

概要

画像処理を行う各段階や較正において、真の物理量からのズレが混入する。このズレをどう定式化し、どう小さくしていくかについて考える。

1 誤差

誤差とは「測定値・理論的推定値また近似計算によって得られた値と、真の値との差(大辞林)」である。「真の値」とは何かが問題である。測定値から、どのような値を求めたいと考えているのか?それを常に心に留めておく必要がある。なお、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」であり、誤差とは別の話である。

光赤外観測で欲しい情報

データを得たとして、そこから「欲しい情報」を最も精度良く取り出すにはどうすればよいか?というのが問題設定である。例えば光赤外の天文データの場合、欲しい情報は、「ある時刻範囲の間に天球上のある位置の範囲(方角)から、ある波長範囲(=エネルギー範囲)で、ある偏光の光子が、手元のある範囲にどれだけの数来ているか」になる。このうち、偏光に関しては「偏光観測」でない撮像や分光観測は、どの偏光成分も足し合わせて観測して、議論する。

時刻の区切り方は、可視の観測では、シャッターを使って、検出器に光が入らないようにすることで区切ることが多く、赤外線観測では、検出器のクリアから次の読み出しのタイミングまでの間で、時間を区切る場合が多い。

時刻、偏光をこのように設定すると、欲しい情報は「天球上のある範囲から、ある波長範囲の光がどれだけの強さで来ているか」であり、この中で「天球上のある範囲」という方向情報を詳しく分けて調べるのが撮像観測になり、一方「ある波長範囲」というエネルギーの情報を詳しく分けて調べるのが分光観測となる。そこで、光の強さ、光量について考える。

光量の誤差

再度書くが、誤差とは真の値からのズレである。では、この「真の値」とは何か?

時間変化しないものは、真の値を言うのは簡単である。例えば真空中の光速度は厳密に 299792458 m/s なので、これと異なる測定値を得た場合は、その差が誤差となる。ところが、光の強さについてはそうはならない。光はマイクロな確率過程を経て、天体から出てくるため「ある1秒に天体から出てくる光の量」は次の1秒のそれと同じではない。つまり、観測値は本質的に一定でありえない。では、ある天体の光量の値とは、何を求めているのかといえば、一般的には「長時間平均」である。この値はある短時間に観測される光量の「確率分布」の「代表値」と見なすことができる。光量の場合の誤差とは、実際に測定あるいは推定した値がどれだけ精度良くその分布の代表値を表しているかという指標である。

喩え話をしてみる。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¹という情報を得たとする。この生データ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」という情報である。

誤差の見積もり

誤差を系統誤差とランダム誤差に分けて考えることは良く行われている。

- 系統誤差:何回やっても同じ量ずれる誤差。例えば機器の目盛りが間違っていて毎回同じ量ずれる場合。
- ランダム誤差:毎回再現性のない誤差。例えば機器が不安定で結果に再現性がない場合。

ある1秒間に受ける光量の平均値からのずれ(バラツキ)は、機器の不安定とは別種だが、ランダム誤差の一種ではある。

今回の話は、まず生の観測量にはどのような誤差が乗っているのかをモデル化し、そのモデルから物理量を求める整約、解析の各段階で、誤差がどのように変化するかを追いかける、という内容である。一般的には、得られたデータから誤差・不確かさを推測する場合も多いが、それはまた別の話となる。

¹これ人じゃなく件です。念のため

2 確率分布

観測ごとに変わる量、観測量 X があったとして、 X が離散的で $X=X_0$ である確率が $p(X_0)$ と書ける場合、この関数 $p(X)$ が確率関数である。 X が連続量の場合、 $X_0 - dx/2 < X < X_0 + dx/2$ という区間に X が入る確率が $p(X_0)dx$ と書ける場合、この $p(X)$ が確率密度関数である。以下、両者まとめて「(確率)分布」と書く。

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」が何らかの方法でわかっているならば「誤差の典型値」を計算できる。この「典型値」として良く使われるのが標準偏差(誤差の二乗平均)で、 σ などと書かれる²。

