Detection

William Sealey Gosset
1876 - 1937

Best known for his “Student’s” t-test, devised for handling small samples for quality control in brewing.

“To many in the statistical world "Student" was regarded as a statistical advisor to Guinness's brewery, to others he appeared to be a brewer devoting his spare time to statistics. ... though there is some truth in both these ideas they miss the central point, which was the intimate connection between his statistical research and the practical problems on which he was engaged. ...“

Any recollection of what I was saying last lecture?

- We looked again at brute-force MC integration.

- We discussed **importance sampling** as a more efficient way for **difficult integrations**.

- This led in to the efficient importance-sampling technique known as **Markov Chain Monte Carlo.** We walked through the algorithm: it is simple, **but there are buts:**

1) The **proposal, which moves us forward in the chain,** requires thought:
   - **'broad' proposals** lead to **low efficiency**, but quick **burn-in** and less correlation
   - **'narrow' proposals** lead to long burn-in and **serious coherence problems**

2) We need to generate several chains for **intercomparison of variance**.

- We illustrated all this with a simple example.

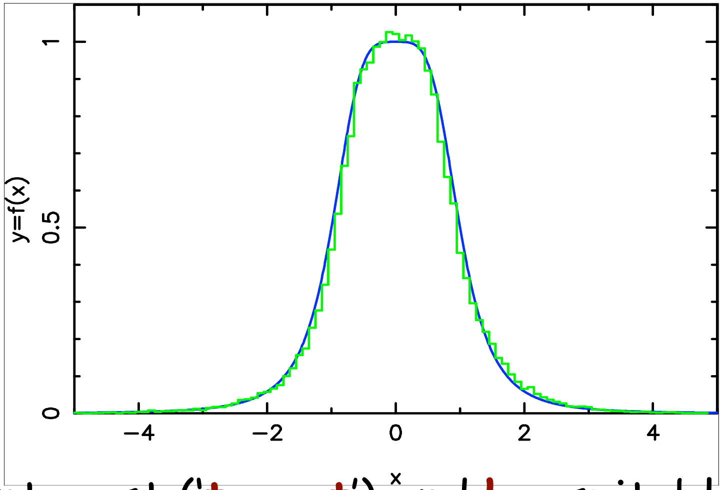
- We considered the multivariate case and introduced the **Gibbs sampler**.

- With this we went back to our **Bayesian integrals**, and considered an example with both MCMC and thermodynamic integration to integrate the **non-normalized multivariate function f/C**.

But before leaving MCMC (temporarily) ....

Remember our first example

$f(x) = 1/(1+x^4)$



And remember the MCMC algorithm:

If **f** is the un-normalized distribution of interest ('**target**') and **h** a suitable transition probability distribution (the '**proposal**') then

