ASSIGNMENT #3: due before FRIDAY 3/11

PROBABILITY I

Chapters 1+2 of Practical Statistics for Astronomers

Every measurement we make, and every parameter or value we derive requires an ERROR ESTIMATE, a measure of range (expressed in terms of probability) that encompasses our belief of the true value of the parameter.

No measure quantity or property is of the slightest use in decision and in science, unless it has a range quantity, i.e., an error, attached to it.

Every measurement we make, and every parameter or value we derive requires an ERROR ESTIMATE, a measure of range (expressed in terms of probability) that encompasses our belief of the true value of the parameter.

No measure quantity or property is of the slightest use in decision and in science, unless it has a range quantity, i.e., an error, attached to it.

Probability
is a numerical formalization of our degree or intensity of belief.

Every measurement we make, and every parameter or value we derive requires an ERROR ESTIMATE, a measure of range (expressed in terms of probability) that encompasses our belief of the true value of the parameter.

No measure quantity or property is of the slightest use in decision and in science, unless it has a range quantity, i.e., an error, attached to it.

Probability
is a numerical formalization of our degree or intensity of belief.

Kolmogorov axioms of probability:
1. any random event A has a probability prob(A) in [0,1]
2. the sure event has prob(A)=1
3. if A and B are exclusive events, then prob(A or B)=prob(A)+prob(B)

Two events A and B are INDEPENDENT if the probability of one is unaffected by what we know about the other: **prob(A and B)=prob(A)prob(B)**

Two events A and B are INDEPENDENT if the probability of one is unaffected by what we know about the other: **prob(A and B)=prob(A)prob(B)**

IF independence does not hold, we should know the CONDITIONAL PROBABILITY, i.e., the probability of A given that we know B:

$$\text{prob}(A \mid B) = \frac{\text{prob}(A \text{ AND } B)}{\text{prob}(B)}$$

Two events A and B are INDEPENDENT if the probability of one is unaffected by what we know about the other: **prob(A and B)=prob(A)prob(B)**

IF independence does not hold, we should know the CONDITIONAL PROBABILITY, i.e., the probability of A given that we know B:

$$\mathrm{prob}(A \mid B) = \frac{\mathrm{prob}(A \text{ AND } B)}{\mathrm{prob}(B)}$$

IF there are several possibilities for event B (i.e., B_i):

$$\mathrm{prob}(A) = \sum_i \mathrm{prob}(A \mid B_i)\,\mathrm{prob}(B_i)$$

Two events A and B are INDEPENDENT if the probability of one is unaffected by what we know about the other: **prob(A and B)=prob(A)prob(B)**

IF independence does not hold, we should know the CONDITIONAL PROBABILITY, i.e., the probability of A given that we know B:

$$\mathrm{prob}(A \mid B) = \frac{\mathrm{prob}(A \text{ AND } B)}{\mathrm{prob}(B)}$$

IF there are several possibilities for event B (i.e., B_i):

$$\mathrm{prob}(A) = \sum_i \mathrm{prob}(A \mid B_i)\, \mathrm{prob}(B_i)$$

Example: “A” might be a cosmological parameter of interest, while “B_i” are not of interest, e.g., instrumental parameters. Knowing prob(B_i), we can get rid of them by summation (or integration), a.k.a., marginalization.

Bayes' Theorem: From prob(B and A) = prob(A and B), one can demonstrate that

$$\mathrm{prob}(\mathrm{B}|\mathrm{A}) = \mathrm{prob}(\mathrm{A}|\mathrm{B})\,\frac{\mathrm{prob}(\mathrm{B})}{\mathrm{prob}(\mathrm{A})}$$

prior

normalizing factor

posterior

likelihood

Bayes' Theorem: From prob(B and A) = prob(A and B), one can demonstrate that

$$\underbrace{\mathrm{prob}(\mathrm{B}|\mathrm{A})}_{\text{posterior}} = \underbrace{\mathrm{prob}(\mathrm{A}|\mathrm{B})}_{\text{likelihood}} \frac{\overbrace{\mathrm{prob}(\mathrm{B})}^{\text{prior}}}{\underbrace{\mathrm{prob}(\mathrm{A})}_{\text{normalizing factor}}}$$

The data, i.e., the event A, are regarded as succeeding (i.e., coming after) B, the state of belief preceding the experiment. prob(B) is the prior probability, which will be modified by experience. This experience is expressed by the likelihood prob(A|B), while prob(B|A) is the posterior probability, i.e., the state of belief after the data have been analyzed.

Bayes' Theorem: From prob(B and A) = prob(A and B), one can demonstrate that

$$\mathrm{prob}(\mathrm{B}|\mathrm{A}) = \mathrm{prob}(\mathrm{A}|\mathrm{B})\,\frac{\mathrm{prob}(\mathrm{B})}{\mathrm{prob}(\mathrm{A})}$$

prior

normalizing factor

posterior likelihood

The data, i.e., the event A, are regarded as succeeding (i.e., coming after) B, the state of belief preceding the experiment. prob(B) is the prior probability, which will be modified by experience. This experience is expressed by the likelihood prob(A|B), while prob(B|A) is the posterior probability, i.e., the state of belief after the data have been analyzed.

Bayes' theorem allows us to make inferences from data, rather than compute the data we would get if we happened to know all the relevant information about the problem.

Bayes' Theorem: From prob(B and A) = prob(A and B), one can demonstrate that

$$\underbrace{\mathrm{prob}(B|A)}_{\text{posterior}} = \underbrace{\mathrm{prob}(A|B)}_{\text{likelihood}} \frac{\overbrace{\mathrm{prob}(B)}^{\text{prior}}}{\underbrace{\mathrm{prob}(A)}_{\text{normalizing factor}}}$$

The data, i.e., the event A, are regarded as succeeding (i.e., coming after) B, the state of belief preceding the experiment. prob(B) is the prior probability, which will be modified by experience. This experience is expressed by the likelihood prob(AIB), while prob(BIA) is the posterior probability, i.e., the state of belief after the data have been analyzed.

Bayes' theorem allows us to make inferences from data, rather than compute the data we would get if we happened to know all the relevant information about the problem.

prob(BIA) = prob(model I data), i.e., the probability of the model B given the data A (or state of the model B given what we know of the data A)

Bayes' Theorem: From prob(B and A) = prob(A and B), one can demonstrate that

$$\underbrace{\mathrm{prob(B|A)}}_{\text{posterior}} = \underbrace{\mathrm{prob(A|B)}}_{\text{likelihood}} \frac{\overbrace{\mathrm{prob(B)}}^{\text{prior}}}{\underbrace{\mathrm{prob(A)}}_{\text{normalizing factor}}}$$

The data, i.e., the event A, are regarded as succeeding (i.e., coming after) B, the state of belief preceding the experiment. prob(B) is the prior probability, which will be modified by experience. This experience is expressed by the likelihood prob(AIB), while prob(BIA) is the posterior probability, i.e., the state of belief after the data have been analyzed.

Bayes' theorem allows us to make inferences from data, rather than compute the data we would get if we happened to know all the relevant information about the problem.

prob(BIA) = prob(model I data), i.e., the probability of the model B given the data A (or state of the model B given what we know of the data A)

prob(AIB) = likelihood = prob(data I model), i.e., the probability of the data A given the model B

Bayes' Theorem: From prob(B and A) = prob(A and B), one can demonstrate that

$$\underbrace{\mathrm{prob(B|A)}}_{\text{posterior}} = \underbrace{\mathrm{prob(A|B)}}_{\text{likelihood}} \frac{\overbrace{\mathrm{prob(B)}}^{\text{prior}}}{\underbrace{\mathrm{prob(A)}}_{\text{normalizing factor}}}$$

The data, i.e., the event A, are regarded as succeeding (i.e., coming after) B, the state of belief preceding the experiment. prob(B) is the prior probability, which will be modified by experience. This experience is expressed by the likelihood prob(AIB), while prob(BIA) is the posterior probability, i.e., the state of belief after the data have been analyzed.

Bayes' theorem allows us to make inferences from data, rather than compute the data we would get if we happened to know all the relevant information about the problem.

prob(BIA) = prob(model I data), i.e., the probability of the model B given the data A (or state of the model B given what we know of the data A)

prob(AIB) = likelihood = prob(data I model), i.e., the probability of the data A given the model B

prob(A) = normalization

to have $\int \mathrm{prob(B|A)} = 1$ ➡ $\mathrm{prob(A)} = \int \mathrm{prob(A|B)prob(B)}$ if continuous

$$\mathrm{prob(A)} = \sum \mathrm{prob(A|B)prob(B)} \quad \text{if discrete}$$

BAYES' THEOREM
Example 1

There are N red balls and M white balls in an urn; we know that N+M=10. We draw T=3 times (putting the balls back after drawing them) and get R=2 red balls. How many red balls are there in the urn?

GOAL: infer how many red balls there are given what we extract

BAYES' THEOREM
Example 1

There are N red balls and M white balls in an urn; we know that N+M=10. We draw T=3 times (putting the balls back after drawing them) and get R=2 red balls. How many red balls are there in the urn?

GOAL: infer how many red balls there are given what we extract

Probability of red ball: $$\frac{\mathrm{red}}{\mathrm{white} + \mathrm{red}} = \frac{N}{N+M}$$

BAYES' THEOREM
Example 1

There are N red balls and M white balls in an urn; we know that N+M=10. We draw T=3 times (putting the balls back after drawing them) and get R=2 red balls. How many red balls are there in the urn?

GOAL: infer how many red balls there are given what we extract

Probability of red ball: $$\frac{\text{red}}{\text{white} + \text{red}} = \frac{N}{N+M}$$

Probability of getting R red balls, i.e., the likelihood:

$$\text{prob}(A|B) = \binom{T}{R}\left(\frac{N}{N+M}\right)^{R}\left(\frac{M}{N+M}\right)^{T-R}$$

bimodal distribution

BAYES' THEOREM
Example 1

There are N red balls and M white balls in an urn; we know that N+M=10. We draw T=3 times (putting the balls back after drawing them) and get R=2 red balls. How many red balls are there in the urn?

