Creating Histogram Grids

April 30, 2021

1 Linear Grids

1.1 General Statements

Given

R_{min} **&** R_{max}
: are the soft bounds of the range to be covered. The actual range must include them, but may be larger.

E_i
: are the bin edges, with edges $[E_{min}, E_{max}]$ covering the range $[R_{min}, R_{max}]$

P
: is the fiducial alignment position

i
: is the bin index, whose origin is defined such that $E_0 \leq P \leq E_1$, with

$$i = \begin{cases} i_{min}, & E_i = E_{min} \\ i_{max}, & E_i = E_{max} \end{cases}$$

w
: is the bin width;

f

is the fractional offset from the alignment position to the left edge of the bin containing it, i.e., $E_0 = P - fw$

n

is the number of bins

Here's what passes for general expressions for the minimum and maximum bin values required to cover the range.

$$\begin{aligned}
E_{min} &< R_{min} \\
E_0 + i_{min} w &< R_{min} \\
E_0 + \left\lfloor \frac{R_{min} - E_0}{w} \right\rfloor w &< R_{min} \\
P - fw + \left\lfloor \frac{R_{min} - (P - fw)}{w} \right\rfloor w &< R_{min} \\
P + \left(\left\lfloor \frac{R_{min} - P}{w} + f \right\rfloor - f \right) w &< R_{min} \qquad (1)
\end{aligned}$$

Similarly,

$$\begin{aligned}
E_{max} &> R_{max} \\
E_0 + i_{max} w &> R_{max} \\
E_0 + \left\lceil \frac{R_{max} - E_0}{w} \right\rceil w &> R_{max} \\
P + \left(\left\lceil \frac{R_{max} - P}{w} + f \right\rceil - f \right) w &> R_{max} \qquad (2)
\end{aligned}$$

Note that these use

$$\begin{aligned}
i_{min} &= \left\lfloor \frac{R_{min} - P}{w} + f \right\rfloor \qquad (3) \\
i_{max} &= \left\lceil \frac{R_{max} - P}{w} + f \right\rceil
\end{aligned}$$

so $i_{max} - i_{min}$ is *not necessarily* n. One must choose either i_{max} or i_{min} as the fiducial index and calculate the other using n. The tricky part is that the the expressions for i_{min} and i_{max} are integral, which makes solving this a bit difficult.

1.2 Aligned bins, Fixed w, variable n

Given: Δ, P, R_{max}, R_{min}, w.

$$\begin{aligned} E_0 &= P - fw \\ i_{min} &= \left\lfloor \frac{R_{min} - E_0}{w} \right\rfloor \\ i_{max} &= \left\lceil \frac{R_{max} - E_0}{w} \right\rceil \\ n &= i_{max} - i_{min} + 1 \end{aligned}$$

1.3 Aligned bins, Fixed n, variable w

Given: Δ, P, R_{max}, R_{min}, n.
Wanted: minimum $w \ni w \geq \frac{R_{min} - R_{max}}{n}$
Because Eqs. 1 and 2 are painful to solve, let's see if we can figure things out another way.
Once we have a bin width such that bin edges $[E_{min}, E_{max}]$ cover our data range, $[R_{min}, R_{max}]$, attending to alignment with the fiducial point P is a simple translation of the bins. P's position relative to its containing bin is given by fw, so it's a periodic condition (it doesn't matter *which* bin it's in) and the maximum we need to translate is exactly one bin. Given n bins, $n-1$ bins will cover the data range, with the extra bin used to accommodate the alignment shift, or

$$w = \begin{cases} \frac{R_{max} - R_{min}}{n} & \text{no alignment} \\ \frac{R_{max} - R_{min}}{n-1} & \text{with alignment} \end{cases}$$

Is it optimal? Is there a smaller w which allows for proper alignment? Why bother?

Well, it'd be nice to have a "nice" value for w (say some exponent of 10, or a rational number), rather than a random sequence of digits, so if we can find a viable range for w there might be a "nice" value in that range.

Eqs. 3 present a problem, but we can simplify things if we restrict w so that they remain a constant.

2 Ratio (geometric series) binning

R
: A soft bound of the range to be covered. The actual range must include it, but may be larger. Only one soft bound is allowed.

