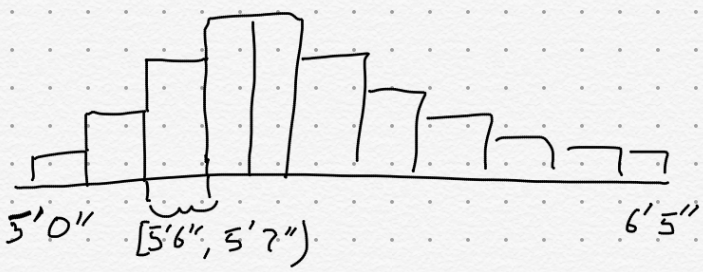
# Approximate Counting

Given $x_1, \ldots, x_n$, such that $x_i \in \Sigma = \{\text{red}, \text{blue}, \ldots\}$

Count frequencies of red, blue, green
- ↳ query($\sigma$) → # of times $\sigma \in \Sigma$ shows up
- ↳ $O(|\Sigma|)$ for 100% acc
- ↳ Approximate with count-min sketch ($O(m)$ $m \ll |\Sigma|$)

What if $\Sigma \to \mathbb{Z}_{(0,D)}$?
- ↳ better to use histogram because of the smoothness structure
  - ↳ values near each other have similar # of occurances

Ex. Heights of students




Query(5'6")
Query(5'7") } relatively closer
Query(6'0")

Aside: Over- and Under- fitting

Overfitting
- ↳ memorizing dataset / learning concepts that won't generalize
- ✗ ↳ Occurs when lots of free parameters

Underfitting → Opposite
- ↳ too few parameters

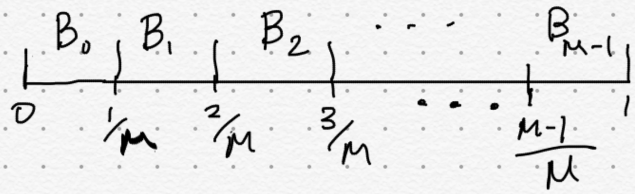
# Histograms

① Observe $x_1, \dots, x_n$

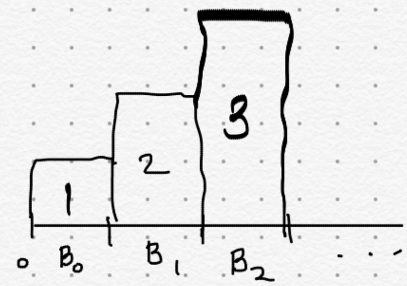
↳ $x_i$'s are iid

↳ $x_i \in [0,1]$ (normalized data)

② "Sketch": Create histogram to $M$ bins




↳ maintain count over $n$ →




↳ Normalize counts to have a valid probability density

↳ $\hat{p}(B_\ell) = \left(\frac{M}{n}\right) |B_\ell|$ ← width of bucket

↳ count in bucket $\ell$

↳ $x_i$'s are sampled from density function $p(x)$

↳ $x_i \sim p(x)$

↳ $\hat{p}(x)$ approximates $p(x)$

↳ $O(M)$ parameters

↳ If $M$ is too small,

↳ Bad estimate because miss features of data (over smoothing)

↳ If $M$ is too big, bad estimate b/c risk noise

↳ Intuition

| | Error |
|---|---|
| Bin is 100% of data | $\frac{\sigma}{\sqrt{n}}$ |
| Bin is 1% of data | $\frac{\sigma}{\sqrt{1\% \text{ of } n}}$ |

# Histograms cont'd

→ $X_1, \dots, X_n$ iid where $x_i \sim p(x)$

→ Assume smoothness → <u>Lipschitz smoothness/continuity</u> → (L-Lipschitz)

↳ $|p(x) - p(y)| \leq L|x-y|$

→ $\hat{p}(x) = \frac{M}{n} \cdot \underbrace{\text{count}(B_{(x)})}_{}$

↳ count in bucket $x$ is in

<u>Goal</u>: compare $\hat{p}(x)$ to $p(x)$

$$E(\hat{p}(x)) = M \cdot P(\underbrace{x \in B_\ell}_{})$$

↳ $x$ in bucket $B_\ell$

$$= M \cdot \int_{\frac{\ell-1}{M}}^{\ell/M} p(v)\, dv$$

$$= M \cdot \left[F(\ell/M) - F\left(\tfrac{\ell-1}{M}\right)\right]$$

→ $F$ is cdf of $p(x)$

$$= \frac{F(\ell/M) - F\left(\frac{\ell-1}{M}\right)}{1/M} = \frac{F\left(\frac{\ell}{M}\right) - F\left(\frac{\ell-1}{M}\right)}{\frac{\ell}{M} - \frac{\ell-1}{M}}$$

→ looks like derivative

By the mean value thm

$= p(x^*)$ for some $x^* \in \left[\frac{\ell-1}{M}, \frac{\ell}{M}\right)$

$$E(\hat{p}(x)) = p(x^*)$$

$$E(\hat{p}(x)) - p(x) = p(x^*) - p(x)$$
$$= L \cdot |x^* - x| \leq \frac{L}{M}$$

→ Bias related to smoothness and number of bins

# Histograms cont'd

Bias is $\frac{L}{M}$

Variance: $Var(\hat{p}(x)) = M^2 \cdot var\left(\frac{1}{n}\sum_{i=1}^{n} I(x_i \in B_\ell)\right)$

$$= M^2 \frac{P(x \in B_\ell)\cdot(1 - P(x \in B_\ell))}{n}$$

Assume roughly uniform distribution $\Rightarrow P(x \in B_\ell) = \frac{1}{M}$

$$Var(\hat{p}(x)) = M^2 \frac{\frac{1}{M}\left(1 - \frac{1}{M}\right)}{n}$$

$$= \frac{M-1}{n}$$

$MSE = Bias^2 + Var$

$f(M) = \frac{L^2}{M^2} + \frac{M-1}{n}$ (for uniform distribution)

$f'(M) = \frac{-2L^2}{M^3} + \frac{1}{N} \rightarrow$ Optimal $M$ is

$$M^3 = 2L^2 n$$

$$M = \sqrt[3]{2L^2 n} = O(\sqrt[3]{n})$$

For roughly uniform $p(x)$, $M_{opt} = \sqrt[3]{2L^2 n} = O(\sqrt[3]{n})$

Assuming nonuniformity,

$var = M^2 \frac{P(x \in B_e)(1 - P(x \in B_e))}{n}$ $\quad Max(P(x \in B_e)(1 - P(x \in B_e)) = \frac{1}{4}$

$var \le \frac{M^2}{4n}$

$\therefore MSE = \frac{L^2}{M^2} + \frac{M^2}{4n}$

$f'(M) = \frac{-2L^2}{M} + \frac{2M}{4N}$

$M_{opt} = \sqrt[4]{4L^2 N} \Longrightarrow O(M_{opt}) \le \underbrace{\frac{13}{10}}_{\approx \sqrt[3]{2}} \sqrt[3]{L^2 N}$ for roughly uniform

$O(M_{opt}) \le \underbrace{\frac{3}{2}}_{\approx \sqrt[4]{4}} \sqrt[4]{L^2 N}$ for very not uniform

Note: Smoothness term $L$ allows for $M_{opt}$ to be cubed / fourth root rather than linear with data.

Intuition: know one point $\rightarrow$ know points near by!!