GOAL: infer how many red balls there are given what we extract

Probability of red ball: $$\frac{\text{red}}{\text{white} + \text{red}} = \frac{N}{N+M}$$

Probability of getting R red balls, i.e., the likelihood:

$$\text{prob}(A|B) = \binom{T}{R}\left(\frac{N}{N+M}\right)^{R}\left(\frac{M}{N+M}\right)^{T-R}$$

bimodal distribution

Number of permutations of the R red balls amongst the T draws

$$= \frac{T!}{(T-R)!R!} = T(T-1)(T-2)...(T-R+1)/R!$$

BAYES' THEOREM
Example 1

There are N red balls and M white balls in an urn; we know that N+M=10. We draw T=3 times (putting the balls back after drawing them) and get R=2 red balls. How many red balls are there in the urn?

GOAL: infer how many red balls there are given what we extract

Probability of red ball: $$\frac{\text{red}}{\text{white} + \text{red}} = \frac{N}{N+M}$$

Probability of getting R red balls, i.e., the likelihood:

$$\text{prob}(A|B) = \binom{T}{R}\left(\frac{N}{N+M}\right)^{R}\left(\frac{M}{N+M}\right)^{T-R}$$

bimodal distribution

Number of permutations of the R red balls amongst the T draws

$$= \frac{T!}{(T-R)!R!} = T(T-1)(T-2)...(T-R+1)/R!$$

Prob that R balls will be red

BAYES' THEOREM
Example 1

There are N red balls and M white balls in an urn; we know that N+M=10. We draw T=3 times (putting the balls back after drawing them) and get R=2 red balls. How many red balls are there in the urn?

GOAL: infer how many red balls there are given what we extract

Probability of red ball: $\frac{\text{red}}{\text{white} + \text{red}} = \frac{N}{N+M}$

Probability of getting R red balls, i.e., the likelihood:

$$\text{prob}(A|B) = \binom{T}{R}\left(\frac{N}{N+M}\right)^{R}\left(\frac{M}{N+M}\right)^{T-R}$$

bimodal distribution

Number of permutations of the R red balls amongst the T draws

$$= \frac{T!}{(T-R)!R!} = T(T-1)(T-2)...(T-R+1)/R!$$

Prob that R balls will be red

Prob that T-R balls will not be red (i.e., will be white)

BAYES' THEOREM
Example 1

There are N red balls and M white balls in an urn; we know that N+M=10. We draw T=3 times (putting the balls back after drawing them) and get R=2 red balls. How many red balls are there in the urn?

GOAL: infer how many red balls there are given what we extract

Probability of red ball: $\frac{\text{red}}{\text{white} + \text{red}} = \frac{N}{N+M}$

Probability of getting R red balls, i.e., the likelihood:

$$\text{prob}(A|B) = \binom{T}{R} \left(\frac{N}{N+M}\right)^{R} \left(\frac{M}{N+M}\right)^{T-R}$$

bimodal distribution

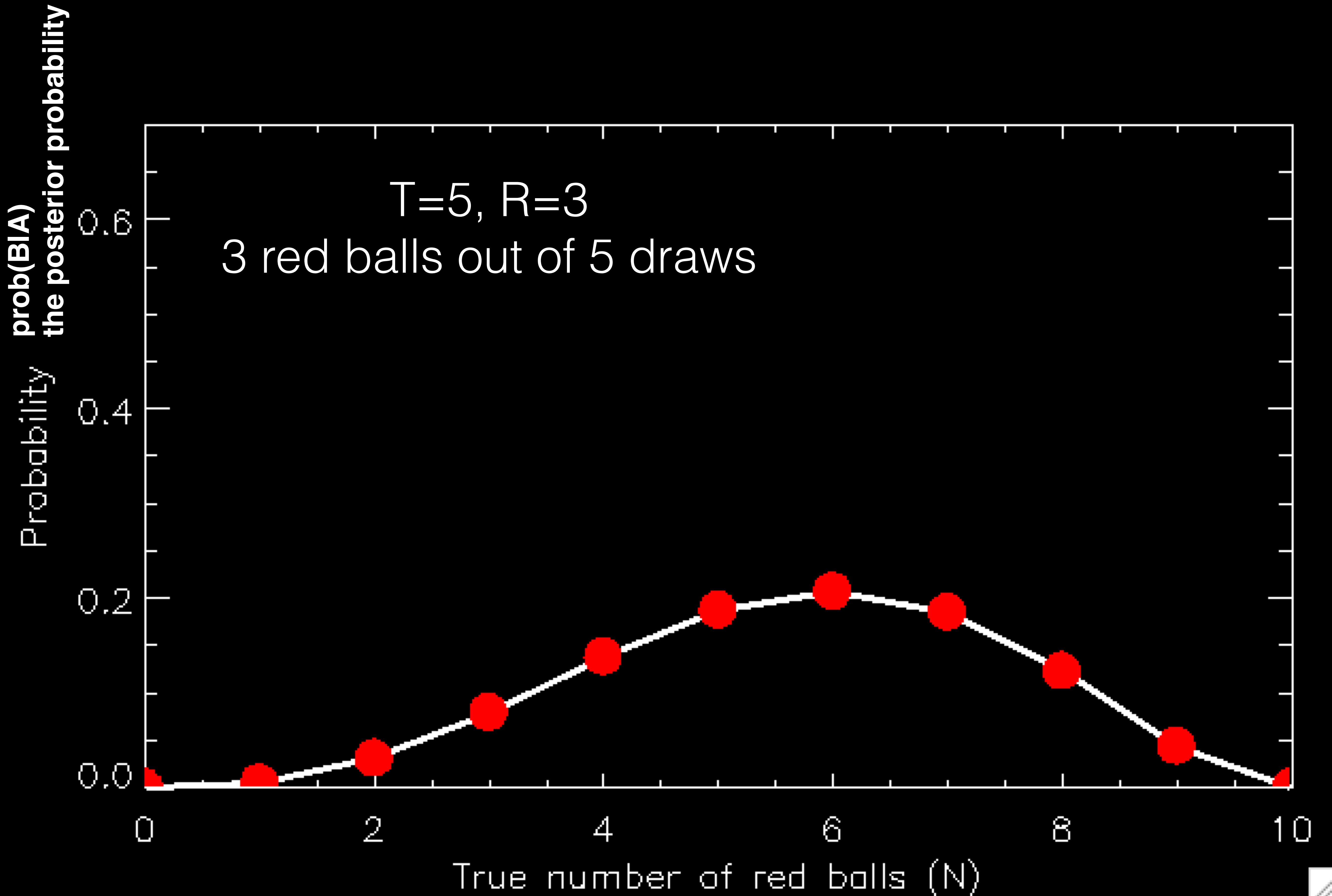
Number of permutations of the R red balls amongst the T draws

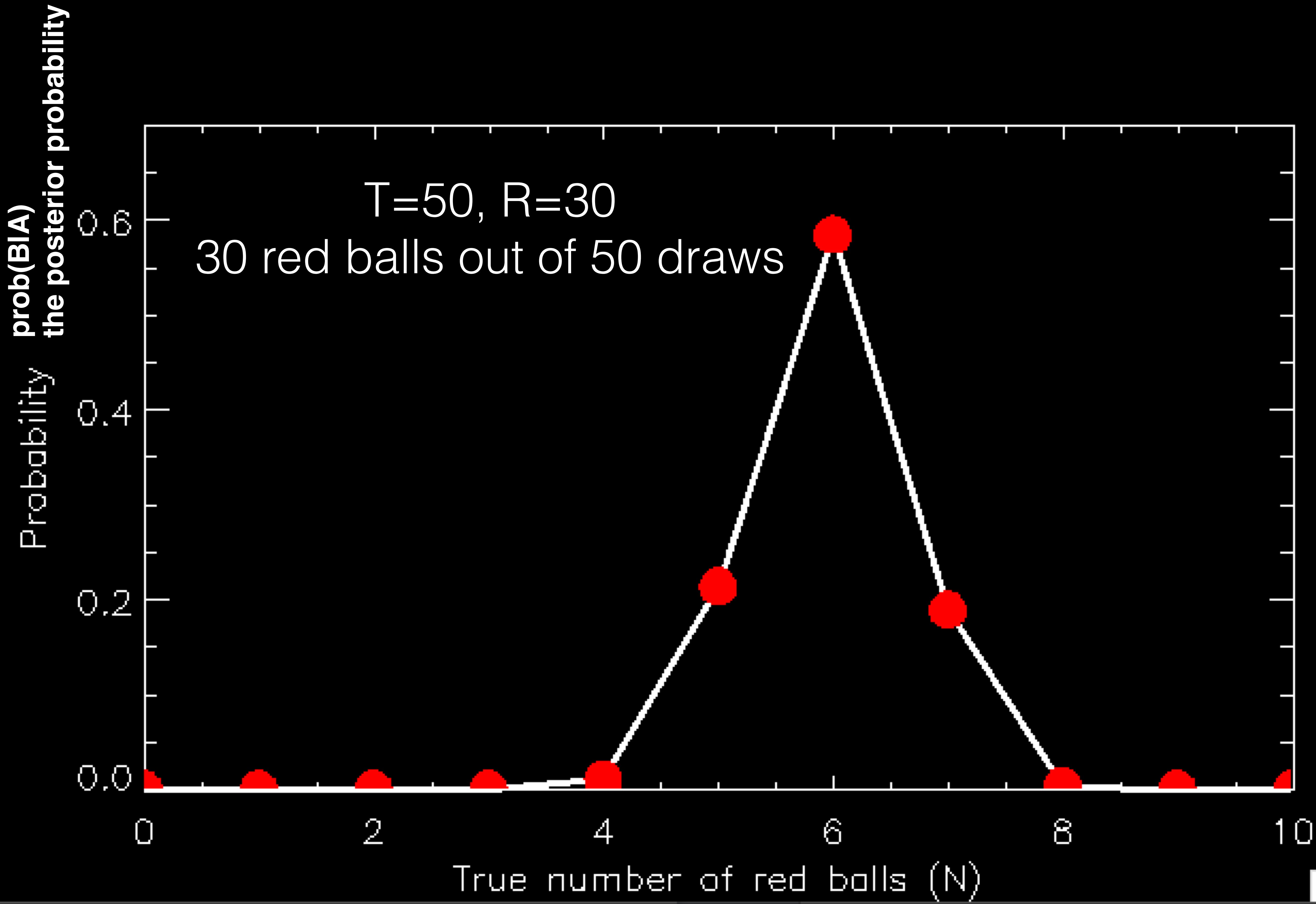
$$= \frac{T!}{(T-R)!R!} = T(T-1)(T-2)...(T-R+1)/R!$$

Prob that R balls will be red

Prob that T-R balls will not be red (i.e., will be white)

prob(B) = prior: start with a uniformly distributed N with N in [0,N+M]


prob(B|A)
the posterior probability
T=5, R=3
3 red balls out of 5 draws
Probability
0.6
0.4
0.2
0.0
0
2
4
6
8
10
True number of red balls (N)



NOTE: as the sample size increases, the distribution becomes narrower ==> las of large numbers

Bayes' theorem + probability as a measure of belief **allows us to answer the question**: given the data, what are the probabilities of the parameters contained in our statistical model?