観測値の母分布がわかる場合は、一回の観測値が典型値からどれだけずれるか(誤差)の分布も推定できて、誤差の典型値も求められる。一方、観測値の母分布の関数形はわかる、あるいは仮定できるが、そのパラメタが分からない場合、観測結果から母分布のパラメタを推定して、そこから誤差を推定することができる。

以下2つの例は、「そもそも誤差はどうなるはずから考え、誤差の伝播を使って観測量の誤差を推定する(限界等級の計算)」と、「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)」である³。

3 例1:(測光の)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、 S/N が3以上の天体が何等かという値である。例えばある観測結果で28等の天体が $S/N=5$ の時、 $S/N=5$ での限界等級が28等、という。等級のランダム誤差 Δm と、 S/N との関係は、

$$\Delta m \sim \frac{1.086}{S/N}$$

である⁴。例えば0.1等以下の精度で測光をしたいのであれば、 $S/N > 10.86$ が必要である。

ここでの限界等級は「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。「では、検出の限界はどう推定するのか?」と質問を受けたが、これを解析的に求めるのは極めて困難である⁵。

S/N の S

この S/N をモデルを仮定して計算してみる。 S/N の S の部分、検出器で検出される信号だとして、露出時間を t 、システムの透過率(大気、光学系、フィルター、素子の量子効率、など全部入り) f 、天体の光度、つまり単位時間当たりの光子数を L とすると、

$$S = f \cdot t \cdot L$$

となる。 S の単位は後で簡単にするため、光電荷数であるとしておく。

S/N の N, その1(読み出し雑音)

次に、 N 、誤差は、2つの項、機器のノイズ(読み出し雑音)と、光量の統計誤差の合成になる、というモデルを今回採用する。

読み出し雑音はとりあえず露出時間によらず一定であると考え、1ピクセルあたりのこの誤差を

$$N_1 = r$$

と書くことにする。単位はこれも電荷数に換算しておく。

この式はあくまで期待値の式である。あるピクセルでの読み出し雑音を X_1 とした場合に、 X_1 の標準偏差が r であるという意味で書かれており、

$$E(X_1) = 0$$

$$V(X_1) = E(X_1^2) = N_1^2 = r^2$$

の意味である。この $E(X)$ は X の平均値(期待値)、 $V(X)$ は X の分散、である。

²但し 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

³後者は時間調整不足で丸々カットだったけど…すみません

⁴ちなみに1.086という数字は $2.5/\ln(10)$

⁵正方形が N 個繋がった図形の数を数えるとかいう数学の未解決問題に関連してたりして、筆者も一度手をつけたまま放置中である

S/N の N, その2(ポアソン雑音)

光量の統計誤差は、観測される電荷の分布の広がり、ポアソン雑音あるいは光子雑音 (photon noise) と呼ばれる。

ポアソン分布

まず、可視や近赤外の場合、天体から出てくる光はポアソン分布に従い⁶、空からの光もポアソン分布に従っていると考え、またダーク電流もポアソン分布に従っていると考える(ことにする)。ポアソン分布とは二項分布で期待値一定のまま確率を 0 に近づけた極限で、期待値を k とし、

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

という分布になる。この平均と分散は k である。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が、定数倍 aX はポアソン分布に従わず、また、差 $X-Y$ もポアソン分布に従うとは限らない。これらの性質をこの後利用する。

