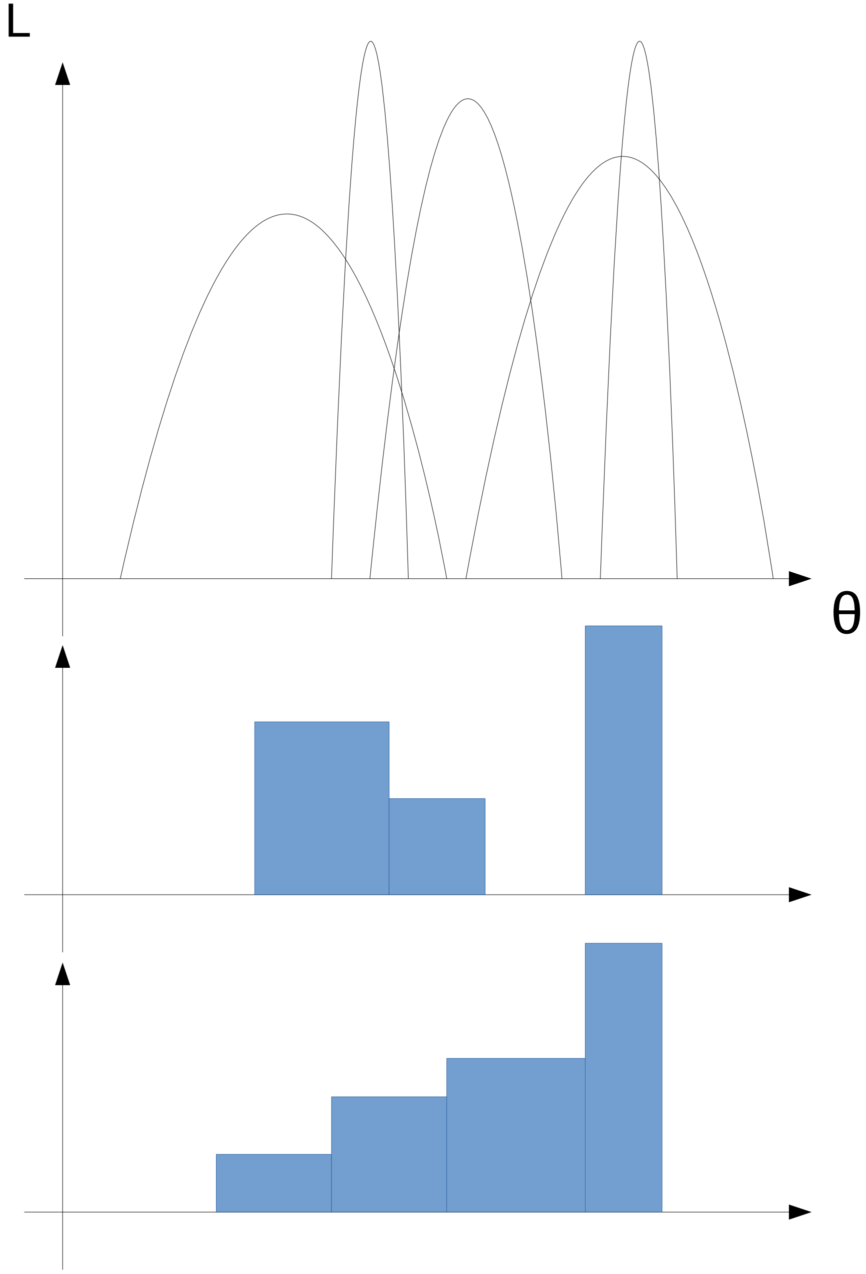
- If away from boundary
- If model is linear
- If ndata → high (symmetric, single gauss)
- If θ is true parameter

→ Monte Carlos simulations (parametric bootstrap)



… p-values

Best fit distributions



Convolution of

True parameter distribution +
Measurement error & analysis method

Confidence intervals

Histogram of best fits

Meaning?
Upper limits?

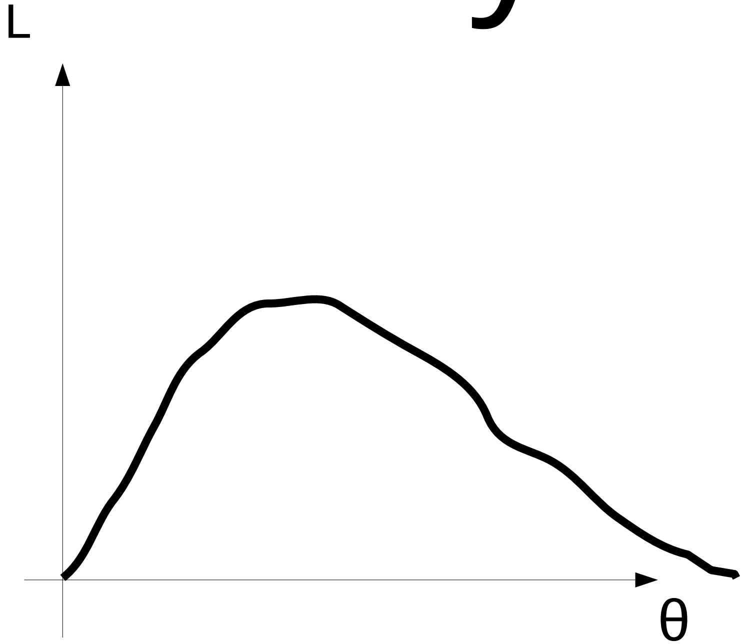
Cumulative distribution

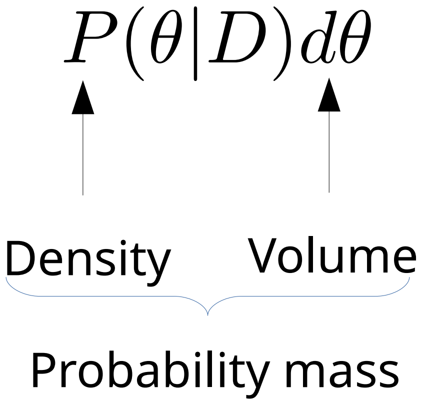
Clean solution:
Hierarchical Bayesian Model (HBM)
for example Baronchelli, Nandra & Buchner (2020)
https://github.com/JohannesBuchner/PosteriorStacker

Sampling

Bayesian posterior

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$





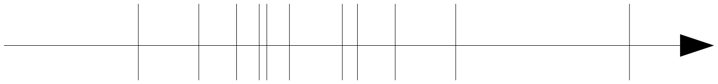
$$P(\theta|D)d\theta$$

Density

Volume

Probability mass

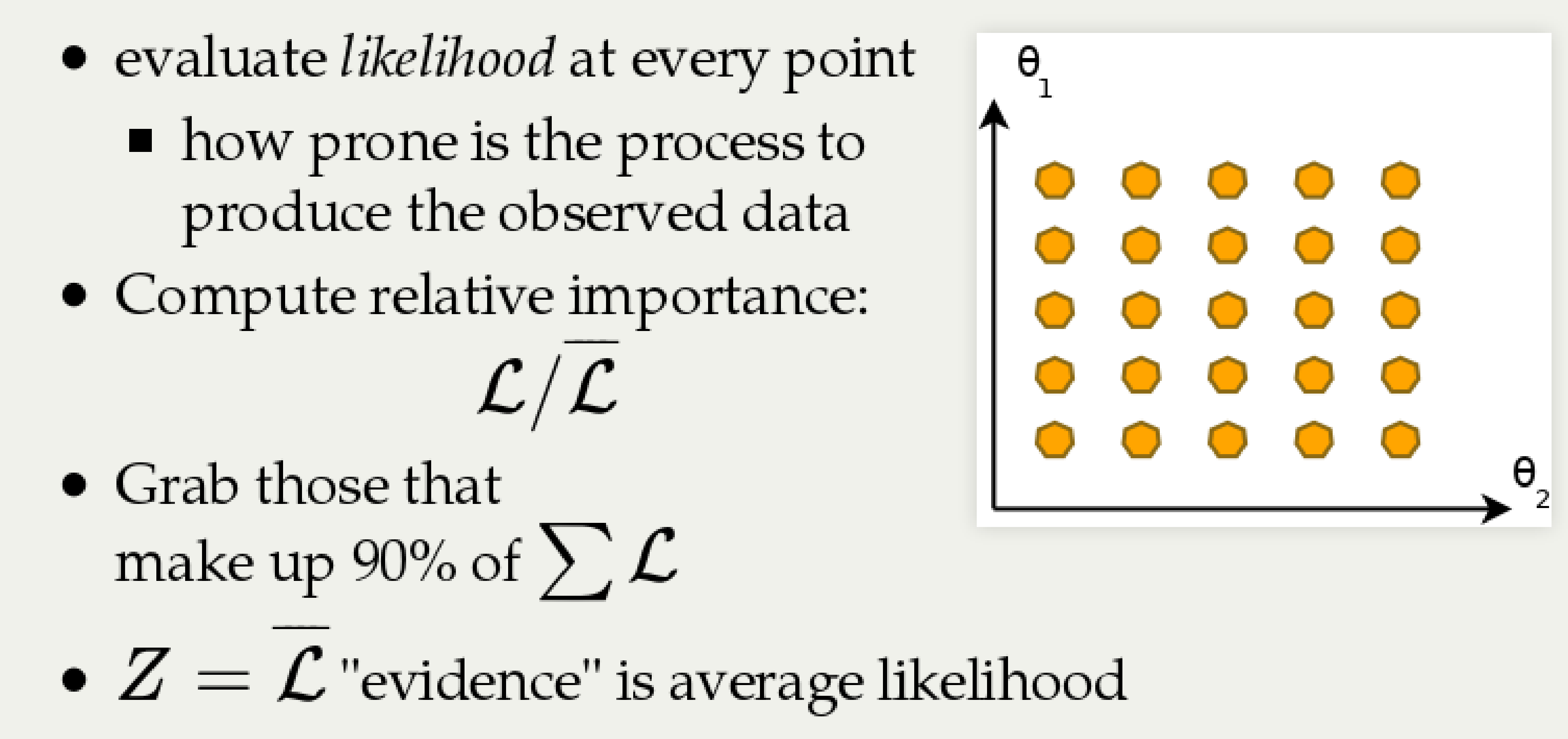
Find regions with high probability mass



Idea: Sample parameter solutions proportionally to their probability

For example with a grid

Posterior grid

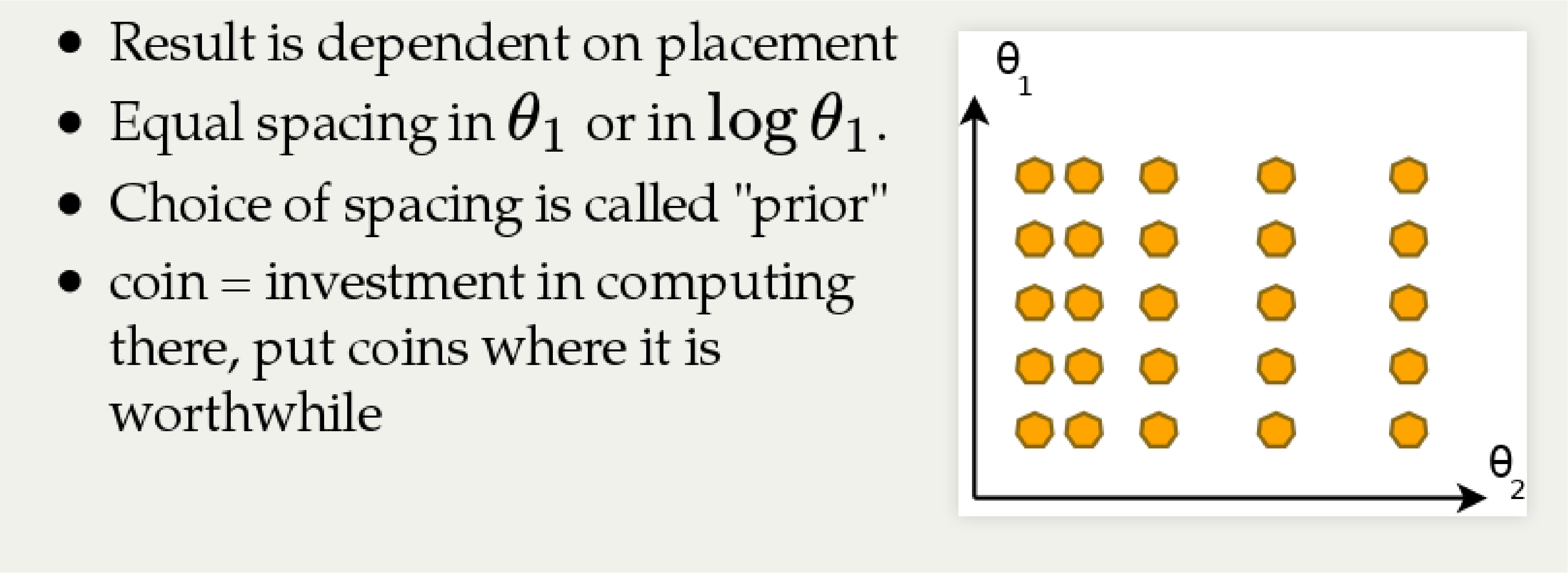


- evaluate *likelihood* at every point
  - how prone is the process to produce the observed data
- Compute relative importance:

$$\mathcal{L}/\overline{\mathcal{L}}$$

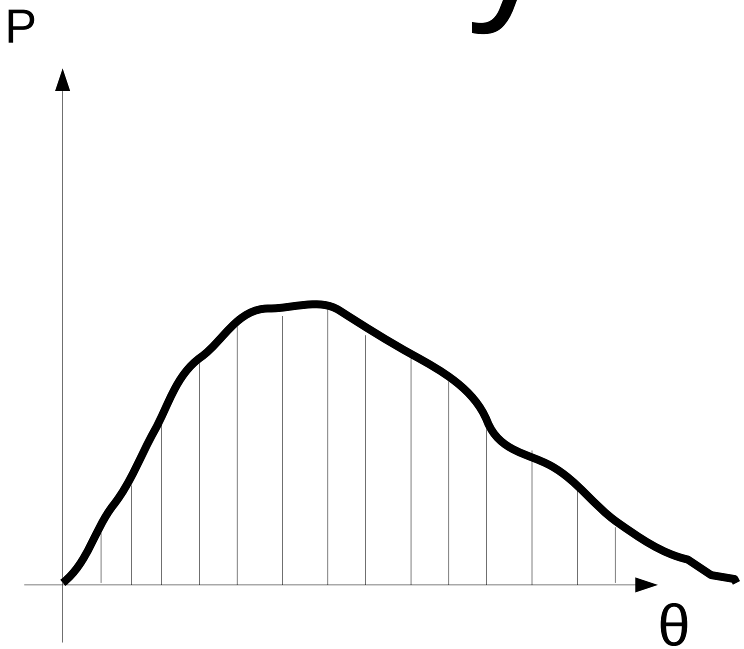
- Grab those that make up 90% of $\sum \mathcal{L}$
- $Z = \overline{\mathcal{L}}$ "evidence" is average likelihood