NOTE: the prior is what we know apart from the data. Sometimes, this can have a dramatic effect on our inferences. Sometimes, for the prior, we even need a “probability of a probability”.

BAYES' THEOREM
Example 2 - effect of prior / peak - mean

Calculate the supernova rate per century (ρ) assuming we observe 4 supernovas on 10 centuries

BAYES' THEOREM
Example 2 - effect of prior / peak - mean

Calculate the supernova rate per century (ρ) assuming we observe 4 supernovas on 10 centuries

$\mathrm{prob}(\rho)$ = uniform between 0 and 1 SNs per century (total ignorance)

BAYES' THEOREM
Example 2 - effect of prior / peak - mean

Calculate the supernova rate per century (ρ) assuming we observe 4 supernovas on 10 centuries

$\mathrm{prob}(\rho)$ = uniform between 0 and 1 SNs per century (total ignorance)

$\mathrm{prob}(\mathrm{data}|\rho)$ = bimodal distribution, as in any century we can either get a supernova or we do not

$$= \binom{10}{4} \rho^4 (1-\rho)^6$$

BAYES' THEOREM
Example 2 - effect of prior / peak - mean

Calculate the supernova rate per century (ρ) assuming we observe 4 supernovas on 10 centuries

$\text{prob}(\rho)$ = uniform between 0 and 1 SNs per century (total ignorance)

$\text{prob}(\text{data}|\rho)$ = bimodal distribution, as in any century we can either get a supernova or we do not

$$= \binom{10}{4} \rho^4 (1-\rho)^6$$

$\text{prob}(\text{data})$ from $\int_0^1 \text{prob}(\rho|\text{data})\, d\rho = 1 \equiv \int_0^1 \binom{10}{4} \rho^4 (1-\rho)^6 \,\text{prob}(\rho)\, d\rho$

BAYES' THEOREM
Example 2 - effect of prior / peak - mean

Calculate the supernova rate per century (ρ) assuming we observe 4 supernovas on 10 centuries

$\mathrm{prob}(\rho)$ = uniform between 0 and 1 SNs per century (total ignorance)

$\mathrm{prob}(\mathrm{data}|\rho)$ = bimodal distribution, as in any century we can either get a supernova or we do not

$$= \binom{10}{4} \rho^4 (1-\rho)^6$$

$\mathrm{prob}(\mathrm{data})$ from $\int_0^1 \boxed{\mathrm{prob}(\rho|\mathrm{data})}\, d\rho = 1 \equiv \int_0^1 \binom{10}{4} \rho^4 (1-\rho)^6 \,\mathrm{prob}(\rho)\, d\rho$

posterior probability

BAYES' THEOREM
Example 2 - effect of prior / peak - mean

Calculate the supernova rate per century (ρ) assuming we observe 4 supernovas on 10 centuries

$\mathrm{prob}(\rho)$ = uniform between 0 and 1 SNs per century (total ignorance)

$\mathrm{prob}(\mathrm{data}|\rho)$ = bimodal distribution, as in any century we can either get a supernova or we do not

$$= \binom{10}{4} \rho^4 (1-\rho)^6$$

$\mathrm{prob}(\mathrm{data})$ from $\int_0^1 \boxed{\mathrm{prob}(\rho|\mathrm{data})}\, d\rho = 1 \equiv \int_0^1 \boxed{\binom{10}{4} \rho^4 (1-\rho)^6}\, \mathrm{prob}(\rho)\, d\rho$

posterior probability

likelihood

BAYES' THEOREM
Example 2 - effect of prior / peak - mean

Calculate the supernova rate per century (ρ) assuming we observe 4 supernovas on 10 centuries

$\mathrm{prob}(\rho)$ = uniform between 0 and 1 SNs per century (total ignorance)

$\mathrm{prob}(\mathrm{data}|\rho)$ = bimodal distribution, as in any century we can either get a supernova or we do not

$$= \binom{10}{4} \rho^4 (1-\rho)^6$$

$\mathrm{prob}(\mathrm{data})$ from $\int_0^1 \boxed{\mathrm{prob}(\rho|\mathrm{data})}\, d\rho = 1 \equiv \int_0^1 \boxed{\binom{10}{4} \rho^4 (1-\rho)^6} \boxed{\mathrm{prob}(\rho)}\, d\rho$

posterior probability

likelihood

prior

BAYES' THEOREM
Example 2 - effect of prior / peak - mean

Calculate the supernova rate per century (ρ) assuming we observe 4 supernovas on 10 centuries

$\mathrm{prob}(\rho)$ = uniform between 0 and 1 SNs per century (total ignorance)

$\mathrm{prob}(\mathrm{data}|\rho)$ = bimodal distribution, as in any century we can either get a supernova or we do not

$$= \binom{10}{4} \rho^4 (1-\rho)^6$$

$\mathrm{prob}(\mathrm{data})$ from

$$\int_0^1 \underbrace{\boxed{\mathrm{prob}(\rho|\mathrm{data})}}_{\text{posterior probability}} d\rho = 1 \equiv \int_0^1 \underbrace{\boxed{\binom{10}{4}\rho^4(1-\rho)^6}}_{\text{likelihood}} \underbrace{\boxed{\mathrm{prob}(\rho)}}_{\text{prior}} d\rho$$

We ascribe a probability distribution to ρ , in itself a probability.

Prior on ρ: a uniform prior is often too agnostic.

$$\Rightarrow \quad \text{prob}(\rho) = \frac{1}{\rho(1-\rho)} \quad \text{A} \qquad \text{JAYNE'S prior}$$

$$\text{prob}(\rho) = \frac{1}{\sqrt{\rho(1-\rho)}} \quad \text{B} \qquad \text{or "Haldane prior"}$$

These two priors reflect the fact that in most experiments, we are expecting a yes or a no answer.

Prior on ρ: a uniform prior is often too agnostic.

$$\Rightarrow \quad \text{prob}(\rho) = \frac{1}{\rho(1-\rho)} \quad \text{A} \qquad \text{JAYNE'S prior}$$

$$\text{prob}(\rho) = \frac{1}{\sqrt{\rho(1-\rho)}} \quad \text{B} \qquad \text{or "Haldane prior"}$$

These two priors reflect the fact that in most experiments, we are expecting a yes or a no answer.

NOTE: Assigning priors when our knowledge is rather vague can be quite difficult. Some obvious priors, uniform from -infinity to +infinity are NOT normalizable, hence they are trouble.

Prior on ρ: a uniform prior is often too agnostic.

$$\Rightarrow \quad \text{prob}(\rho) = \frac{1}{\rho(1-\rho)} \quad \text{A} \qquad \text{JAYNE'S prior}$$

$$\text{prob}(\rho) = \frac{1}{\sqrt{\rho(1-\rho)}} \quad \text{B} \qquad \text{or "Haldane prior"}$$

These two priors reflect the fact that in most experiments, we are expecting a yes or a no answer.

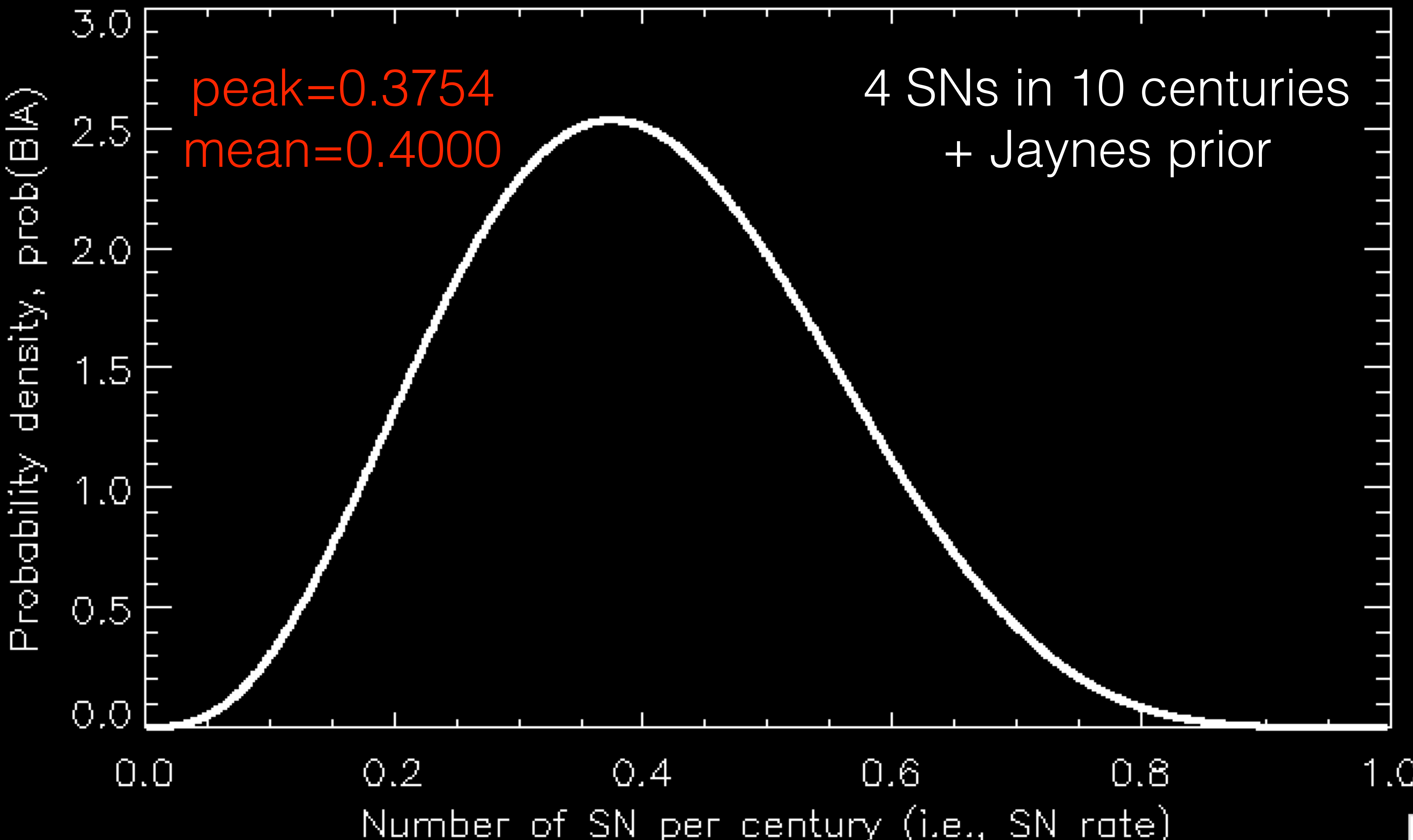
NOTE: Assigning priors when our knowledge is rather vague can be quite difficult. Some obvious priors, uniform from -infinity to +infinity are NOT normalizable, hence they are trouble.