素子に溜まる総電荷数を n とし、天体が入るピクセル数を A とすると、まず、総電荷数の期待値は

$$n = n_{object} + n_{sky} + n_{dark} = (ftL) + (ftAs_{sky}) + (tAs_{dark})$$

と書ける。 s_{sky} は 1 ピクセル当たりのバックグラウンドスカイの面輝度 (光子数/秒/ピクセル) で、 s_{dark} は単位時間当たり 1 ピクセル当たりの暗電流の量 (電荷数/秒) である。従って、

$$N_2 = \sqrt{n} = \sqrt{(ftL) + (ftAs_{sky}) + (tAs_{dark})}$$

となり、この N_1, N_2 の合成が誤差 N になる。

この N_2 の式も要注意で、すべて期待値でしかない。厳密に書けば、天体から来た光に由来する電荷数、背景のスカイに由来する電荷数、暗電流に由来する電荷数をそれぞれ、 $X_{object}, X_{sky}, X_{dark}$ と置いて、その総和 X_2 の標準偏差が N_2 である。

$$\begin{aligned} E(X_{object}) &= n_{object} \\ E(X_{sky}) &= n_{sky} \\ E(X_{dark}) &= n_{dark} \end{aligned}$$

であり、

$$V(X_2) = E(X_2^2) = (N_2^2) = n$$

である。

ポアソン統計に従うもの従わないもの

ここで、例えば X_{object} の期待値は ftL だが、 X_{object} はポアソン分布に従う L を ft 倍した変数ではない。まず、天体から 1 秒当たり 1 ピクセルに来る光は、期待値が L のポアソン分布に従う。ここは仮定というか前提である。これを透過率 f のシステムで t 秒積分した結果の光電荷数が X_{object} だが、透過率 f のシステムを通すことも、 t 秒積分することも、いずれも確率的に行われることで、比喻で言えばサイコロを振りなおして足し合わせることになるので、 X_{object} もポアソン分布に従うと考えられるのである。サイコロの比喻を続けると、あるポアソン分布に従う変数を定数倍する(電荷数を FITS のカウント数に変換する)のと、変数を確率的なフィルターを通すこと(光子がある確率で電荷になる)の違いは、サイコロ 1 つを振った出目を 100 倍することと、サイコロ 100 個を振った出目の総和を取る事の違いである。前者は離散的な一様分布だが、後者は正規分布に近づく。この違いは十分に理解し納得しておくことが重要である。

なお、更に追記しておく、素子に溜まる光電荷の量がポアソン分布と仮定してもよいのは、そもそも天体から来る光がポアソン分布であると仮定できることに依存している。例えば月面から光子を一定数、例えば毎秒 100 個なり 10000 個なりすばる望遠鏡めがけて打ち出して、それを観測した場合、観測される光電荷数がポアソン分布とはなるとは必ずしも限らず、特に CCD での撮像の場合は二項分布にはなってもポアソン近似はできないと予想される。なぜならシステム透過率がとてもよいからである。「観測結果の光電荷数が一般的にポアソンだと思えるのは、天体からの光がポアソン分布だと考えられるからである」と再度強調しておく。

⁶中間赤外辺りだと縮退の影響(光の波の性質)のせいでポアソンに従わないと以前話を聞いた記憶があるのだが、出典がどこだったか見失っている

正規分布

別名ガウス分布。誤差はこの分布に従うとよく仮定され、読み出し雑音もこれに従うと仮定される。平均 m 、標準偏差 σ の場合、正規分布 $N(m, \sigma^2)$ は、

