

TA Seminar Notes

Astronomical Observation and Lab at Seoul National University

Yoonsoo P. Bach

Contents

1	Introduction	5
2	Statistics	7
2.1	The $n\text{-}\sigma$	7
2.2	The Meaning of \pm Sign	8
2.3	Caveat on the “ $n\text{-}\sigma$ CI” Notation	9
2.4	Central Limit Theorem (CLT)	10
2.5	Meaning of the Confidence Interval	13
2.6	Answer to the Question	15
2.7	Poisson Noise	16
3	Idea of Photometry	19
3.1	The Point-Spread Function (PSF)	19
3.1.1	Diffraction	19
3.1.2	Seeing	20
3.1.3	Summary of PSF	20
3.1.4	Seeing Disc Size	21
3.2	Centroid	23
3.3	Aperture Sum	23
3.4	Sky Estimation	24
3.4.1	Simple Parametric Sky Estimation	24
3.4.2	Other Ways for Sky Estimation	26
3.4.3	Annulus Sky	26
3.4.4	Source Extractor Sky	27

Chapter 1

Introduction

In math, people prefer to use \log_e or just \log rather than \ln . But due to its simplicity (especially when typing), I will stick to the usage of \ln which is identical to \log_e . But to avoid any confusion, I will always specify the base 10 (\log_{10}).

In astronomy, people use \equiv as “is defined as” or any kind of “trivial by the definition” cases. But I will use $:=$ for the definition and use \equiv to mean “equality holds trivially by the definition”. For example, under 1-D linear homogeneous isotropic matter, linear electric polarization is written as $P \equiv \epsilon_0 \chi E$ or $\chi := P/\epsilon_0 E$. The latter is the *definition* of electric susceptibility, while the former is *trivially true by the definition*.

In this note, I used some math-like notations like theorem (Thm). However, the theorem here is not necessarily the same as that of mathematics. I tried to put some important assumptions and summary notes to theorem

The reference “Walpole” is Walpole et al. 2013, Essentials of Probabilities and Statistics.

Chapter 2

Statistics

Let's start with a question.

Ex 1

For the **same star**, consider the following three scenarios:

- Case A: Researcher 1 says $m_1 = 14.0^m \pm 0.1^m$; Researcher 2 says $m_2 = 15.0^m \pm 1.0^m$.
- Case B: Researcher 1 says $m_1 = 14.0^m \pm 0.1^m$; Researcher 2 says $m_2 = 15.0^m \pm 0.3^m$.
- Case C: Researcher 1 says $m_1 = 14.0^m \pm 0.1^m$; Researcher 2 says $m_2 = 15.0^m \pm 0.1^m$.

For each of the three cases: Are the two studies coincide? If so, with how much confidence would you say so? Wait, but what does that “ \pm ” sign means in rigorous mathematical sense?

Some of the previous course takers answered that they coincide, because the “3-sigma rule” says, e.g., for case A, $m_1 = 14.0^m \pm 0.3^m$ & $m_2 = 15.0^m \pm 3.0^m$ overlaps with each other. Crudely speaking, this makes sense, but it is of course not the “publication level” reasoning.

To give you the answer: It is an ill-defined question. The answer can change based on the number of observations each researcher used for the determination of the magnitude. If we assume that number is 5 for both researchers, for example, we can give some meaningful answers: We *cannot reject* $m_1 = m_2$ for case A & B with 90 % confidence, and we *can reject* $m_1 = m_2$ for case C with 90 % confidence. Depending on the confidence level, the answer may change. Also the expression that *can/cannot reject* is very important! You shouldn't say you *accept* $m_1 = m_2$. Let me explain why throughout this chapter.

2.1 The n - σ

Maybe you are familiar of the 1- σ , 2- σ , and 3- σ words (cf. fig. 2.1). For example, mean ± 1 - σ contains 68.27...% of the total area. Similarly 3- σ contains 99.73...% of it, so it is very unlikely to get a sample outside of the mean ± 3 - σ .

In this sense, the “ n - σ ” is defined as the x -axis value which gives the integrated area of certain value. More precisely, we have to call this confidence interval (CI):

Def 1 (n - σ Confidence Interval; CI)

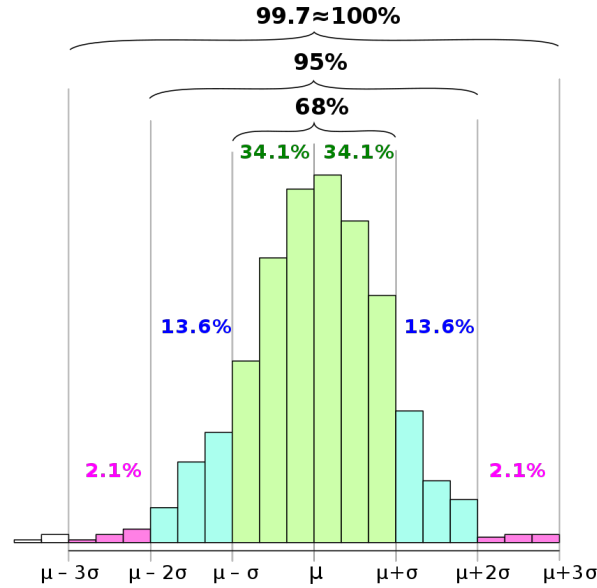


Figure 2.1: A test sampling from Gaussian (normal) distribution (from Wikimedia).

1- σ CI is the integrated area is $0.6827\dots$ of the total area. Similarly, 2- σ CI is that with $0.9545\dots$, 3- σ CI is that with $0.9973\dots$, etc.

Mathematically it is more correct to call, e.g., 1- σ CI as the $68.27\dots\%$ CI. We also define the significance level α as 100% minus this percentage, i.e., the 1- σ CI is of significance level $\alpha = 1 - 0.6827\dots = 0.3173\dots$:

$$1\text{-}\sigma \text{ CI} \equiv 68.27\dots\% \text{ CI} \equiv \text{CI of significance level } 0.3173\dots \quad (2.1)$$

In natural sciences we often use $n\text{-}\sigma$ CI, but in mathematics and other branches of sciences and engineering, the 90 %, 95 %, and 99 % CIs, i.e., CIs with significance level $\alpha = 0.10, 0.05$, and 0.01 are used more often.

Because when we say “ $n\text{-}\sigma$ ”, we are just omitting the term “CI” at the end of it, this is not n times the standard deviation (σ): What we mean by “ $n\text{-}\sigma$ ” is actually “ $n\text{-}\sigma$ CI” and is *not necessarily* $n \times \sigma$.

They are the same for, e.g., Gaussian distribution. For Gaussian (normal) distribution, which is a distribution that appears very often in natural sciences, such a 2- σ is nothing but 2 times the standard deviation (σ). For almost all of the distributions other than Gaussian, this is not the case.

2.2 The Meaning of \pm Sign

Simple answer: The \pm sign means the 1- σ confidence interval (CI). It is also simply called as the “error-bar”.

Honestly speaking, the \pm sign, i.e., the 1- σ CI, does *not* fully describe the uncertainty. The best

way to show or describe the uncertainty is to show a graph of probability distribution function (pdf or p.d.f.) as in fig. 2.2. In many but not all publications or books, people do not show this, because (1) the distribution is very similar to Gaussian or certain widely known distribution and it is clearly written in the text or assumed as all the readers know that, (2) detailed calculation is not of interest and/or does not affect the final result, (3) the writer is lazy.

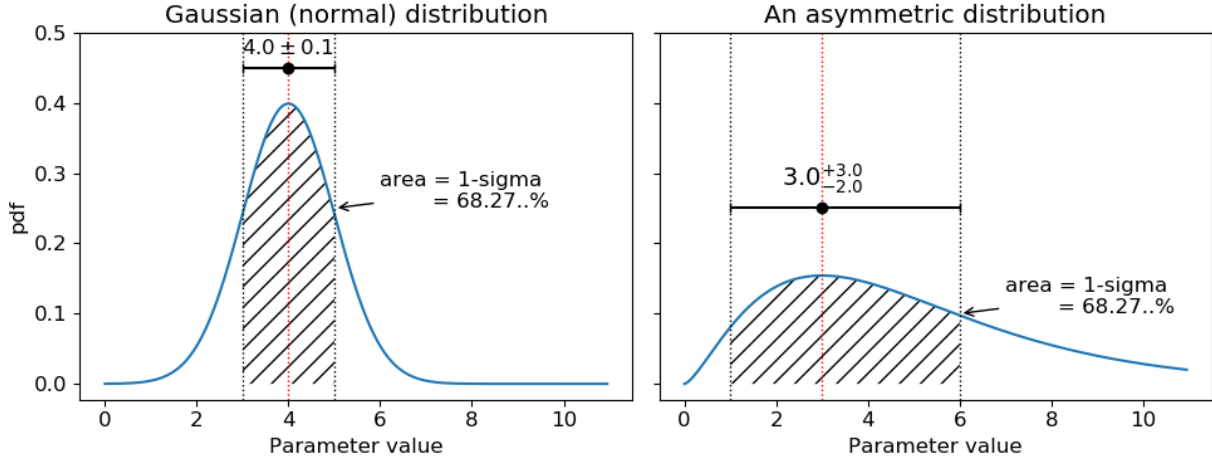


Figure 2.2: The probability distribution function (pdf) of two examples of distributions, namely, Gaussian and an asymmetric distributions. I used a chi-square distribution for the latter for plotting purpose.