Posterior grid



- Result is dependent on placement
- Equal spacing in θ_1 or in $\log \theta_1$.
- Choice of spacing is called "prior"
- coin = investment in computing there, put coins where it is worthwhile

Bayesian posterior



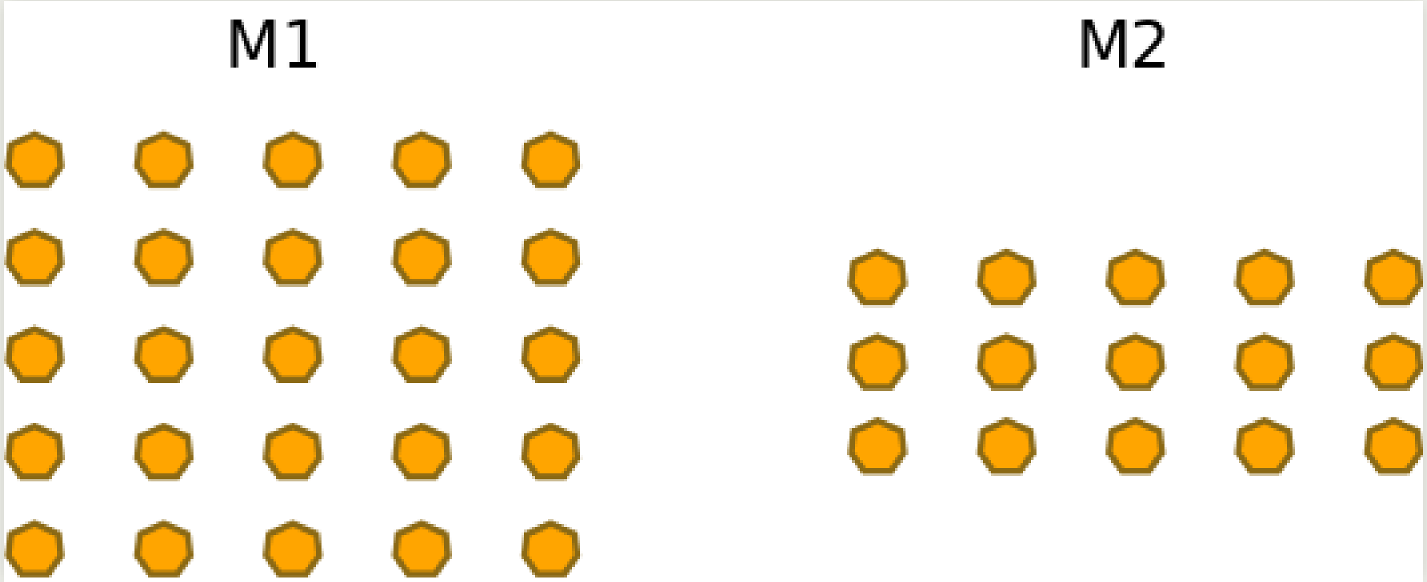
parameter solutions weighted by their probability

Credible intervals

Definitions:

Density → cumulative → quantiles

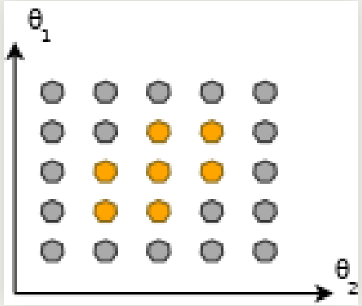
Two models

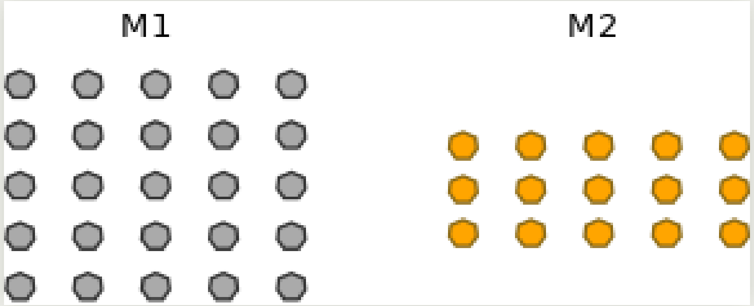


- Compare two parameter spaces by
$$\sum \mathcal{L}\Big|_{M1} \Big/ \sum \mathcal{L}\Big|_{M2}$$
- How many coins to put in M1, M2?
- model prior

Parameter Estimation vs. Model Comparison

- Remove coins contributing less than 10%.
- Under Bayesian inference, same problem:
  - comparing bags of hypotheses





- prior is measure, rule of averaging, deformation of space to "natural variables", investment in/weighting of sub-regions
- most common priors: uniform, log-uniform.
- model priors are relative size of spaces

Curse of dimensionality

- k^d grid → infeasible
- Sample θ

  θ_1 θ_2 θ_3 ….

  w_1 w_2 w_3 ….

  (Posterior chains)

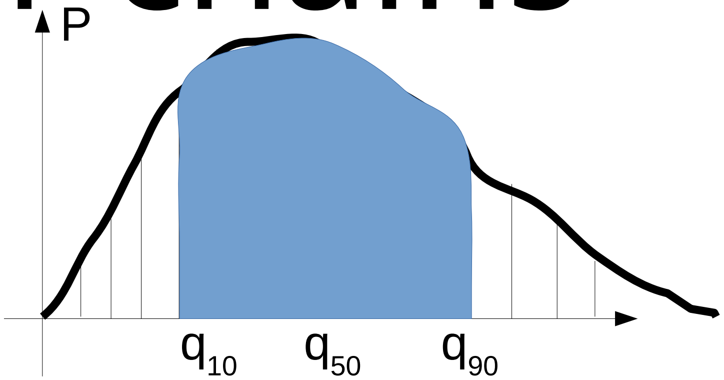
- Techniques:
  - Importance sampling
  - MCMC
  - Nested sampling

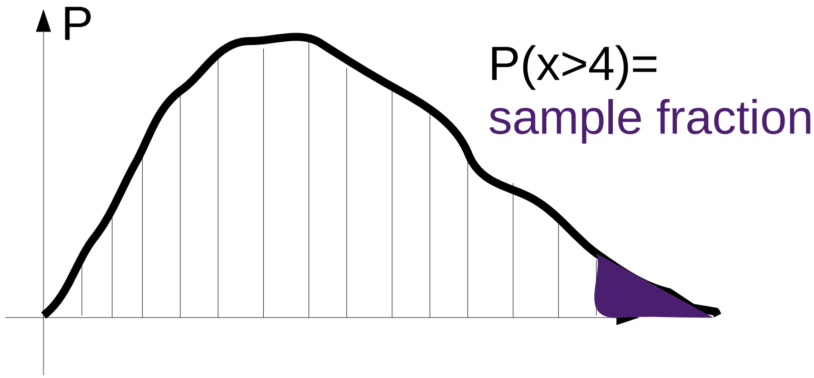
Using posterior chains

- Posterior chain

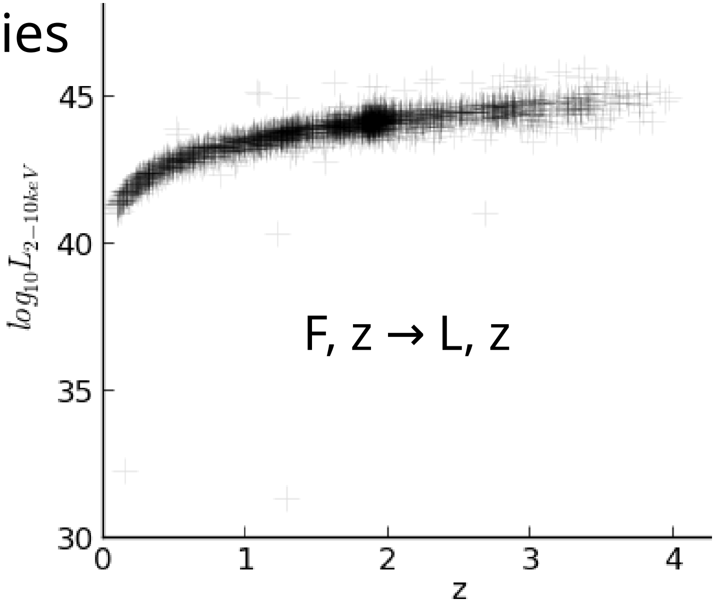
  $\theta_1\ \theta_2\theta_3$ ….

- Find regions with high prob
- Compute prob. of regions
- Posterior predictions
- Derived quantities

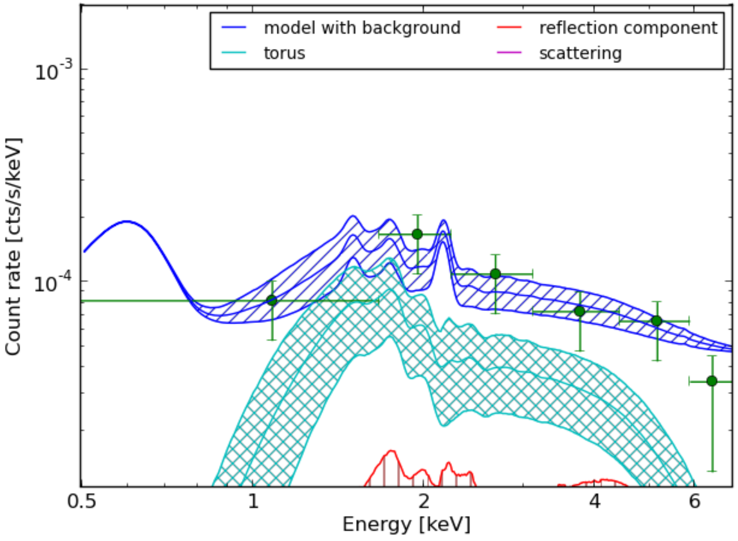




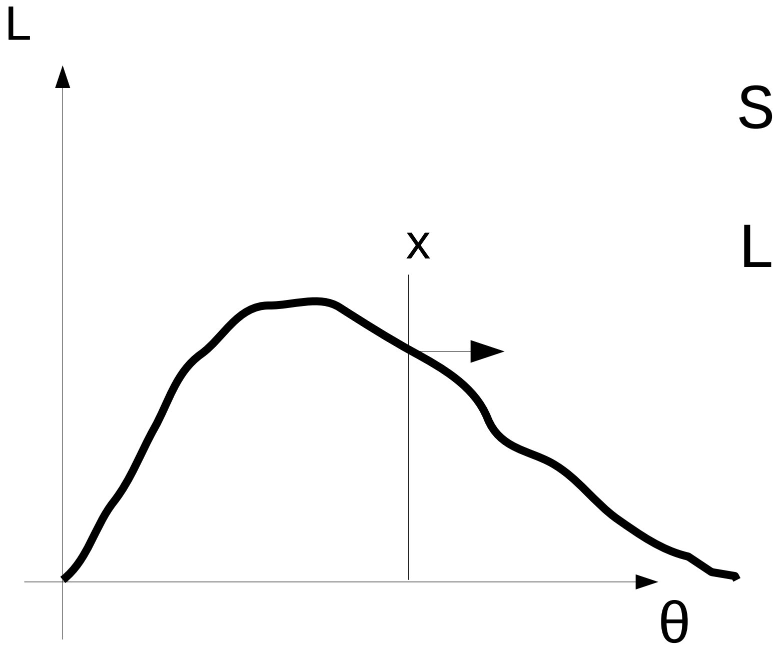
P(x>4)=
sample fraction

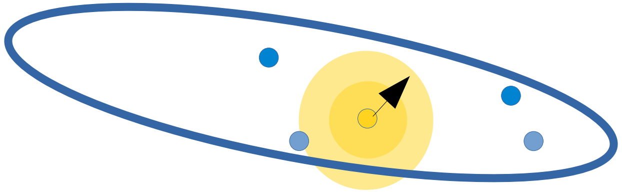


F, z → L, z



Markov Chain Monte Carlo

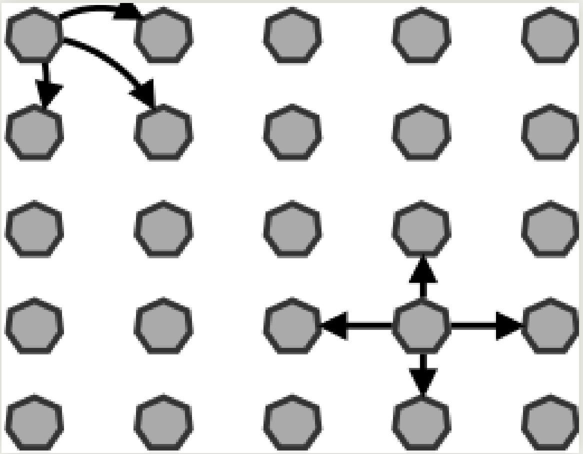




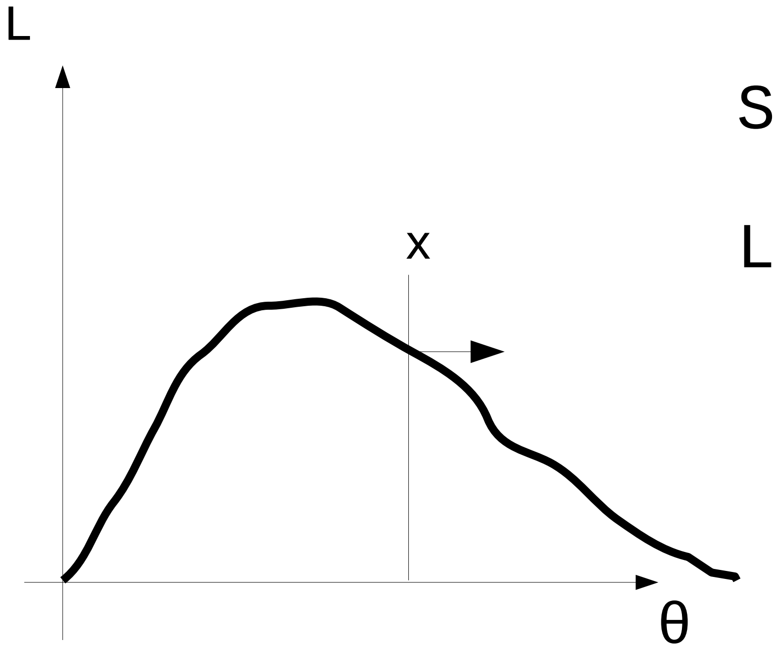
```
Starting point θ

Loop forever:
   θ' = Normal(θ, sigma_p)
   if P(θ'|D)/P(θ|D) > U():
      θ = θ'
   add θ to chain
```