1. **Draw a random number X_i from h**

2. **Draw a random number U_i, uniformly distributed 0.0 to 1.0**

3. **Compute α, the minimum of 1.0 and $f(X_i)/f(X_{i-1})$**

4. **if $U_i < \alpha$ then accept X_i**

5. **Otherwise set $X_i = X_{i-1}$**

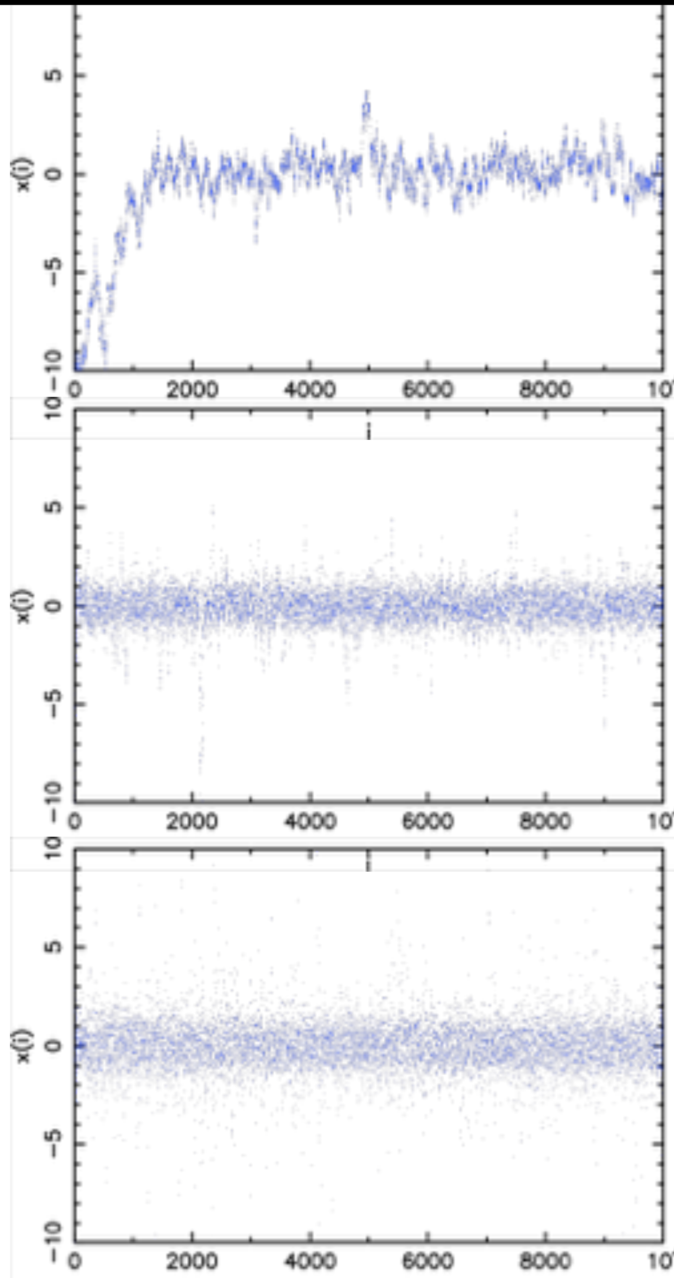
This may look simple to program – but there is a trap

Before leaving MCMC (temporarily) continued ....

func is $1/(1+x^4)$

proposal is random pick from a Gaussian

```
idum=-1
sigma=10.0
j=1
x(1)=-10.0
fn(1)=func(x(1))
ifail=0
do i=2,100000
xx=proposal(x(j),sigma,idum)
u=ran1(idum)
f(j)=func(xx)
r=f(j)/f(j-1)
alpha=amin1(1.0,r)
if(u.lt.alpha)then
j=j+1
x(j)=xx
else
ifail=ifail+1
endif
enddo
```



sigma = 0.2

burn-in ~2000

7% rejected

coherence

sigma = 1.0

burn-in ~ 0?

34% rejected

coherence

sigma = 10.0

burn-in = 0?

89% rejected

no coherence?

Detection - what do we mean?

- **preliminary** to much else that happens in astronomy, whether it means locating a spectral line, a faint star or a gamma-ray burst.

- we take it here as **locating + confident measurement** of some sort of **feature** in a fixed region of an image or spectrum.

- elusive objects or **features at the limit of detectability** tend to become the focus of interest in any branch of astronomy.

Detection - what do we mean? continued…

- **non-detections are important** because they define how representative any catalogue of objects may be. This set of non-detections can represent **vital information in** deducing the properties of a population of objects.

- if something is **never detected**, it's a datum, and can be exploited statistically; **every** observation potentially contains information, constrains emission energy.

- it is possible to deduce **distributions of parameters**, e.g. luminosity function, from the detections/catalogue approach, or directly from **fairly raw sky data**.

- we consider both – but **detections first**.

Detection - a Model Fitting Process 1

When we say "We've got a detection" we generally mean **"We have found what we were looking for"** – this is OK at 'reasonable' signal-to-noise.

E.g. comparison of model **point-spread functions** with the data - but in the case of **extended objects**? – **wider range of models** needed.

A clear statistical model is required. The noise level (residuals from the model) may follow **Poisson ($\sqrt{N}$) statistics → Gaussian for N > 10**.

The statistics depend on more than the physical and instrumental model. **How were the data selected**?

E.g. picking out the **brightest** spot means a **special** set of data. The **peak pixel** will follow the distribution appropriate to the maximum value of a set of, say, Gaussian variables. **Adjacent pixels** will follow a less well-defined distribution.

Detection - a Model Fitting Process 2

Simulation – “Model sources" are strewn in the real map, and the reduction software (detection-finding algorithm) is given the job of telling us **what fraction is detected**.

