Empirical Probability vs true probability

- Definition of probability: Given the probability of event A is Pr(A). As the number of samples we take approaching infinity, the ratio $\frac{\#event\ A}{\#all\ outcomes}$ will approach the Pr(A).

  Ex: coin flip. Given Pr(head) = 0.7, $\frac{\#\ head}{\#head + \#tail} \rightarrow 0.7$ as total time toss goes to infinity
- Real Life: no infinity sample: limited sample
- Observed ratio $\frac{\#\ head}{\#head + \#tail}$ on limited number of samples is the empirical probability

Discuss Participation Question from Week 1

1. type of domain adaptation is the color image change example given in OT for ML:
   a. Exist type of domain adaptation: supervised, unsupervised, semi supervised and weakly supervised
   b. Color image shift: Given distribution of the image in data time and sunset time. No label telling the exact correspondence relationship between the color points from distribution of day image and of sunset image → unsupervised. The OT in this case only focuses on distribution of color but not the position of the color in the image or shapes in the image.
   c. Image Editing example: supervised: given the mask to mark the religion of facial features of source image and Mona Lisa: Gives what color each region should be and performs color transform.

2. Image color shift: linear: generally not linear.

Real Life application: For satellite image, a single pixel contains lots of information, successfully matching the color of the pixel may be able to tell what object the pixel represents.

3. Why do random forests use a randomly chosen subset of the features when learning each node?

   Fix an input feature vector **x.**
   Consider repeatedly running a random forest with n trees. Let #vote denote the number of votes for class 1. Clearly #vote is between 0 and n. Each time you run the random forest algorithm you get a different value for #vote for the label of the example **x**. Suppose you run the random forest algorithm many times, this would give you a distribution over #votes, the curves below represent the histogram for #vote. a Curves

Bias-Variance Tradeoff

Example1: polynomial regression model: Model with higher order of polynomial can fit to current samples better (lower bias). However, the model trained from one dataset won't perform well on a new dataset. The variance of the model increases compared with the model with lower order of polynomial.

Example2: Decision Tree:

- Increasing the number of nodes and depth of the tree can reduce bias. In extreme cases, each leaf node could end with only one sample. The bias of the overfitted tree decreases but the variance increases.

Example3: bagging

- Averaging over many trees does not change the bias, but it reduces the variance.