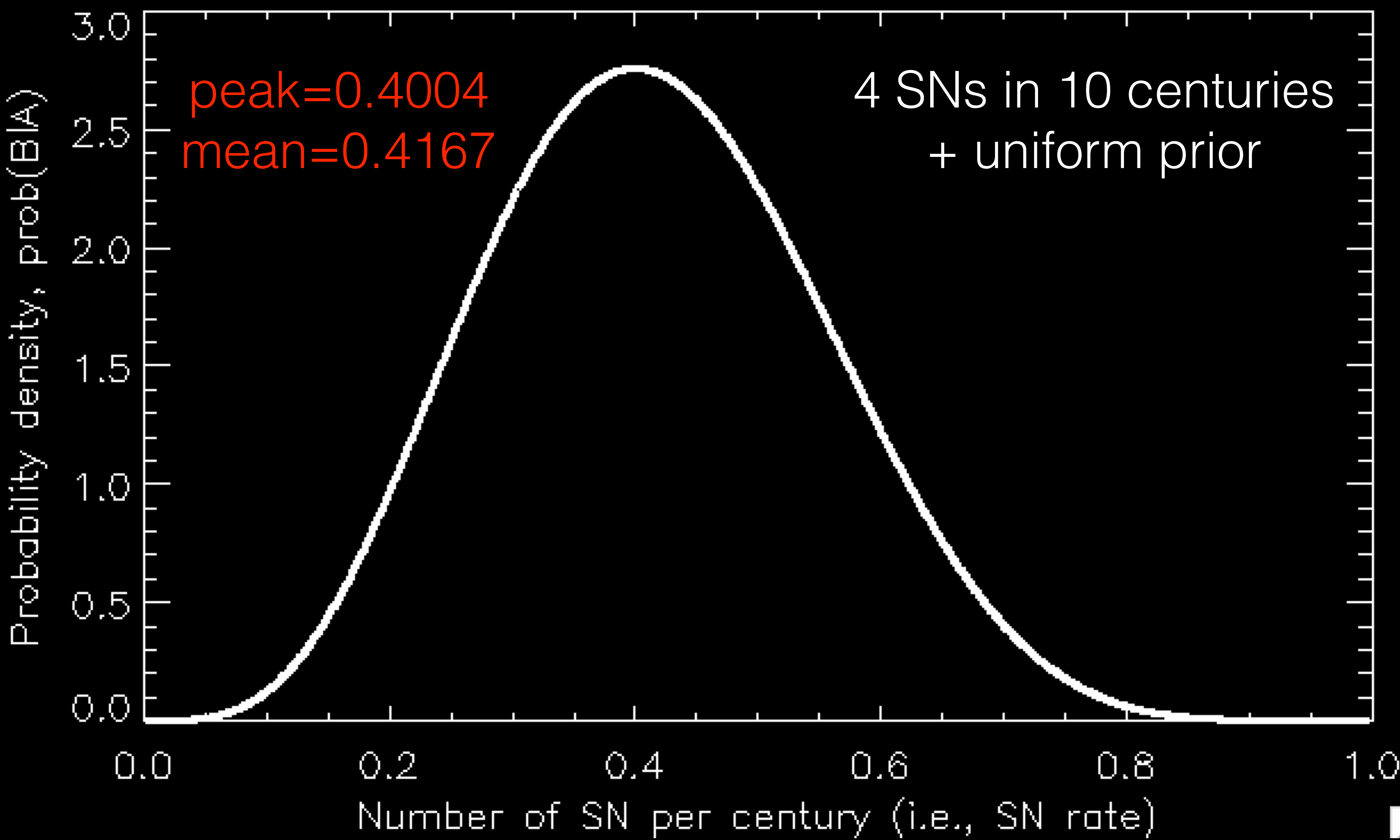
NOTE: How to characterize the posterior probability by a single number

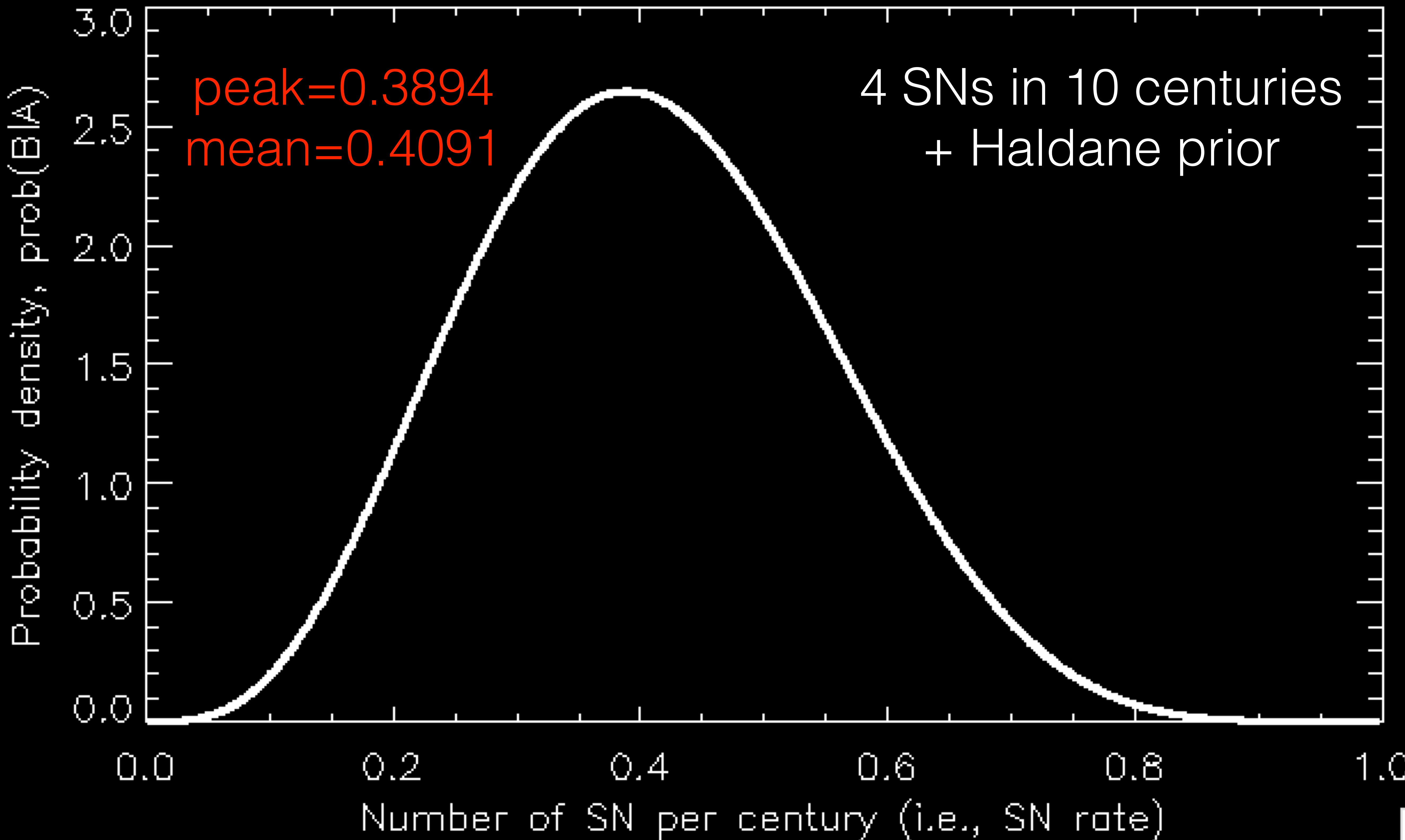
1. Peak of the posterior probability (max[prob(B|A)])