These essential large-scale techniques are very necessary for handling the detail of how the observation was made, **because we may be hazy on this!**

Example: radio astronomy synthesis images: noise level at any point depends on:

- gains of all antennas,
- noise of each receiver,
- sidelobes from whatever sources happen to be in the field of view,
- map size, tapering parameters, ionosphere, cloud…..

These input data are either known incompletely or not known at all!

Detection - the simpler classical approach I

What do we really want from the survey?

- Are we more concerned with **detecting as much as possible (completeness)** ?
- Are we more worried about **false detections (reliability)** ?
- **What are we going to do with the detections ? E.g. :**
  (1) will we publish a complete set of posterior probabilities of observed parameters everywhere? (unlikely)
  (2) or just the covariance matrix, as an approximation?
  (3) or marginalized signal-to-noise, integrating away nuisance parameters?

From the classical point of view, if we are trying to measure a parameter α, the **likelihood** sums up what we have achieved:

$$\mathcal{L} = \mathrm{prob}(\mathrm{data} \mid \alpha).$$

Detection - the simpler classical approach 2

Suppose that ***α*** is a ***flux density*** and we wish to set a **flux limit for a survey**. We record catalogue detections only when our data exceed this limit s_{lim}. Then two properties of the survey useful to know are ***reliability*** and ***completeness***:

1. The false-alarm rate is the chance that pure noise will produce data above the flux limit:

$$\mathcal{F}(\mathrm{data}, s_{\mathrm{lim}}) = \mathrm{prob}(\mathrm{data} > s_{\mathrm{lim}} \mid \alpha = 0).$$

   The *reliability* is $1 - \mathcal{F}$, *i.e.* $\mathcal{F} = 5/100$ gives 95 per cent reliability. The may sound good, but note that it is the infamous 2σ result.

2. The *completeness* is the chance that a measurement of a real source will be above the flux limit:

$$\mathcal{C}(\mathrm{data}, s_{\mathrm{lim}}, S) = \mathrm{prob}(\mathrm{data} > s_{\mathrm{lim}} \mid \alpha = S).$$

Classical Detection - Example

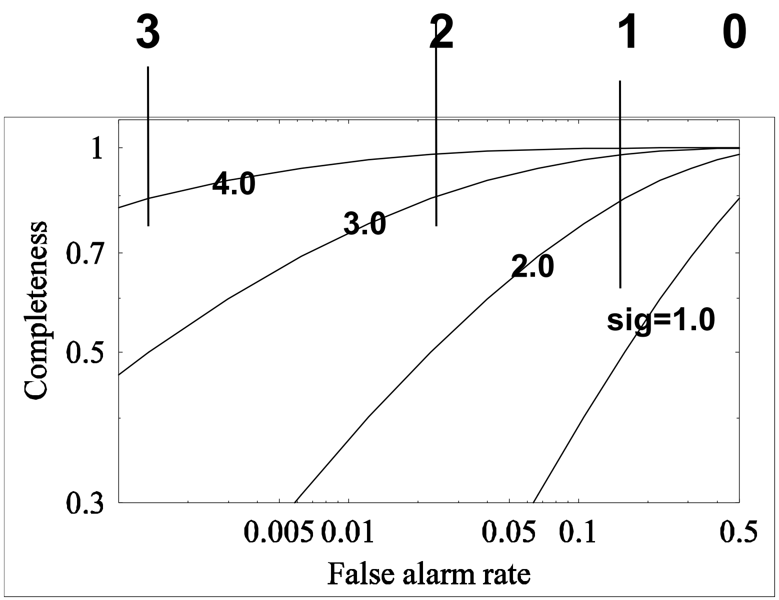
Suppose our measurement is of ***s*** and the noise on the measurement is Gaussian, of unit std dev. The source has a “true" flux density s_0, measured in units of the std dev.