The uncertainty ranges shown in the figure (4.0 ± 0.1 and $3.0^{+3.0}_{-2.0}$) are determined such that the lower/upper bounds include $68.27 \dots \%$ of the total area. It is trivial for a Gaussian distribution: mean $\pm 1\sigma$. But if the distribution is asymmetric (right panel of the figure), there are few choices to set such bounds. First and the most widely used one is to set the bounds such that the distance between upper/lower bounds are minimized (but include $68.27 \dots \%$ of the total area). Second is to make it symmetric while include $68.27 \dots \%$ of the total area, so that you can use \pm sign for simplicity ($3.0 \pm 2.XX$ for example). Third choice would be something like FWHM: find x -axis values such that the pdf value is $0.5 \times \text{pdf}_{\text{max}}$. The last two bounds are simple to calculate but not as accurate as the first choice.

Very assymmetric probability distributions appear in cosmological sciences* and exhaustive statistical analyses are conducted on such models to reject cosmological models. The reason is that, in cosmology, we have only one single sample (our universe), and the statistical analyses given the observational data is of utmost importance. Of course similar exhaustive statistical analyses should be conducted on any research field if the data/object/target is of such importance.

2.3 Caveat on the “ n - σ CI” Notation

Let me emphasize again: What we mean by “ n - σ ” is actually “ n - σ CI” and is *not necessarily* $n \times \sigma$.

It maybe is tempting to convert the 1- σ CI to the 3- σ CI. For instance, from fig. 2.2, 4.0 ± 0.1

*Google “posterior distribution cosmological constants”.

and $3.0^{+3.0}_{-2.0}$ to 4.0 ± 0.3 and $3.0^{+9.0}_{-6.0}$. This is true for the former (Gaussian) but wrong for the latter (non-Gaussian). The lower bound of the second parameter now become negative ($3.0 - 6.0 = -3.0$), which is not even physically correct if this parameter is defined to be positive. What you have to do is, get the 3- σ CI which should contain 0.9973... of the total area from the distribution given in fig. 2.2. There can be at least three choices to do it, which I mentioned in the previous section.

2.4 Central Limit Theorem (CLT)

The Central Limit Theorem (CLT) is at the heart of all the observational or experimental sciences. It is stated as*

Thm 1 (*Central Limit Theorem; CLT*)

Consider a random sample (observation, measurement, etc) with size n and the mean value of this sample is \bar{X} . Then $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ approaches the standard normal distribution (\mathcal{Z}) as $n \rightarrow \infty$, where μ and σ^2 are the (finite) mean and variance of the population:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{Z} \quad (\text{as } n \rightarrow \infty) . \quad (2.2)$$

There is one more important theorem, which we usually skip to mention:

Thm 2 (*Sample Variance Distribution*)

Consider a random sample (observation, measurement, etc) with size n and the sample variance is S^2 . Then $(n-1)S^2/\sigma^2$ follows a chi-squared distribution with degrees of freedom $(n-1)$, where σ^2 is the variance of the population:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)} \quad (2.3)$$

Also the definition of the Student t -distribution[†]:

Def 2 (*Student t -distribution*)

Let Z be a standard normal random variable and V a chi-squared random variable with v degrees of freedom. If Z and V are independent, then the distribution of the random variable T , where

$$T = \frac{Z}{\sqrt{V/v}} \quad (2.4)$$

is given by the density function

$$h(t) = \frac{\Gamma[\frac{v+1}{2}]}{\Gamma[\frac{v}{2}]\sqrt{\pi v}} \left(1 + \frac{t^2}{v}\right)^{-\frac{(v+1)}{2}}, \quad -\infty < t < \infty . \quad (2.5)$$

This is known as the Student t -distribution with v degrees of freedom.

*For strict definitions of random sample, population, independence, the statement of CLT in mathematical senses, etc, I recommend mathematical statistics textbooks.

[†]This definition is directly copied from Walpole et al. p.177

We can see that the t -distribution approaches the standard normal distribution as the degrees of freedom $v \rightarrow \infty$, because the power term takes the $e^{-t^2/2}$ form.

Thm 3 (Practical Usage of the CLT)

Consider a random sample (observation, measurement, etc) with size n large enough and the mean value of this sample is \bar{X} . If μ , σ^2 , and S^2 are the (finite) mean and variance of the population, and the sample variance, respectively:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim T_{(n-1)} . \quad (2.6)$$

Sometimes people empirically say $n \geq 30$ is enough. This really depends on the situation and I will skip this issue here. The sketch of the proof of the theorem is simple. By the definition of the t -distribution and Thm 2,

$$T = \frac{Z}{\sqrt{V/v}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \quad (2.7)$$

so the theorem is plausible. For your information, the sample variance is defined as

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (2.8)$$

and the sample standard deviation S is the square root of this.

Note here the difference between Thm 1 and Thm 3: σ is changed to S , and the standard normal distribution is changed to the t -distribution of $(n-1)$ degrees of freedom. Since the true variance σ^2 is unknown, it is difficult to use the CLT (Thm 1) directly. But thanks to Thm 2, we can use the variance of the sample, S^2 , which is measurable, and utilize the t -distribution, which is slightly bothersome than the standard normal distribution but still useful.

Also be aware that CLT says the *expectation value of the mean* is normally distributed, not the *sample* is so.

Ex 2 (Coin Tossing and the CLT)

Consider a coin-tossing experiment and assign ± 1 to heads and tails. Each experiment will give you ± 1 , but never 0. Meanwhile, we know a fair coin should have $\mu = 0$ (same probability of heads and tails). After $n = 100$ experiments, say you obtained $\bar{X} = 0.01$ and $S = 0.1$, so

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{0.01 - \mu}{0.1/\sqrt{100}} = 1 - 100\mu .$$

Since $T \sim t_{99}$, the significance level $\alpha = 0.05$ confidence interval, i.e., the confidence interval containing $100(1 - \alpha)\% = 95\%$, can be

$$t_{99,0.025} < T < t_{99,0.975} \quad \rightarrow \quad -1.9842 < 1 - 100\mu < 1.9842 \quad \rightarrow \quad \mu \in [-0.0098, 0.0298] .$$

The notation $t_{\nu,x}$ means the input argument of the t -distribution with the degrees of freedom ν such that $\int_{-\infty}^{t_{99,x}} h(t)dt = x$. Thus, to get the two-tail CI of significance level α , we need to calculate $t_{\nu,\alpha/2}$ and $t_{\nu,1-\alpha/2}$. For a symmetric distribution like t here, $t_{\nu,1-\alpha/2} = -t_{\nu,\alpha/2}$. This is very widely used standard notation.

If you just calculated without pondering about the meaning of the calculation, you may be surprised that your next experiment gives either $+1$ or -1 , while the expectation is $\mu \in [-0.0098, 0.0298]$. This is because the result from CLT is, as described, about the *mean* of the samples, not the *single sample*. Thus, the error-bar from the CLT is not necessarily predicting the possible range of *future experiments*, but it just confines the *position of the mean*.

The range which predicts the future experiments is called the prediction interval. You may learn the prediction interval (PI) to clarify this difference.

The two-tail significance level α interval of the t -distribution calculable in python by

```
from scipy.stats import t
alpha = 0.95
nu = 99

# Note that the alpha below is NOT the significance level,
# but the area under the curve!
lo, hi = t.interval(alpha=alpha, df=nu)
print(f"Confidence Interval: [{lo:.4f}, {hi:.4f}]")
```

Note: Uncertainty of Median

The $1\text{-}\sigma$ error-bar of the median is more difficult to handle than that of the mean (CLT: Thm 1). But we can get a result from simplifying assumptions, which are not necessarily true for real observations:

Thm 4 (*Uncertainty of Median*)

Consider $n(\gg 1)$ samples are independently drawn from $X \sim \mathcal{G}(\mu, \sigma^2)$, a general distribution with finite mean and variance μ and σ^2 . The uncertainty of the median estimator is

$$\Delta_{\text{med}} \approx \sqrt{\frac{\pi/2}{n-1}} s = \sqrt{\frac{\pi}{2} \frac{n}{n-1}} \Delta_{\text{mean}} \approx 1.25 \Delta_{\text{mean}}, \quad (2.9)$$

where s is the sample standard deviation and $\Delta_{\text{mean}} = s/\sqrt{n}$ is the error-bar of the mean from Thm 3.

To prove it, say the true median is ν_0 and samples $\{x_1, \dots, x_n\}$ are sorted as increasing order. For an integer m , set $n = 2m + 1$, because $n \gg 1$. The sample median is then $\nu = x_{m+1}$.

The probability of having sample median ν is calculable by noting that we have to sample m of sample with $x_i \leq \nu$ and the other m with $x_i > \nu$:

$$f(\nu) = \frac{n!}{m!m!} q^m (1-q)^m,$$

where $q = q(\nu) = \mathbb{P}\{X_i \leq \nu\} = \int_{-\infty}^{\nu} p(x) dx$. $p(x)$ is the pdf of the normal distribution.