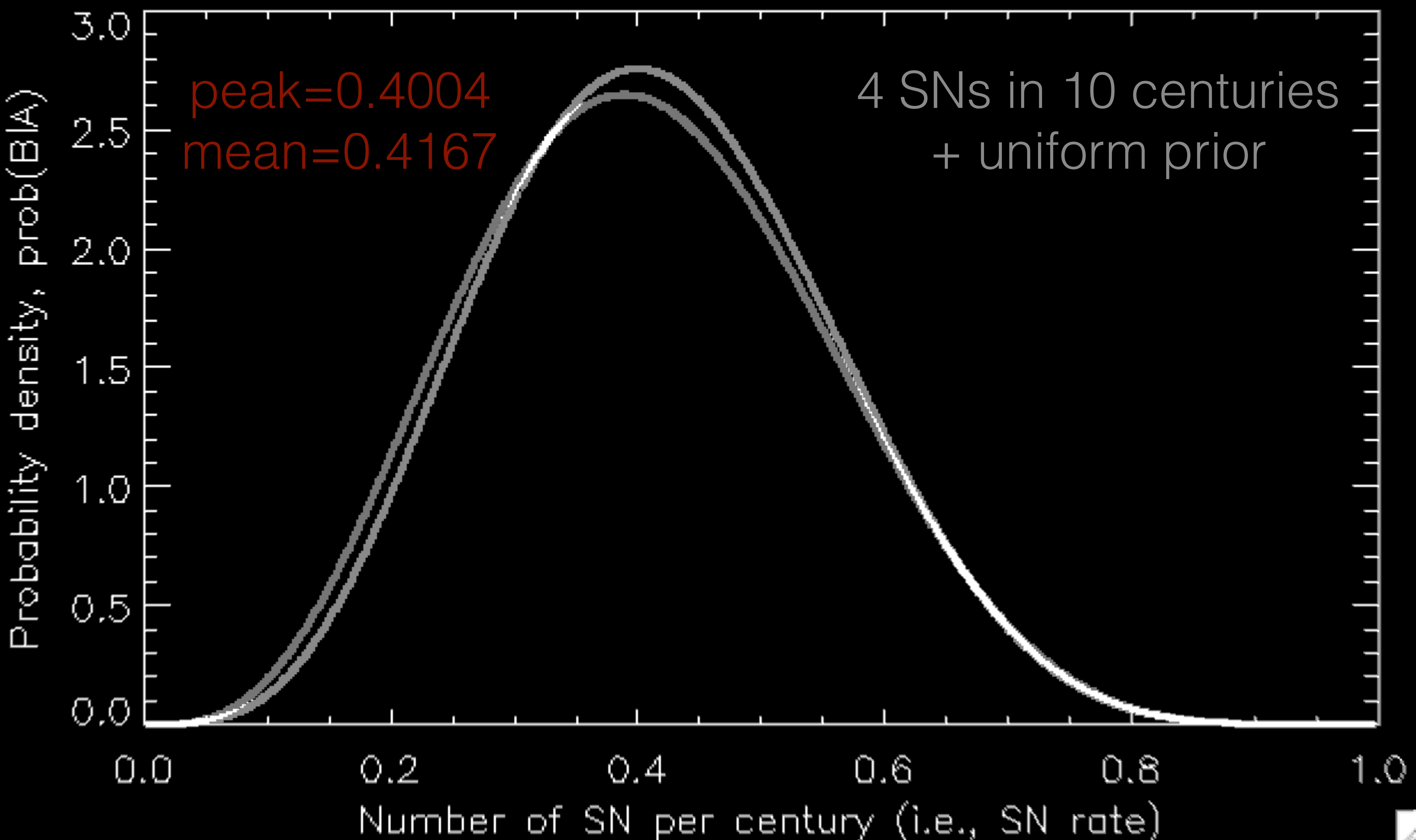
2. Posterior mean $<\rho>=\int_0^1 \rho \ \text{prob}(\rho|\text{data}) d\rho$

3. Unless the posterior distributions are very narrow, attempting to characterize them by a single number is misleading. How to best characterize them depends on what is to be done with the answer, which, in turn, depends on having a carefully posed question.


peak=0.3754
mean=0.4000
4 SNs in 10 centuries
+ Jaynes prior
Probability density, prob(B|A)
Number of SN per century (i.e., SN rate)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0


peak=0.4004
mean=0.4167
4 SNs in 10 centuries
+ uniform prior
Probability density, prob(B|A)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.0
0.2
0.4
0.6
0.8
1.0
Number of SN per century (i.e., SN rate)


peak=0.3894
mean=0.4091
4 SNs in 10 centuries
+ Haldane prior
Probability density, prob(B|A)
Number of SN per century (i.e., SN rate)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0



peak=0.4004
mean=0.4167
4 SNs in 10 centuries
+ uniform prior
Probability density, prob(B|A)
Number of SN per century (i.e., SN rate)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Probability Distributions: A function describing the expectation of concurrence of event x.
It can be discrete or continuous

Probability Distributions: A function describing the expectation of concurrence of event x.
It can be discrete or continuous

f(x) probability density

—> f(x)dx = probability of getting a number near x within a tiny range dx

Probability Distributions: A function describing the expectation of concurrence of event x.
It can be discrete or continuous

f(x) probability density

—> f(x)dx = probability of getting a number near x within a tiny range dx

If x is a continuous random variable, then f(x) is its probability density function, a.j.a., probability distribution, when:

1. prob[a<x<b] = $\int_a^b f(x)dx$
2. $$\int_{-\infty}^{+\infty} f(x)dx = 1$$
3. f(x) is a single-valued non-negative number for all real x

Probability Distributions: A function describing the expectation of concurrence of event x. It can be discrete or continuous

f(x) probability density

—> f(x)dx = probability of getting a number near x within a tiny range dx

If x is a continuous random variable, then f(x) is its probability density function, a.j.a., probability distribution, when:

1. prob[a<x<b] = $\int_a^b f(x)dx$

2. $$\int_{-\infty}^{+\infty} f(x)dx = 1$$

3. f(x) is a single-valued non-negative number for all real x

Cumulative probability distribution function: $F(x) = \int_{-\infty}^{x} f(y)dy$

Probability Distributions: A function describing the expectation of concurrence of event x. It can be discrete or continuous

f(x) probability density

—> f(x)dx = probability of getting a number near x within a tiny range dx

If x is a continuous random variable, then f(x) is its probability density function, a.j.a., probability distribution, when:

1. prob[a<x<b] = $\int_a^b f(x)dx$

2. $$\int_{-\infty}^{+\infty} f(x)dx = 1$$

3. f(x) is a single-valued non-negative number for all real x

Cumulative probability distribution function: $F(x) = \int_{-\infty}^{x} f(y)dy$

Mean $\mu := \int_{-\infty}^{\infty} x f(x) dx$ ie: the center

Variance $\sigma^2 := \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$ ie: the spread

Standard deviation $\sigma = \sqrt{\text{Variance}}$

UNIFORM Distribution:

$$f(x; a, b) = \begin{cases} 0 & \text{for } x < a \text{ or } x > b \\ \dfrac{1}{(b-a)} & \text{for } a \le x \le b \end{cases}$$

$$\mu = \frac{a+b}{2} \qquad \sigma^2 = \frac{(b-a)^2}{12}$$

BIMODAL Distribution: The bimodal distribution gives the chance of n successes in N trials, where the probability of a success at each trial is the same (ρ) and successive trials are independent.

$$\text{prob}(n) = \underbrace{\binom{N}{n}}_{\text{\# of distinct ways of choosing } n \text{ items out of } N \text{ trials}} \rho^{\overbrace{n}^{\text{\# of successes}}} \underbrace{(1-\rho)}_{\text{probability of failure}}{}^{\overbrace{N-n}^{\text{\# of failures}}}$$

\# of distinct ways of choosing n items out of N trials $= \dfrac{N!}{n!\,(N-n)!}$

$$\mu = N\rho \quad \left(\equiv \sum_{n=0}^{N} n \cdot \text{prob}(n) \right)$$

$$\sigma^2 = N\rho\,(1-\rho) \quad \left(\equiv \sum_{n=0}^{N} (n - N\rho)^2\, \text{prob}(n) \right)$$

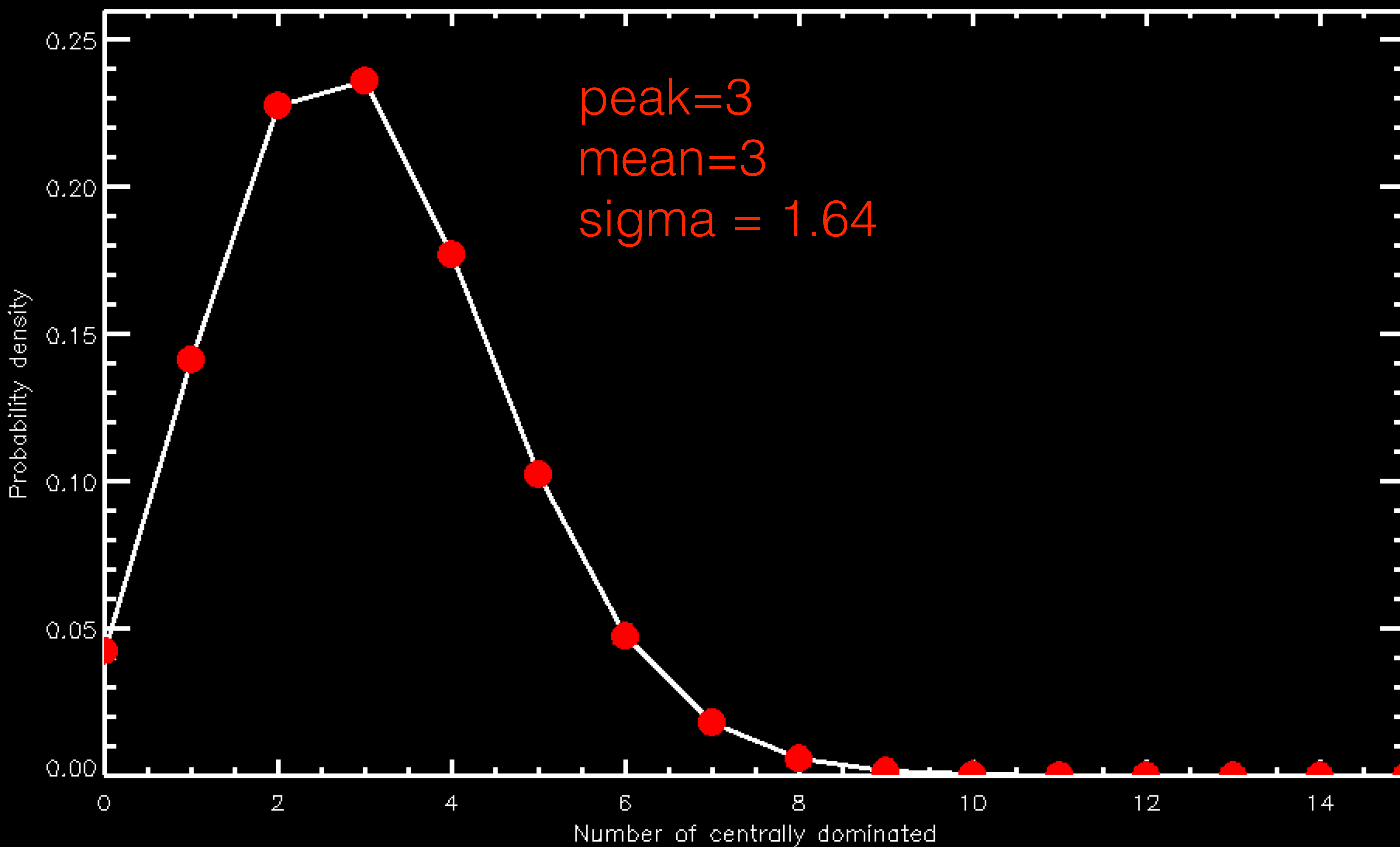
NOTE: The bimodal distribution is the parent population of the Poisson and Gaussian distributions.

