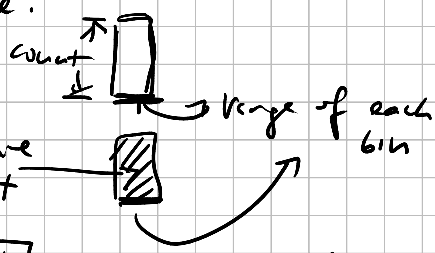
stationarity (time series) — counterpart of identical dist in TS setting.

(later: independence?)

histogram
- freq on the vertical axis — count
- density on the vertical axis — relative count

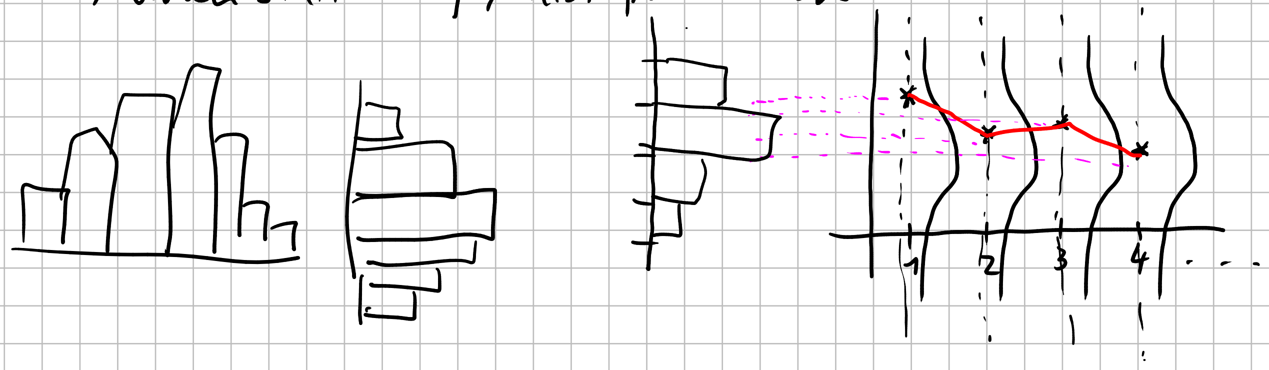


range of each bin

area = length of each bin * height

" rel. count — $\underbrace{\text{height}}_{\text{density}}$

Why can I visualize TS data using a histogram?
- → not at all fine (does not make sense if you have nonstationarities)
- → under stationarity, histograms are ok



kernel density estimator

- kernel function $K(\cdot)$ — user-specified (defaults available)

  in practice, this choice does not matter so much

  but there are requirements for a valid kernel function ^tech.

  but there are exceptions!

- bandwidth (b) — user-specified (defaults available)

  in practice, this is the crucial component.

dummy/notation

$f(y)$ density

$\hat{f}(y)$ estimator of density

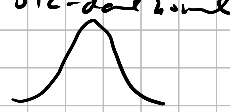
inputs

$$\hat{f}(y) = \frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{y-Y_i}{b}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \boxed{\frac{1}{b}K\left(\frac{y-Y_i}{b}\right)}$$

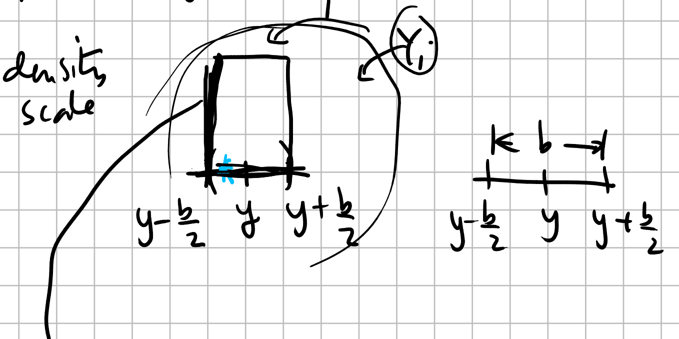
Y_i's are data

typically standard normal



relate to histograms

density scale



indicator

$$\mathbb{1}\left(y-\frac{b}{2} \le Y_i \le y+\frac{b}{2}\right) \begin{cases} 1 \\ 0 \end{cases}$$

histogram estimate of density $= \boxed{\hat{f}(\boxed{y})} = \frac{1}{nb}\sum_{i=1}^{n} \boxed{\underbrace{\mathbb{1}\left(y-\frac{b}{2} \le Y_i \le y+\frac{b}{2}\right)}_{\text{Is } Y_i \text{ inside the bin?}}}$

histogram estimate of density

$$\hat{f}(y) = \frac{1}{nb}\sum_{i=1}^{n} \mathbb{1}\left(y-\frac{b}{2} \le \underline{Y_i} \le y+\frac{b}{2}\right)$$

<u>IID</u>

$$\underline{\mathbb{E}(\hat{f}(y))} = \frac{1}{nb}\sum_{i=1}^{n} \mathbb{P}\left(y-\frac{b}{2} \le Y_i \le y+\frac{b}{2}\right)$$

$$= \frac{1}{b}\mathbb{P}\left(y-\frac{b}{2} \le Y_i \le y+\frac{b}{2}\right)$$

$$= \frac{1}{b}\int_{y-\frac{b}{2}}^{y+\frac{b}{2}} \underline{f(s)}\, ds$$

true density

F cdf, f density

$F' = f$

$$= \frac{1}{b}\left[F\left(y+\frac{b}{2}\right) - F\left(y-\frac{b}{2}\right)\right]$$

$$\mathrm{Var}(\hat{f}(y)) = \frac{1}{n^2b^2}\sum_{i=1}^{n}\mathrm{Var}\left(\mathbb{1}\left(y-\frac{b}{2} \le Y_i \le y+\frac{b}{2}\right)\right)$$

$$= \frac{1}{nb^2}\mathrm{Var}\left(\mathbb{1}\left(y-\frac{b}{2} \le Y_i \le y+\frac{b}{2}\right)\right)$$

$$= \frac{1}{nb^2}\mathbb{P}\left(y-\frac{b}{2} \le Y_i \le y+\frac{b}{2}\right)\left(1-\mathbb{P}\left(y-\frac{b}{2} \le Y_i \le y+\frac{b}{2}\right)\right)$$

$$= \frac{1}{nb^2}\left(F\left(y+\frac{b}{2}\right)-F\left(y-\frac{b}{2}\right)\right)\left(1-\left(F\left(y+\frac{b}{2}\right)-F\left(y-\frac{b}{2}\right)\right)\right)$$

You want $\mathbb{E}(\hat{f}(y)) = f(y)$ but this not possible

Could show that $\underbrace{\mathbb{E}(\hat{f}(y)) - f(y)}_{\text{bias of } \hat{f}(y)} \le \boxed{L}\,\boxed{b}$ (no n is involved)

related to derivative of f

$$\mathrm{Var}(\hat{f}(y)) = \frac{1}{nb}\left(\underline{f}(y^*)(1-\underline{f}(y^*))\right) \qquad y^* \in \underline{\text{interval}}$$

(n is involved)

<u>MSE</u> = <u>bias</u>^{2} + <u>variance</u>