We then have respectively for the **pd of the data given the source**, and for the **pd of the data when there is no source:**

$$\mathrm{prob}(s \mid s_0) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(s-s_0)^2}{2}\right]$$

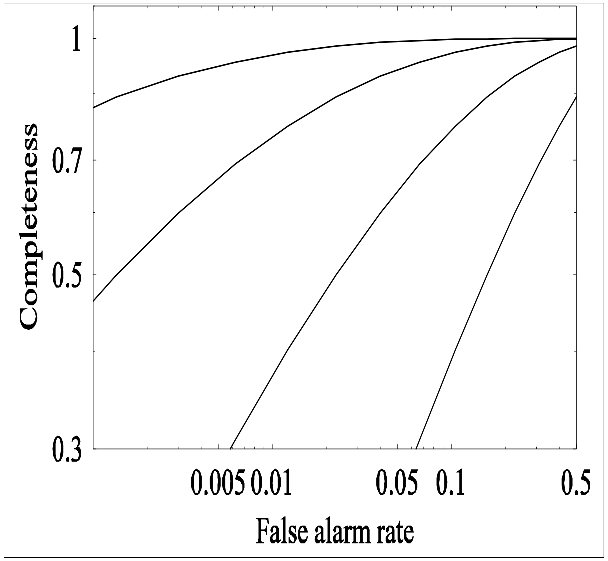
$$\mathrm{prob}(s \mid s_0 = 0) = \frac{1}{\sqrt{2\pi}} \exp\frac{-s^2}{2}$$

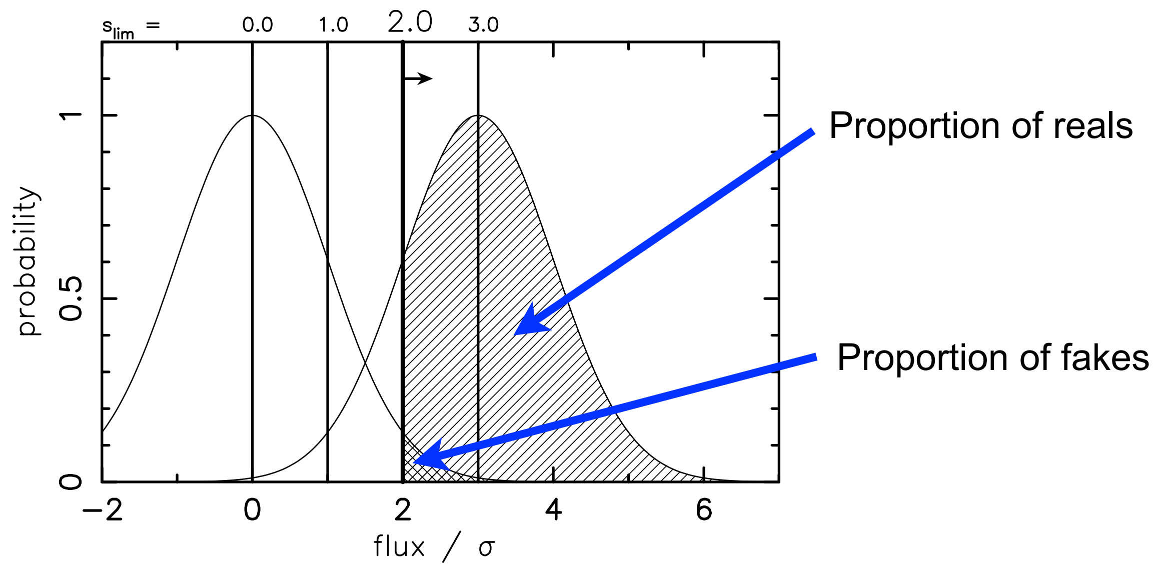
Integrating these functions from ***0*** to ***s*** makes it easy to plot up the **completeness against the false-alarm rate**, taking the flux limit as a parameter.



For source flux densities in units of σ_{noise} ranging from 1 unit (lower right) to 4 units (upper left). Flux limits are indicated by dots, from 0 on right to 3 on left. A 4σ source and a 2σ flux limit give a false-alarm rate of 2% and a completeness of 99%.

Classical Detection - Example continued





So how does this work? How were these curves computed? Let us say that the noise is 1σ, and that our survey detection limit is set at 2σ. (This is extremely low as we shall see.) **Focus on the line for sources of 3σ in size**.

(a) How complete is our survey for sources that are truly 3σ high? **About 80%.**

(b) How reliable are sources detected in this way? Or conversely, what is the false alarm rate? **About 7%.**

High completeness goes hand in hand with a high false alarm rate. However there are **satisfactory combinations** for flux limits and source intensities of a few standard deviations. Problem - outliers!

