誤差について

撮像データ解析周りの誤差について、定式化の仕方と実例をいくつか解説する。

八木雅文

誤差とは？

真の値からの測定値のズレ

ここで「真の値」とは何かが問題。
測定値から、どのような値を求めたいと考えているのか？

ちなみに測定値や推定値から
どの範囲に真の値があるかの
指標は「不確かさ」

系統誤差とランダム誤差

系統誤差:
　何回やっても同じ量ずれる誤差
ランダム誤差:
　毎回再現性のない誤差

例えば機器の目盛りが間違っていて
　毎回同じ量ずれる⇒系統誤差

例えば機器が不安定で
　結果に再現性がない⇒ランダム誤差

真の値？

誤差は真の値からの測定値のズレ

変化しないものは、真の値を言うのは簡単。
例えば真空中の光速度は厳密に
毎秒 299792458 m
一方、光の強度の場合、光子はミクロな
確率過程を経て天体から出てくるため、
ある1秒間に天体から出てくる光の量は
次の1秒と同じではない。

つまり観測値は本質的に一定ではない。

光量の誤差とは？

その上で、天文の測光の場合、
求めたい「真の値」は、観測される光量の
「確率分布の代表値」
確率分布が時間変化しない場合には
これは「長時間平均」であると考えてよい。

光量の誤差は、測定あるいは推定した値が
どれだけ精度良くその確率分布の代表値を
求められたかという指標。

たとえ話

東京都の1日の交通事故数を調べたいと考え、
ある日が150件、次の日は140件という
情報を得たとする。

この生データ値を可能な限り正しく求めるのも
重要だが、最終的に欲しいのは
「典型的な発生件数」なので、150とか140という
数字自体の精度だけをひたすら追求するのは
あまり意味がない。

欲しいのは、例えば「平均 142.0 件/日」
という情報。

たとえ話の蛇足

直前のスライドは2011年版の再利用なので
例が交通事故でコロナではないのだった、
とかいう背景はさておき、
コロナ陽性者数の場合は時間変化する。

このような時間変化する量の場合も、
欲しいのはその時の厳密な発生数ではなく、
その時の確率分布の典型値。

天文の場合、いくつ光子が来たかではなく
来るはずだったか、を知りたい。

目指すところ

光量の推定の場合、光子数のポアソン誤差は
原理的に減らすことができないので、
目指すのは、ポアソン誤差以外の誤差を
ポアソン誤差に比べて十分無視できるくらいまで
小さくすること。

なお、HSCの観測では多くの場合、
目標天体以外の天体、主に夜空からの光子の
ポアソン誤差が一番大きい。
空が暗い時に観測したい理由の1つはこれ。

確率分布

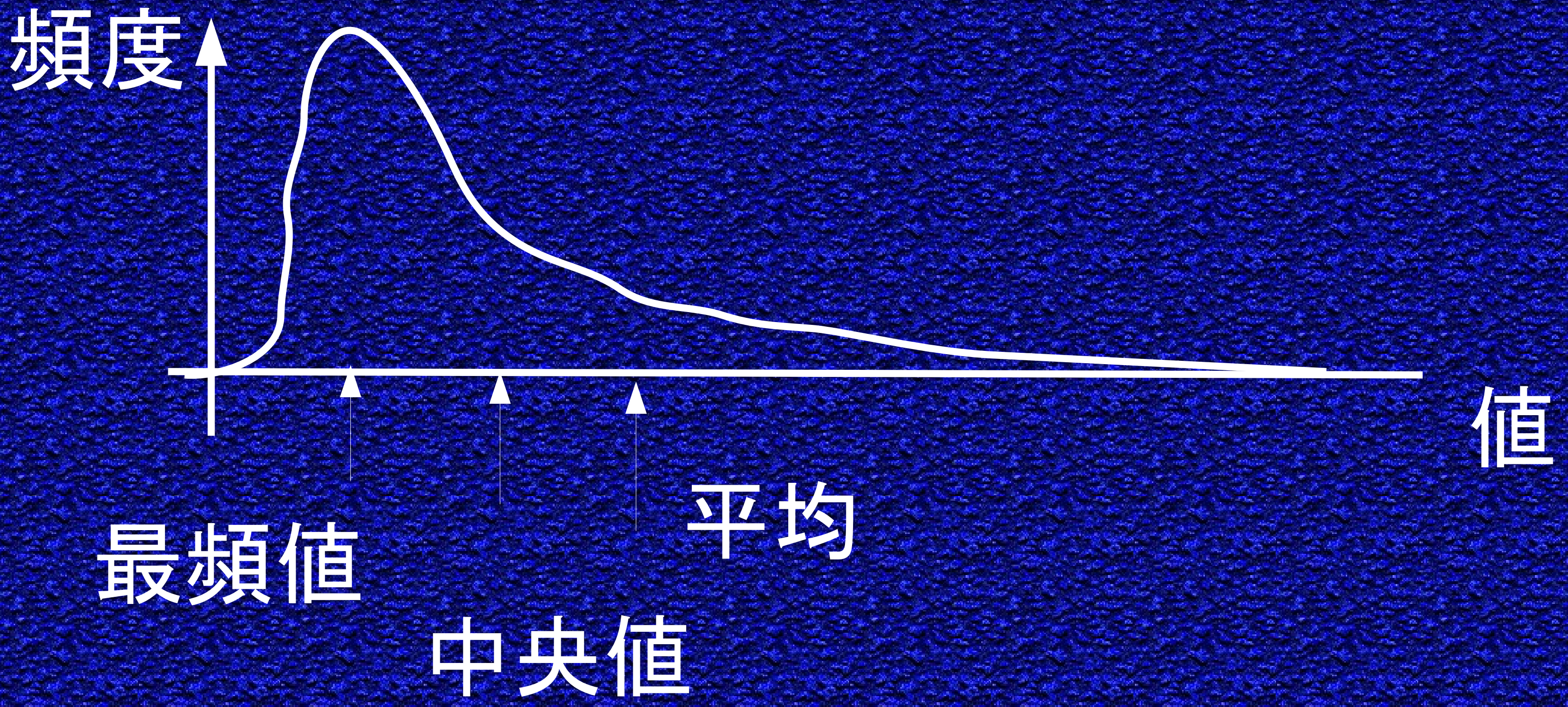
観測量X:観測ごとに変わる量
があったとして、ある試行で
x0である確率を p(x0)と書ける場合、
関数p(x)が確率関数。