E_0
: is the fiducial bin edge, with

$$E_0 = \begin{cases} E_{min}, & w > 0 \\ E_{max}, & w < 0 \end{cases}$$

E_n
: is at the opposite extremum from the fiducial bin edge, with

$$E_n = \begin{cases} E_{max}, & w > 0 \\ E_{min}, & w < 0 \end{cases}$$

ΔR
: is the actual range covered, $E_n - E_0$

i
: is the bin index, with

$$i = \begin{cases} i_{min}, & E_i = E_{min} \\ i_{max}, & E_i = E_{max} \end{cases}$$

E_i
: are the bin edges, such that $[E_{min}, E_{max}]$ covers the range, which may be either $[E_{min}, R]$ or $[R, E_{max}]$

w

: is the width of the fiducial bin, e.g. $w = E_1 - E_0$; w may be negative, indicating that bin widths increase towards $-\infty$

r

: is the ratio of each bin relative to its neighbor. $r > 0, r \neq 1$

n

: is the number of bins

Geometrically binned grids follow the scheme:

$$
\begin{aligned}
E_1 &= E_0 + w \\
E_2 &= E_1 + wr \\
E_3 &= E_2 + wr^2 \\
&\cdots \\
E_n &= E_{n-1} + wr^{n-1} \\
&= E_0 + \sum_{i=0}^{n-1} wr^i
\end{aligned}
$$

For $r \neq 1$,

$$
\begin{aligned}
(E_k - E_0) &= \sum_{i=0}^{k-1} wr^i \\
r(E_k - E_0) &= \sum_{i=0}^{k-1} wr^{i+1} \\
(E_k - E_0) - r(E_k - E_0) &= \sum_{i=0}^{k-1} wr^i - \sum_{i=0}^{k-1} wr^{i+1} \\
(E_k - E_0)(1-r) &= w + \left(\sum_{i=1}^{k-1} wr^i - \sum_{i=0}^{k-2} wr^{i+1} \right) - wr^k \\
&= w - wr^k + \left(\sum_{i=1}^{k-1} wr^i - \sum_{i=1}^{k-1} wr^i \right) \\
&= w(1 - r^k) \\
E_k &= E_0 + w\frac{(1-r^k)}{(11-r)} \qquad (4)
\end{aligned}
$$

$r = 1$ implies a linear grid, so we'll ignore it.
For $|r| < 1$,

$$
\begin{aligned}
E_\infty &= E_0 + \frac{w}{1-r} \qquad (5) \\
|\Delta R_{max}| &= \left| \frac{w}{1-r} \right| \qquad (6)
\end{aligned}
$$

Given a range, $[E_0, E_n]$, what is n? From Eq. 4,

$$\begin{aligned}
(E_n - E_0) &= w\frac{1-r^n}{1-r} \\
\Delta R &= w\frac{1-r^n}{1-r} \\
\frac{(1-r)\Delta R}{w} &= 1 - r^n \\
r^n &= \frac{w-(1-r)\Delta R}{w} \\
n &= \ln\left(\frac{w-(1-r)\Delta R}{w}\right) / \ln(r)
\end{aligned} \tag{7}$$

If one of the extrema is soft, e.g.

$$\Delta R = \begin{cases} R - E_0 \\ E_n - R \end{cases}$$

Then

$$n(\Delta R) = \left\lceil \ln\left(\frac{w-(1-r)\Delta R}{w}\right) / \ln(r) \right\rceil \tag{8}$$

2.1 E_0, w, r, n

This one is easy, generate the bin edges with Eq. 4. Just note that if $w < 0$ then the edges will be generated in decreasing order, *i.e.*, $E_{i+1} < E_i$.

2.2 $[E_{min}, R]$, w, r

The grid must cover $[E_{min}, R]$. The sign of w indicates whether E_{min} is the fiducial bin edge:

$$
E_{min} \equiv \begin{cases} E_n & w < 0 \\ E_0 & w > 0 \end{cases}
$$

If $|r| < 1$, it is possible that the prescribed grid cannot cover the range (Eq. 5).

The number of bins, n, can be determined from Eq. 8. If $E_{min} \equiv E_n$,then E_0 may be determined from Eq. 4, which in any case provides the remaining bin edges.

2.3 $[R, E_{max}]$, w, r

Similar to the last section, just note that

$$
E_{max} \equiv \begin{cases} E_0 & w < 0 \\ E_n & w > 0 \end{cases}
$$

2.4 E_0, $[R_{min}, R_{max}]$,w, r

If $E_0$is not at one of the grid extrema, i_{min} and i_{max} can be determined from Eq. 8, with

$$
\begin{aligned}
i_{min} &= n(R_{min} - E_0) - 1 \\
i_{max} &= n(R_{max} - E_0)
\end{aligned}
$$