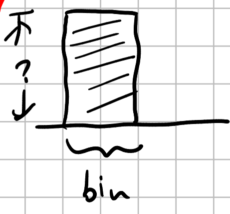
stationarity (time series) — counterpart of identical distribution (in IID setting)

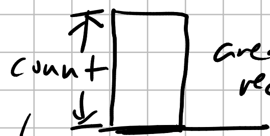
independence is harder to justify in time series

histogram
- frequency is on the vertical axis
- density is on the vertical axis — more useful



how long is the bin/interval?

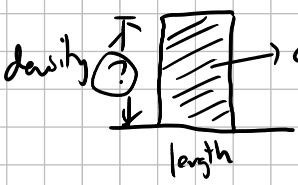
Frequency version



area of rectangle has no meaning

↳ how many observations belong to the bin?

density version



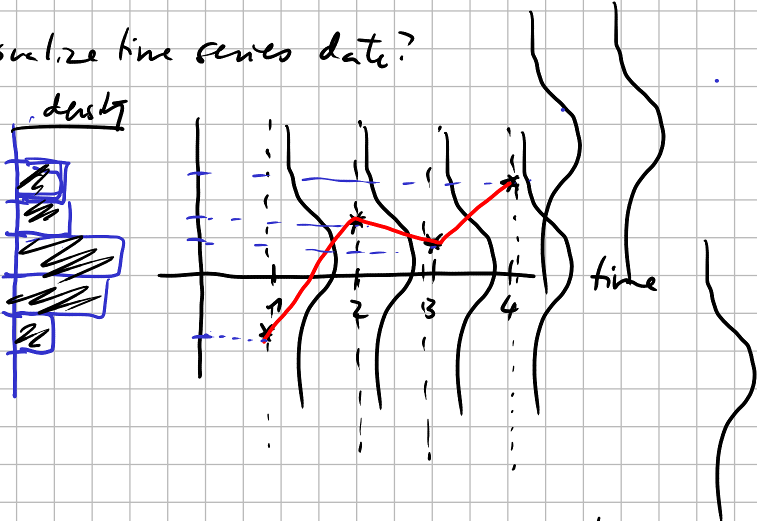
area = relative count

prop. of obs. that belong to the bin

area = (length of interval) * (height of rectangle)

Why can we use a histogram to visualize time series data?

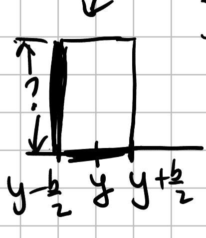
- stationarity
- we may be situations where a histogram is not appropriate

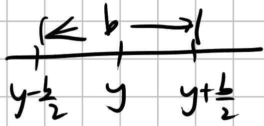


kernel density estimators
- kernel function: user-specified function which must satisfy some technical conditions. (default is usually standard normal density function)
  - not very crucial, but there are exceptions!
- bandwidth: user-specified (defaults available in software)
  - very crucial

histogram — also a density estimator

Y_i data

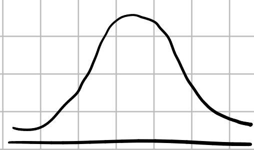




$$\hat{f}(y) = \frac{1}{nb}\sum_{i=1}^{n} \underset{\substack{\uparrow \\ \text{indicator}}}{\mathbb{1}}\left(y-\frac{b}{2} \le Y_i \le y+\frac{b}{2}\right) \begin{cases} 1 \\ 0 \end{cases}$$

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

e.g. standard normal



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

Y_i ↳ r.v.

To estimate $f(y)$

Y_i's IID

$$\mathbb{E}(\hat{f}(y)) = \mathbb{E}\left(\frac{1}{nb}\sum_{i=1}^{n} \underbrace{\mathbb{1}\left(y-\frac{b}{2} \le Y_i \le y+\frac{b}{2}\right)}_{\text{Bernoulli r.v.}}\right)$$

$$= \frac{1}{nb}\boxed{\sum_{i=1}^{n}} \underbrace{\mathbb{P}\left(y-\frac{b}{2} \le Y_i \le y+\frac{b}{2}\right)}_{\text{same across all } i}$$

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

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

$$= \frac{1}{b}\left[F\left(y+\frac{b}{2}\right) - F\left(y-\frac{b}{2}\right)\right] \qquad \text{not exactly } f(y)$$

would be shown that

$$\underbrace{\mathbb{E}(\hat{f}(y)) - f(y)}_{\text{bias of density estimator}} \le L\,b$$

(n has no role at all!)

L depends only on the first deriv of f

even if b is 0 (which is impossible), bias does not disappear

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

$$\vdots = \frac{1}{nb}\, f(y^*)(1-f(y^*)) \qquad y^* \in \text{interval}$$

(n plays a role in decreasing variance)

MSE = bias^2 + variance

↑ mean squared error

choose b to min MSE