Xがとる値xが連続量の時、
x0-dx/2<x<x0+dx/2 に入る確率が
p(x0)dx になるようなp(x)が、
確率密度関数。

これらを本資料中では(確率)分布と
呼ぶことにする。

分布と代表値

分布の代表値として良く使われるのは、
平均、中央値（メディアアン）、
最頻値（モード）など。
今回は平均を代表値と考える。



誤差の評価

一回の測定での誤差は、
真の値がわからないと求まらないが、
測定値の「分布」が仮定されていたり
何らかの方法でわかっていたりすれば
「誤差の典型値」を計算できる。

例えば良く使われるのが標準偏差
（誤差の二乗平均）σなどと書かれる。
※σは正規分布の標準偏差に
　相当する別の量の場合もある。

誤差の推定

1）そもそも誤差はどうなるはずか
　から考え、誤差の伝播を使って
　観測量の誤差を推定する
　（測光の限界等級）

2）観測量の統計を考えた上で、
　解析での誤差の広がりを推定する
　（画像などの並行移動）

を例として挙げる。

二項分布

ある確率pで事象が起きる試行をN回行ったとき、X回事象が起きる確率分布を二項分布と呼ぶ。書き方には色々あるが、ここではB(N,p) と書くことにする。

$$B(N,p)(x) = {}_N C_x \, p^x (1-p)^{N-x}$$

$${}_N C_x = \frac{N!}{x!(N-x)!}$$

例えば N=5, p=1/6 の時、(6面サイコロを5個振った時に1の目の出る数の確率分布)はB(5,1/6) と書ける。B(5,1/6)(0)～0.402

二項分布の平均は Np, 分散は Np(1-p)

ポアソン分布

二項分布で Np を一定のまま p→0 N→∞に飛ばした極限がポアソン分布。
k=Np と置いて、p(k)は以下になる。

平均 E(X)=k
分散 V(X)=k

$$p(k)(x)=\frac{k^X e^{-k}}{x!}$$

二項分布では平均は Np、
分散は Np(1−p)だったが、
ポアソン分布は p→0 の極限なので、1−p→1で
k=Np=Np となる。

標準偏差は√k

ポアソン分布する変数の和

ポアソン分布に従うX、Yという2つの
変数があったとき、和　X＋Y　も
ポアソン分布に従う。証明略。

一方、二項分布の和は、pが等しければ
二項分布になるが、pが違うと
二項分布には必ずしもならない。

また、差、X－Yはポアソン分布には
従わないので要注意。

ポアソン分布する変数の定数倍

ポアソン分布に従う変数の定数倍は
ポアソン分布に従わない。
ここも要注意。

電荷数はポアソン分布に従うが、
アンプを通して出てきた値は
ゲインで割り算されているので、
ポアソン分布には従っていない。

変数に定数を足しても
ポアソン分布に従わなくなる。

ポアソンを確率で間引く

例えば単位時間当たりの個数がポアソン分布に従う光子が、透過率 f の系を透過して来た結果もポアソン分布に従う

二項分布にまで戻って考えればB(N,p) の確率部分pが p で発生した後 f の確率で生き残るので確率が fpになったのと同じ。ポアソンでも期待値kや分散kがfkになる。

ポアソンを確率で間引く

解析的にも確認しておく。間引かれた結果zを観測し、間引かれたのがyだった場合、その確率は

$$\frac{(y+z)!}{y!\,z!}p^{z}(1-p)^{y}\frac{k^{(y+z)}e^{-k}}{(y+z)!}$$

これを足していけば

$$\sum_{y=0}^{\infty}\frac{k^{(y+z)}e^{-k}}{y!\,z!}p^{z}(1-p)^{y}=\frac{(kp)^{z}e^{-k}}{z!}\sum_{y=0}^{\infty}\frac{(k(1-p))^{y}}{y!}$$

