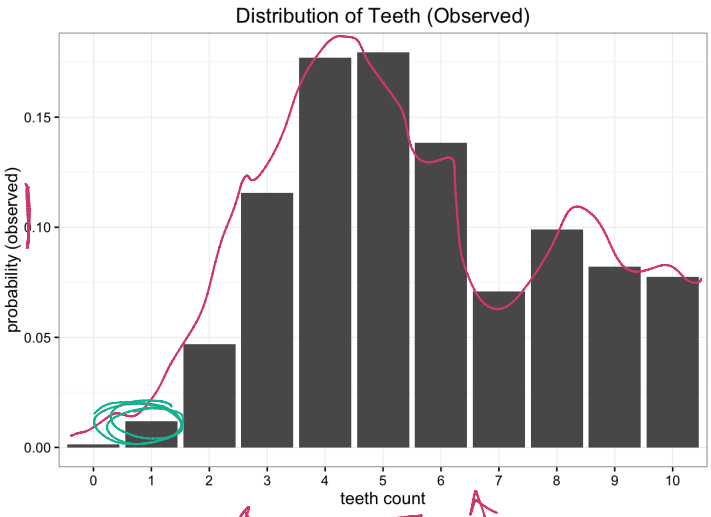
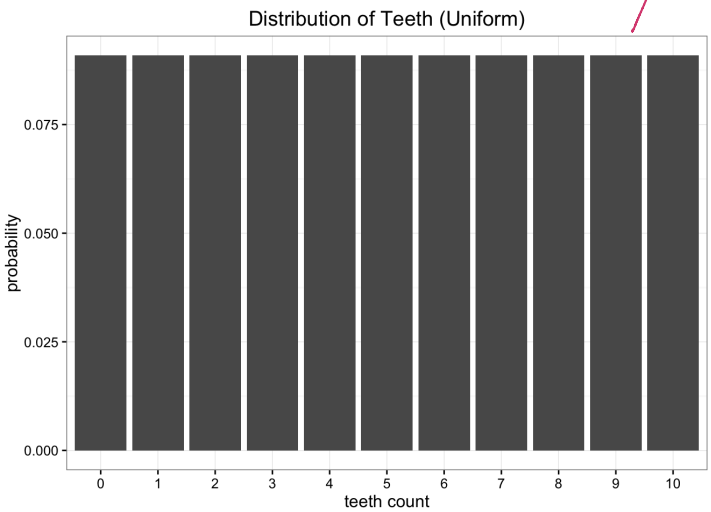
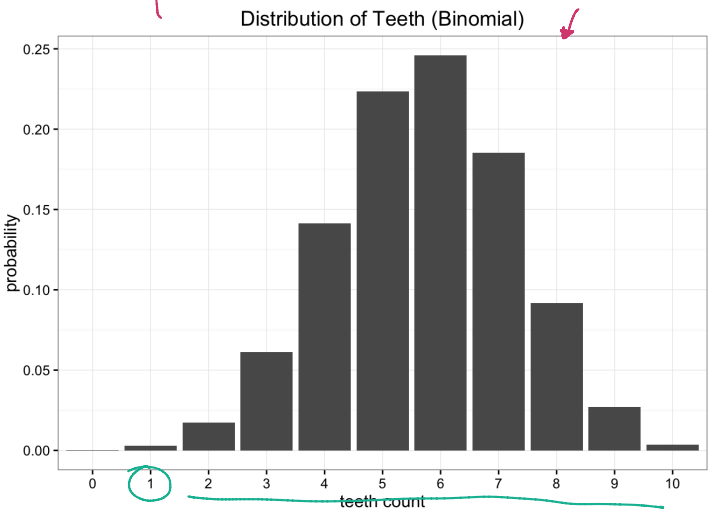
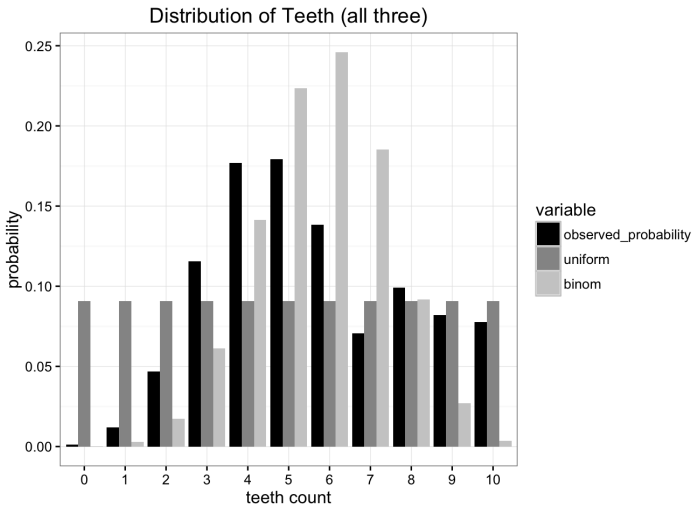
Estimating Distributions







$$\sum_{x_i} 0.002 \log\left(\frac{0.002}{0.001}\right)$$



entropy

$$H = -\sum_{i=1}^{N} p(x_i) \cdot \log p(x_i)$$

information

$-\lg p(x) \rightsquigarrow$ info in p (observed)

$-\lg q(x) \rightsquigarrow$ info in q (approximated) → loss

"Expected info difference". (loss)

$$\sum_{i=1}^{N} p(x_i) \quad -\lg q(x_i) - (-\lg p(x_i))$$

$$= \sum_i p(x_i) \left[\lg p(x_i) - \lg q(x_i) \right]$$

$\log(a) - \log(b) = \log \frac{a}{b}$

KL-Divergence

$$D_{KL}(p||q) = \sum_{i=1}^{N} p(x_i) \cdot log \frac{p(x_i)}{q(x_i)}$$

$$D_{kl}(\text{Observed} \,||\, \text{Uniform}) = 0.338$$

$$D_{kl}(\text{Observed} \,||\, \text{Binomial}) = 0.477$$

"larger info. loss"