- Missing ingredient: transition kernel
- tune to the problems
- Fraction of visits ~ converges to ~ probability of hypothesis
- Where does chain spend 90% of its visits



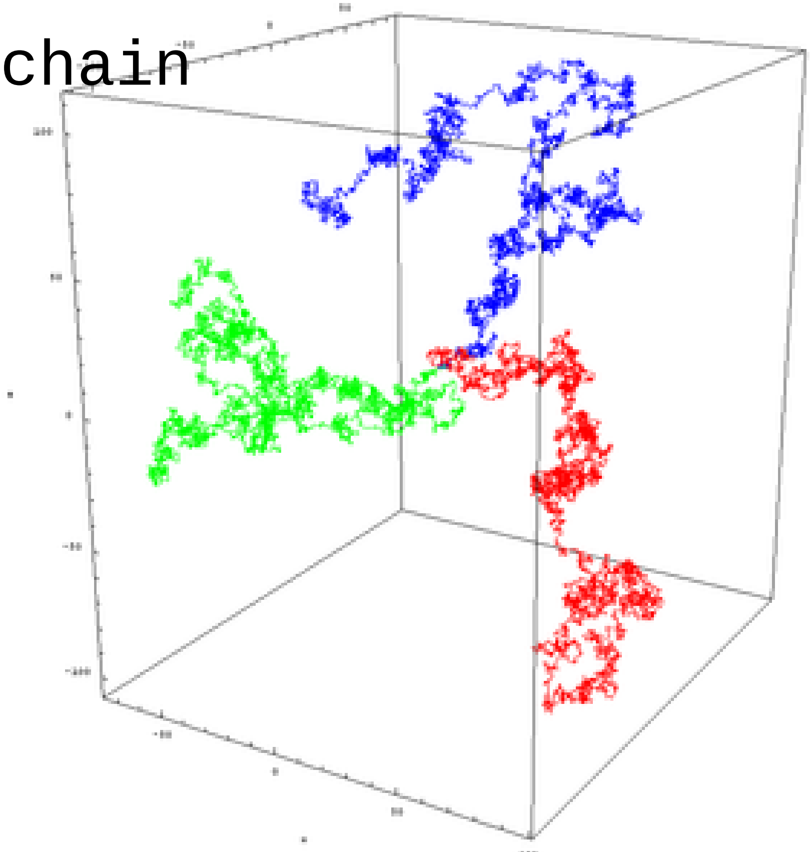
MCMC



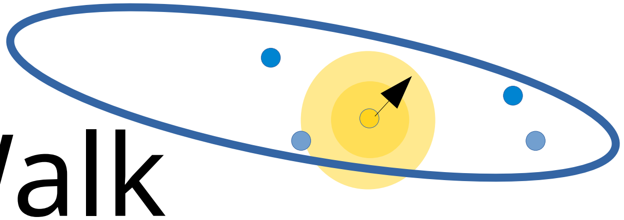
```
Starting point $\theta$

Loop forever:
   $\theta' = Normal(\theta, sigma\_p)$
   if $P(\theta'|D)/P(\theta|D) > U()$:
      $\theta = \theta'$
   add $\theta$ to chain
```

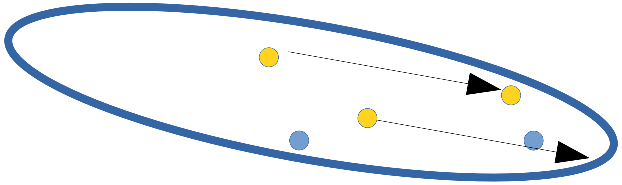
Emerging behaviour:



MCMC proposals



- Metropolis + Random Walk



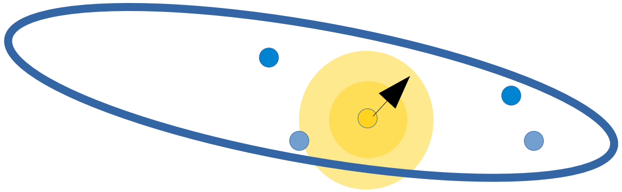
- Goodman-Weare (emcee)
- HMC (Hamiltonian Monte Carlo)

→ animation

https://chi-feng.github.io/mcmc-demo/app.html

Random walk, HMC

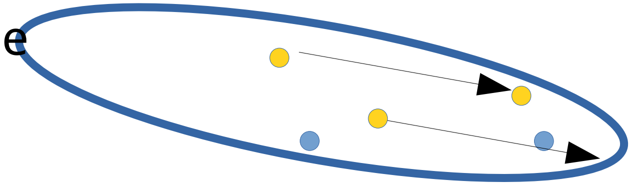
MCMC proposals



- Metropolis Random Walk
  - Adv: simple
  - Disadv: poor mixing

Goodman & Weare (2010)

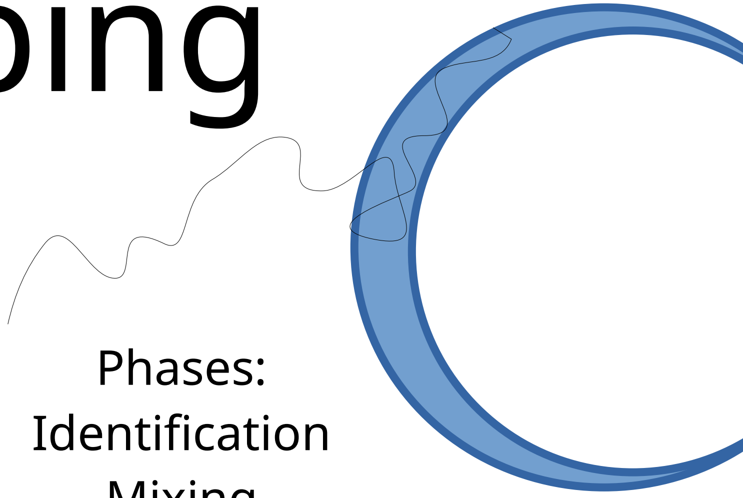
emcee



- Affine-invariant ensemble
  - Adv: auto-tuning for gaussian L
  - Disadv: poor mixing in bananas, collapses in high-d (Huijser+15)
- HMC (Hamiltonian Monte Carlo)
  - Adv: tunes itself to surface
  - Disadv: need gradients of models

MCMC stopping

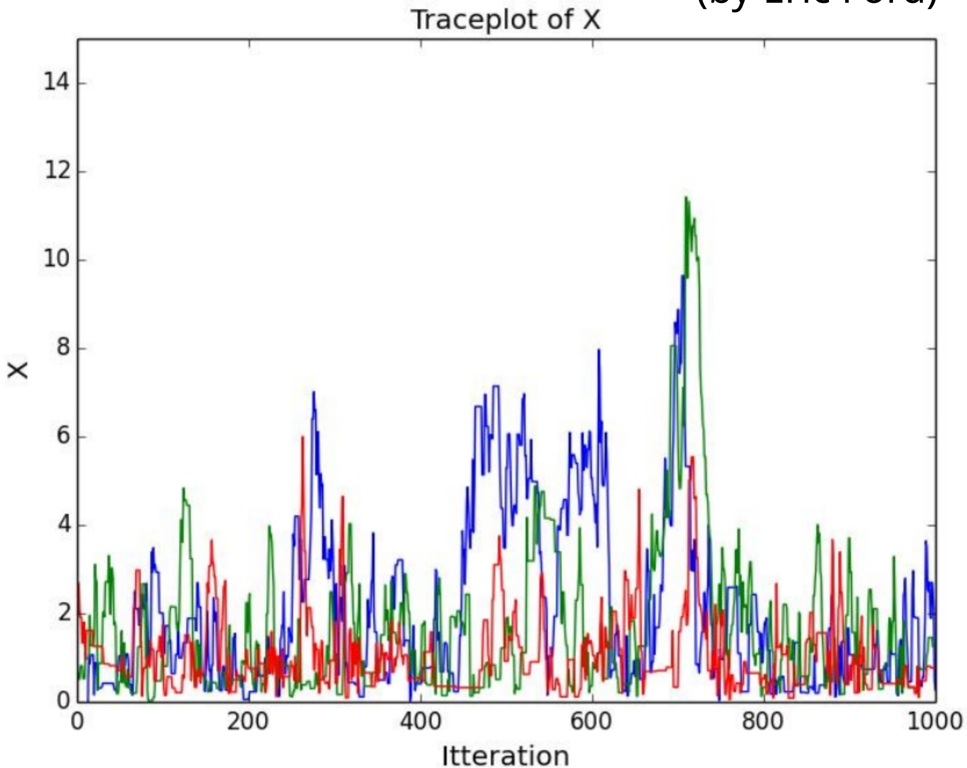
- MCMC theory: n→inf
- Trace plots
- Autocorrelation length
- Convergence tests
  - Detect if unreliable
  - Gelman-Rubin diagnostic
  - (many more)



Phases:
Identification
Mixing

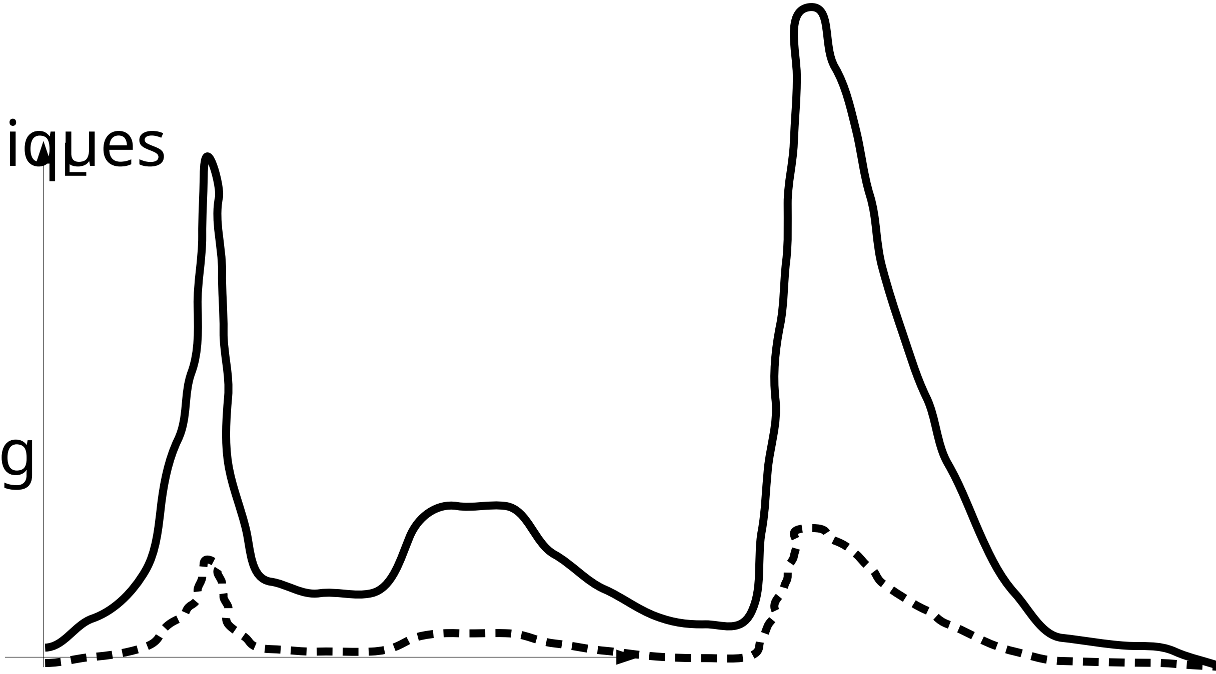
(burn-in)

(by Eric Ford)



Escaping local maxima: strategies

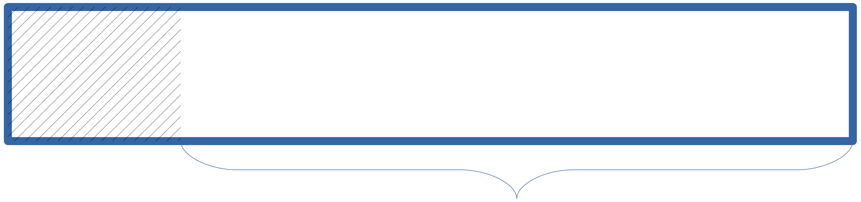
- Multiple random start positions
  - Augment local techniques
- Make surface easier
  - Tempering/Annealing
- Walker population
  - GW
  - Genetic algorithms (DE)



Model comparison

Model comparison Buchner+14

- Empirical models
  - Information content
  - Prediction quality
- Component presence
  - Regions of practical equivalence
- Physical effects
  - Bayesian model comparison
  - Priors often well-justified



https://arxiv.org/abs/1506.02273
Betancourt (2015)

Information criteria

- Akaike information criterion

Akaike (1973)

- Is more complex worth storing?