$$=\frac{(kp)^{z}e^{-k}}{z!}e^{k(1-p)}=\frac{(kp)^{z}e^{-kp}}{z!}$$

と、平均 kp のポアソン分布にちゃんとなる。

正規分布

別名ガウス分布。
平均m、標準偏差σの場合、

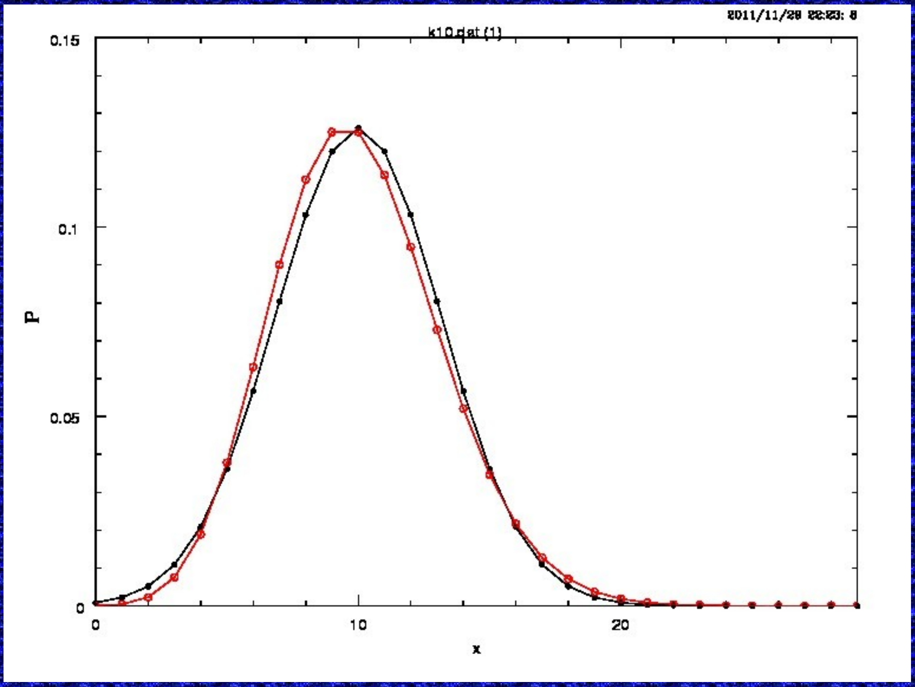
$$N(m,\sigma^2)(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-m)^2}{2\sigma^2}}$$

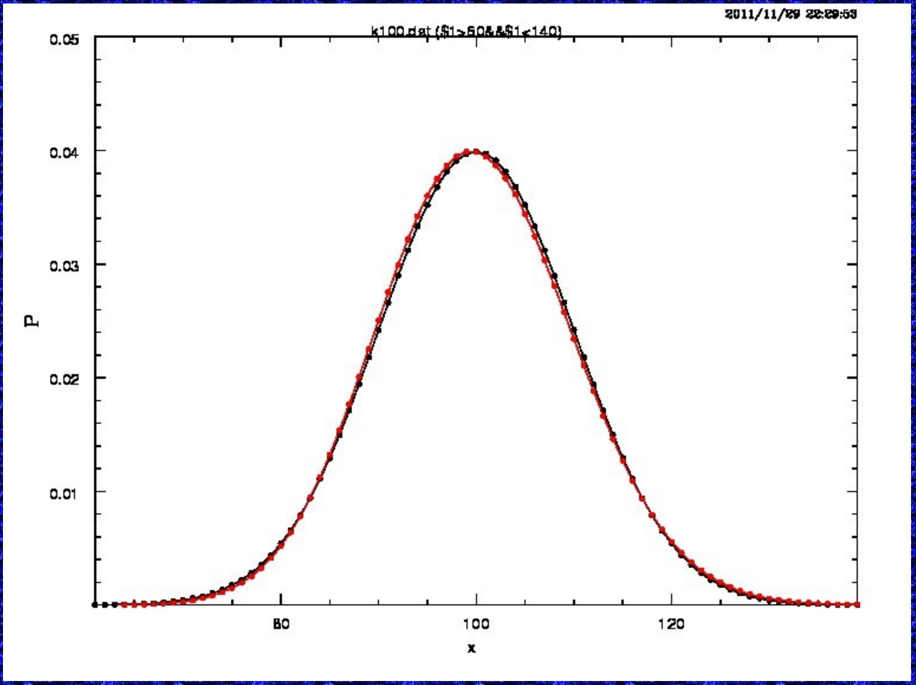
ポアソン分布は「kが大きければ」
正規分布 N(k,k) で良く近似できる。

多くの誤差は正規分布だと仮定される
ことが多い。

kが大きければ？

どれだけkが大きければ、正規分布はポアソン分布の良い近似になるのか？これ、何をもって「良い近似」とするか、色々考え方ありますが、k=10と100をグラフにすると、こんな感じ。





誤差の計算

確率変数、X，Yと関数f(X,Y)が
あった場合、f(X,Y)の誤差は
概ね以下の方針で求められる。

1）確率変数Xと誤差δXを求める。
Yも同様にδYを求める。
2）期待値E(f(X,Y))を求める
3）分散V(f(X,Y))=E((f(X,Y)−E(f(X,Y))^2)
つまり観測値から期待値を引いた
二乗平均を求める。
この平方根が誤差の典型値。

誤差の伝播の例

X, Yの平均をμ_X, μ_Yと書き、
標準偏差を　σ_X, σ_Yとする。
この時、$E(X+Y)=\mu_X+\mu_Y$

$$V(X+Y)=E\left(\left((X+Y)-(\mu_X+\mu_Y)\right)^2\right)$$
$$=E\left(\left((X-\mu_X)+(Y-\mu_Y)\right)^2\right)$$
$$=E\left((X-\mu_X)^2\right)+2E\left((X-\mu_X)(Y-\mu_Y)\right)$$
$$+E\left((Y-\mu_Y)^2\right)$$
$$=\sigma_X^2+\sigma_Y^2+2E\left((X-\mu_X)(Y-\mu_Y)\right)$$

なお、一般的にE(aX+bY)=aE(X)+bE(Y)

共分散

ここで出てきた$E\left(\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right)$
を共分散と呼んで $Cov\left(X,Y\right)$ と書く。
共分散が0の時、X、Yは独立という。
この時は、$V\left(X+Y\right)=\sigma_x^2+\sigma_y^2$

誤差の伝播で変数が独立かどうかは
極めて重要な点なのだが、
実際のデータでは独立かどうか
判断するのは、必ずしも簡単ではない。

追記:あとこれも有用。$V\left(aX\right)=a^2\sigma_x^2$
Vが定義できる分布なら常に成立する。

バイアスの統計

例えばバイアスを補正する場合、
データ1枚のバイアスが N(b0,r^2),
バイアスパターンはn枚合成して
N(b0,r^2/n) となっている場合、独立ならば
差し引いた後の統計は N(0,r^2(1+1/n))