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

である。

ポアソン分布は k が大きければ、正規分布でよく近似できるので、実は N_1+N_2 は正規分布の和と考えてもよい。なお、正規分布は線形性が成り立ち、

$$X_1 \sim N(m_1, \sigma_1^2)$$

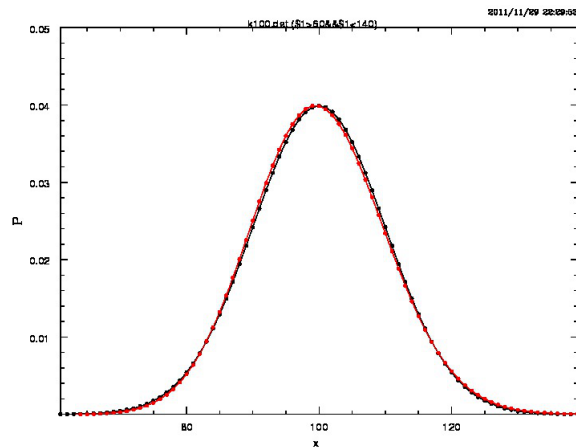
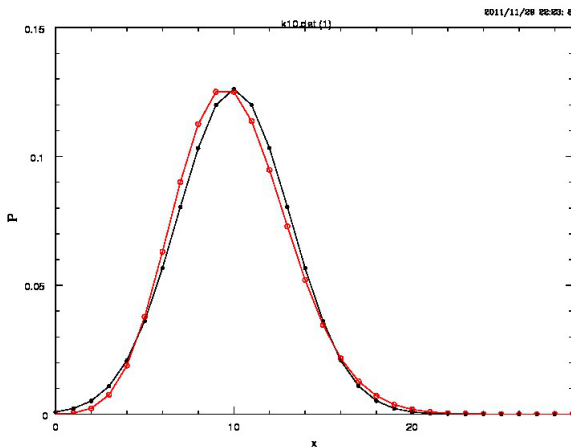
$$X_2 \sim N(m_2, \sigma_2^2)$$

のとき、

$$aX_1 + bX_2 \sim N(am_1 + bm_2, (a\sigma_1)^2 + (b\sigma_2)^2)$$

である⁷。

この「 k が大きければ」の参考までに、 $k=10$ と $k=100$ の時のポアソン分布と正規分布 $N(k,k)$ とをプロットしてみたのが以下である。



左が $k=10$ 、右が $k=100$ で、赤がポアソン、黒が正規分布である。正規分布で近似してよいかは、分布にどの程度の精度を期待するかによる。

誤差の伝播

N_2 がポアソンなのは OK として、 N_1 は違う分布である。そんな、2種の誤差源があった場合、その合成に用いられるのが「誤差の伝播の式」で、一般的には分散（標準偏差）が演算によってどのように変わるかをまとめたものである。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、

1. 変数 X と誤差 δX を求めておく。 Y も同様に δY を求めておく。
2. 期待値 $E(f(X, Y))$ を計算する。
3. 分散 $V(f(X, Y)) = E((f(X, Y) - E(f(X, Y))))^2$ を計算し、その平方根を取る。つまり観測値から期待値を引いた二乗平均を求める。

という方針で求められる。また、この際、期待値や分散の性質、

$$E(aX + bY) = aE(X) + bE(Y)$$

$$V(aX) = a^2V(X)$$

⁷ちなみにこの証明はかなり面倒くさかったような記憶がある。我こそと思う人は挑戦してみてください

などを使う。なお、これらの性質は分布の平均や分散の定義

$$E(X) = \sum Xp(X)$$

$$V(X) = \sum (X - E(X))^2 p(X)$$

から導くことができる。

$f(X, Y) = X + Y$ について試してみる。まず X, Y の平均を、 μ_X, μ_Y と書き、標準偏差を σ_X, σ_Y と書く。この時、

$$E(X + Y) = \mu_X + \mu_Y$$

であり、

$$\begin{aligned} V(X + Y) &= E(((X + Y) - (\mu_X + \mu_Y))^2) \\ &= E(((X - \mu_X) + (Y - \mu_Y))^2) \\ &= E((X - \mu_X)^2) + 2E((X - \mu_X)(Y - \mu_Y)) + E((Y - \mu_Y)^2) \\ &= \sigma_X^2 + \sigma_Y^2 + 2E((X - \mu_X)(Y - \mu_Y)) \end{aligned}$$

