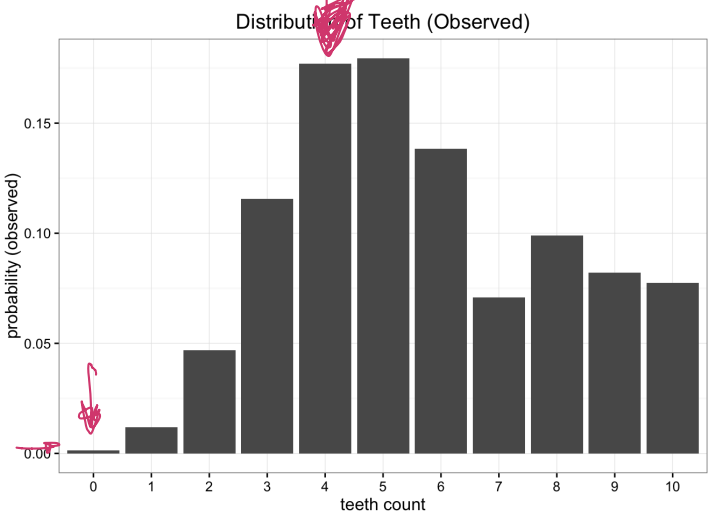
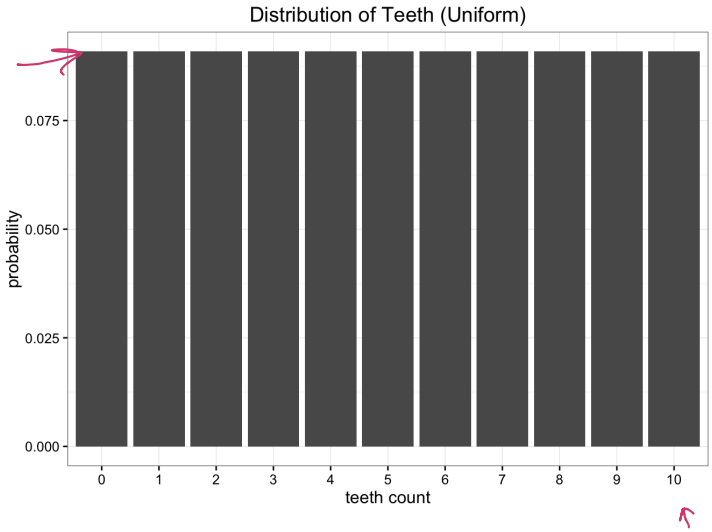
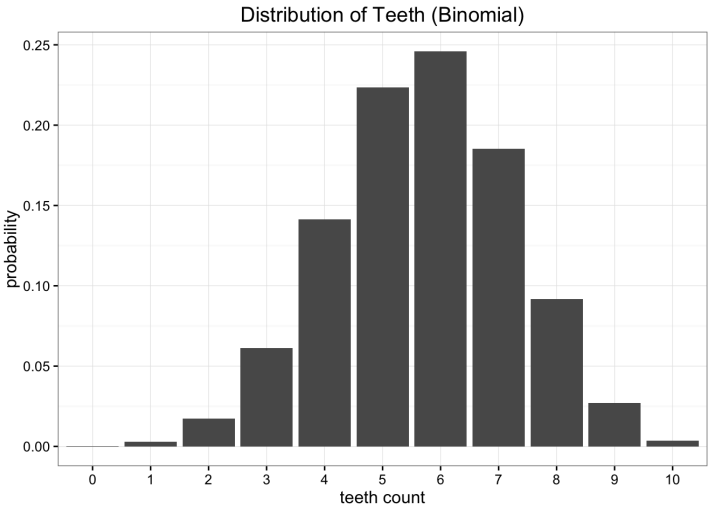
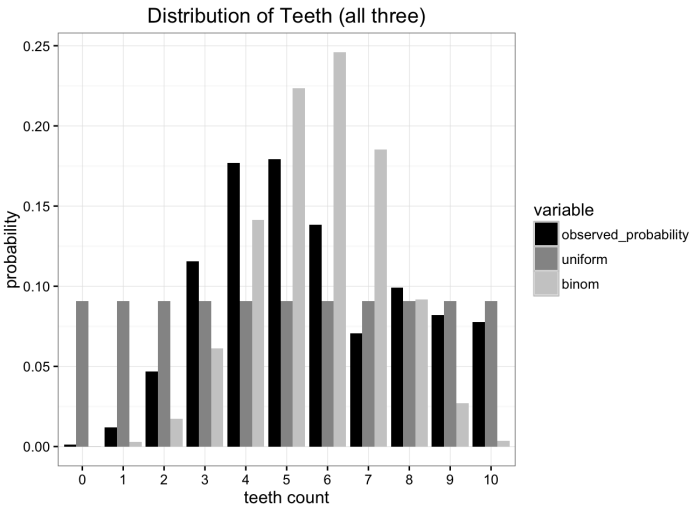

Distribution of Teeth (Observed)
probability (observed)
0.15
0.10
0.05
0.00
0 1 2 3 4 5 6 7 8 9 10
teeth count


Distribution of Teeth (Uniform)
probability
0.075
0.050
0.025
0.000
0 1 2 3 4 5 6 7 8 9 10
teeth count


Distribution of Teeth (Binomial)
probability
0.25
0.20
0.15
0.10
0.05
0.00
0 1 2 3 4 5 6 7 8 9 10
teeth count



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

neg. info

$-\lg p(x)$ ⇝ info in p (observed)

$-\lg q(x)$ ⇝ info in q (approximated)

"Expected info difference". (loss)

$$\sum_{i=1}^{N} p(x_i) \left[-\lg q(x_i) - (-\lg p(x_i)) \right]$$

$$= \sum_{i} p(x_i) \left[\lg p(x_i) - \lg q(x_i) \right]$$

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

KL Divergence

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

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

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

(Neural networks, variational autoencoders ...)