Bayesian Detection - Example 1

As intro to Bayes and detection, consider first an example of the simple case in which we ignore the possibility that no source might be present in our field.

1. Radio telescope, randomly-selected position in the sky.

2. The data are ***D***, namely the single measured flux densities ***f*** : Gaussian distribution about true flux density ***S*** with a variance ***σ²***.

3. The extensive body of radio source counts also tells us the **prior for *S***; approximate this by the power law ***prob(S) = K $S^{-5/2}$***. (***K*** normalizes the counts to unity; there is presumed to be one source in the beam at some level.)

4. Probability of observing ***f*** when the true value is ***S*** : $\exp[-\frac{1}{2\sigma^2}(f-S)^2]$.

5. From Bayes : with ***n*** independent flux measurements ***f_i*** then

$$\mathrm{prob}(S \mid D) = K'' \exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(f_i - S)^2] S^{-5/2}.$$

Bayesian Detection - Example 1 concluded

Suppose that

- the source counts were known to extend from ***1*** to ***100*** units,

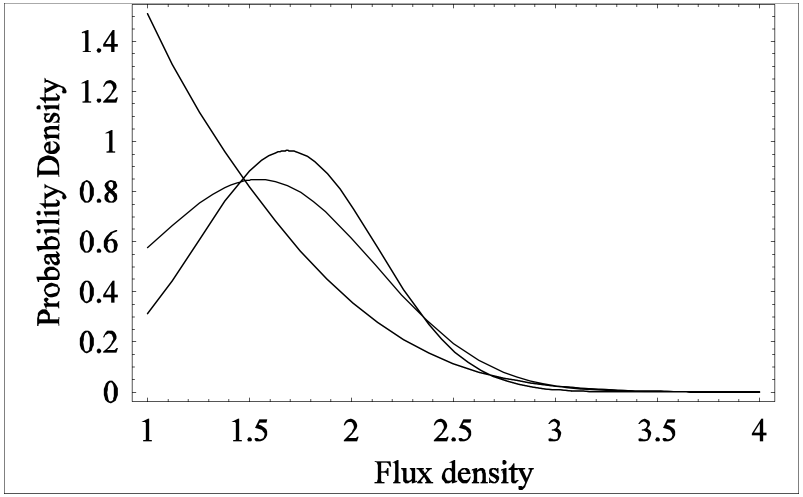
- the noise level was $\boldsymbol{\sigma = 1}$, and

- the data were ***2.0, 1.3, 3.0, 1.5, 2.0*** and ***1.8***.

The Figure shows posterior probabilities for the first 2, then 4, then 6 measurements.

The increase in data gradually overwhelms the prior,

But the prior affects conclusions markedly (as it should) when there are few measurements.



Power-law prior, Gaussian error dist. Posterior probability distribution for 2, 4, 6 measurements shows that the form approaches Gaussian as number of data increase.

True value of mean = 1.93.

Bayesian Detection - the fuller picture

Now consider a more realistic description of detection.

We have

prob(data | a source is present, of flux density *s*)

and **prob(data | no source is present).**

(1) Take the prior probability that a source, intensity ***s***, is present in the measured area to be ***εN(s)***, where ***N(s)*** is a normalized distribution, the probability that a single source will have a flux density ***s***.

(2) The prior probability of no source is ***(1- ε) δ(s)***; ***δ*** is a Dirac delta function.

Then the posterior probability

prob(a source is present, brightness s | data)

is given by

$$\frac{\epsilon\,\mathrm{prob}(\mathrm{data} \mid s)N(s)}{\epsilon \int \mathrm{prob}(\mathrm{data} \mid s)N(s)\,ds + (1-\epsilon)\int \mathrm{prob}(\mathrm{data} \mid s=0)}.$$

Integrating this expression over ***s*** gives the probability that a source is present, for given data.