The Taylor expansion of q at $\nu = \nu_0$:

$$\begin{aligned} q(\nu) &= q(\nu_0) + \frac{1}{1!} q'(\nu_0)(\nu - \nu_0) + O((\nu - \nu_0)^2) \\ &\approx \frac{1}{2} + q'(\nu_0)(\nu - \nu_0) \\ &\equiv \frac{1}{2} + p(\nu_0)(\nu - \nu_0). \end{aligned}$$

Then

$$\begin{aligned}
 f(\nu) &\approx \frac{n!}{m!m!} \left[\frac{1}{2} + p(\nu_0)(\nu - \nu_0) \right]^m \left[\frac{1}{2} - p(\nu_0)(\nu - \nu_0) \right]^m \\
 &= \frac{n!}{m!m!} \left[\frac{1}{4} (1 - 4p(\nu_0)^2(\nu - \nu_0)^2) \right]^m \\
 &= \frac{n!}{m!m!4^m} \left[1 - \frac{4mp(\nu_0)^2(\nu - \nu_0)^2}{m} \right]^m \\
 &\approx \frac{n!}{m!m!4^m} e^{-4mp(\nu_0)^2(\nu - \nu_0)^2} .
 \end{aligned}$$

This has the form of Gaussian distribution ($\propto e^{-(x-\mu)^2/2\sigma^2}$) with mean ν_0 and variance $\frac{1}{4(n-1)p(\nu_0)^2}$ since $m = (n-1)/2$. Thus,

$$\nu \sim \mathcal{N}\left(\nu_0, \frac{1}{4(n-1)p(\nu_0)^2}\right) \rightarrow \Delta_{\text{med}} \approx \frac{1}{2p(\nu_0)\sqrt{n-1}} . \quad (2.10)$$

Now if $p(x)$ is a Gaussian, median is equal to mean, so the pdf becomes $p(\nu_0) = p(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$, and

$$\nu \sim \mathcal{N}\left(\nu_0, \frac{\pi/2}{n-1}\sigma^2\right) .$$

If we denote $A = \sqrt{(n-1)/(\pi/2)}$, the statistic $Z = A\frac{\nu - \nu_0}{\sigma}$ will follow the standard normal distribution. On the other hand, $V = \frac{(n-1)s^2}{\sigma^2}$ follows the chi-squared distribution of degrees of freedom $(n-1)$ by Thm 2. Then

$$T = \frac{Z}{\sqrt{V/(n-1)}} = A\frac{\nu - \nu_0}{s}$$

will follow a t -distribution with degrees of freedom $(n-1)$ by Def 2. T is nearly a Gaussian when n is large (rule-of-thumb: when $n \gtrsim 30$), so

$$\nu \sim \mathcal{N}\left(\nu_0, \frac{\pi/2}{n-1}s^2\right) \rightarrow \Delta_{\text{med}} \approx \sqrt{\frac{\pi/2}{n-1}}s . \quad (2.11)$$

Q.E.D.

2.5 Meaning of the Confidence Interval

I circumvented the definition of the CI so far. It is defined as

Def 3 (Confidence Interval; CI)

A confidence interval of significance level α of a parameter X is the interval of X value such that if we conduct the identical parameter estimation process in many *parallel universes* (i.e., ensemble) which have the identical true value X_{true} , fraction of $(1 - \alpha)$ of such universes could have calculated a CI such that the X_{true} is included in that CI.

This is not a practically meaningful definition, but philosophically important. Let me give an example to elaborate the meaning of this.

Ex 3 (Meaning of the Confidence Interval)

Imagine there are many *parallel universes* and the same observer observes the same star in each universe. Because of random errors (which we will learn later but that includes Poisson photon noise, readout noise, etc), the observer at each universe will obtain slightly different observational results. The observers will observe the star for n times, and calculate their own mean and error-bar as \bar{X} and S/\sqrt{n} from their observations. Note again that S/\sqrt{n} is the uncertainty of the *mean* value.

For visualization, the true magnitude of the star is $\mu = 14.0^{\text{m}}$ and true standard deviation of the observation $\sigma = 0.2^{\text{m}}$ with $n = 9$ are used for the generation of fig. 2.3. The universe ID 0 of both the upper and lower panels show roughly $m = 14.0^{\text{m}} \pm 0.1^{\text{m}}$. That is, when the true magnitude and true standard deviation are given as 14.0^{m} and 0.2^{m} , an observer at one of the universes (in our case universe ID 0) got $m = 14.0^{\text{m}} \pm 0.1^{\text{m}}$, and that is our universe.

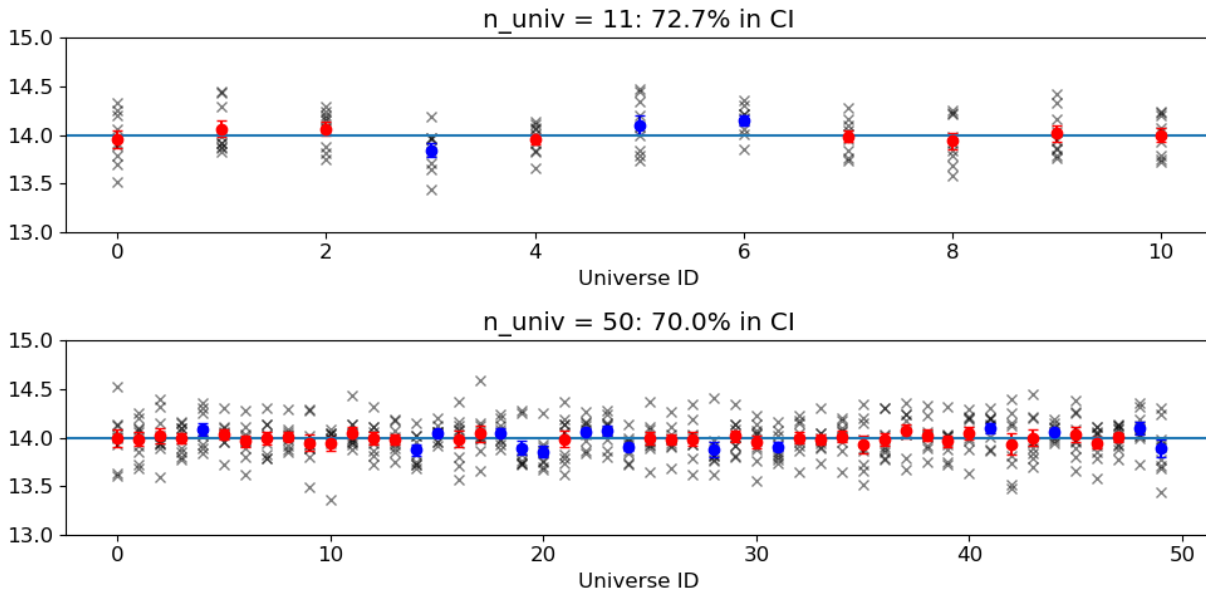


Figure 2.3: Simulation to show the concept of CI and CLT. Black crosses are the observation in each universe, and the circles are the mean and its uncertainty from the CLT S/\sqrt{n} . Red circles are the universes where the confidence interval contains the true mean $\mu = 14.0^{\text{m}}$ and blue circles are the others. In the title, the fraction of the universes which contains the true mean within their own confidence intervals out of n_{univ} simulated universes is shown. This fraction approaches $0.6827 \dots$ as the number of universe gets infinity.

I will close this section with an excerpt from Walpole p.234:

The interpretation of a CI is often misunderstood. It is tempting to conclude that the parameter falls inside the CI with probability of(, e.g.,) 0.95. A CI merely suggests that if the experiment is conducted and data are observed again and again, about 95% of such intervals will contain the true parameter.

2.6 Answer to the Question

In statistics, the *null hypothesis*, denoted H_0 , is the one that is the simplest, and that has value if it is *rejected* (not *accepted*). The *alternative hypothesis*, H_1 , is another possibility if H_0 is not true. In the simplest case, H_1 is the complementary of H_0 . For example, $H_0 : \mu = 0$ and $H_1 : \mu \neq 0$.

Let's consider the opening question: $m = 14.0^m \pm 0.1^m$. Now you understand that 14.0^m means that the mean value* from the observation, and $\pm 0.1^m$ means the uncertainty or the error-bar of the Gaussian probability distribution of that mean value, S/\sqrt{n} . If it were not Gaussian, the author should have given more information. Note that the sample standard deviation is \sqrt{n} times the error-bar.

To answer the question, we set the hypotheses†:

$$H_0 : m_1 = m_2 \quad ; \quad H_1 : m_1 \neq m_2 , \quad (2.12)$$

where $m_1 = 14.0^m \pm 0.1^m$ and m_2 is the other (from researcher 2). m_1 and m_2 here means the “true” magnitude of the star, not the measured value! If it were the measured value, they are just different, since one is 14.0 and the other is 15.0. The two true means may be different because (1) researchers may have mistakenly observed different star, (2) the star is maybe a variable (including binary), (3) exoplanet occulted some part of the star, or any other scenario is possible.

Now we have to choose which formula to use. When two measurements with $\bar{x}_1 \pm s_1$ and $\bar{x}_2 \pm s_2$ from the number of observations n_1 and n_2 are given, and μ_1 and μ_2 are the true means of two sampling distributions that will be tested under the hypothesis testing, there are two different formulae:

- The true variances are unknown and different:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \quad (2.13)$$

follows the t -distribution with degrees of freedom

$$\nu \approx \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}} . \quad (2.14)$$

- The true variances are unknown but the same:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{1/n_1 + 1/n_2}} \quad (2.15)$$

follows the t -distribution with degrees of freedom $\nu = n_1 + n_2 - 2$ and

$$s_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{\nu} . \quad (2.16)$$

*It can actually be the median value, most probable value, or whatever the representative value.