読み出しノイズ相当の量は sqrt(1+1/n) 倍
だけ増えている。n=10 の場合、5%増。
読み出しノイズを5%増やしてでも
バイアスを引いたほうがいいのかは
考える必要がある。
(なお通常はポアソンノイズが圧倒するので
読み出しノイズが5%増えても気にならない)

追記：平均と分散

$$E(aX+aY)=a\mu_X+b\mu_Y$$
$$V(X+Y)=\sigma_X^2+\sigma_Y^2+2Cov(X,Y)$$
$$V(aX)=a^2\sigma_x^2$$

などは、平均や分散が定義できれば、
X,Yが正規分布でなくても成立する
強い式なので使いでがある。

「定義できれば」とか書いているのは、
コーシー分布のように平均や分散が
定義できない分布もあるため。

限界等級

撮像データの測光の限界等級とは、
天体の測光値(Signal)が、測光誤差
(Noise)の何倍かという比(S/N) が
ある一定値、例えば、10以上の天体が
何等か、という値。

例えば28等の天体が S/N=10 の時、
「S/N=10 での限界等級が28等」
と言う。

限界等級

この場合の限界等級は「測光精度」の
意味での限界等級であって、
検出の限界の意味での限界等級
ではない。

測光精度の意味での限界等級が
同じであっても、天体検出の割合が
同じであるとは全く限らない。
全然別の話である。

検出限界

同じ等級でも星のような点光源では、
ほとんどの光が小さな範囲、例えば
半径5pixくらいに収まるのに対し、
銀河ではもっと広い範囲に
光が淡く広がる。

このため、銀河の方が検出が難しく、
特に大きく広がったものの検出は
どんどん難しくなる。
（測光のS/Nも悪くなる）

検出限界

検出限界は私の知る限りでは、
解析的に求めるのは極めて困難。
（もし解析モデル知ってる方
　おられましたら御教示ください）
多分、ある確率で画素が
閾値を超えるような画像で、
閾値を超えたピクセルがN個つながる
確率は？、という問題になる、はず。

実用上は、疑似画像を埋め込んで
検出確率を求めることが多い、と思う。

等級の系統誤差とS/N

等級の場合、まず単位時間あたりの光子数がポアソン分布し、フラックスは正規分布と仮定できる。このときフラックスを F, 等級をmと置くと、等級の期待値は

$$E(m) = E(-2.5\log(F) + const)$$

$$E(m) = -\frac{2.5}{\ln(10)} E(\ln(f_0 + \delta F)) + const$$

$$E(\ln(f_0 + \delta F)) = E(\ln(f_0) + \ln((1 + \frac{\delta F}{f_0})))$$

$$\ln\left(1 + \frac{\delta F}{f_0}\right) = \frac{\delta F}{f_0} - \frac{1}{2}\left(\frac{\delta F}{f_0}\right)^2 + O\left(\frac{\delta F}{f_0}\right)^3$$

等級の系統誤差とS/N

Fは正規分布だと思っているのでδFの平均は0

$$E\left(\left(\delta F\right)^2\right)=\sigma^2$$

と置けば、 $E\left(\ln\left(1+\frac{\delta F}{f_0}\right)\right)=-\frac{1}{2}\left(\frac{\sigma}{f_0}\right)^2=-\frac{1}{2}\left(\frac{1}{(S/N)}\right)^2$

したがって、 $m_0=-2.5\log\left(f_0\right)+const$
としたとき、

$$E\left(m\right)=m_0+\frac{2.5}{2\ln\left(10\right)}\left(\frac{1}{\left(S/N\right)}\right)^2\sim m_0+\frac{0.54}{\left(S/N\right)^2}$$

S/Nが7くらいより小さいと、0.01等以上の
「系統誤差」が等級に乗る、と計算できる。

等級の系統誤差とS/N

・・・書いていて本当にそうなるのか
不安になってきたので一応確認してみた。

S/N=5 の場合を想定して、平均5標準偏差1、
N(5,1^2)の乱数を10000回振る。
その結果を −2.5logして求めた等級の
平均をとって、−2.5log(5) と比較してみると、
確かに 0.022 等暗い側にズレしている。
0.54/5^2=0.022 なので、計算は合っている。