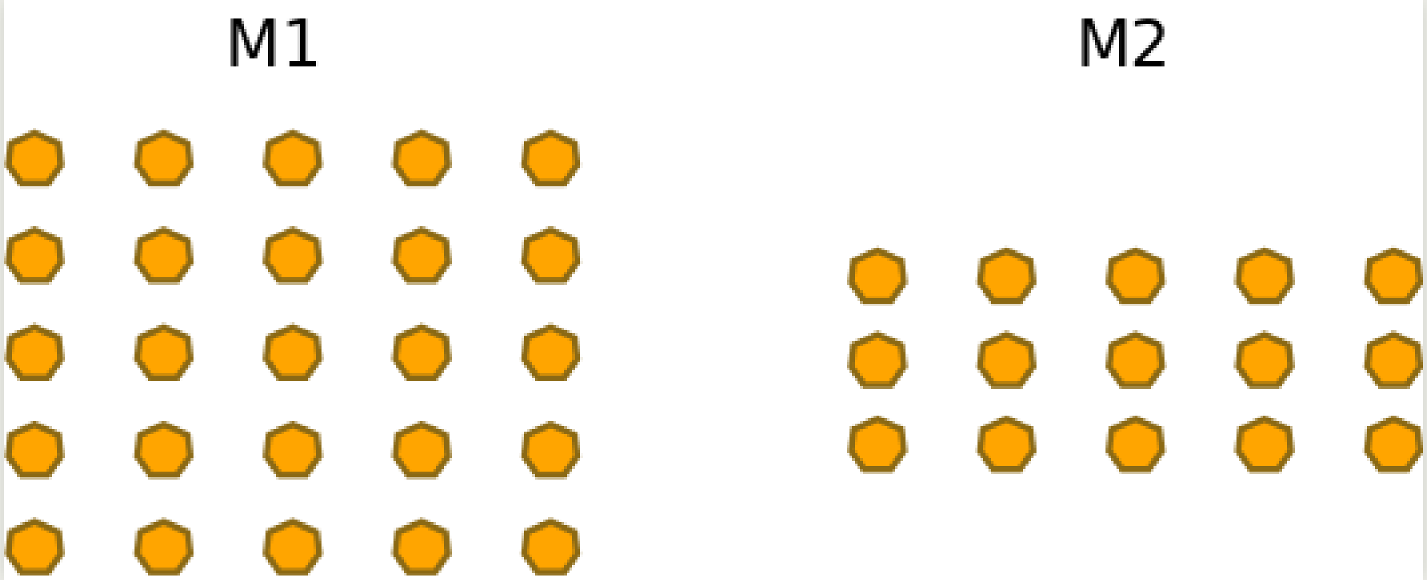
$AIC = 2 * d – 2 * L_{max}$

$AIC = 2 * d + CStat$

Advantages:
- rooted in information theory
- independent of prior

Disadvantages:
- No uncertainties, thresholds unclear
- ...

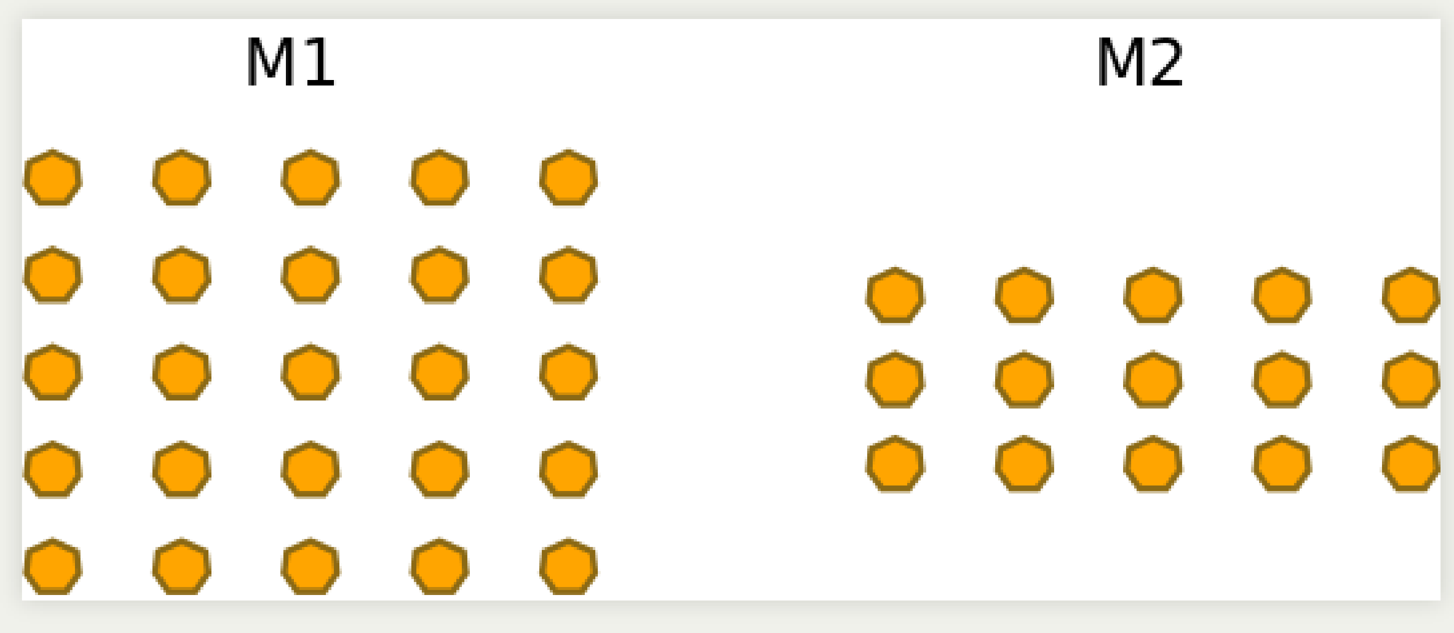
Two models



- Compare two parameter spaces by
$$\sum \mathcal{L}\Big|_{M1} / \sum \mathcal{L}\Big|_{M2}$$
- How many coins to put in M1, M2?
- model prior

Punishing prediction diversity

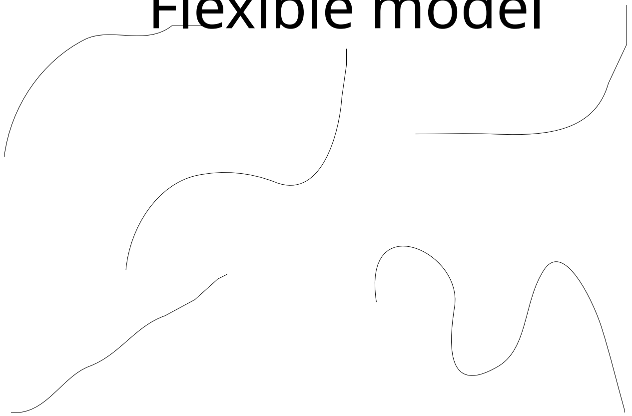
(not number of parameters)

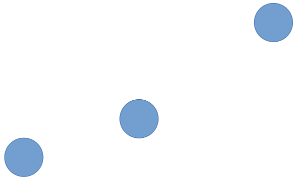


Flexible model

Inflexible model

Data





L high, V tiny

L medium, V medium

What to do with Z

- Z1, Z2

$$\underbrace{\frac{p(M1|D)}{p(M2|D)}}_{\text{Posterior odds ratio}} = \frac{Z1 \cdot p(M1)}{\underbrace{Z2}_{\text{Bayes factor}} \cdot \underbrace{p(M2)}_{\text{Prior odds ratio}}}$$

Buchner+14

What to do with Z

- Z1, Z2

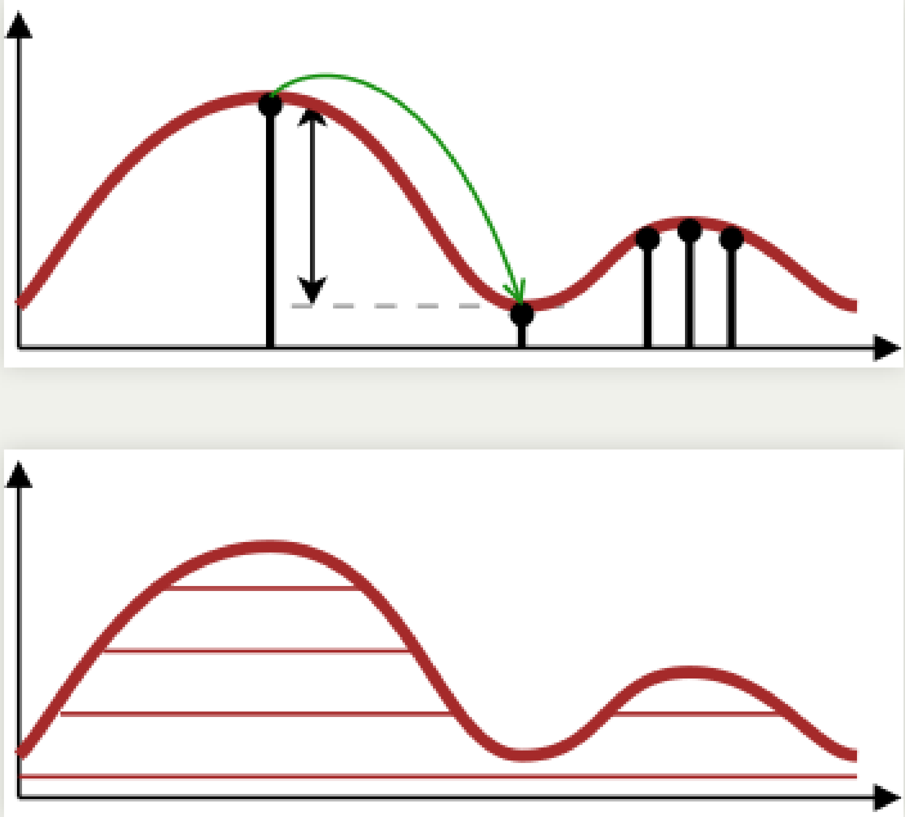
$$\frac{p(M1|D)}{p(M2|D)} = \frac{Z1 \cdot p(M1)}{Z2 \cdot p(M2)}$$

$$\frac{p(M_1|D)}{\sum p(M_i|D)} = \frac{Z_1 \cdot p(M_i)}{\sum_i Z_i \cdot p(M_i)}$$

- model priors: leave to reader or motivated by theory
- Discard highly improbable model or marginalise
- Does $\frac{p(M1|D)}{p(M2|D)} = 3/1$ mean M2 is correct in a quarter of the cases?

Global sampling

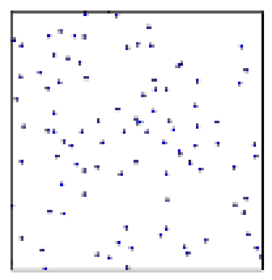
nested sampling idea



- MCMC: only consider likelihood ratios. Integration by vertical slices
- nested sampling: compute geometric size at various likelihood thresholds
- orthogonal, unique re-ordering of volume by likelihood

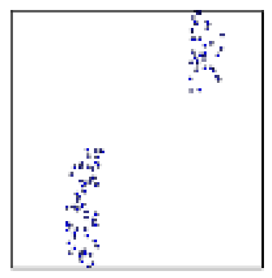
$$\sum \underbrace{\text{Shrinkage} \times \text{Likelihood}}_{\text{Importance of shell}} = Z$$

nested sampling algorithm



- Start with volume 1, draw randomly uniformly 200 points

- remove one, volume shrinks by 1/200.



- draw a new one excluding the removed volume
- Unique ordering of space required: via likelihood

draw a new uniformly random point,
with higher likelihood
(the crux of nested sampling)

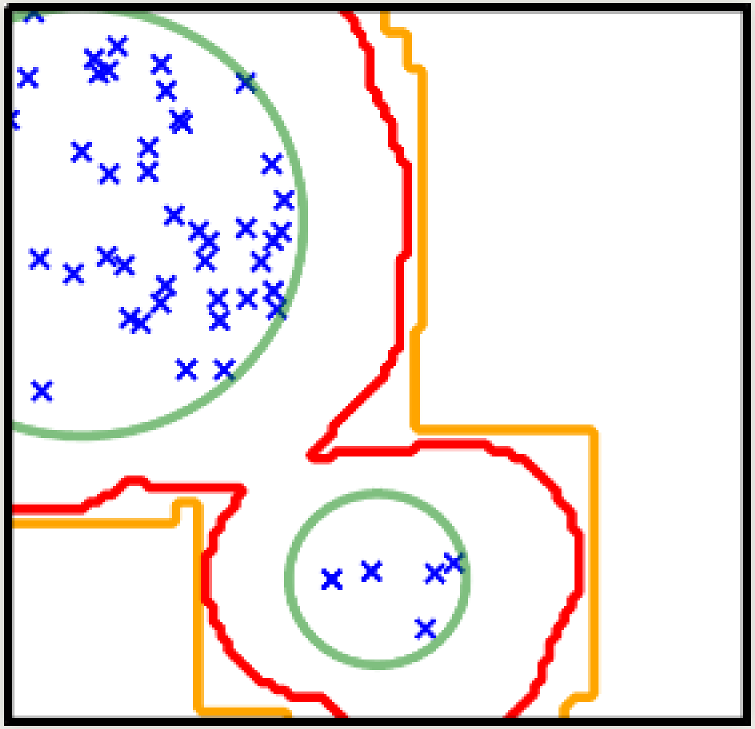
- Scanning up vertically, done at some point
- converges (flat at highest likelihood)

Missing ingredients

- MCMC: Insert tuned transition kernel
- NS: Insert constrained drawing algorithm
  - General solutions: MultiNest, MCMC, HMCMC, Galilean, RadFriends, PolyChord

RadFriends / MultiNest

- **Use existing points** to guess contour
- Expand contour a little bit
- Draw uniformly from contour
- Reject points below likelihood threshold
- RadFriends: Compute distance at which every point has a neighbor. Bootstrap (Leave out) for safety.
- MultiNest clusters and uses ellipses



Animation:

https://johannesbuchner.github.io/mcmc-demo/app.html#RadFriends-NS,stan

What to do with Z

- Z1, Z2

$$\frac{p(M1|D)}{p(M2|D)} = \frac{Z1 \cdot p(M1)}{Z2 \cdot p(M2)}$$

$$\frac{p(M_1|D)}{\sum p(M_i|D)} = \frac{Z_1 \cdot p(M_i)}{\sum_i Z_i \cdot p(M_i)}$$

- model priors: leave to reader or motivated by theory
- Discard highly improbable model or marginalise
- Does $\frac{p(M1|D)}{p(M2|D)} = 3/1$ mean M2 is correct in a quarter of the cases?