BIMODAL DISTRIBUTION
Example

Out of 100 clusters of galaxies, 10 contain a dominant central galaxy. We plan to check a different sample of 30 galaxies' clusters selected with X-ray observations. How many do we expect to have a central dominant galaxy?

$$\text{prob}(n) = \binom{30}{n} \underbrace{0.1}_{10/100}^{\,n} \; 0.9^{30-n}$$

= probability of getting n dominant central galaxies.


peak=3
mean=3
sigma = 1.64
Probability density
Number of centrally dominated
0.00
0.05
0.10
0.15
0.20
0.25
0
2
4
6
8
10
12
14

POISSON Distribution: From the bimodal distribution, with $\rho << 1 \ (i.e., \rho \to 0)$ (i.e., very rare independent events) and a large number of trials

Appropriate to describe small samples from large populations

$$\text{prob}(n) = \frac{\mu^n}{n!} e^{-\mu}$$

w/ $\mu = N\rho$ mean

$\sigma^2 = \mu$

POISSON Distribution: From the bimodal distribution, with $\rho << 1 \; (i.e., \rho \to 0)$ (i.e., very rare independent events) and a large number of trials

Appropriate to describe small samples from large populations

$$\text{prob}(n) = \frac{\mu^n}{n!} e^{-\mu} \qquad \text{w/ } \mu = N\rho \quad \text{mean}$$
$$\sigma^2 = \mu$$

The Poisson distribution plays its biggest role in the lives of astronomers via the photons with which we measure emission from our chosen objects.

Poisson statistics governs the number of photons arriving during an exposure. The probability of a photon arriving in a fixed internal of time is often small. The arrival of successive photons are independent, hence the Poisson distribution applies.

POISSON Distribution: From the bimodal distribution, with $\rho << 1$ $(i.e., \rho \to 0)$
(i.e., very rare independent events) and a large number of trials

Appropriate to describe small samples from large populations

$$\text{prob}(n) = \frac{\mu^n}{n!} e^{-\mu}$$

w/ $\mu = N\rho$ mean

$\sigma^2 = \mu$

The Poisson distribution plays its biggest role in the lives of astronomers via the photons with which we measure emission from our chosen objects.

Poisson statistics governs the number of photons arriving during an exposure. The probability of a photon arriving in a fixed internal of time is often small. The arrival of successive photons are independent, hence the Poisson distribution applies.

t integration time

λ rate of photons (#/sec) $\rightarrow$ $\mu = \lambda t$ $\rightarrow$ $\sigma = \sqrt{\mu}$

POISSON Distribution: From the bimodal distribution, with $\rho << 1$ $(i.e., \rho \to 0)$ (i.e., very rare independent events) and a large number of trials

Appropriate to describe small samples from large populations

$$\text{prob}(n) = \frac{\mu^n}{n!} e^{-\mu}$$

w/ $\mu = N\rho$ mean

$\sigma^2 = \mu$

The Poisson distribution plays its biggest role in the lives of astronomers via the photons with which we measure emission from our chosen objects.

Poisson statistics governs the number of photons arriving during an exposure. The probability of a photon arriving in a fixed internal of time is often small. The arrival of successive photons are independent, hence the Poisson distribution applies.

t integration time

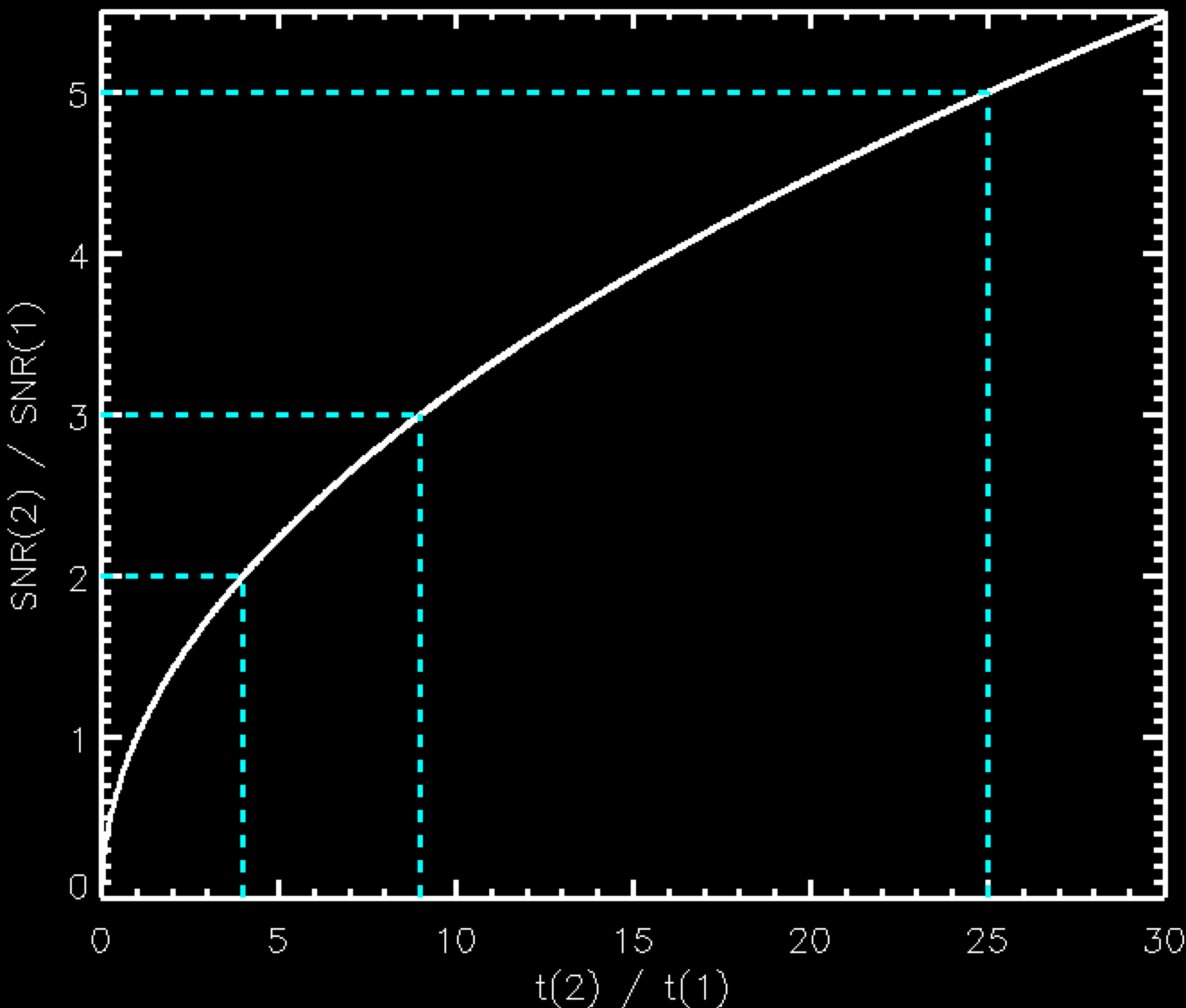
λ rate of photons (#/sec) $\rightarrow$ $\mu = \lambda t$ $\rightarrow$ $\sigma = \sqrt{\mu}$

Assume photons are arriving only from the object measured

$\rightarrow$ $\mu = \lambda t$ & scatter on μ is (Poisson) $\sigma = \sqrt{\lambda t}$

With long exposures, the photon-limited case is $\sigma \propto \sqrt{t}$ w/ signal $\propto t$

$\Rightarrow$ $S/N \propto \sqrt{t}$


SNR(2) / SNR(1)
t(2) / t(1)
0
1
2
3
4
5
0
5
10
15
20
25
30

GAUSSIAN Distribution:

$$\text{prob}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

w/ mean μ
variance σ^2

GAUSSIAN Distribution:

$$\text{prob}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$ w/ mean μ, variance σ^2

For binomial w/ sample size very large where p is the probability $\Rightarrow$ $$\text{prob}(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(n-\mu)^2}{2\sigma^2}\right]$$

w/ $\mu = Np$ $\quad \sigma^2 = Np(1-p)$

(appropriate for smooth symmetric distributions)

GAUSSIAN Distribution: $\text{prob}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$ w/ mean μ, variance σ^2

For binomial w/ sample size very large where p is the probability $\Rightarrow$ $\text{prob}(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(n-\mu)^2}{2\sigma^2}\right]$

w/ $\mu = Np$ $\quad \sigma^2 = Np(1-p)$

(appropriate for smooth symmetric distributions)

The probability that the value of a random observation will fall within the interval dx around x is prob(x)dx $\Rightarrow$ $\int_{-\infty}^{\infty} \text{prob}(x)\, dx = 1$

GAUSSIAN Distribution:

$$\text{prob}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad \text{w/ mean } \mu \text{, variance } \sigma^2$$

For binomial w/ sample size very large where p is the probability (appropriate for smooth symmetric distributions) $\Rightarrow$

$$\text{prob}(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(n-\mu)^2}{2\sigma^2}\right]$$

w/ $\mu = Np \qquad \sigma^2 = Np(1-p)$

The probability that the value of a random observation will fall within the interval dx around x is prob(x)dx

$$\Rightarrow \int_{-\infty}^{\infty} \text{prob}(x)\, dx = 1$$

At $x = \mu \pm \sigma$, the height of the curve is reduced to $e^{-1/2}$ of its value at the peak, ie:

$$\text{prob}(\mu \pm \sigma) = e^{-1/2}\, \text{prob}(\mu)$$

GAUSSIAN Distribution: $\text{prob}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$ w/ mean μ, variance σ^2

For binomial w/ sample size very large where p is the probability $\Rightarrow$ $\text{prob}(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(n-\mu)^2}{2\sigma^2}\right]$

(appropriate for smooth symmetric distributions)

w/ $\mu = Np$ $\quad \sigma^2 = Np(1-p)$

The probability that the value of a random observation will fall within the interval dx around x is prob(x)dx

$$\Rightarrow \int_{-\infty}^{\infty} \text{prob}(x)\, dx = 1$$

At $x = \mu \pm \sigma$, the height of the curve is reduced to $e^{-1/2}$ of its value at the peak, ie: $\text{prob}(\mu \pm \sigma) = e^{-1/2}\, \text{prob}(\mu)$

The Gaussian distribution displays the characteristic bell shape and symmetry about the mean.