なお、このトイシミュレーションでは
等級のメディアンは 0.001 等しかずれなかったが
理論的にもそうなるのかは未確認。

等級のランダム誤差

一方、等級のランダム誤差は
E((m−m0)^2) で評価できる。

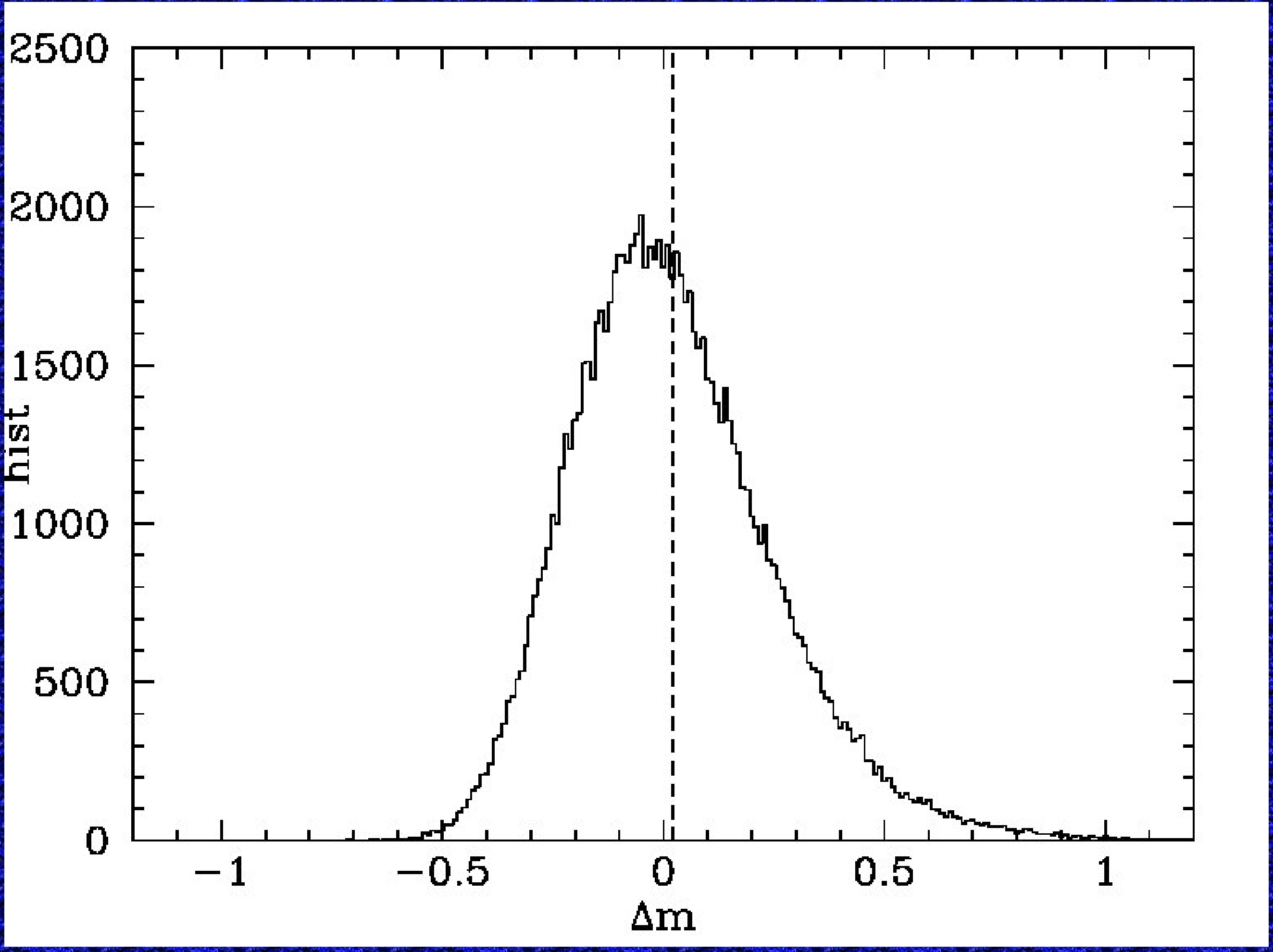
$$m \sim m0 - \frac{2.5}{\ln(10)}\left(\frac{\delta F}{f_0} + O\left(\frac{\delta F}{f_0}\right)^2\right)$$ なので、

$$V(m) \sim \left(\frac{2.5}{\ln(10)}\right)^2 \frac{1}{(S/N)^2} \qquad \sigma(m) \sim \left(\frac{2.5}{\ln(10)}\right)\frac{1}{S/N} \sim \frac{1.086}{S/N}$$

例えば 0.1 等以下の精度の測光を
したいのであれば、S/N>10.86 が必要。

等級の系統誤差とS/N

S/N=5 のときNに正規分布乱数を
100000回振って書いた等級誤差のヒストグラム



足し合わせの方法

標準偏差で誤差を評価した場合、
S/N を最も高くする足し合わせ方は
重みづけ平均。

しかし、重みづけ平均は外れ値に弱い。
そこで、複数の露出の足し合わせには
sigma-clipped mean や、メディアンが
用いられる。

メディアン

例えば、平均1OO、分散20の
観測値のうち1つに非常に大きな
ノイズが乗った場合、
113, 92, 66, 10000, 104, 68, 90
これを単純平均を取ると ～1500。
一方、median は 92,
MAD でσを推定した上での、
3σ-clipped mean は、89
単純平均よりは遥かに良い推定

メディアンの誤差

しかし、メディアンは単純平均よりも
誤差が大きくなる。
分散1の正規分布の平均とメディアンを
比べると、Nが十分大きいところでは、
標本標準偏差は約25%悪化する。

$$s(mean)=\frac{1}{\sqrt{N}}$$

$$s(median)=\frac{1.25}{\sqrt{N}}$$

メディアンの誤差

N	s(median)	s(1-clipped)
3	1.16	1.22
4	1.09	1.15
5	1.20	1.12
6	1.14	1.10
7	1.21	1.08

単純平均の標準偏差(s(mean))との比。
この値が大きいほど誤差が大きい。
１つだけclip される場合とメディアンでは
N>4 では clipped の方が誤差が小さい。
なお偶数 median で誤差が小さいのは
中央２つの平均を取っているから。

系統誤差の影響

１σ程度の系統誤差、例えば
画像の足し合わせの際の
バックグラウンドの引き残りやゴースト
などには、
メディアンは引きずられにくいが、
clipped mean は clip がうまくできないと
引きずられる。

一方、外れ値が常に大きい側に偏る
場合などは、メディアンは
系統的にそちら側にずれる。

メディアンのずれ

具体的には例えば N=2m+1 個の
データのメディアンで、
1個大きい側に外れ値が混じる場合、
本来は大きい方から m+1 番目を
とるはずが、m 番目をとることになる。

このことでmedian xは、元のデータが
外れ値以外は N(μ,σ^2)に従う場合、
p(x)=(2m)!/((m−1)!(m+1)!)
(1−erf(x))^(m−1)gauss(x)erf(x)^(m+1)
に従うことになる。