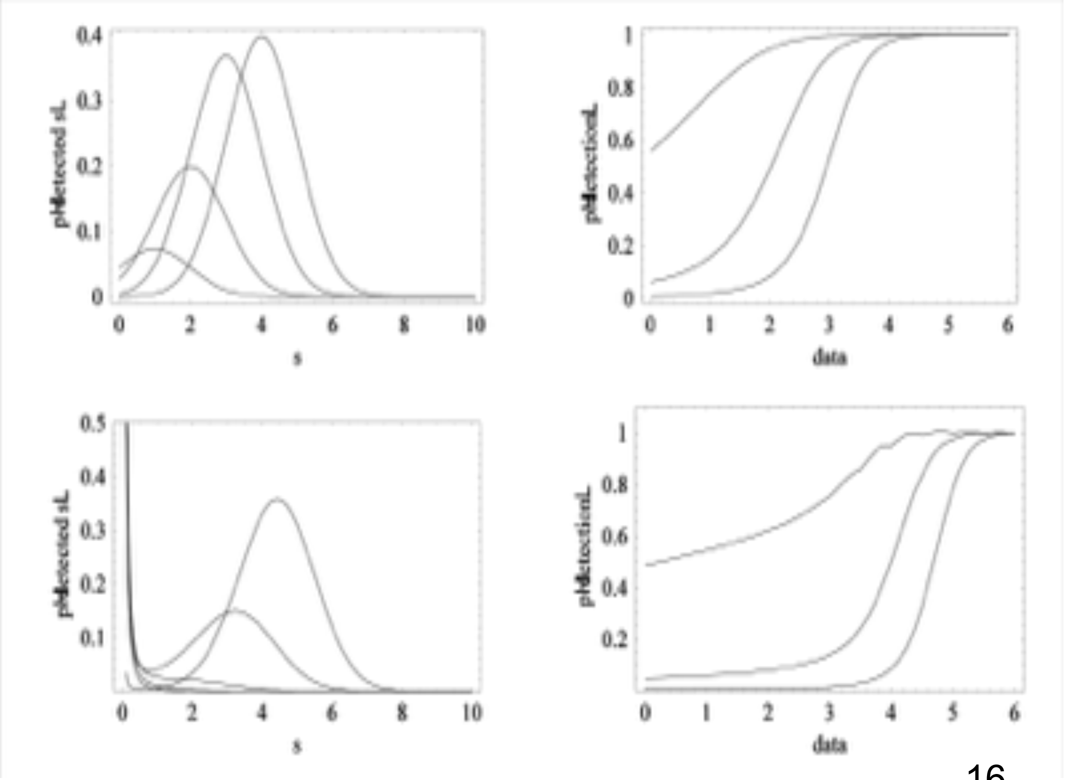
Bayesian Detection - Example 2

Take the **noise distribution to be Gaussian** and take a flat prior ***N(s)***.

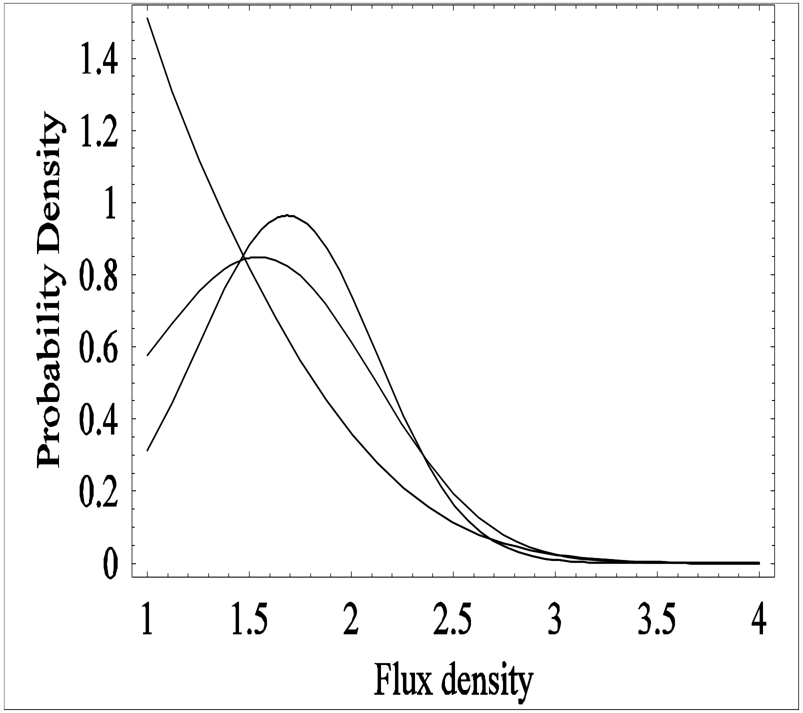
The value of *ε* reflects our initial confidence that a source is present at all, and so in many cases will be small.