GAUSSIAN Distribution: $$\text{prob}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$ w/ mean μ, variance σ^2

For binomial w/ sample size very large where p is the probability $\Rightarrow$ $$\text{prob}(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(n-\mu)^2}{2\sigma^2}\right]$$

(appropriate for smooth symmetric distributions)

w/ $\mu = Np$ $\quad \sigma^2 = Np(1-p)$

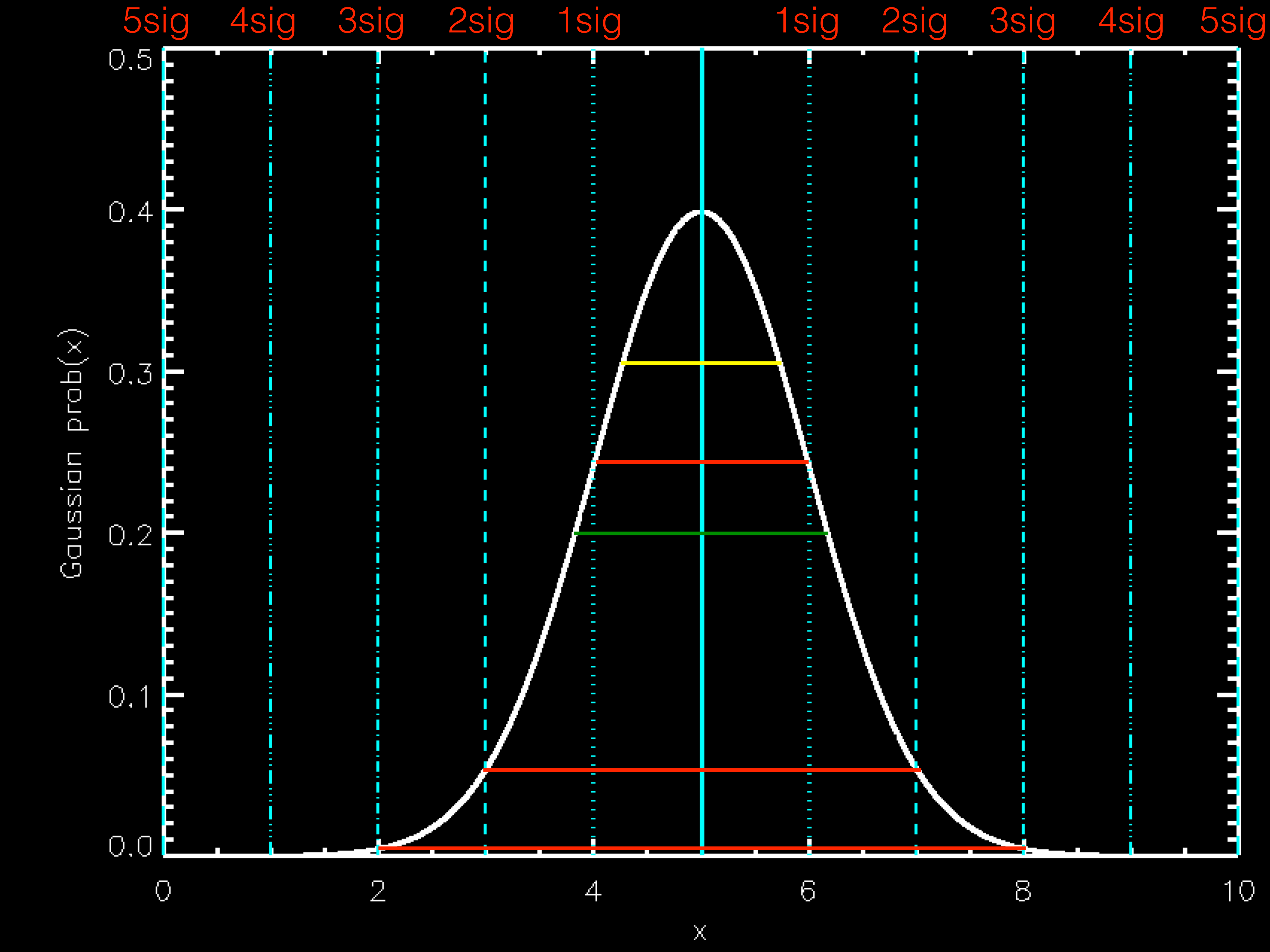
The probability that the value of a random observation will fall within the interval dx around x is prob(x)dx $\Rightarrow$ $$\int_{-\infty}^{\infty} \text{prob}(x)\,dx = 1$$

At $x = \mu \pm \sigma$, the height of the curve is reduced to $e^{-1/2}$ of its value at the peak, ie: $$\text{prob}(\mu \pm \sigma) = e^{-1/2}\,\text{prob}(\mu)$$

The Gaussian distribution displays the characteristic bell shape and symmetry about the mean.

Full width @ half maximum Γ defined as
$$\text{prob}\left(\mu \pm \tfrac{1}{2}\Gamma\right) = \tfrac{1}{2}\,\text{prob}(\mu)$$

$\Rightarrow \Gamma = 2.354\,\sigma$



5sig
4sig
3sig
2sig
1sig
1sig
2sig
3sig
4sig
5sig
0.0
0.1
0.2
0.3
0.4
0.5
Gaussian prob(x)
0
2
4
6
8
10
x

Using the dimensionless variable $z = \frac{x-\mu}{\sigma}$

$\Rightarrow$ $\text{prob}(z)\,dz = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$ standard gaussian distribution

Probability that a measurement will deviate from the mean by a specified amount Δx

$$A(\Delta x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-\Delta x}^{\mu+\Delta x} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$\hookrightarrow$ probability that any random value of x will deviate from the mean by LESS THAN $\pm\Delta x$

$\Rightarrow$ $1 - A(\Delta x)$ is the probability that a measurement will deviate from the mean by MORE THAN Δx

$$A(\Delta x) \longrightarrow \underset{\Delta z = \frac{\Delta x}{\sigma}}{A(\Delta z)} = \frac{1}{\sqrt{2\pi}} \int_{-\Delta z}^{\Delta z} e^{-z^2/2} dz$$