と計算できる。最後の項、 $E((X - \mu_X)(Y - \mu_Y))$ を共分散と呼び、 $Cov(X, Y)$ と書く。共分散が 0 の時、「 X, Y は独立」と言い、この時は

$$V(X + Y) = \sigma_X^2 + \sigma_Y^2$$

であり、誤差は

$$\sigma_{X+Y} = \sqrt{V(X + Y)} = \sqrt{\sigma_X^2 + \sigma_Y^2}$$

となる。誤差の伝播で変数が独立かどうかは極めて重要なのだが、実際のデータでは独立かどうか判断するのは、簡単とは限らない⁸。

S/N まとめ

今回の S/N 計算では、各ピクセルの読み出し雑音とポアソン雑音はすべて独立と考えられるので、目標天体の光を含むピクセルの数を A と置くと、誤差 N は、

$$N = AN_1 + N_2 = \sqrt{Ar^2 + (ftL) + (ftAs_{sky}) + (tAs_{dark})}$$

となり、

$$\frac{S}{N} = \frac{ftL}{\sqrt{Ar^2 + (ftL) + (ftAs_{sky}) + (tAs_{dark})}}$$

となる。

t が十分小さい場合、S/N は、

$$\frac{S}{N} \sim \frac{fL}{r} t$$

と、 t に比例、 t が十分大きい場合は、

$$\frac{S}{N} \sim \frac{fL}{\sqrt{f(L + As_{sky}) + As_{dark}}} \sqrt{t}$$

となる。露出時間が長い場合は S/N は露出時間の平方根に比例するので、S/N を倍にしたい場合は、露出時間は 4 倍必要である。追記しておく、等級の誤差は S/N に比例するので、等級の誤差を例えば 0.1 等から 0.05 等にした場合も、やはり 4 倍の露出時間が必要となる。

⁸例えば、今回も口頭でちょっと触れられていたが、赤外アレイでピクセル間に相関があったなどの調査報告もある

4 例2:画像の足し合わせと誤差

1回の観測での S/N は前節のように求められるのだが、実際には露出時間には、サチリ、ガイドの精度、宇宙線の増加などの問題で上限があるため、露出時間をいくつか分割して撮ることになり、これらを足し合わせた結果は、必ずしも総露出時間の平方根に比例して S/N が上がるとは限らない。これを考えてみる。

複数回の露出を足し合わせる際の誤差は、本来は位置ずれや P S F の変動の影響など様々な要素を考えなければならないのだが、「あるピクセルの中に来る flux の分布は一定」という仮定で、mean、median、clipped mean を比較してみる。

平均と中央値と clipped mean

ここでは(講義内容とは若干ズレるが) プラス方向に偏った外れ値入りの正規分布で、パラメタ推定を行なうことを考える。これはまあ宇宙線が良く入る CCD の観測データのモデルだと思ってよい。具体的には X が平均 0、標準偏差 1 の正規分布 $N(0,1)$ に従う N 個と、大きな外れ値 x(例えば a=100) H 個を含むようなモデルを考える。確率密度関数は $h=H/(N+H)$ を使って、

$$f(x) = \frac{(1-h)}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + h\delta(x-a)$$

と書ける。標本平均は $E(X)=ah$ である。証明略。

標本の中央値は、正規分布から N 個サンプルして、その大きい方から $(N-H+1)/2$ 番目を取ることに相当する。離散効果を無視すると誤差関数を含む以下の方程式を解くことになる。

$$\operatorname{erf}(x) = \frac{1}{2(1-h)}$$

誤差関数の定義は以下を用いている。

$$\operatorname{erf}(x) = \int_{-\infty}^x e^{-\frac{z^2}{2}} dz$$

例えば h をいくつか例にとって解くと、以下の表ようになる。

H	x
0.001	0.00089
0.01	0.0090
0.05	0.047
0.1	0.099
0.2	0.23
0.3	0.40

例えば 10 枚に 1 枚高い側の外れ値が乗る場合、単純に 10 枚の中央値を取ると、およそ 0.1σ ほど高いほうに偏る結果となることがわかる。

次に推定値の分散を計算してみる。議論を簡単にするため、一旦外れ値を除いて考えると、X の分散は定義より 1。標本の標準偏差を計算すると、N 個の標本平均 $E(X)$ は