メディアンのずれ

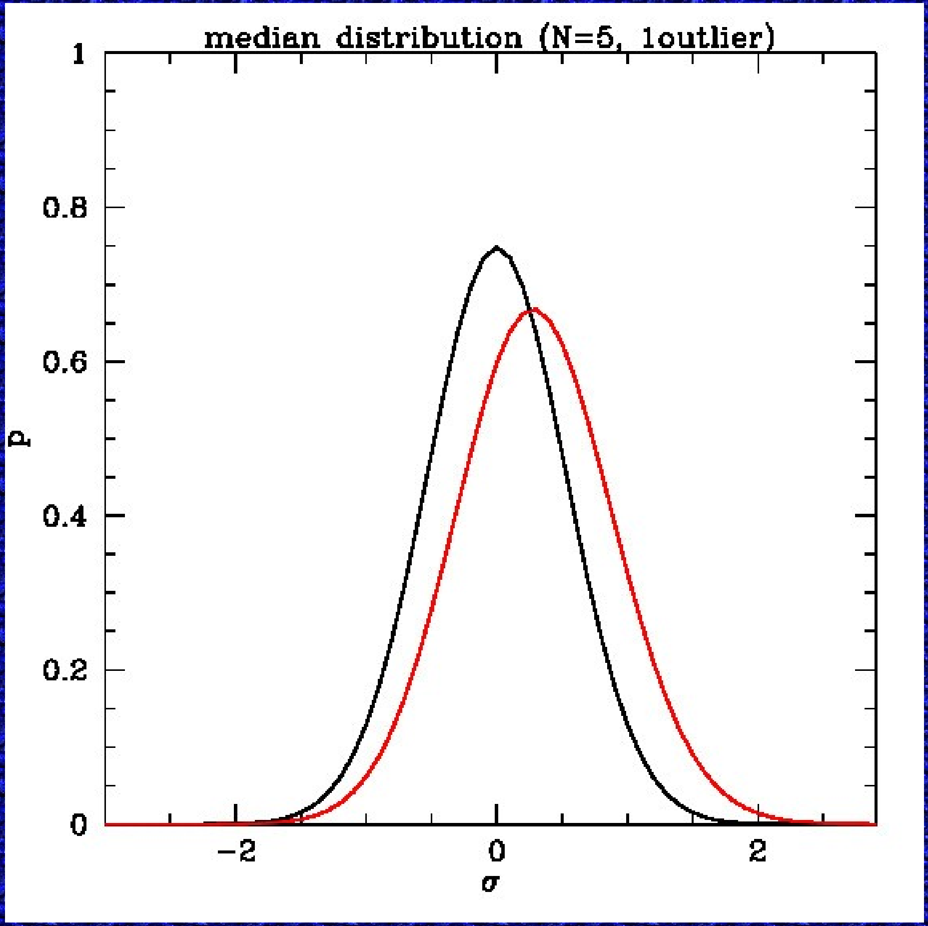
定量的には、元のデータが
外れ値以外 N(μ, σ^2)に従うとして、
サンプル数Nが十分大きく、大きい側の
外れ値の混じる割合が h (h<0.1程度)
の場合、経験的には〜1.25hσほど
プラス側にズレるようである。

Nが小さい場合、1つ大きな外れ値が
混じった場合、N=3,5,7 では
0.5σ,0.3σ,0.2σほど、
平均値がプラス側にずれる。

メディアンのずれ

外れ値が1つ入った場合の分布は
こんな感じになる。N=5 で、1つ外れ値が
混じると、
0.3σほどずれる。

この分布のxの
一次のモーメントが
外れ値入りの場合
のメディアンの
期待値(平均値)



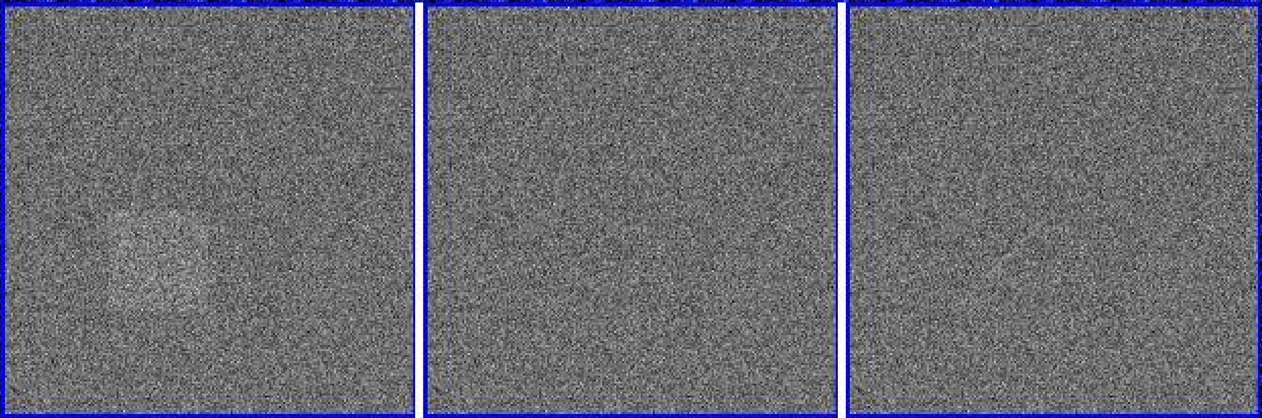
メディアン

したがって、常に外れ値がプラスにしか入らないような場合には、
メディアンは系統的にずれる。

これを考えると、
「もしうまくクリップができるのであれば」
クリップ平均の方が結果が偏らない。

追記：やってみた

講義後、宇宙線が入ったらプラスに偏るのか？
という質問をいただいたので「そうです」と答えつつ
念のため実際に試してみた。
bias ５枚からあるチップあるチャネルの適当な範囲
を切り出して、そのうち１枚のある領域に+10000
して、残り4枚とmedian stack した結果３種。