†There are some “standard” ways to set the hypotheses, and it may seem very sudden if you are not familiar with these. Please refer to basic applied statistics textbooks for more examples, e.g., p.246 and p.264 of Walpole et al.

Which one will you choose? To answer our original question, it is more reasonable to assume the true variances are not necessarily identical, because they may have used different instruments at different sky conditions. So I will use the first choice.

To apply the formula, you need the number of observations from each researcher, i.e., the n_1 and n_2 values. As I mentioned, it is an ill-defined question since these numbers are not given. So let me assume $n_1 = n_2 = n = 5$. The question can be reformulated as $\bar{x}_1 = 14.0$, $\bar{x}_2 = 15.0$, $s_1 = 0.1\sqrt{n}$, and $s_2 = 1.0\sqrt{n}$, $0.3\sqrt{n}$, $0.1\sqrt{n}$, where I dropped the magnitude sign for brevity. Under the null hypothesis ($H_0 : \mu_1 = \mu_2$), $\mu_1 - \mu_2 = 0$. For the three s_2 values,

$$\begin{aligned} t &= \frac{(14.0 - 15.0) - 0}{\sqrt{(s_1^2 + s_2^2)/n}} & \rightarrow & \quad t = \quad -0.995, \quad -3.162, \quad -7.07 \\ \nu &= (s_1^2 + s_2^2)(n - 1) & \rightarrow & \quad \nu = \quad 4.08, \quad 4.32, \quad 8.0 \\ & & & \approx \quad 4, \quad 4, \quad 8 \end{aligned} \tag{2.17}$$

The values are in the order for $s_2 = (1.0, 0.3, 0.1)\sqrt{n}$ cases. The (two-tail) significance level α test is listed in table 2.1.

Table 2.1: The confidence interval calculation for the example question.

CI %	significance level α and $\frac{\alpha}{2}$	$\nu = 4$ case	$\nu = 8$ case
90 %	0.10 & 0.05	$[-2.1318, 2.1318]$	$[-1.8595, 1.8595]$
95 %	0.05 & 0.025	$[-2.7764, 2.7764]$	$[-2.3060, 2.3060]$
99 %	0.01 & 0.005	$[-4.6041, 4.6041]$	$[-3.3554, 3.3554]$

The answers to the questions are, assuming $n_1 = n_2 = 5$,

- Case A ($t = -0.995$ with $\nu = 4$): The t value is inside of all three CIs. Thus, we cannot reject H_0 under 90 %, 95 %, and 99 % CI criterion (cannot reject under the significance level $\alpha = 0.10, 0.05, 0.01$).
- Case B ($t = -3.162$ with $\nu = 4$): The t value is inside of 90 % and 95 % CI but outside of the 99 % CI. Thus, we reject H_0 with $\alpha = 0.10$ and 0.05 , but cannot reject it with $\alpha = 0.01$.
- Case C ($t = -7.07$ with $\nu = 8$): We reject H_0 with all the three $\alpha = 0.10, 0.05, 0.01$.

By *rejecting the null hypothesis*, $H_0 : m_1 = m_2$, we mean that it is unlikely that the true magnitude value of the two studies are identical. By *failing in rejecting the null hypothesis*, we mean that it is impossible to conclude whether m_1 and m_2 are the same under the given significance level.

The results are as expected: For Case A, the error-bar of m_2 is too large, which means it is almost impossible to reject the claim that $m_1 = m_2$, so we always fail to reject the null hypothesis. For Case C, the error-bars are too small and thus m_1 is of course vastly different from m_2 , so the null hypothesis is always rejected.

2.7 Poisson Noise

The Poisson process is a fancy naming for some special “counting”. In astronomy, we count photons (well, actually the CCD counts the electrons) over time, in social sciences we can count the death

toll over time, in experiments we could count the lattice of a randomly shaped crystal over the distance. A formal definition is*

Def 4 (*Poisson Process 1*)

The Poisson process $N(t)$ for $t \geq 0$ (t can be time, distance, or similar things), with rate λ is defined by

1. $N(0) = 0$
2. $N(t)$ has independent increment
3. $N(t_2) - N(t_1)$ follows Poisson distribution of rate $\lambda(t_2 - t_1)$ for $t_1 < t_2$.

It can be proven that it is identical to

Def 5 (*Poisson Process 2*)

The Poisson process $N(t)$ for $t \geq 0$ (t can be time, distance, or similar things), with rate λ is defined by

1. $N(0) = 0$
2. $N(t)$ has independent increment
3. $\mathbb{P}(N(h) = 1) = \lambda h + o(h)$ for small h .
4. $\mathbb{P}(N(h) \geq 2) = o(h)$ for small h .

Here $o(h)$ is any function such that $\lim_{h \rightarrow 0} o(h)/h = 0$. The last two items in this second definition can be re-phrased like this:

3. For a very small interval h , the probability of one single “counting” (Poisson process) occurs in that interval is proportional to the length of h .
4. For a very small interval h , the probability of two or more “counting” (Poisson process) occurs in that interval is nearly 0.

These must be independent of whether the counting happened outside of the interval, because the second item of the definition states the increment is independent.

Ex 4 (*Photon Counting and Poisson Process*)

The photon counting process of astronomy resembles the Poisson Process. If $N(t)$ is the number of photons or electrons under the exposure time t , it is trivial that the first two items of the definition is satisfied. Next, use the rephrased version of the next items. The number of photons hitting CCD with infinitesimal time interval $h = dt$ must be 0 or 1, but not larger than 1 (e.g., if a pixel count is 10,000 after 10 sec of exposure, we can take $h \ll 10 \text{ s}/10,000 = 1 \text{ ms}$ to meet this condition). Thus, photon counting is a Poisson process.

The Poisson process follow the Poisson distribution:

Def 6 (*Poisson Distribution*)

The Poisson process of rate λ has the following pdf

$$p(x; \lambda t) = \frac{(\lambda t)^x}{x!} e^{-\lambda t} \quad (2.18)$$

for non-negative integer x .

*See, e.g., http://dept.stat.lsa.umich.edu/~ionides/620/notes/poisson_processes.pdf and https://www.probabilitycourse.com/chapter11/11_1_2_basic_concepts_of_the_poisson_process.php

Thm 5 (Poisson Distribution Mean and Variance)

The mean, variance, and the standard deviation of the Poisson distribution $p(x; \lambda t)$ are

$$\text{mean} = \lambda t \quad ; \quad \text{var} = \lambda t \quad ; \quad \text{std} = \sqrt{\lambda t} . \quad (2.19)$$

The reason we use λt not just λ is clear from Thm 5: the mean changes as t changes, and we want to express $p(x; \text{mean})$ rather than $p(x; \text{mean}/t)$. For instance, if the photon influx is $\lambda = 10$ photons/s, the mean of photon count is a function of time as λt , so for $t = 10$ s, the pdf is $p(x; 100 \text{ photons})$. If the particles are distributed with mean density along the line of sight $\lambda = 0.1$ particle/m, the mean of particles along the line of sight is a function of the reaching distance λd , so the pdf for $d = 5$ m is $p(x; 0.5 \text{ particles})$.

Thm 6 (Poisson to Gaussian)

As $x \rightarrow \infty$, the Poisson distribution approaches normal distribution $\mathcal{N}(\lambda t, \lambda t)$.

Although I will omit here, the proof uses Stirling's formula $x! \approx \sqrt{2\pi x} x^x e^{-x}$ and $\ln[(1+\varepsilon)^{\lambda t(1+\varepsilon)+1/2}] \approx \lambda t \varepsilon + \lambda t \varepsilon^2/2$, where $x = \lambda t(1+\varepsilon)$ with $\lambda t \gg 1$ and $\varepsilon \ll 1$. Stirling's formula has relative error of less than 1 % when $x = 1$ and is almost negligible for any x of interest in many cases. Considering all the approximations in the proof, we can safely assume that Poisson distribution is quite Gaussian (normal) in many astronomical contexts, unless x is very small ($\lesssim 10..?$).

Ex 5 (Photon Counting and Poisson Noise)

We saw the photon or electron counting is a Poisson process. Thus, we can use eqs. (2.18) and (2.19). For example, if we collected 10,000 electrons, the $1\text{-}\sigma$ uncertainty ($1\text{-}\sigma$ CI) or the standard deviation of this counting is $\sqrt{10,000} = 100$, so we denote $N = 10,000 \pm 100$. Since the N is large enough, Thm 6 says this is just a Gaussian distribution of mean 10,000 and standard deviation 100 (1 % of the mean). Since magnitude is $(\text{const}) - 2.5 \log_{10}(\text{count})$, differentiation gives a first-order estimation of the uncertainty in the magnitude $\Delta m = -\frac{2.5}{\ln 10} \frac{\Delta(\text{count})}{\text{count}} = 0.01^{\text{m}}$ (error-bar), which is small enough for some scientific purposes.