$$E(X) = \frac{1}{N} \sum_{i=1}^N X_i$$

であり、各 X_i は独立なので、

$$V(E(X)) = \frac{1}{N^2} \sum_{i=1}^N V(X_i) = \frac{1}{N}$$

なので、標準偏差は $1/\sqrt{N}$ となる。

中央値の場合、分散を求めるのはまたも数値計算になってしまうのだが、まず、中央値の確率分布を求める。一般に $f(x)$ に従う変数 X に関し、標本数 N が奇数の場合の中央値 y の分布 $g(y)$ は 1 つが y $(N-1)/2$ 個が y 以上、 $(N-1)/2$ 個が y 以下になる確率になるので、 $k=(N-1)/2$ と置いて、

$$g(y) = {}_N C_k {}_{N-k} C_k f(y) \left(\int_{-\infty}^y f(x) dx \right)^k \left(\int_y^{\infty} f(x) dx \right)^k$$

と書ける。この分布について分散を求めると、 N が十分大きいところで、 $\pi/2N$ に漸近する。従って、中央値の標準偏差は

$$\sqrt{\frac{\pi}{2N}} \sim \frac{1.25}{\sqrt{N}}$$

となって、平均よりも 25%ほど大きくなる。この比は N が小さくなると若干緩和する。

このように平均は外れ値に弱く、中央値は標準偏差（バラツキ）が悪くなるのだが、この両方の欠点を補うような方法が clipped mean である。これは最初に代表値を推定し、そこから大きく外れた値は除去 (clip) して、残ったデータで平均を取る手法で、うまく clip できれば、外れ値の影響は 0、標準偏差も $1/\sqrt{N(1-h)}$ となる。clip される標本が 1 つの場合と、median での標準偏差を、単純平均との比として表にすると以下ようになる。

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

median の標準偏差が偶数のところで小さくなっているのは、偶数の median は中央の 2 つの平均をとっているからである。このように大雑把に言って $N > 4$ では、median よりも clipped mean の方がランダム誤差が少なくなる。

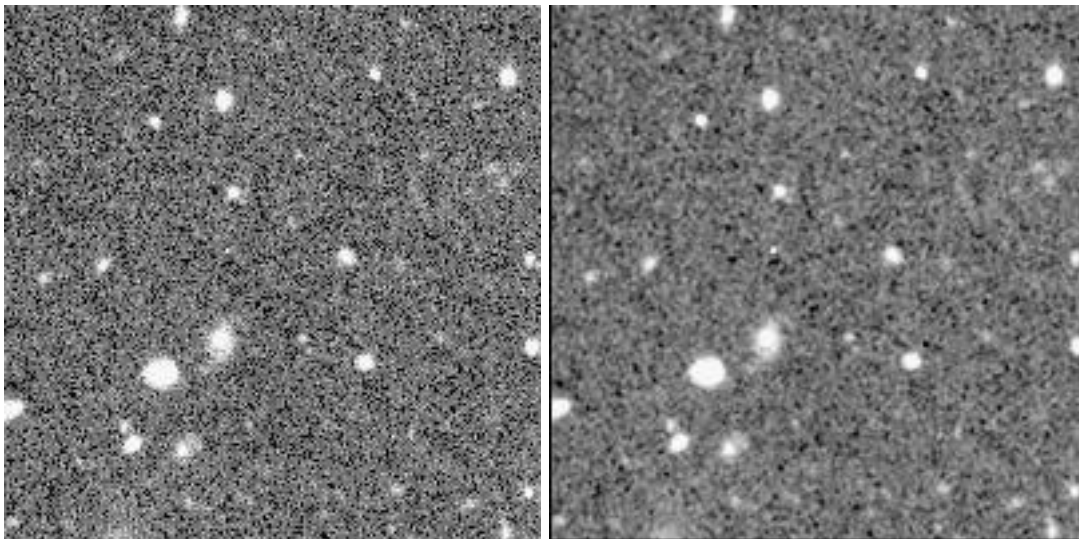
問題は「最初の代表値の推定」と、「大きく外れた」という、しきい値の設定方法であり、どの値を基準にどの範囲の平均を取って clipped mean とするか、その「標準的な方法」というのは、多分ない。clip がうまくいかない場合、 1σ 程度以下の系統誤差、例えば画像の足し合わせの際のバックグラウンドの引き残りなどに対しては、median は引きずられにくい、clipped mean は引きずられる。