What to do with Z

- Z1, Z2

$$\underbrace{\frac{p(M1|D)}{p(M2|D)}}_{\substack{\text{Posterior}\\\text{odds ratio}}} = \frac{Z1 \cdot p(M1)}{\underbrace{Z2}_{\substack{\text{Bayes}\\\text{factor}}} \cdot \underbrace{p(M2)}_{\substack{\text{Prior}\\\text{odds ratio}}}}$$

Buchner+14

What to do with Z

- Z1, Z2

$$\frac{p(M1|D)}{p(M2|D)} = \frac{Z1 \cdot p(M1)}{Z2 \cdot p(M2)}$$

$$\frac{p(M_1|D)}{\sum p(M_i|D)} = \frac{Z_1 \cdot p(M_i)}{\sum_i Z_i \cdot p(M_i)}$$

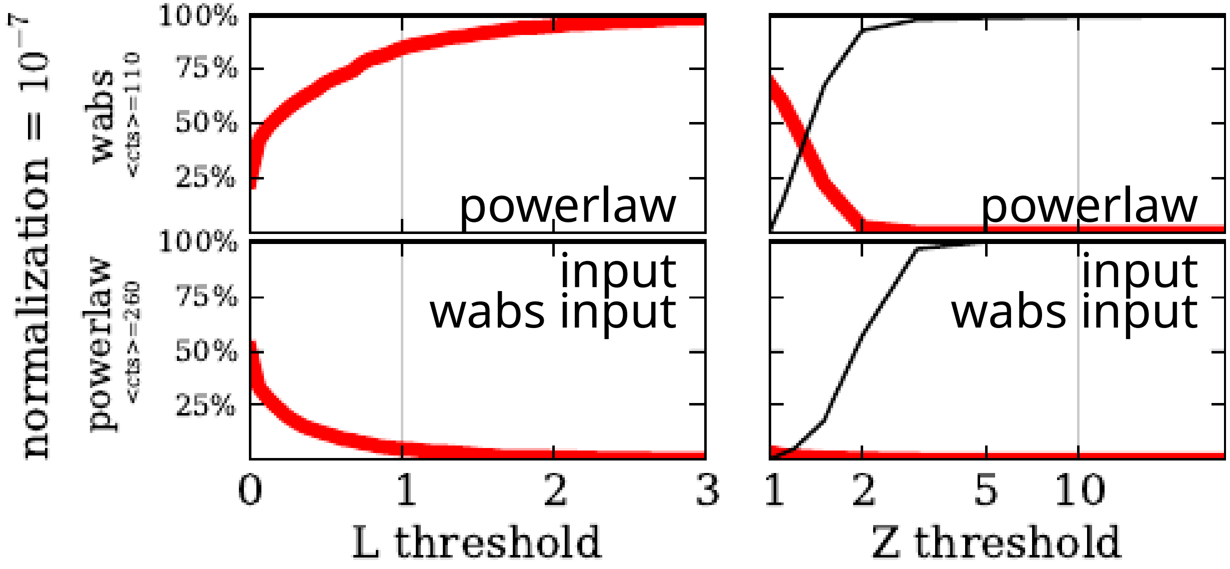
- model priors: leave to reader or motivated by theory
- Discard highly improbable model or marginalise
- Does $\frac{p(M1|D)}{p(M2|D)} = 3/1$ mean M2 is correct in a quarter of the cases?

Calibrating model decisions

- Model probabilities → decisions
- False decision rate (false positives/negatives)
  - Monte Carlo simulations (parametric bootstrap)

Buchner+14

Calibrating model decisions



Buchner+14

False negatives
Non-decisions

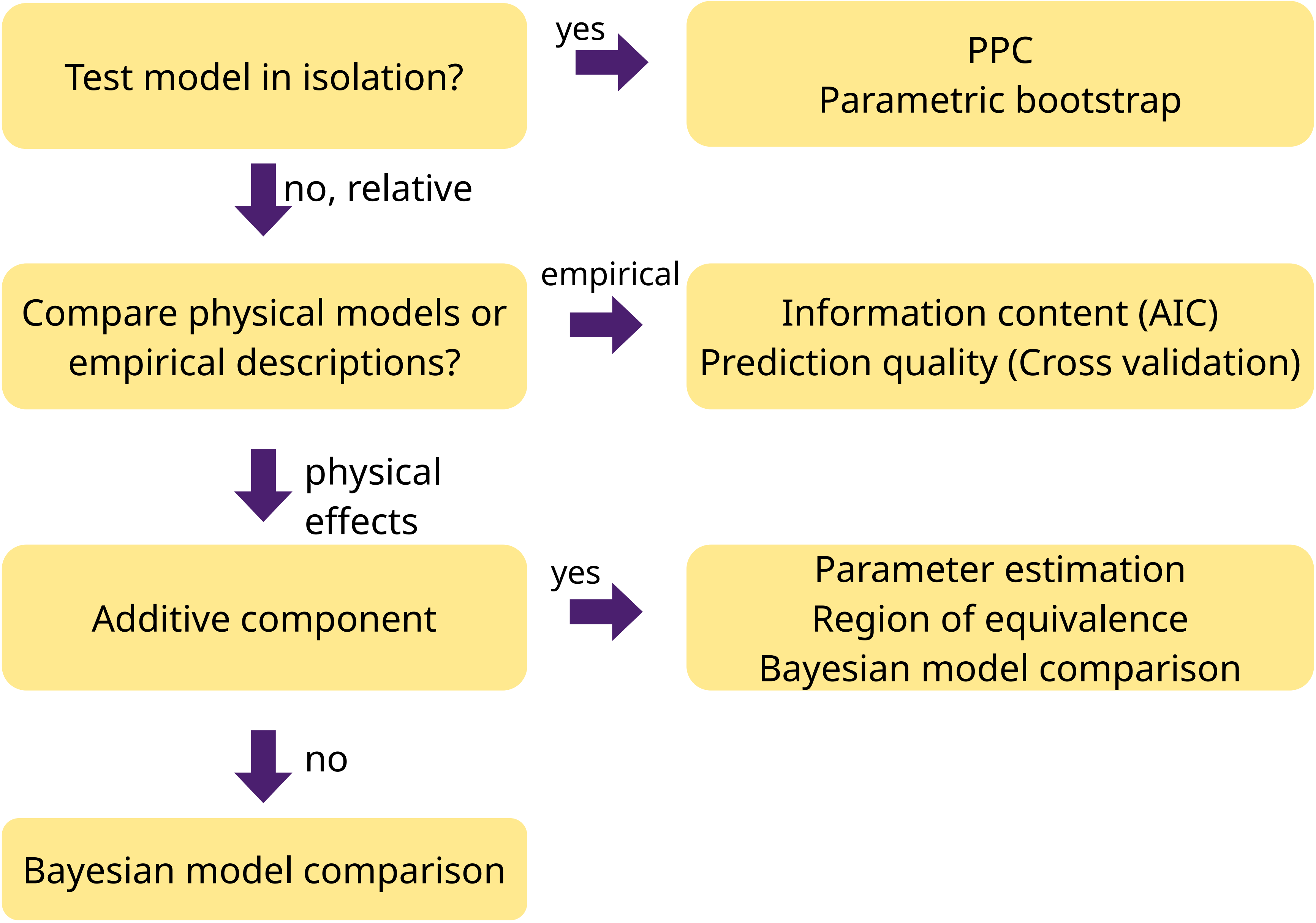
Advantages:
- Get rid of parameter prior dependences
- Have frequentist properties of Bayesian method
- Completely Bayesian treatment + decisions

Disadvantages:
- Can be computationally expensive

Frequentist properties of Bayesian methods

- Make decisions
  - Is parameter greater than C?
  - Is this model “better” than the other?
- Parametric bootstrap
  - Monte Carlo simulation allow arbitrary complexity

Model comparison



Test model in isolation?

yes

PPC
Parametric bootstrap

no, relative

Compare physical models or empirical descriptions?

empirical

Information content (AIC)
Prediction quality (Cross validation)

physical effects

Additive component

yes

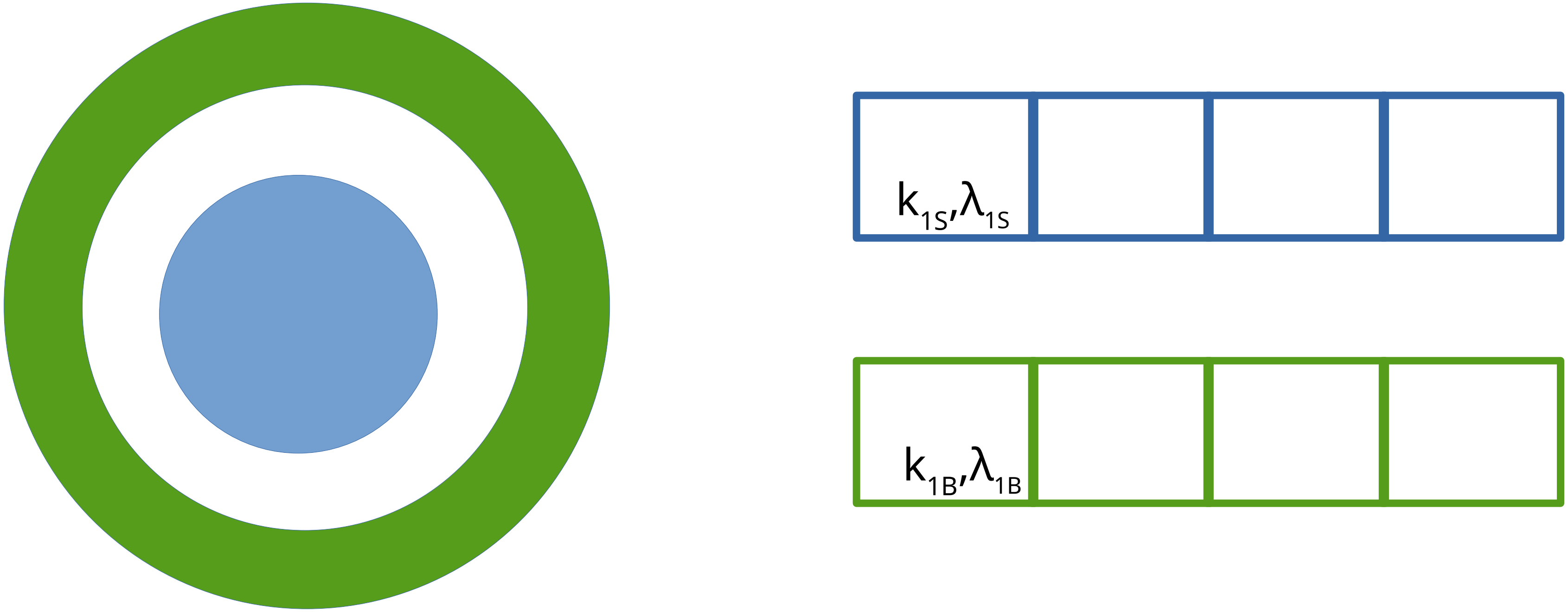
Parameter estimation
Region of equivalence
Bayesian model comparison

no

Bayesian model comparison

Backgrounds

Backgrounds

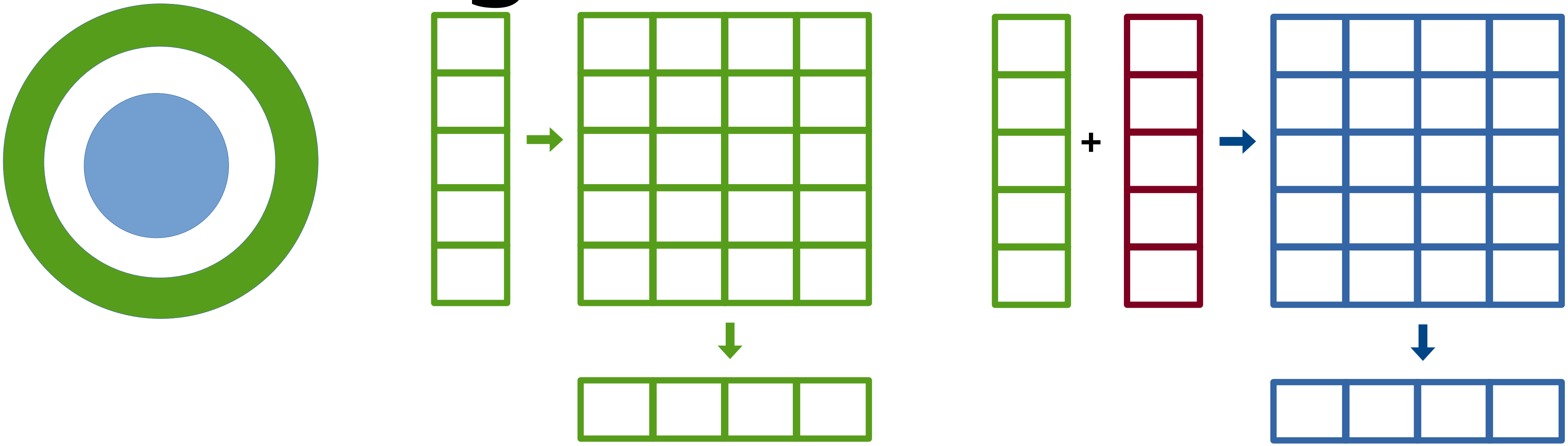


Assume time, location-independence

$k_{src}, \lambda_{src}, t_{src}, A_{src}$

$k_{bkg}, \lambda_{bkg}, t_{bkg}, A_{bkg}$

Background + Source



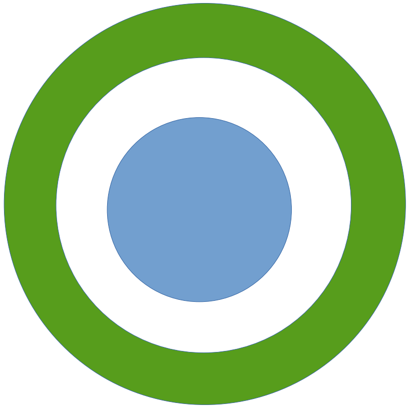
$$\overrightarrow{\lambda}_{\text{src}} = \overrightarrow{F}_{\text{src}} \cdot \underline{R}_{\text{src}} + \overrightarrow{F}_{\text{bkg}} \cdot \underline{R}_{\text{src}}$$
$$\overrightarrow{\lambda}_{\text{bkg}} = \overrightarrow{F}_{\text{bkg}} \cdot \underline{R}_{\text{bkg}}$$