The next example is not directly related to Poisson process, but related to the binomial distribution (Bernoulli process) which has the pdf of $b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$ for getting x outcomes from n trials with p probability to have outcome. The pdf of binomial process approaches Poisson pdf with $p(x; np)$ when $n \rightarrow \infty$, $p \rightarrow 0$ but np is finite. Thus, the mean, variance, and standard deviation of such binomial distribution is calculable using Thm 5.

Ex 6 (Babies in Hospital: Provided by Prof. Jae-Kwang Kim at KAIST)

There are two hospitals A and B. New babies are born as either sex chromosome XX/XY with probability 50:50. Everyday A gives birth to 45 babies and B to 15 babies. Each hospital recorded the number of days when more than 60 % of babies are girl.

1. Which hospital will have more such days?
2. With how much confidence can you say so?

Chapter 3

Idea of Photometry

3.1 The Point-Spread Function (PSF)

Stars must be point sources. But from our experience, we know that they are never a point, but a usually nearly circular extended sources on the CCD. The image on the CCD when a single perfect point source is imaged, is called the **point spread function (PSF)**. There are basically two mechanisms responsible for this: diffraction and seeing. Although diffraction can also be a part of seeing, it is somewhat special than others: It is just impossible to eliminate the diffraction, while other sources of seeing can be removed (see below).

3.1.1 Diffraction

When a light ray passes through a finite-sized aperture, diffraction must occur. Due to this, a point source is blurred to a finite size after it passed through an aperture (e.g., telescope or DSLR camera aperture).

Ex 7 (Diffraction)

For $D = 1$ m in visible (550 nm wavelength), the diffraction limit is $1.22 \frac{\lambda}{D} = 0.14$ arcsec. You may be familiar with this formula. Using the proportionality, you can memorize it as $\theta_{\min}^{(550 \text{ nm})} = \frac{0.14}{D[\text{m}]} \text{ arcsec}$.

It is more complicated in reality: the mirrors and other optical instruments located in the optical path affect the final shape of a point source by diffractions. See fig. 3.1 for the examples of simulated PSFs from optics, i.e., when the seeing effects are neglected.

As you can see in fig. 3.1, although the detailed PSF are different (mid row), the long-range features are weaker compared to the central nearly circular features (bottom row) in cases except for the last column, which is not of our interest.

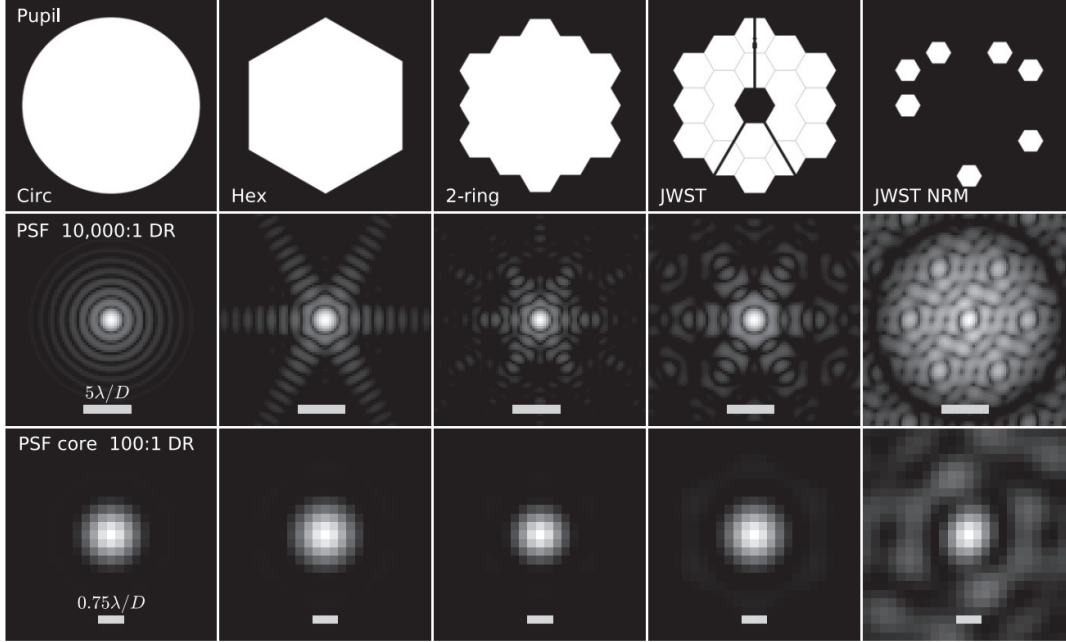


Figure 2. Comparison of the point spread functions (PSFs) of different shaped apertures. The top row shows apertures as labeled. The middle row shows the corresponding PSFs with 10,000:1 dynamic range at consistent stretch and pixel scale. The bottom row shows the cores of the same PSFs with 100:1 dynamic range. Note the relatively small core and deep null surrounding the *JWST* NRM PSF.

Figure 3.1: Example of PSFs. For circular aperture, the PSF shows concentric diffraction pattern as we learned in general physics course. For *JWST* case, you can see the secondary mirror and the mirror holding pipe (black things) makes the PSF different compared to the 2-ring case. The bottom row is not just a zoom-in of the PSF, but also the stretch (contrast) is different from the mid row (DR means the dynamic range). Direct excerpt from Figure 2 of FordKES+, 2014, ApJ, 783, 73.

3.1.2 Seeing

Even if we ignore the diffraction, a point source is blurred due to two main reasons: (1) air turbulence and (2) discontinuous movement of the telescope since the telescope gear is not a perfectly smooth continuous ideal machine. The blurring caused by these are called *natural (atmospheric)* and *telescopic seeing*, respectively. The atmospheric seeing makes a point source an extended circular source (for details, search for the theory of, e.g., Andrei Kolmogorov), while the telescope seeing *may* make it very elongated or irregular depending on the quality of the tracking system. If the tracking system is well established on a telescope, the final telescope seeing effect is quite circular.

All these effects can be removed if we move our telescope in space. The atmospheric turbulence is now gone, and the telescope can be very stable such that the telescopic seeing is ignorable. This is what is called the diffraction-limited case.

3.1.3 Summary of PSF

To summarize:

$$\begin{aligned}\text{PSF} &= \text{diffraction} + \text{seeing} \\ \text{seeing} &= \text{atmospheric seeing} + \text{telescopic seeing}\end{aligned}$$

and practically speaking,

$$\text{PSF shape} \approx \text{circular} .$$

This is why we can talk about the circular aperture photometry. Also

$$\begin{aligned} &\text{diffraction} \ll \text{other sources of seeing} \\ &(\text{unless adaptive optics or space telescope used}) . \end{aligned}$$

Actually, for many ground-based observations, it is not necessarily important to distinguish diffraction from other seeing effects. I feel like people tend to use the word “PSF” when the PSF is measured rather accurately (e.g., PSF photometry), and use the word “seeing” when we assumed circular PSF (e.g., aperture photometry).

3.1.4 Seeing Disc Size

As mentioned above, when we assume circular PSF (as we will do frequently in aperture photometry), we use the word “seeing” for the PSF. *Seeing disc* is the word for the circular shape of the PSF, and the *seeing disc size* is a measure of the size of that circle. The definition of the “size” is not trivial, but many people use the FWHM of the stellar radial profile*. This seeing disc size measurement is used to find the focal position and give a sense to the data quality.

The seeing disc FWHM is dependent on wavelength, weather, airmass, etc. At Seoul National University observatories in Seoul, it is roughly 1 to 3 arcsec at Bldg. 46 and 3 to 6 arcsec at Bldg. 45. For Subaru telescope at Hawai’i, for example, it is less than 1 arcsec†. Compared to the diffraction size θ_{\min} , these are much larger. It can go even below if we use adaptive optics (then diffraction \approx total seeing and non-circular PSF may be important).

The fact that the seeing disc size (i.e., size of the circular PSF) on the CCD is normally much larger than the diffraction scale hint that atmospheric and telescope seeing are very important. Since the telescope seeing is roughly constant over time for continuous tracking of a good telescope, the change in seeing can also strongly dependent on the weather condition. So astronomers use seeing disk size (usually rough estimate of FWHM) as a proxy of the weather condition.

It can also be used as a data-quality-indicator.

Ex 8 (*Contamination by Other Object*)

If you have your faint target near a bright stellar source with 1 arcsec separation and if the seeing disc FWHM is roughly 1 arcsec, your faint target is contaminated by the nearby bright star. A possibility is to do photometry of “your target + bright star” and subtract the flux of the “bright star” from catalog or future/previous observations. In this case, although you can get the flux of the target, its uncertainty may increase significantly.

*Radial profile means the pixel values of the star image as a function of the distance from the center of the star: $I(r)$. From many free/commercial products, it is rather easily drawn, such as **ginga**, **SAO ds9**, or **Maxim DL 6**. They also automatically calculates FWHM for you. The details on how to find the “center” will be studied later. Here, just think about something like Gaussian curve.

†<https://www.naoj.org/Observing/Telescope/ImageQuality/Seeing/>

Ex 9 (Slit Width Determination)

For spectroscopy, seeing disc size may determine the slit width you have to use to collect the target's light into the slit. It will be dealt in the spectroscopy chapter.