deviations in units of the standard deviation σ

$< 1\sigma \longrightarrow$ probability $= 0.68268$

$< 2\sigma \qquad = 0{,}9545$

$< 3\sigma \qquad = 0{,}9973$

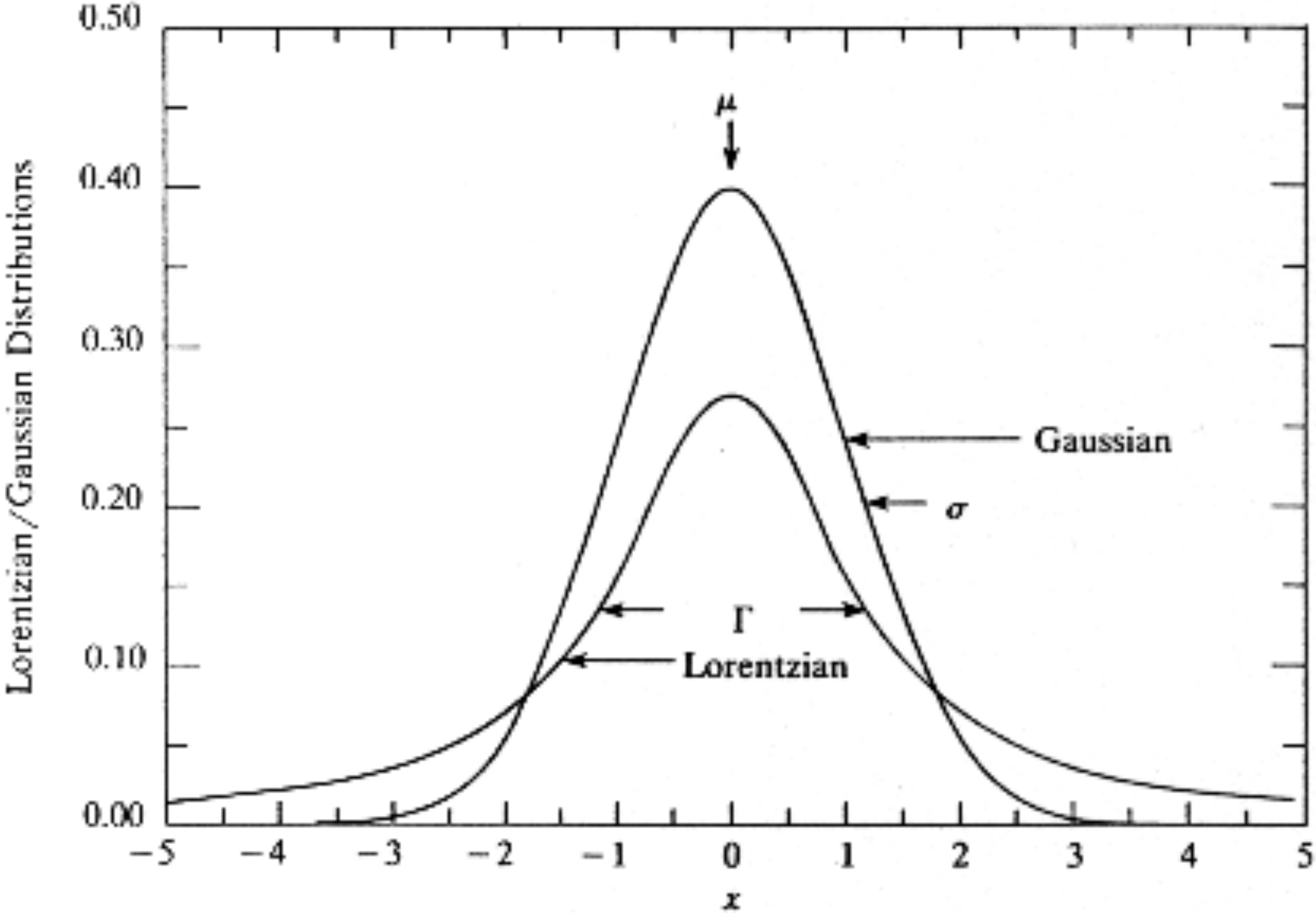
$< 4\sigma \qquad = 0{,}999937$

$< 5\sigma \qquad = 0{,}9999994$

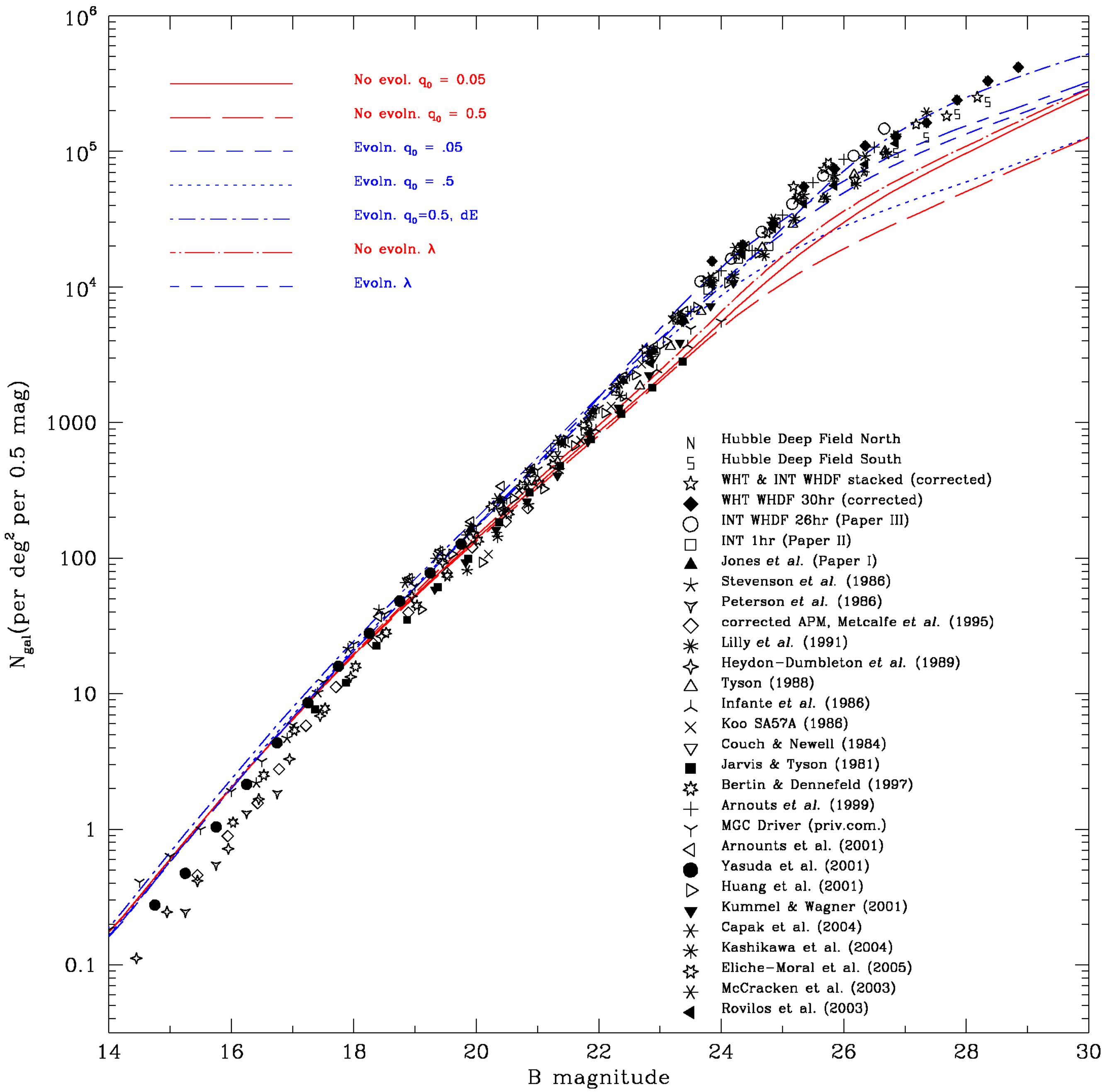
$< \sigma_{pe} = 0.67449\sigma \qquad = 0.50$

└ probable error by definition

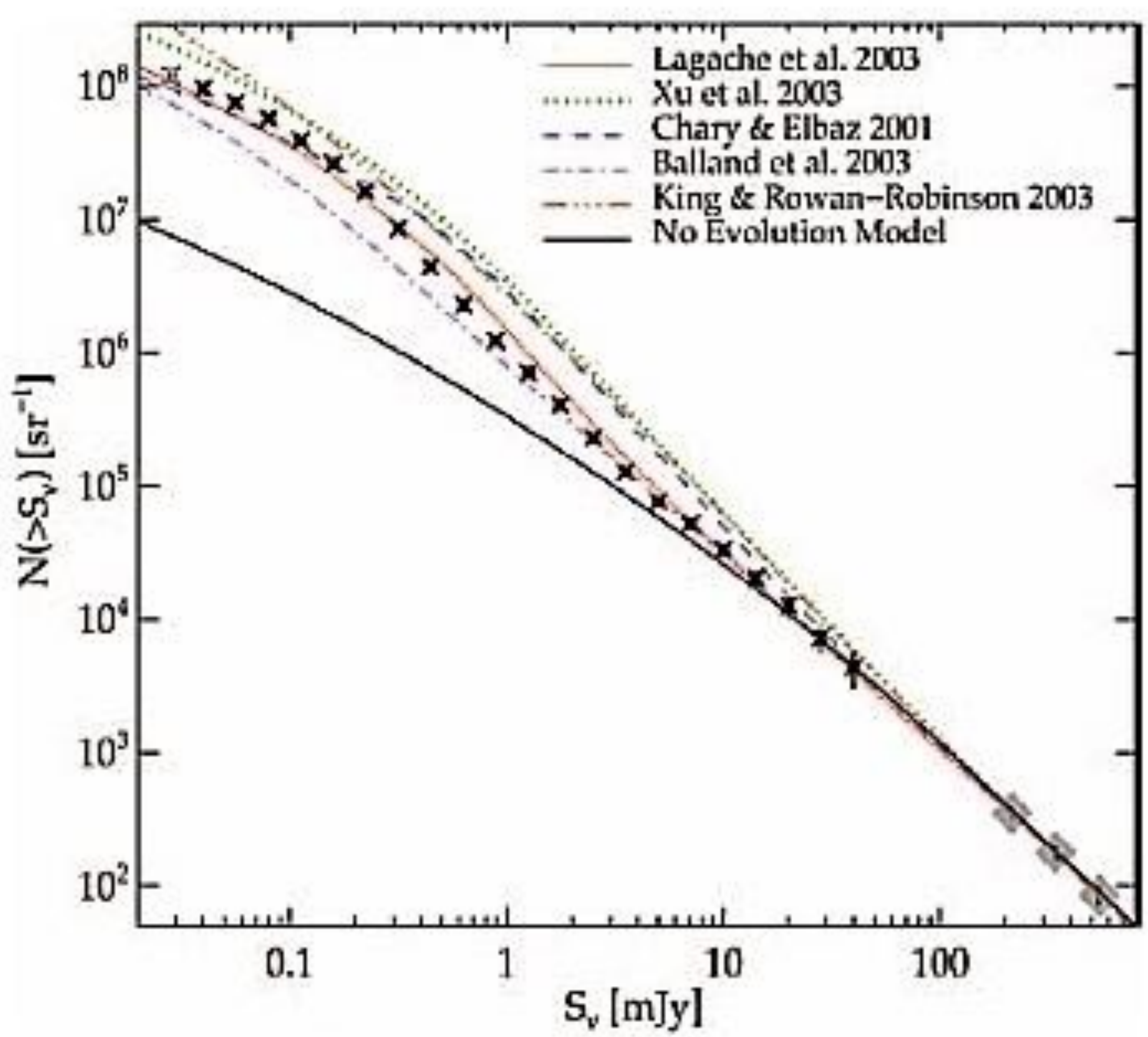
Lorentzian Distribution


Lorentzian/Gaussian Distributions
0.50
0.40
0.30
0.20
0.10
0.00
μ
Gaussian
σ
Γ
Lorentzian
−5
−4
−3
−2
−1
0
1
2
3
4
5
x

Power Law distribution



No evol. q_0 = 0.05
No evoln. q_0 = 0.5
Evoln. q_0 = .05
Evoln. q_0 = .5
Evoln. q_0=0.5, dE
No evoln. λ
Evoln. λ
Hubble Deep Field North
Hubble Deep Field South
WHT & INT WHDF stacked (corrected)
WHT WHDF 30hr (corrected)
INT WHDF 26hr (Paper III)
INT 1hr (Paper II)
Jones et al. (Paper I)
Stevenson et al. (1986)
Peterson et al. (1986)
corrected APM, Metcalfe et al. (1995)
Lilly et al. (1991)
Heydon-Dumbleton et al. (1989)
Tyson (1988)
Infante et al. (1986)
Koo SA57A (1986)
Couch & Newell (1984)
Jarvis & Tyson (1981)
Bertin & Dennefeld (1997)
Arnouts et al. (1999)
MGC Driver (priv.com.)
Arnounts et al. (2001)
Yasuda et al. (2001)
Huang et al. (2001)
Kummel & Wagner (2001)
Capak et al. (2004)
Kashikawa et al. (2004)
Eliche-Moral et al. (2005)
McCracken et al. (2003)
Rovilos et al. (2003)
N_{gal}(per deg^2 per 0.5 mag)
B magnitude



Lagache et al. 2003
Xu et al. 2003
Chary & Elbaz 2001
Balland et al. 2003
King & Rowan-Robinson 2003
No Evolution Model
$N(>S_\nu)$ [sr^{-1}]
S_ν [mJy]