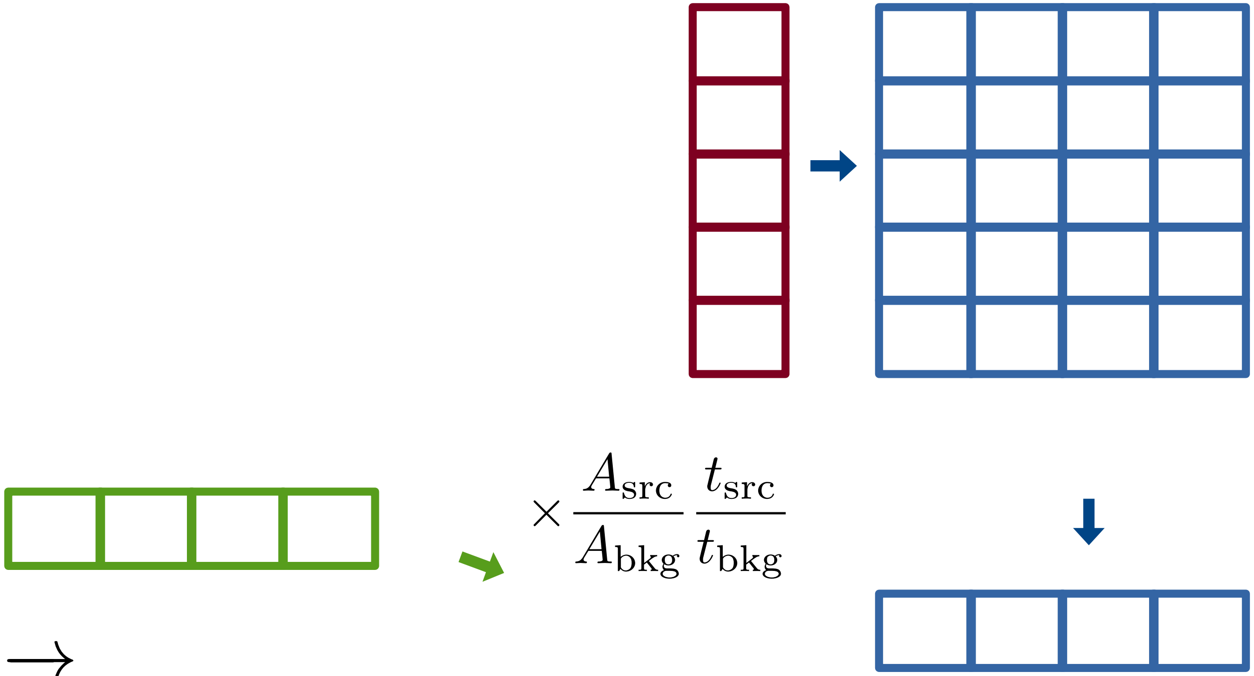
$$\begin{aligned} C = \quad & 2\,\overrightarrow{\lambda}_{\text{src}} \cdot \overrightarrow{\lambda}_{\text{src}} - 2\,\overrightarrow{k}_{\text{src}} \cdot \log \overrightarrow{\lambda}_{\text{src}} + \\ & 2\,\overrightarrow{\lambda}_{\text{bkg}} \cdot \overrightarrow{\lambda}_{\text{bkg}} - 2\,\overrightarrow{k} \cdot \log \overrightarrow{\lambda}_{\text{bkg}} \end{aligned}$$

Assumptions:
- area energy-independent
- rate constant with area, time, location

Remember:
λ=number / cm^2 / s / keV * dE * dt * dA

Background + Source





$$\times \frac{A_{\rm src}}{A_{\rm bkg}} \frac{t_{\rm src}}{t_{\rm bkg}}$$

$$\vec{\lambda}_{\rm src} = \vec{F}_{\rm src} \cdot \underline{R}_{\rm src} + \vec{F}_{\rm bkg} \cdot \underline{R}_{\rm src}$$

$$\vec{\lambda}_{\rm bkg} = \vec{F}_{\rm bkg} \cdot \underline{R}_{\rm bkg}$$

$$\begin{aligned} C = \quad & 2\,\vec{\lambda}_{\rm src} \cdot \vec{\lambda}_{\rm src} - 2\,\vec{k}_{\rm src} \cdot \log \vec{\lambda}_{\rm src} + \\ & 2\,\vec{\lambda}_{\rm bkg} \cdot \vec{\lambda}_{\rm bkg} - 2\,\vec{k} \cdot \log \vec{\lambda}_{\rm bkg} \end{aligned}$$

Assumptions:
- area energy-independent
- rate constant with area, time, location

Remember:
λ=number / cm^2 / s / keV * dE * dt * dA

Background + Source

- src+bkg Gauss → Gauss (subtractable, flats/darks)
- src+bkg Poisson → Poisson
    - High counts (>100) in every single src and bkg bin → Gauss + Subtract with bkg variance propagation
    - Subtract & model with Skellam distribution
    - Do the right thing and model both as Poisson

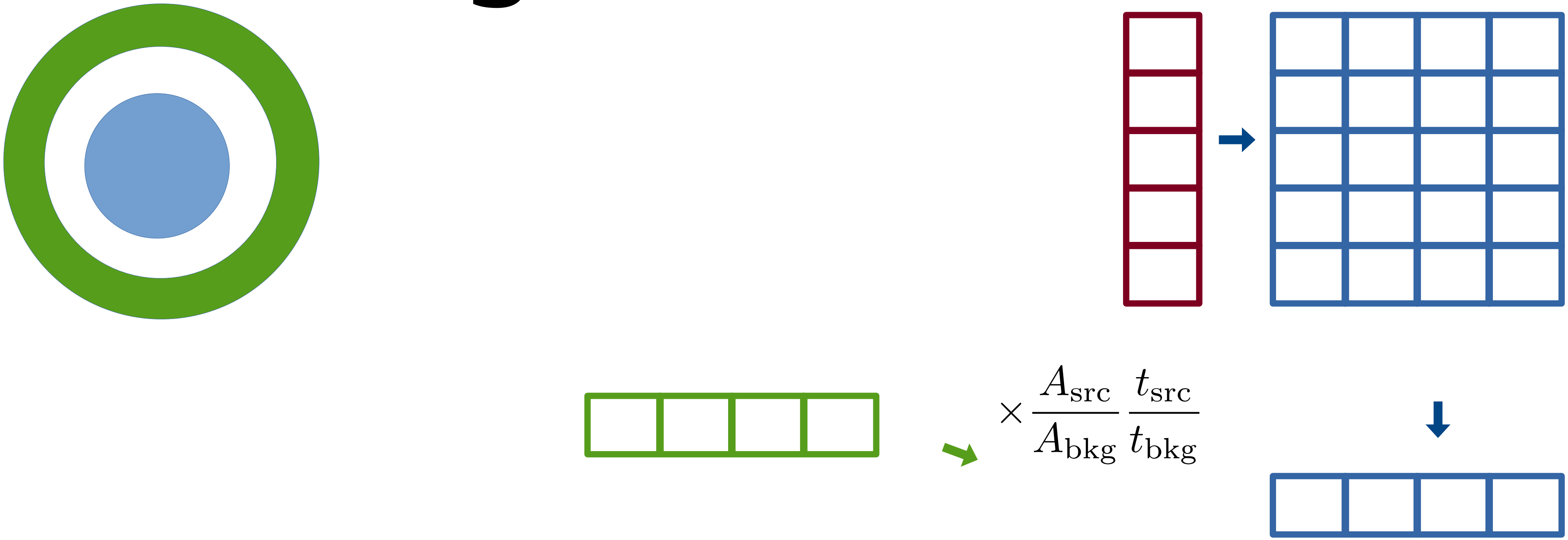
Background + Source

- src+bkg Gauss → Gauss (subtractable, flats/darks)
- src+bkg Poisson → Poisson
  - High counts (>100) in every single src and bkg bin → Gauss + Subtract with bkg variance propagation
  - Subtract & model with Skellam distribution
  - Do the right thing and model <u>both</u> as Poisson
    - Poisson estimate of rate in each bin, independently
    - Function approximation of background
      - In counts (empirical model)
      - Physical background flux model

      - Fit simultaneously with source
      - Fit background model first, use best-fit background shape for source fit

(“WStat”, default in xspec
if you set statistic “cstat”)

Background + Source



$$\times \frac{A_{\mathrm{src}}}{A_{\mathrm{bkg}}} \frac{t_{\mathrm{src}}}{t_{\mathrm{bkg}}}$$

$$\overrightarrow{\lambda}_{\mathrm{src}} = \overrightarrow{F}_{\mathrm{src}} \cdot \underline{R}_{\mathrm{src}} + \overrightarrow{F}_{\mathrm{bkg}} \cdot \underline{R}_{\mathrm{src}}$$

$$\overrightarrow{\lambda}_{\mathrm{bkg}} = \overrightarrow{F}_{\mathrm{bkg}} \cdot \underline{R}_{\mathrm{bkg}}$$

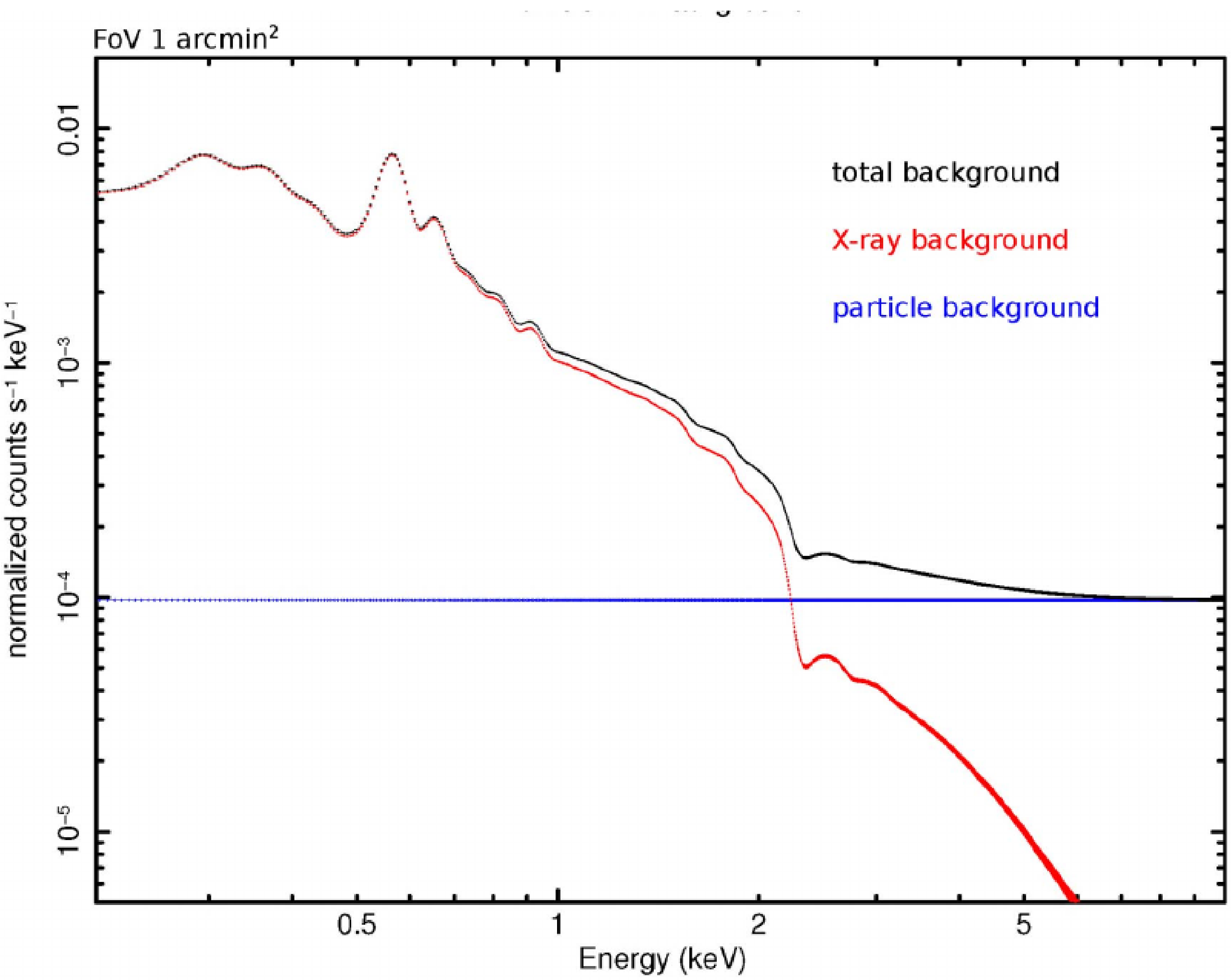
$$\begin{aligned} C = \quad & 2\,\overrightarrow{\lambda}_{\mathrm{src}} \cdot \overrightarrow{\lambda}_{\mathrm{src}} - 2\,\overrightarrow{k}_{\mathrm{src}} \cdot \log \overrightarrow{\lambda}_{\mathrm{src}} + \\ & 2\,\overrightarrow{\lambda}_{\mathrm{bkg}} \cdot \overrightarrow{\lambda}_{\mathrm{bkg}} - 2\,\overrightarrow{k} \cdot \log \overrightarrow{\lambda}_{\mathrm{bkg}} \end{aligned}$$