Ex 10 (Finding Focal Position from Seeing FWHM)

First find a random bright star that do not vary over short period of time (e.g., about 1 hour). You then expose and check the FWHM of the PSF (assuming it is circular). Now tune the focal position, e.g., by moving the position of the secondary mirror, and expose again, and check the FWHM again. Doing this repeatedly, you can get the FWHM as a function of the focal position. When the FWHM becomes the minimum, that is *the* focal position (fig. 3.2). Normally this curve is roughly second order polynomial.

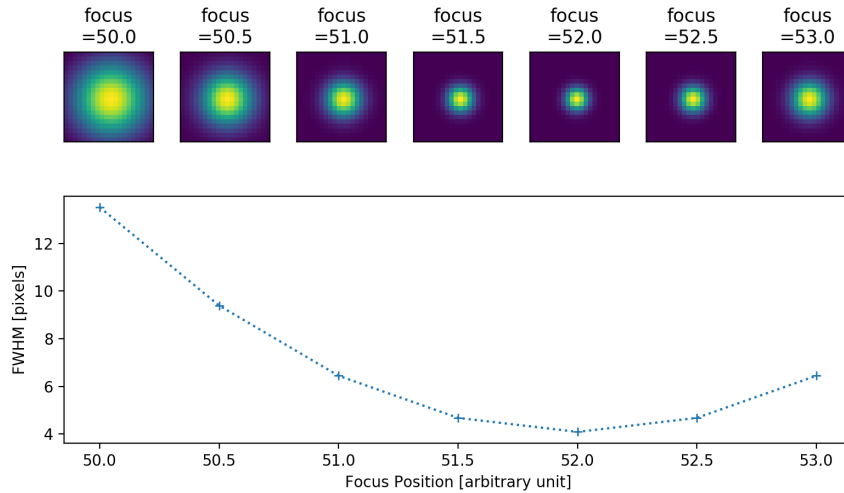


Figure 3.2: Finding focal position. While tuning the focal position value (x-axis), check the FWHM of the stellar profile. In this fake simulation, it is found that the focal position of 52.0 is *the* focal position with FWHM ~ 4 pixels. Multiplying with the pixel scale will give the FWHM in angular units (e.g., if the pixel scale is 0.8 arcsec/pixel, FWHM after the focusing is $4 \text{ pixel} \times 0.8 \text{ arcsec/pixel} = 3.2 \text{ arcsec}$)

Our goal in observation is to “gather all the photons from the target”. For a faint source, you have to tune the focal position very accurately, so that the signal from the source gathers to as small number of pixels as possible. Otherwise, the signal will be buried in sky values or other noise sources (dark current or readnoise) and our goal may not be achieved. The non-circular PSFs for high-quality data shares this same philosophy: the “sprays” of PSF should be analyzed. But for a very bright source, you may even intentionally defocus before the exposure. This is because we can avoid saturation and blooming from the central pixels while the signal is so strong that the noise is overwhelmed by the source signal. Then our goal is achieved even if we forget about the accurate focal position or accurate PSF.

3.2 Centroid

Even though we know the accurate PSF, we cannot use it for photometry if we cannot determine the center or the origin. There are various ways to determine this “origin”, but the most widely used one is centroiding (center of mass).

Let I_i for $i = 1, 2, \dots, N$ means the i -th pixel with $(x, y) = (x_i, y_i)$ of the original $N_x \times N_y$ 2-D pixel array so that $N = N_x N_y$. Say we set a threshold I_0 (e.g., background value, signal-to-noise ratio of 3, etc). Then change $I_i = 0$ if $I_i < I_0$. For $\xi = x$ or y , the centroid is by definition

$$\xi_c = \frac{\sum_i I_i \xi_i}{\sum_i I_i} . \quad (3.1)$$

The uncertainty of this is, considering ξ_i 's are fixed with no uncertainty,

$$(\Delta \xi_c)^2 = \left(\frac{\partial \xi_c}{\partial I_1} \right)^2 (\Delta I_1)^2 + \dots + \left(\frac{\partial \xi_c}{\partial I_N} \right)^2 (\Delta I_N)^2 . \quad (3.2)$$

But

$$\frac{\partial \xi_c}{\partial I_k} = \frac{\partial}{\partial I_k} \left(\frac{I_1 \xi_1 + \dots + I_N \xi_N}{I_1 + \dots + I_N} \right) = \frac{\xi_k}{\sum_i I_i} - \xi_c , \quad (3.3)$$

so

$$(\Delta \xi_c)^2 = \sum_k \left(\frac{\xi_k}{\sum_i I_i} - \xi_c \right)^2 (\Delta I_k)^2 = \frac{\sum_k (\xi_k - \xi_c)^2 (\Delta I_k)^2}{\sum_i I_i} . \quad (3.4)$$

Pixels near the object are likely to have high pixel values, so we approximate the uncertainty ΔI_i is just the Poisson noise of the pixel value (which is source + sky + read noise), i.e., in eq. (2.19), $\Delta I_i = \sqrt{I_i}$. Then expanding the square and using the definition of ξ_c , we get

$$\Delta \xi_c = \sqrt{\frac{\sum_k (\xi_k - \xi_c)^2 I_k}{\sum_i I_i}} = \sqrt{\frac{s_c^2}{\sum_i I_i}} \quad (3.5)$$

where

$$s_c^2 = \frac{\sum_k \xi_k^2 I_k}{\sum_i I_i} - \xi_c^2 . \quad (3.6)$$

There are some other techniques which involves Gaussian fitting: marginalize the image onto the x - and y -axes, and do the 1-D Gaussian fitting to each of them. Also the sigma-clipping inside the centroiding box (cbox) can be used (e.g., Ma+ 2009, Opt.Exp., 17, 8525).

3.3 Aperture Sum

We need first to find the total flux inside the aperture. The sky subtraction will be done later. Consider a very simple case in fig. 3.3: The centroid is calculated following eq. (3.1), and it is trivial that the centroid is at the center of the shown image. If we put circular aperture of radius 1 pixel (middle panel), the pixel contribution from the pixel with value 10 (right panel) is

$$\text{pixel value} \times \frac{\text{area in aperture}}{\text{total area}} = 10 \times \frac{\pi/4}{1} = 7.854 . \quad (3.7)$$

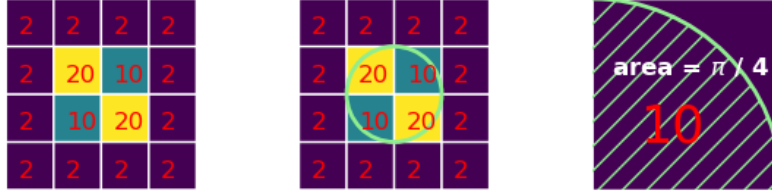


Figure 3.3: Left: An example pixel values near the source. Middle: An aperture of radius 1 pixel shown as green circle. Right: A zoom in of one pixel with pixel value 10.

Doing the same calculation for all the 4 pixels inside the aperture, you will get $N_{\text{apsum}} = 47.12$, where apsum means the “aperture sum”.

One thing to note is the uncertainty in position. In this 4 by 4 pixel example, $\sum_i I_i = 84$ and $\xi_c = (x_c, y_c) = (1.5, 1.5)$ under the 0-indexing. For Δx_c , we consider only the x -coordinate for eq. (3.5) calculation. Note that in that case, $\sum_k (x_k^2 I_k)$ for all the $4 \times 4 = 16$ pixels is nothing but $\sum_{k=1}^4 (x_k^2 I'_k)$ where $I' = [8, 34, 34, 8]$ is the marginalized intensity along the x direction from fig. 3.3. Then $\Delta x_c = 0^2 \times 8 + 1^2 \times 34 + 2^2 \times 34 + 3^2 \times 8 = 242$. This must be the same for the y -axis, and thus $s_c^2 = 242/84 - 1.5^2 = 0.63$, so $\Delta x_c = \Delta y_c = \sqrt{0.63/84} = 0.087$ pixel.

Here, I did not touch the topics such as (1) how to select the aperture radius, r_{ap} and (2) in reality, we need to fit some function to the radial profile, but which function should we use? These will be dealt in the next chapter.

Since this aperture sum includes not only the source but also the sky values (sky value is always non-zero for ground-based observations), we need to subtract the sky.

3.4 Sky Estimation

Estimation of the sky is one of the most trickiest part in observational astronomy. There are two most widely used methods: annulus and 2D interpolation. I will first explain few formulae, and then explain these two methods.

3.4.1 Simple Parametric Sky Estimation

There have been some parametric methods to estimate correct sky value. The most widely used ones include the “MMM” relation, or the “mean-median-mode” relation, and regard the mode (most frequent value) as the constant sky value. This has been used by virtually all astronomical softwares, including IRAF, IDL, SExtractor, and python packages. It is the simplest and does not require much computational speed, while gives reasonable results in many cases.

Why mode? The mission in determining the *constant* sky value is to find a *robust*, i.e., trustful constant value. Average (mean) is very vulnerable to few outliers even though we did sigma-clipping, and the median is less robust than mode, because if faint object with S/N ratio ~ 1 is in the sky

region, the median can also be overestimate the sky. Thus, mode is the most widely used simple statistic for sky estimation.

The determination of the modal value is very difficult, because the it is not simple to define it for the *sampled data*. If the distribution is not a mathematically continuous one, i.e., if it is discrete as we always encounter after sampling process, the mode is the largest value when we draw a *histogram*. The histograms are basically sensitive to the subjective selection of bins, and you can change the modal value to an unexpected value depending on the bin you choose (see fig. 3.4).