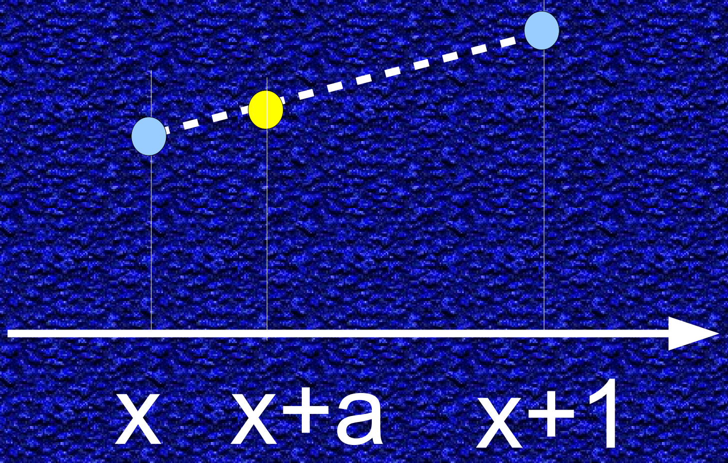


四角はわかる。幅1の縦線はわからない（と思う）
幅3の斜め線はかろうじてわかるか。
と、目で見てわからなくても、このように偏る。

画像の並行移動

回転等でも類似の問題があるのだが、ここでは
並行移動を考える。画像・データを
非整数画素だけ並行移動させたい。
例えば位置を合わせて足し合わせる場合など。
方法として、線形補間のリサンプリングを考える。
例えばx軸方向に −a (0<a<1) シフトさせるときは

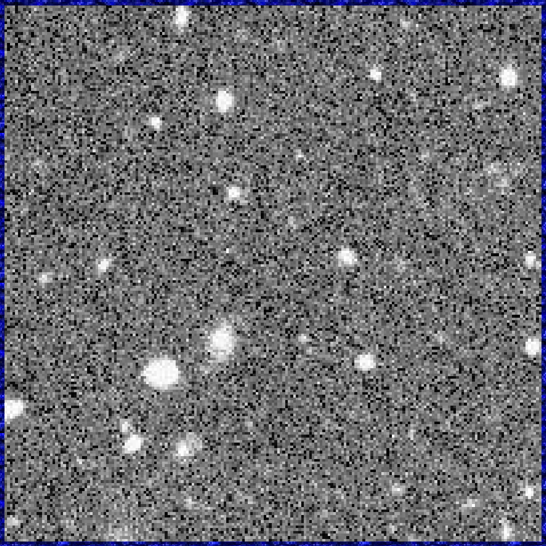
$$v'(x') = (1-a)v(x) + a\,v(x+1)$$

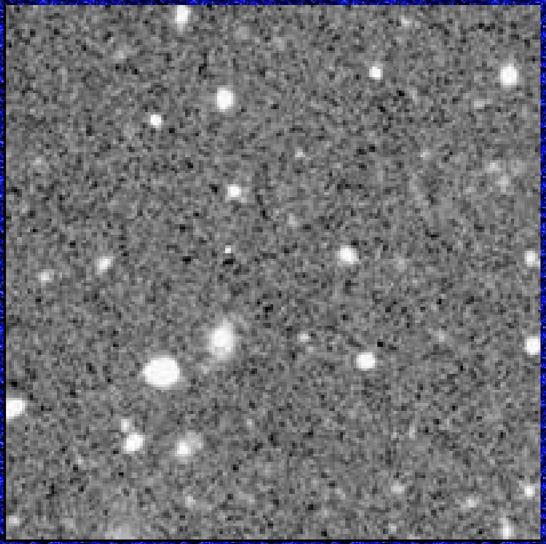


この結果、1画素あたりの
分散は $((1-a)^2 + a^2)\sigma^2$
$= (1-2a(1-a))\sigma^2 < \sigma^2$
と小さくなる。

実際にシフトさせてみた

左がシフト前、右はx+0.5, y+0.5 シフト





見た目でもざらつきが減っているのだが、
実際測ってみると、画素あたりのノイズは
確かに減っている。左67ADU, 右43ADU

すると並行移動でS/Nが上げられる！？

もちろんそんなことはない

シフト後は画素間が独立ではなくなる。
このため、ある範囲の画素を足した合計の
分散は、各画素での誤差の二乗（分散）の
和に加えて、更に共分散を足す必要がある。

具体的に[0,N-1] の範囲の総和の誤差を
計算してみる。
元の誤差は独立なN点の値の和なので分散は

$$V\left(v(0)+v(1)+\ldots v(N-1)\right)=N\sigma^2$$

平行移動後

$$v'(0)=(1-a)v(0)+av(1)$$
$$v'(1)=(1-a)v(1)+av(2)\ldots$$

誤差は減っていない

$$V(v'(0)+v'(1)+\dots v'(N-1))$$
$$=V((1-a)v(0)+v(1)+\dots v(N-1)+av(N))$$
$$=((1-a)^2+(N-1)+a^2)\sigma^2=(N-2a(1-a))\sigma^2$$

Nが大きければほとんど元と変わらないまま。

このような1画素あたりの分散が見かけ上
小さくなることは画像処理の結果よく起きる。

このため、実際のデータの測光誤差の推定は
画素ごとではなくある程度の範囲のカウントの
和のバラツキで通常は推定する。

観測モデル

モデルとして以下を考える。

$$DATA = \frac{(OBJECT + SKY) \times flat \times t_{obs} + DARK \times t_{dark}}{gain} + BIAS$$