Assumptions:
- area energy-independent
- rate constant with area, time, location

Remember:
λ=number / cm^2 / s / keV * dE * dt * dA

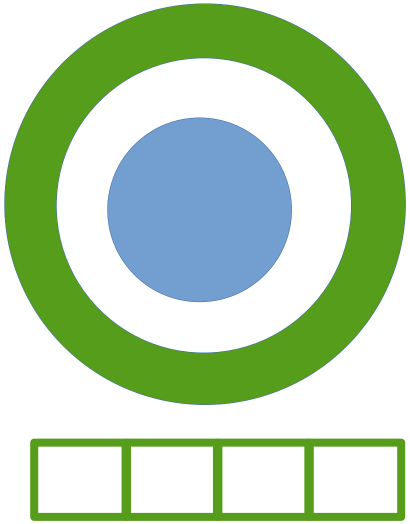
eROSITA background

- Diffuse emission
  - Local hot bubble
  - Galactic disk
  - Galactic halo
- Cosmic background
  - Unresolved AGN
- High-energy particle background



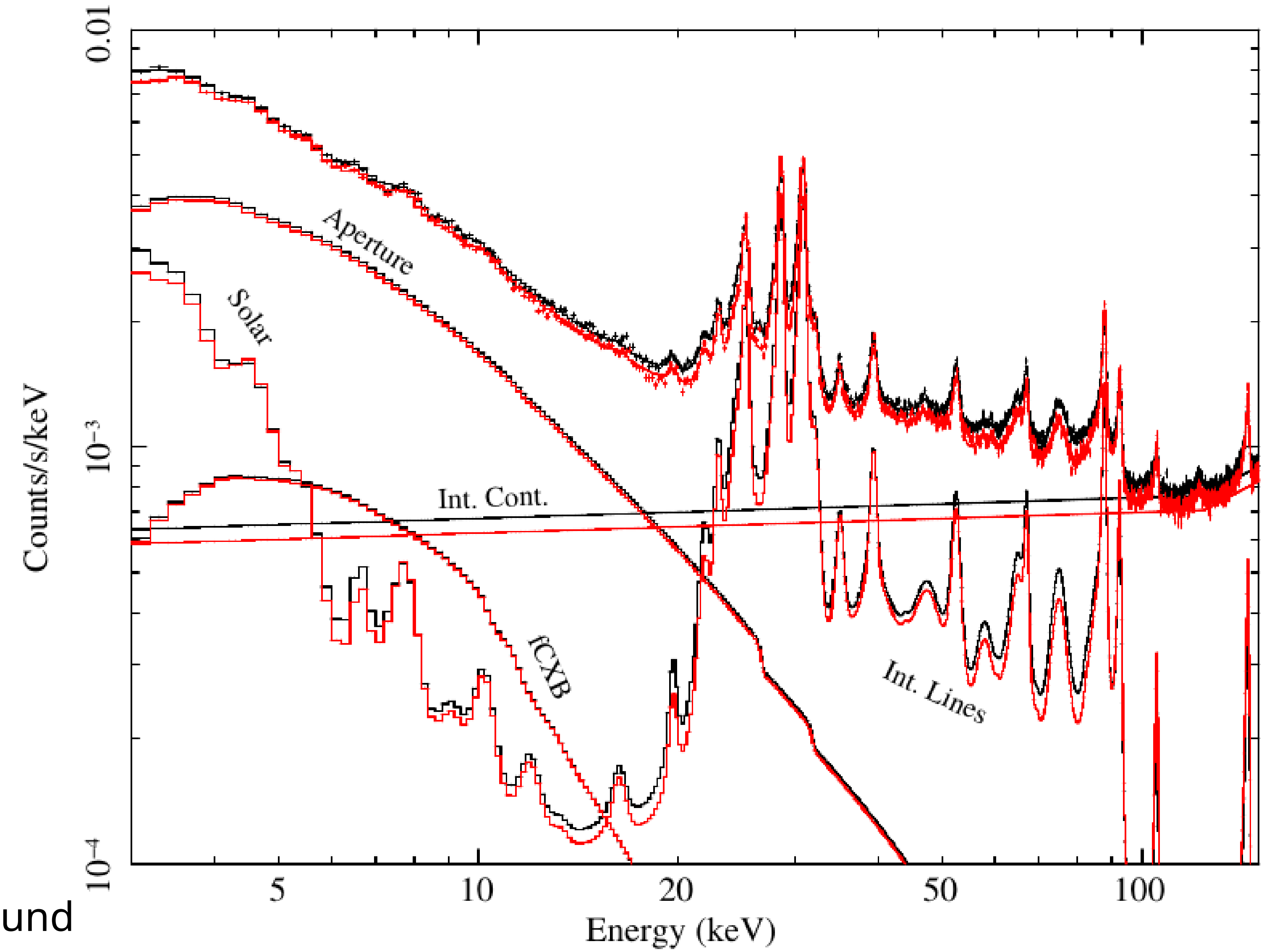
https://wiki.mpe.mpg.de/eRosita/ScienceRelatedStuff/Background

Semi-physical background models



Maximize poisson likelihood at all bins →shape

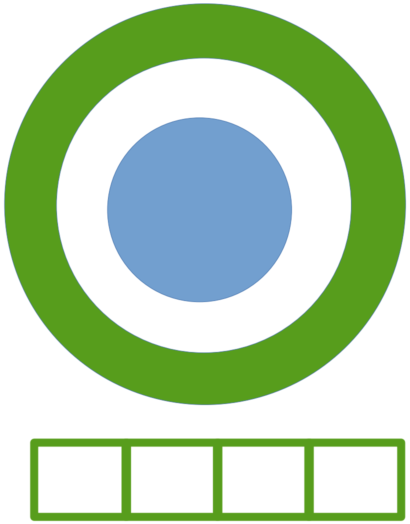
Particle background
Cosmic background
Instrumental background
…
Location & time-dependent



NuSTAR (Wik+14)

→ especially important for extended source

Empirical background



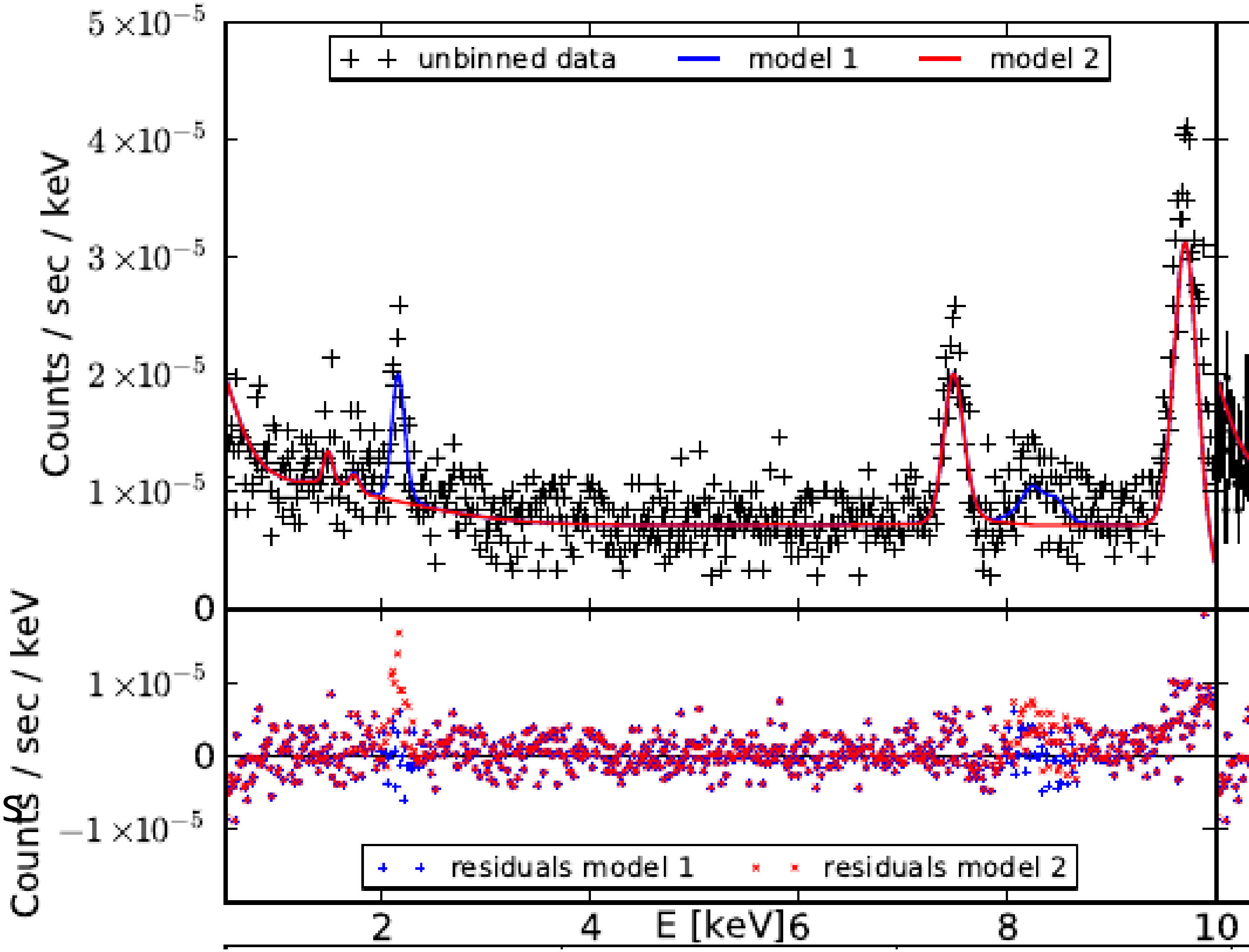
Maximize poisson likelihood in each bin →shape

Pros:

- Can contain physical knowledge & smoothness
- Small uncertainties
- 0 bin counts ok

Cons:

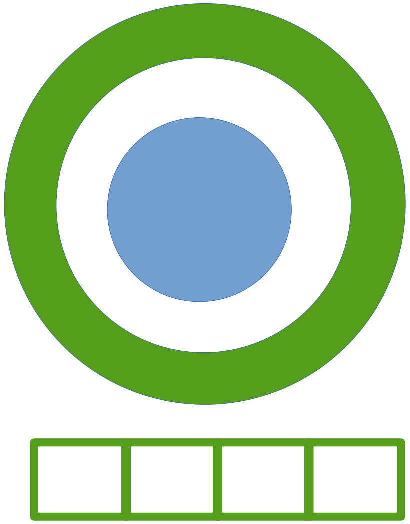
- Need to specify model
- Fit can be poor



Chandra

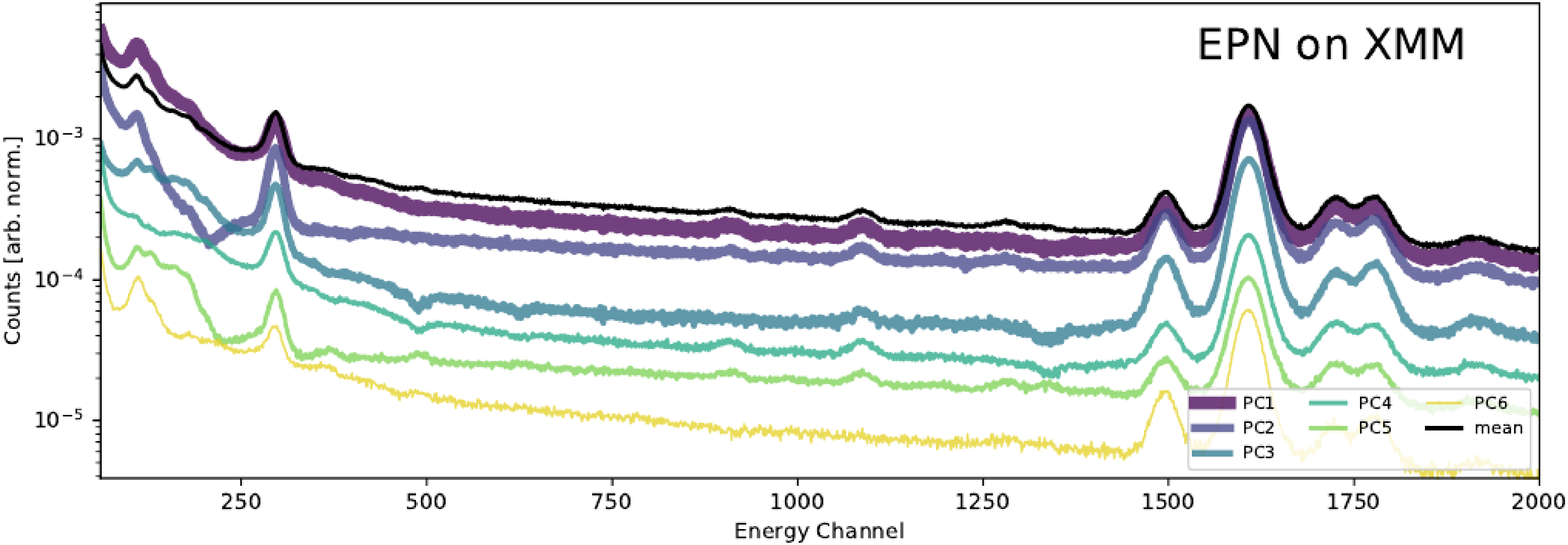
(XMM, Chandra, Swift/XRT models in BXA)