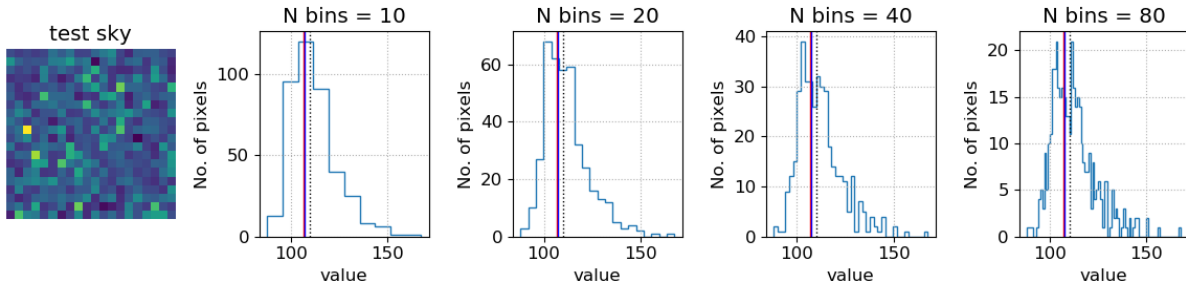


Figure 3.4: Left: A fake sky of 20 by 20 pixels generated by $100 + \mathcal{N}(0, 5^2) + 5\chi_2^2 + 3\mathcal{U}(0, 1)$ (normal distribution, chi-squared distribution, and uniform distribution, respectively). Others: The histogram of the pixel values with different bin numbers with fixed bin width. The black dotted lines is the median, and the red and blue solid lines are the mode estimation from eqs. (3.8) and (3.9), respectively (almost indistinguishable). As can be seen, the mode value does not lie at the same position, and it even become bimodal when the number of bins is 80.

Since 1800s, people empirically knew that the mode has the following relationship which holds for a moderately asymmetric distribution:

$$\text{mode} \approx 3 \times \text{median} - 2 \times \text{mean} , \quad (3.8)$$

or equivalently, *the distance between the mode and median is two-thirds of that between mode and mean*. DoodsonAT (1917, Biometrika Trust, 11, 425) first gave mathematical proof of it using some mathematical tricks including Laplace's and De Morgan's (1836), and it is recognized by many statisticians including KendallMG (1943; or the second edition of it in 1945, The Advanced Theory of Statistics 2e, Vol 1, p.35). People now refer to more recent references like KendallMG+StuartK (1977, The Advanced Theory of Statistics, Vol. 1. Charles Griffin & Co., London).

SExtractor (BertinE+ArnoutsS, 1996, A&AS, 117, 393) uses a variant of this:

$$\text{mode} \approx 2.5 \times \text{median} - 1.5 \times \text{mean} . \quad (3.9)$$

There is no clear reason provided in the original paper nor the SEx user manual, but it is just descibed that it was found to be more accurate when they tested. The estimation process works like this: Get the mean, median, and sample standard deviation (mean, med, std, respectively) from the sky pixel values after sigma-clipping. If $\frac{\text{mean} - \text{med}}{\text{std}} > 0.3$, use $\text{mode} \approx \text{med}^*$. Otherwise, use eq. (3.9).

*Although the publication and SEx manual says they use mean if $\frac{\text{mean} - \text{med}}{\text{std}} > 0.3$, but the SEx software actually returns med, not mean (described in photutils v0.6, Background Estimation (photutils.background))

IDL MMM uses eq. (3.8). IRAF also uses eq. (3.8) internally, but it gives the mean value if $\text{mean} < \text{med}$. In astropy-related packages (e.g., `photutils`), it is the user's choice. In all cases, you can set the sigma-clipping factors, e.g., a such that data outside $\text{med} \pm a \times \text{std}$ can be rejected, and n , the maximum iterations this process should be done until no more data are rejected. Whether to use $\text{med} \pm a \times \text{std}$ or $\text{mean} \pm a \times \text{std}$, whether to give asymmetric a values (upper and lower sigma-clipping), etc are up to the user.

I recommend to follow what SE_x does, with sigma-clipping of $a \sim 3$ and $n \geq 5$, because that became a virtual standard in parametric sky estimation.

3.4.2 Other Ways for Sky Estimation

As can be seen from fig. 3.4, the sky histogram is mostly skewed towards right. Because of this, BijaouiA (1980, A&A, 84, 81) first introduced to fit the sky histogram with Gaussian multiplied by Laplace:

$$p(I) = \frac{e^{\sigma^2/2a^2}}{a} e^{-(I-s)/a} \text{erf}_c\left(\frac{\sigma}{a} - \frac{I-s}{\sigma}\right) \quad (3.10)$$

where

$$\text{erf}_c(x) := \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-t^2/2} dt. \quad (3.11)$$

IrwinMJ (1985, MNRAS, 214, 575) argued that this method for modal estimation (Bayesian maximum likelihood is used) is better compared to mean, median, and Gaussian fitting. But because of the complexity of the calculation, Irwin used simpler method, i.e., get the smoothed histogram near the estimated mode and fit a Gaussian. BeardSM+ (1990, MNRAS, 247, 311) used slightly different method: smooth the (sky pixel) frequency histogram with moving box filter, find the mode estimate with it, sample the pixels nearby it which has pixel values larger than half of it. Then it does a cubic polynomial fit to the unsmoothed original histogram and finds the peak from this polynomial as the mode estimate.

For more, you may refer to Appendix A of AkhlaghiM and IchikawaT (2015, ApJS, 220, 1) and section 3.1 of MasiasM+ (2012, MNRAS, 422, 1674).

3.4.3 Annulus Sky

The simplest way is to use circular annulus. This method assumes the sky is nearly constant near the target of interest, PSF is circular, and there is no nearby celestial object. Under these simplifying assumptions, the sky must be azimuthally symmetric and a function of radius centered on the centroid of the target. If we set the inner and outer radii r_{in} and r_{out} much larger than the seeing disc size to define the circular annulus, the pixel values within that annulus are now the “sample” of sky values, and should not have clear tendency along the radial and azimuthal direction. fig. 3.5 shows the result of centroiding and sky estimation.

Consider we obtained the sky value m_{sky} somehow, e.g., median of sky, eq. (3.9) after sigma-clipping, etc. There can be many ways to estimate the uncertainties of this m_{sky} . Let me call this Δm_{sky} . The fluctuation of sky value can critically affect the final result of the photometry. But

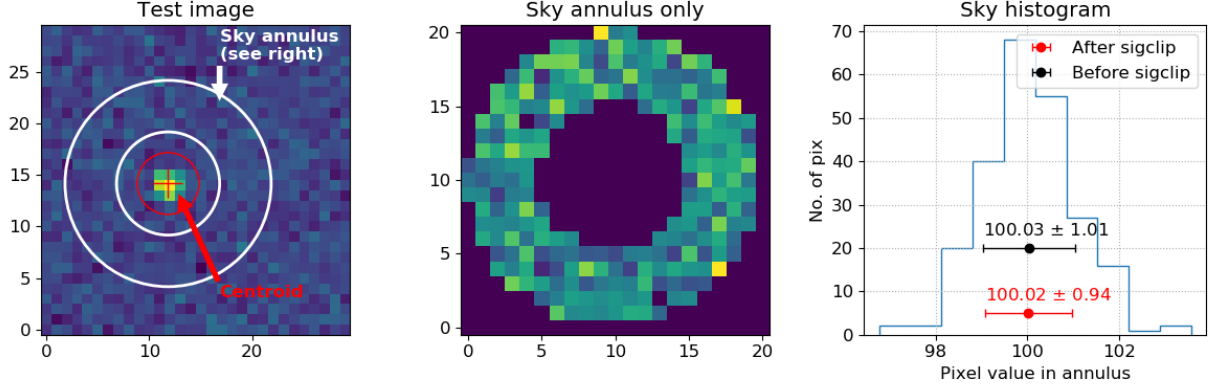


Figure 3.5: Left: The test image. Centroid is determined by eq. (3.1) and annulus is defined as $r_{\text{in}} = 5$ and $r_{\text{out}} = 10$ pixels, respectively. Middle: Plotting the sky annulus region only, to just visualize the random fluctuations in the sky region. Clearly, there is no radial trend and it seems quite symmetric along the azimuth. Right: The histogram of the sky pixel values in the annulus. The red/black markers are the median and sample standard deviation of the sky values after and before sigma-clipping, respectively. The sigma clip was done with 3-sigma 50-iterations.

once a robust way to estimate the sky is set and if the target is bright enough, the uncertainty due to the sky value is not so large. Thus, people do not care too much about the accurate pdf of the m_{sky} value. The most widely used way is to use the practical CLT (Thm 3) to m_{sky} , treating as if it is the mean of the sky samples (although we almost never use the mean for m_{sky}). Then

$$\Delta m_{\text{sky}} = \frac{s_{\text{sky}}}{\sqrt{n_{\text{sky}}}}. \quad (3.12)$$

Here, s_{sky} and n_{sky} are the sample standard deviation and the number of the sky pixels survived after the sigma-clipping, respectively. So we are assuming here that the mode has similar uncertainty as the mean.