OBJECT,SKY,DARK,BIAS は確率変数で、
観測量の DATA も確率変数。
flat, tobs, tdark, gain は定数。

モデルとしては1回の観測では分布に従った
確率変数 OBJECT,SKY,DARK,BIAS から
DATA が得られると考える。

仮定

恐らくは、画素1つ当たりの光電荷の数は
ポアソン分布に従い、そこを読み出した
カウントは正規分布（ガウス分布）に従うだろう、
と「仮定」する。

独立性に関しては、あるピクセルでの光電荷
の数は隣のピクセルの光電荷の数とは、
独立だろうと仮定する。

※ただし、実際それぞれの装置で、画素間の
独立性がどこまで成り立っているかは、
私は未確認です。

S/N のモデル計算

露出時間をt、システムの透過率をf、
天体の光度をLとすると、検出器で
検出される信号は

$$DATA = \frac{(OBJECT + SKY) \times flat \times t_{obs} + DARK \times t_{dark}}{gain} + BIAS$$

のうち $OBJECT \times flat \times t_{obs}$ の部分。
単位は光電荷数、としておく。
単位時間当たりの光子数（相当の量）に
時間と透過率を掛け算。

追記：flatの誤差

講義で、flat の誤差について質問いただき、
「実際にはflatの誤差はすごく効く」と答えたのですが、
その辺りを追加解説。前のページの式はあくまで
　我々は実際には知らないけど、理想的には
　ある定数としての flat(透過率),gain があって
　それは試行を繰り返しても変わらないので
　確率変数ではなくモデル中では誤差もない
というモデルです。強いて書くなら真の値 flat0, gain0
とでも書くべき値でした。
実際の解析では観測データから推定した
flat=flat0+Δflat を使うことになるのですが、この
flat の誤差Δflat がDATAからOBJECTを求める際に
すごく効くはずです。ただし、全く同じ条件で標準天体で
較正できれば、この誤差はキャンセルして消えます。

追記：flatの誤差

横道にそれますが flat の誤差Δflat は、
ドームフラットの各ピクセルのポアソン誤差、
フラット板や光源の非一様性、観測時の姿勢
と違うことによる周辺減光パターンの違い、
ほとんど効かないけど読み出し雑音や
ダークの誤差に加えて、全体をcoaddして
足し合わせる前の各データの規格化の誤差
などなど、考えるべき要素は
あれこれとあるので結構大変です。
しかも多分1%未満のレベルまで気にして
調査する必要があるのでそれも大変です。

S/N のモデル計算

$OBJECT \times flat \times t_{obs}$

のOBJECT はポアソン分布に従う変数
式は定数を掛けているように見えるが、
これは変数の定数倍ではない。
flat は、確率での間引き、
tobs は最初の1秒の光量のtobs倍
ではなく、整数で考えれば独立な
tobs 回の光量の和。
OBJECT の期待値を object と書いて

$S = object \times flat \times t_{obs}$

S/N のモデル計算

次に、S/N の N の方。誤差。
測光での大きな誤差源の一つは機器のノイズ（読み出し雑音）。読み出し雑音は露出時間によらず一定であると、とりあえず考えてよい。
1画素あたりのこの誤差を r として、

$$E(N_1)=0, V(N_1)=r^2$$

天体が A 画素に入っている場合、
読み出し雑音は画素ごと独立だと仮定して

$$E(N'_1)=0, V(N'_1)=Ar^2$$

と、分散はA倍。

S/N のモデル計算

もう一つの大きな誤差要因は
測定される電荷の確率的な誤差。
（ポアソン雑音）
目標天体の画素数をAとすると、
ポアソンの期待値は、OBJECT,DARK
の期待値を object, dark と書いて、

$$(object + sky \times A) \times flat \times t_{obs} + dark \times A \times t_{dark}$$

となる。分散（誤差の二乗）は、

$$V(N_2) = (object + sky \times A) \times flat \times t_{obs} + dark \times A \times t_{dark}$$

測光のS/N

読み出し雑音とポアソン雑音は
全て独立と考えられるので、
雑音の分散は合わせて

$$V(N) = Ar^2 + (object + sky \times A) \times flat \times t_{obs} + dark \times A \times t_{dark}$$

誤差Nは

$$N = \sqrt{Ar^2 + (object + sky \times A) \times flat \times t_{obs} + dark \times A \times t_{dark}}$$

$$\frac{S}{N} = \frac{object \times flat \times t_{obs}}{\sqrt{Ar^2 + (object + sky \times A) \times flat \times t_{obs} + dark \times A \times t_{dark}}}$$

測光のS/N

$$\frac{S}{N}=\frac{object\times flat\times t_{obs}}{\sqrt{Ar^2+(object+sky\times A)\times flat\times t_{obs}+dark\times A\times t_{dark}}}$$

tが小さい場合、分母の平方根の中はAr^2が残って、$\frac{S}{N}\sim\frac{object\times flat}{r\sqrt{A}}\times t_{obs}$

tが大きい場合、$t_{obs}\sim t_{dark}$として、

$$\frac{S}{N}\sim\frac{object\times flat}{\sqrt{(object+sky\times A)\times flat+dark\times A}}\times\sqrt{t_{obs}}$$

つまり

露出時間tが小さい場合は、$\frac{S}{N} \propto t$

大きい場合は、$\frac{S}{N} \propto \sqrt{t}$

露出時間が長い場合、S/N は
露出時間の平方根に比例。

S/N を倍にしたい場合は
露出時間は4倍必要、
精度を1桁上げるには100倍必要。

まとめ

このようにモデルを立てて、順に計算すれば
誤差がどうなるか推定できるケースが
それなりにある。

しかし、一方で解析的な推定が難しい問題も
また多く存在する。
そのような場合、例えば検出限界のように
疑似データを加えて実際に検出を試して
その結果から推定するなど、
シミュレーションから誤差を求める場合も多い。