Empirical background models

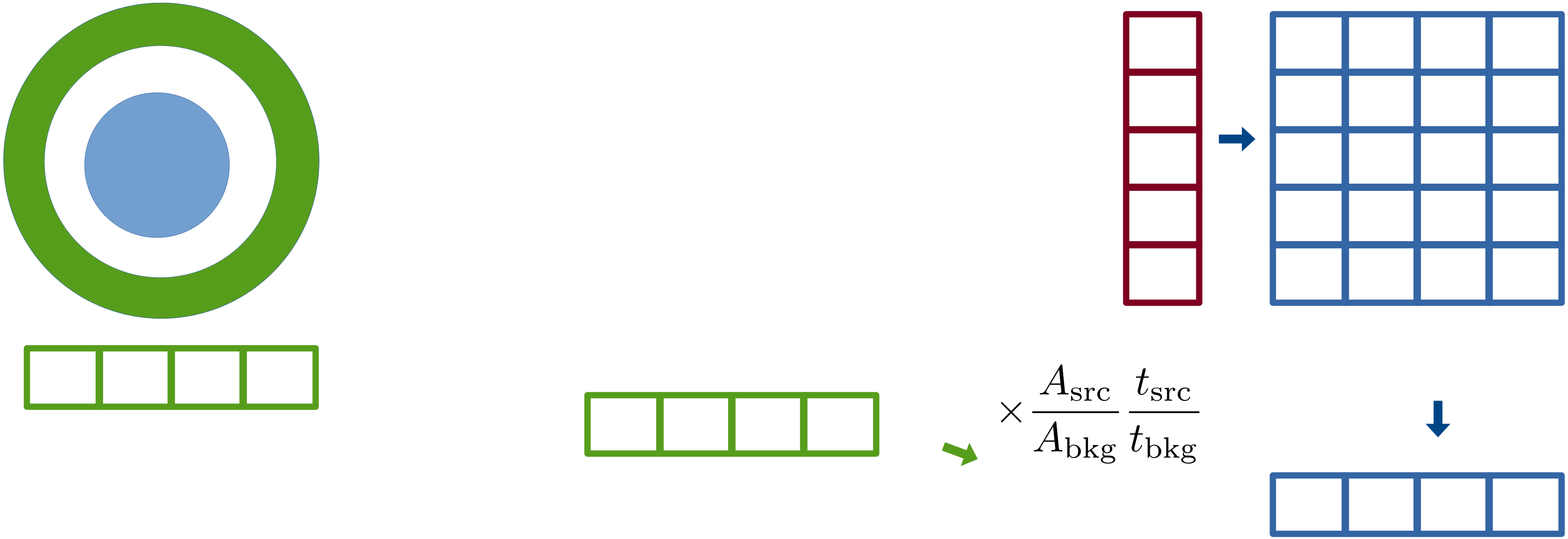


Automated shape finding
Simmonds, Buchner et al. (2017)

XMM/PN,MOS, Chandra/ACIS, NuSTAR,
Suzaku, RXTE, Swift/XRT



Background: Individual bins



$$\times \frac{A_{\mathrm{src}}}{A_{\mathrm{bkg}}} \frac{t_{\mathrm{src}}}{t_{\mathrm{bkg}}}$$

Estimate most likely background rate in each bin

Add scaled to source region counts

(wstat, Xspec default if set to cstat with no background model)
pgstat

Pros:

- no model specification needed

Cons:

- no continuity
- unnecessarily large uncertainties
- need >0 counts per bin
- Need >3 counts per bin, otherwise biased!
- https://giacomov.github.io/Bias-in-profile-poisson-likelihood/

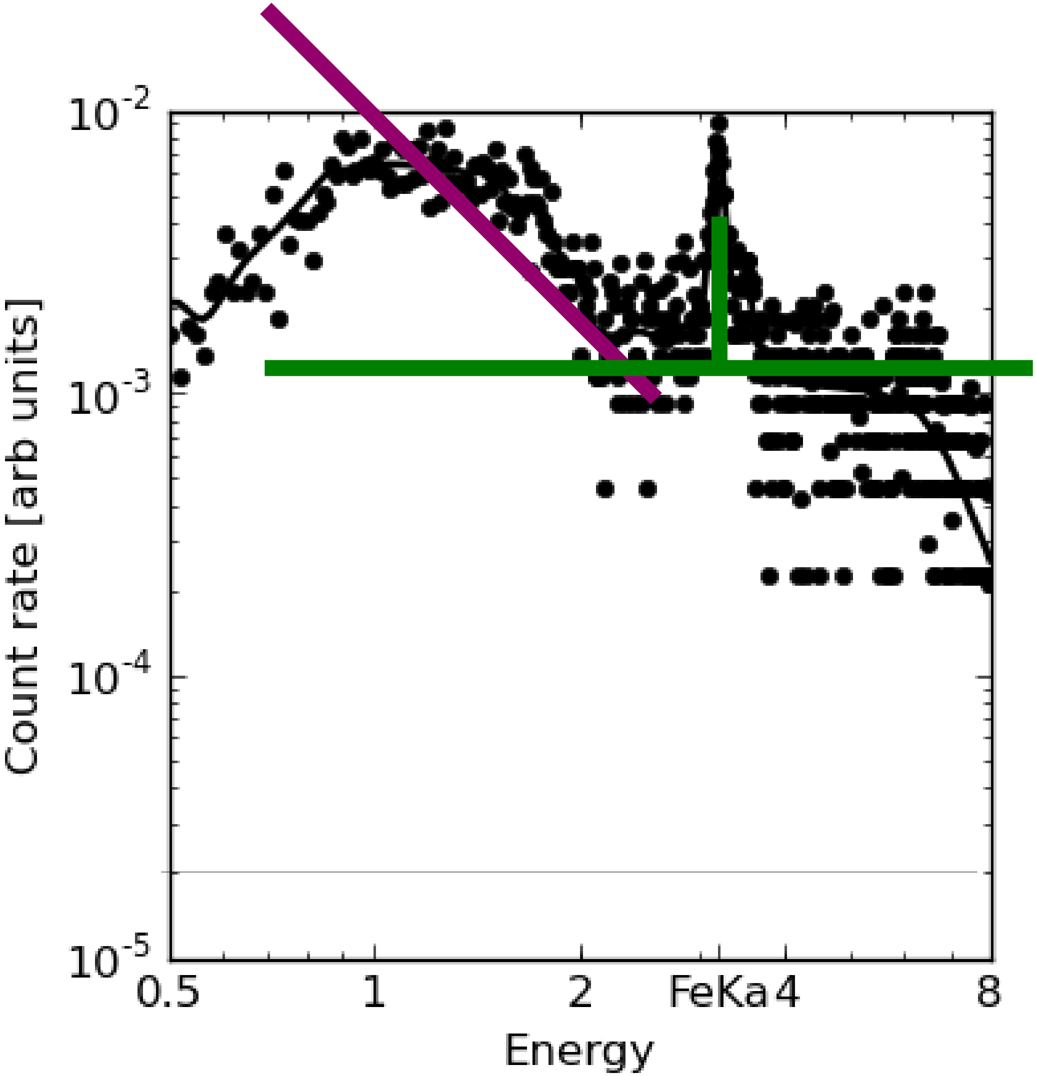
Spectra with few counts

- Are nothing special
- Poisson likelihood + good background handling

- 0 counts
- Think in terms of allowed regions

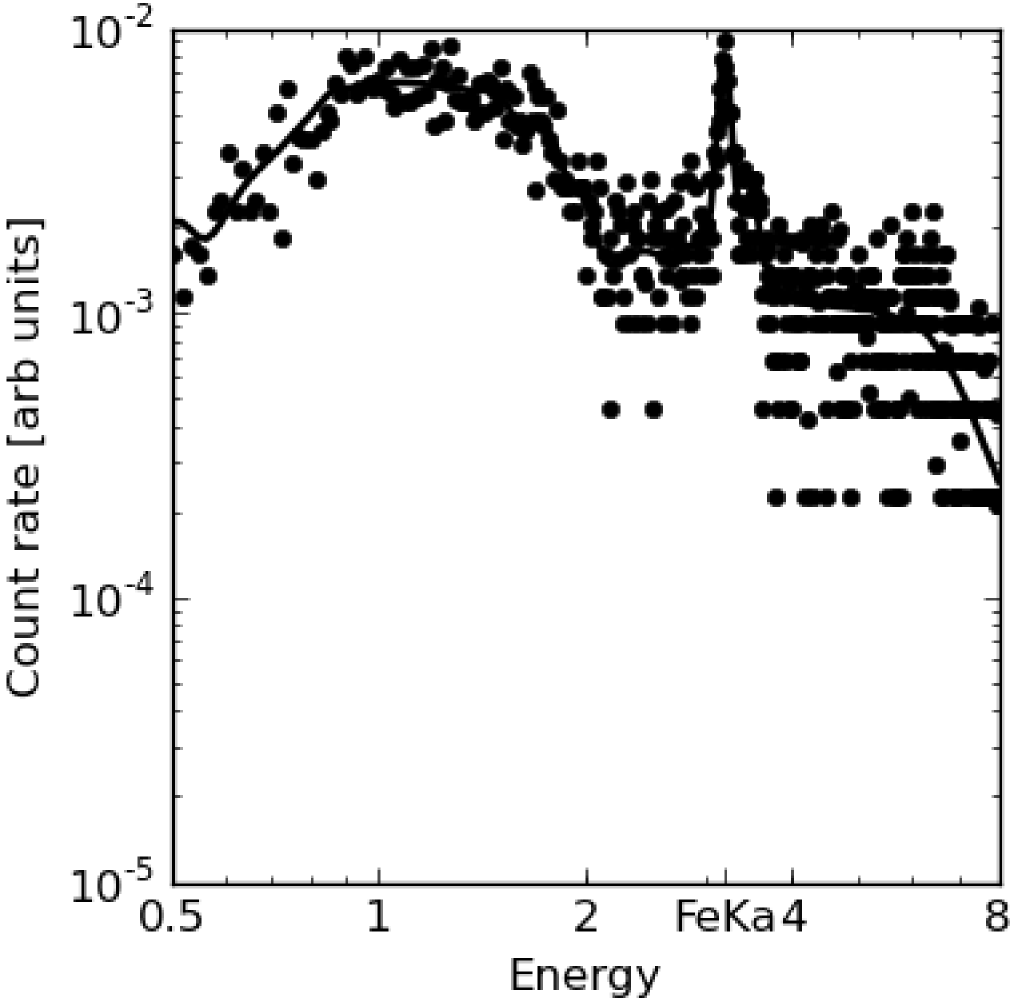
L, N_H from X-ray spectrum

Scattered Powerlaw component

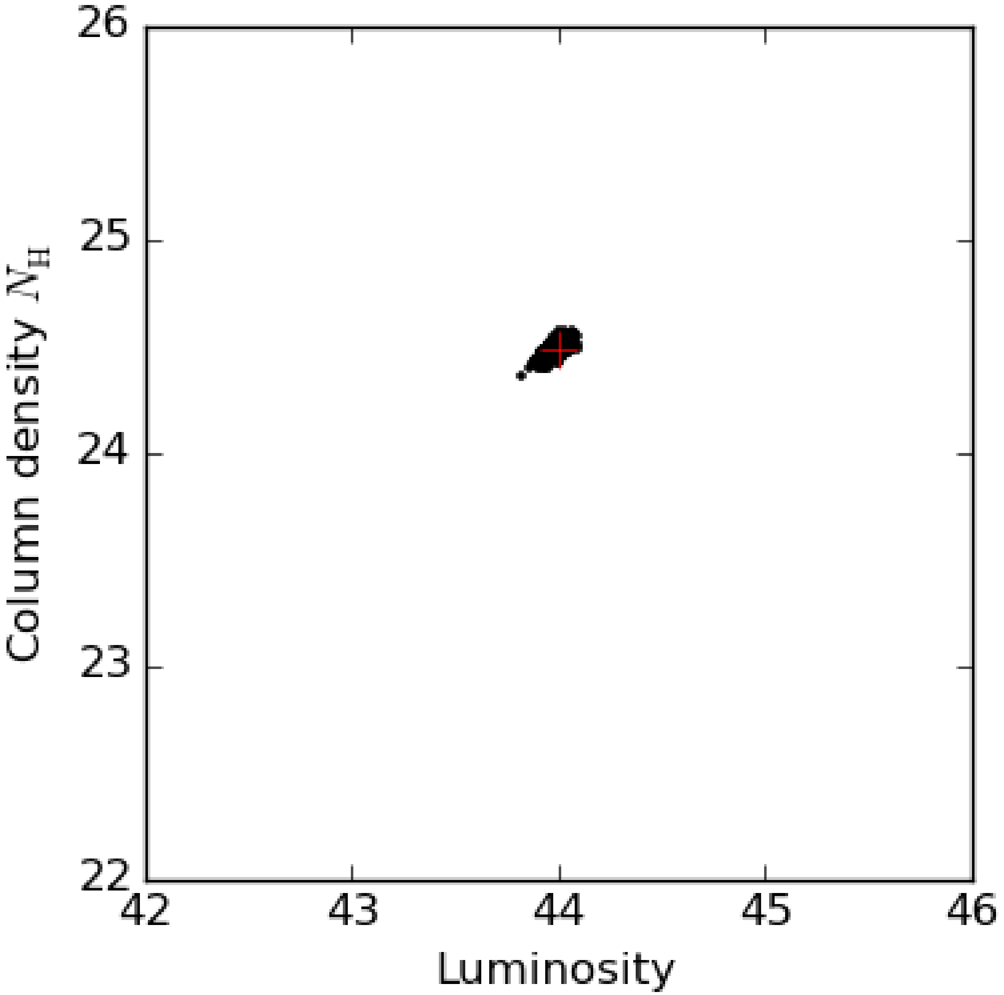


CTK flat spectrum +
FeK line

L, N_H from X-ray spectrum



Probability cloud



Practical advice

- You can do this in any package!
- State what you are doing
- CStat (Poisson)
- Background with functions (check fit)
- Visualise, visualise, visualise
- Show posterior distributions & fits in data space
- Vary priors & assumptions
- Use nested sampling, MCMC with care
- Make simulations
- Ask for help

Packages

- XSpec (NASA)
- Sherpa (SAO, Chandra)
- ISIS (MIT)
- Spex (SRON, high-res)
- and others ...

can do good and bad analyses with any of them

→ understand what you are doing & assuming

Resources

- Longer workshops, videos & slides

  https://johannesbuchner.github.io/BXA/tutorials.html

- Statistical Aspects of X-ray Spectral Analysis

  https://arxiv.org/abs/2309.05705

- X-ray primer

  https://cxc.cfa.harvard.edu/cdo/xray_primer.pdf

Contact points for questions

- Ask a colleague
- Astrostatistics Facebook group
- XSPEC Facebook group
- Today’s tutors
- Email me