3.4.4 Source Extractor Sky

We will not cover Source Extractor (SE or SEx or SExtractor) deeply. But simply put, it works like the description below. Say we have the image of 1000 by 1000 pixel.

1. **Meshing:** Chop the image by given box size. For example, if you set the box size as (50, 25), 20 by 40 (in total 800) “pads” or “meshes” will be made. Each pad will of course have 50 by 25 (in total 1250) pixels.
2. **Filtering:** Determine the size of the median filter. If the size is (3, 3), for instance, a 3 by 3 median filter will be used to reduce sharp noised pixels in each pad (similar to Gaussian convolution). The filter size should be determined as a function of seeing disc size.
3. **Rejecting Meshes:** In this median-filtered pad, the sigma-clipping is done. If too many pixels inside a pad is rejected, that pad is rejected for the next background estimation step. Such a threshold, say exclusion percentile, must be given by the user. For example, if that percentile is 10 %, any of the 800 mesh which rejected more than 10 % of its pixels from

sigma-clipping (10 % out of 1250 pixels, i.e., 125 pixels) will be regarded as “bad region” for the sky estimation.

4. **Sky at each Mesh** (bottom middle panel of fig. 3.6): Now come back to the original unfiltered meshes. From the pixel values, estimate the sky value and its uncertainty following eq. (3.9) and the description below the equation. For the meshes which are flagged as “bad” from previous step are not used for this process. As a result, you will have 20 by 40 array of sky values, while some elements may be empty or NaN, if some meshes are flagged as “bad”.
5. **Interpolation** (top middle panel of fig. 3.6): From the estimated single sky value at each mesh, i.e., the 20 by 40 array, we do interpolation to estimate the sky at all the 1000 by 1000 pixels. The estimation uncertainties should also be calculated for each pixel.

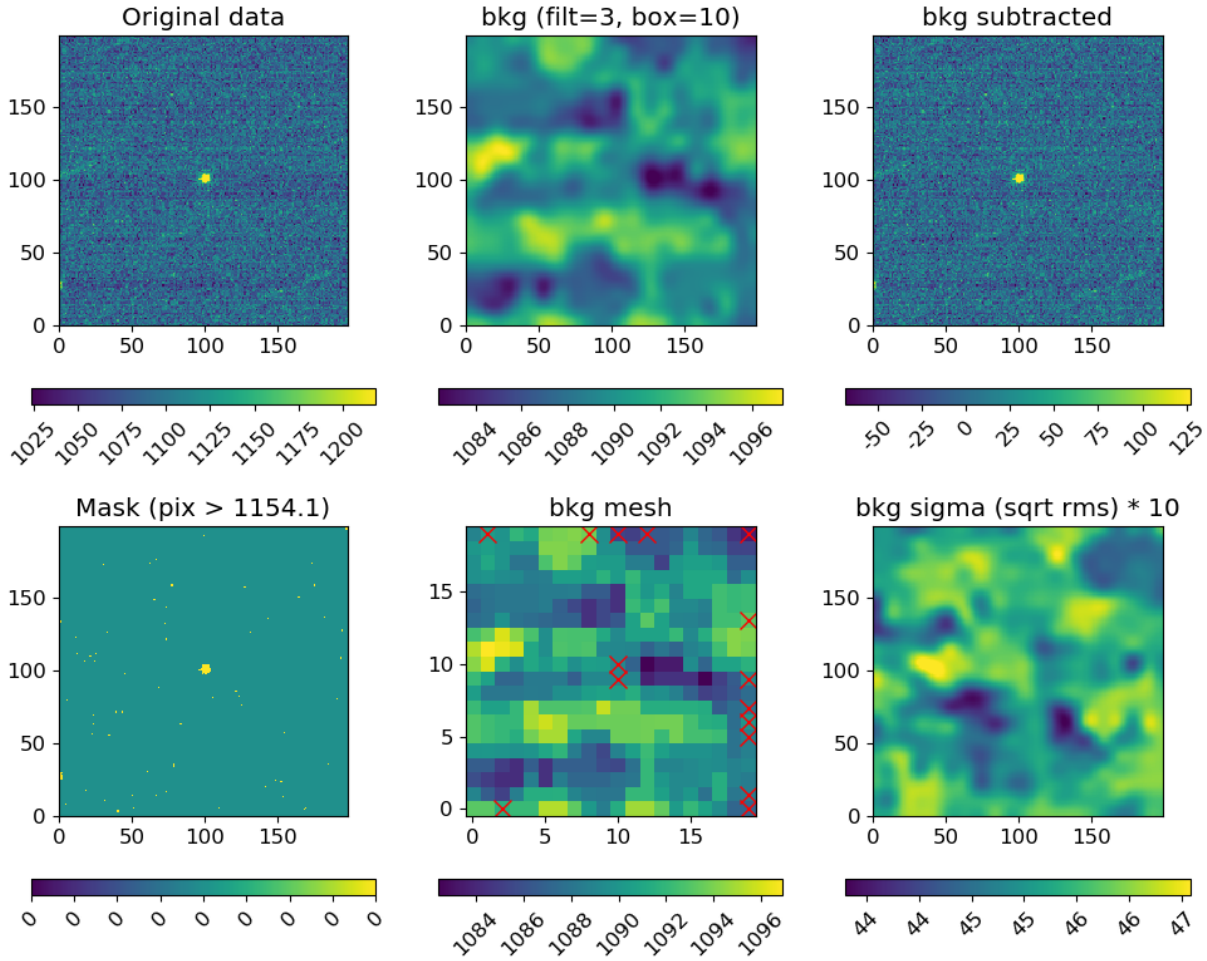


Figure 3.6: Example of the SExtractor background estimation usage. The original data is the target-to-centric combined image from 13 KMTNet images from SAAO observatory. The pixels with value 1154.1, which is the med + 3std after the sigma-clipping (3-sigma, 5-iter clipping) are masked because those can be stars (bottom left). The background meshes with the estimated sky values (eq. (3.9)) are shown in bottom center, and the interpolated one is given in top middle. Note that the sky level changes at maximum around 12 ADU out of the pixel value 1000+. Normally this should be smaller like 5 ADU or 2 ADU within this small region (200 by 200 pixel). The background sigma shown in bottom right is around 4 ADU (note the figure is 10 times the sigma), but many times the sky sigma is smaller than this.

After the sky subtraction, I plotted the Box plot of the image with x-axis 70 to 90 and y-axis 50 to 150 (python indexing) in fig. 3.7.

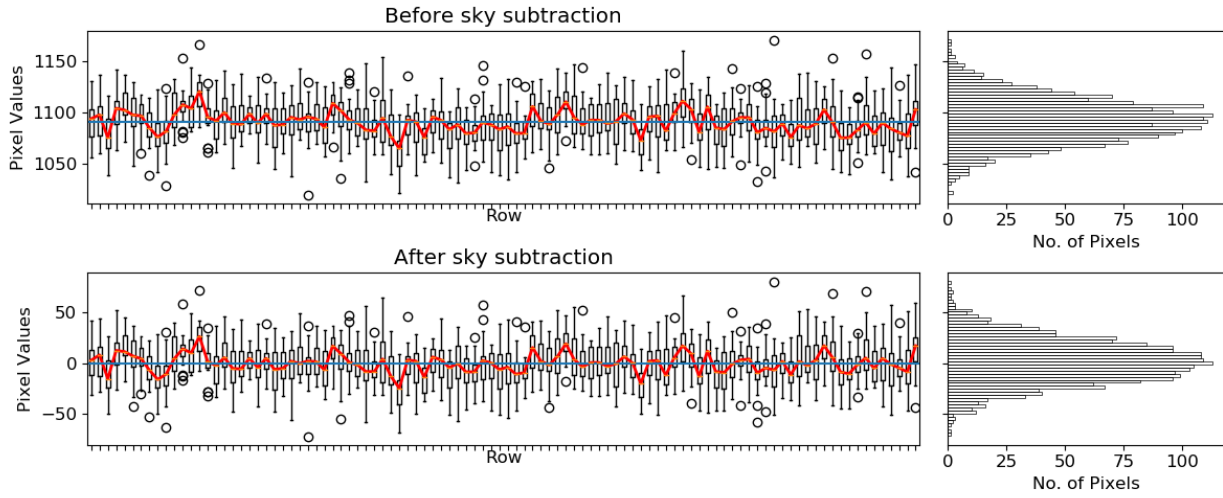


Figure 3.7: The Box plot and pixel value histogram of before/after the sky subtraction, of a region defined by `data[50:150, 70:90]` (python indexing). The horizontal axes of Box plots correspond to the “row” or the y index. Blue horizontal lines are the median of all the pixels in that region, and red solid lines are the medians

As can be noticed, SEx background estimation*, unlike the annulus sky estimation technique, is estimating the sky values at all the pixels, not for each source. Thus, it takes much longer time. Also you can see that lots of parameters (rejection percentile, box size, filter size, interpolation method and related parameters, etc) should be set for SEx to work! In some cases, astronomers run SExtractor again and again to do get expected results, and find best combination of such parameters. But these parameters of course depend on the instruments, sky conditions, etc, so if the results are suspicious, some fine tuning must be made or you have to find different ways to do photometry than SExtractor.

*We tend to call “background” rather than “sky” when we are dealing with SExtractor.