5 例 3:画像の平行移動

画像変形に伴う S/N の変化に関しては、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⁹。画像・データを非整数ピクセル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、

$$v'(x') = (1-a)v(x) + av(x+1)$$

である。実際にシフトさせてみたのが図である。



⁹この節は時間切れで全然話せなかったのであるよ

左がシフト前、右は $x+0.5, y+0.5$ シフトである。見た目でもざらつきが減っているのだが¹⁰、実際測るとピクセルあたりのノイズはシフト前 67ADU、シフト後 43ADU と、確かに減っている。すると並行移動で S/N が上げられる！と一瞬思うのは無理もない。

元画像での各点での標準偏差を σ とする。X 方向の移動だけ考えて

$$v'(x') = (1 - a)v(x) + av(x + 1)$$

とする。もし元画像で各点の誤差が独立であれば、並行移動後の各点での分散は

$$V((1 - a)v(x) + av(x + 1)) = (1 - a)^2\sigma^2 + a^2\sigma^2$$

となり、例えば $a=0.5$ の時、 $0.5\sigma^2$ となるので、確かに元画像の σ^2 よりも小さい。

しかし、並行移動操作の結果、隣り合った点の値は既に独立ではなくなっているため、広い範囲で見た場合の S/N は増えないのである。

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

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

次に、並行移動後の各点の値は

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

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

...

$$v'(N - 1) = (1 - a)v(N - 1) + av(N)$$

最初の 2 点の共分散を計算すると、

$$\text{Cov}(v'(0), v'(1)) = \text{Cov}((1 - a)v(0) + av(1), (1 - a)v(1) + av(2))$$

ここで、共分散の線形性などを利用して展開する。

$$\text{Cov}(v'(0), v'(1)) = (1 - a)^2\text{Cov}(v(0), v(1)) + a^2\text{Cov}(v(1), v(2)) + a(1 - a)\text{Cov}(v(1), v(1)) + a(1 - a)\text{Cov}(v(0), v(2))$$

ここで、 $v(x)$ は独立なので、 $\text{Cov}(v(1), v(1)) = V(v(1)) = \sigma^2$ だけが残って、

$$\text{Cov}(v'(0), v'(1)) = a(1 - a)\sigma^2$$

この結果、

$$V(v'(0) + v'(1)) = V(v'(0)) + V(v'(1)) + 2\text{Cov}(v'(0), v'(1)) = 2(1 - a(1 - a))\sigma^2$$

となる。更に幅を広げて同様に計算できるが、

$$v'(0) + \dots + v'(N - 1) = (1 - a)v(0) + v(1) + \dots + v(N - 1) + av(N)$$

と書いてやれば、各項は独立なので、この分散は

$$((1 - a)^2 + (N - 1) + a^2)\sigma^2$$

と簡単に計算できる。この結果、 N が大きくなると平行移動の効果が全くない事がわかる。並行移動で画像が綺麗になっても、ある範囲で積分した値、つまり天体で見た時の限界等級は上がらない。S/N は上がってないのに、見かけ上、測定したピクセルあたりの RMS は小さくなるのであった。

一般的に、何らかの複数ピクセルを跨ぐような操作を行なった後の画像から測定した RMS から元の画像のピクセルあたりの誤差を推定するのは困難である。解析ソフトウェア等で、ピクセルあたりのノイズをパラメータとして与える事が求められたりするが、どのような数値が求められているのか？を正しく理解したうえで使わないと道を誤る。

このように、足し合わせや並行移動により、各ピクセルの誤差やある領域内の誤差は変わりうる。特に、バラツキ増加（ランダム誤差増加）ではなく、系統誤差が生じる場合は、データからは判断できない場合があるので、このようなモデルからの検証を行ない、どの操作がどう影響するかは把握し理解しておくことが望ましい。

¹⁰見える？