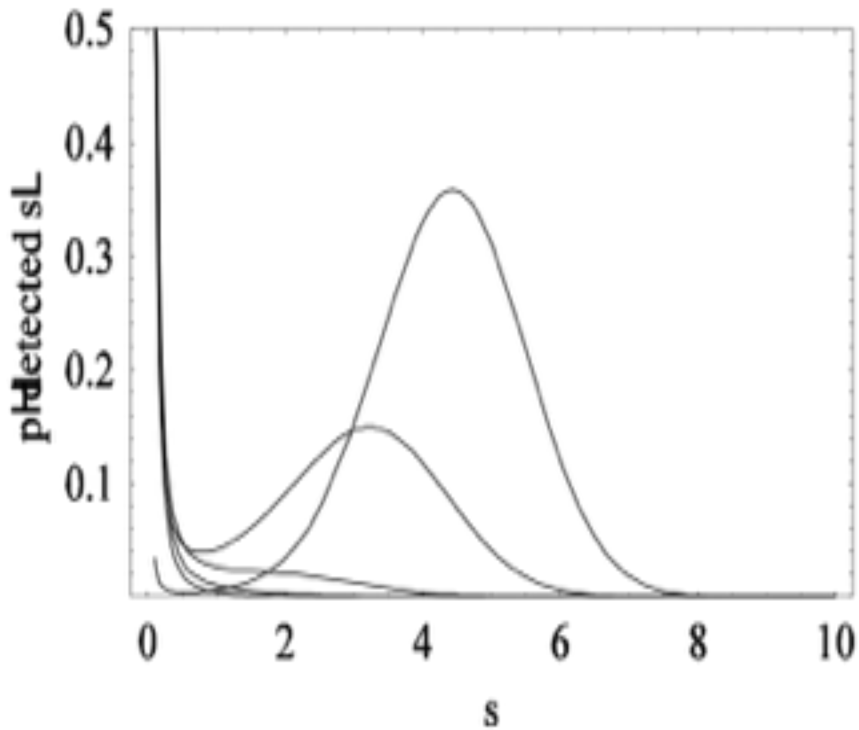
The Fig. shows that the posterior distribution of flux density ***s*** peaks at the value of the data, as expected; the role of *ε* is **to suppress our confidence of a detection in low s/n cases**. ***4-σ*** data points mean detection with high probability.

The top left panel shows the probability of a detected source of flux density s; the curves correspond to measurements of 1 to 4 units (unit = 1 noise std dev) A prior ε =0.05 was used. On the top right these curves are integrated to give the probability of detection at any positive flux density, as a function of the data values. The curves are for ε =0.5,0.05 and 0.005. The bottom panels show the results for the power-law prior, truncated at 0.1 unit.



Bayesian Detection - Example 2 concluded





Using a power-law prior ***N(s) = k s***$^{-5/2}$ gives results (right) similar to the previous example (left), but :

recognizes the possibility that no source might be present.

The rarity of bright sources in this prior now means that we need a **better s/n to achieve the same confidence** that we have a detection.

Detection - summary I

1. **Bayesian treatment of detection gives a direct result**; we may read off a suitable flux limit that will give the desired probability of detection.

2. But – the **confusion, and confusion limit** - images or spectral lines **crowd together**, **overlap** as we reach fainter. **Several** different objects may contribute to to the total flux at any coordinate. Even if only one 'object' is present, with a steep ***N(s)*** it will be more likely that the flux results from a **faint source plus a large upward noise excursion**, rather than vice-versa. (The 'Eddington bias'.)

3. Then we only expect to measure population properties -- parameters of the flux-density distribution ***N(s)*** or its spectral equivalent. We are getting in to the **confusion regime**, a concept we'll consider later. Beware of the return of **hyperparameters!**

Detection - summary 2

4. Detection is a modelling process:

- it depends on what we are looking for,
- how the answer is expressed depends on what we want to do with it next.

5. The **simple idea of a detection**, making a measurement of something that is really there, **applies when signal-to-noise is high** and individual objects can be isolated from the general signal. At low s/n, measurements can constrain population properties, with the notion of “detection" disappearing, in two senses:

We drop into the **confusion level**, and/or we deal with **censored data**.