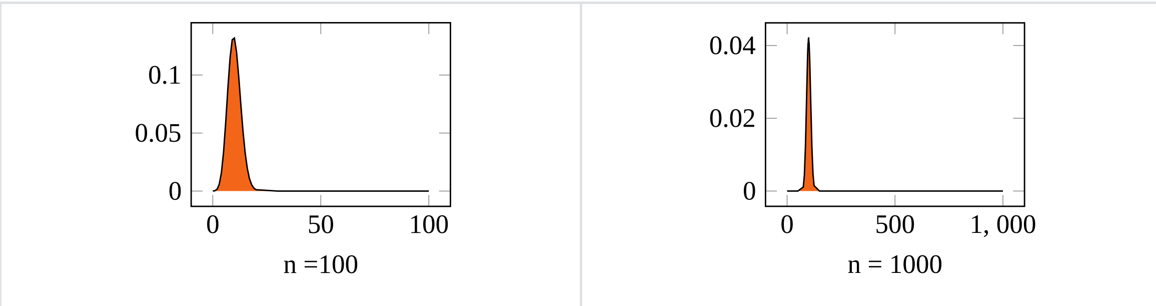
Definition: Typical Set

Example: Biased coin flip

Consider flipping a biased coin, with probability of heads being $P_{X}(\texttt{h}) = 0.1$ and probability of tails being $P_{X}(\texttt{t}) = 0.9$, and counting the number of heads that come up. The random variable Y describing this number is distributed according to the binomial$(n,P_{X}(\texttt{h}))$ distribution, where n is the number of coin flips. Below, the distribution of Y is plotted for $n = 100$ and $n = 1000$:



We see that the variance of the sample mean, $\text{Var}[\frac{1}{n}Y] = \frac{0.1(1-0.1)}{n}$, decreases as the number of samples increases. The weight of the distribution becomes centered around an increasingly narrow set of outcomes. Sequences of coin flips that result in an unusual number of heads become increasingly rare: almost all sequences are "typical".

The above example leads us to defining the following subset of the image of a random variable:

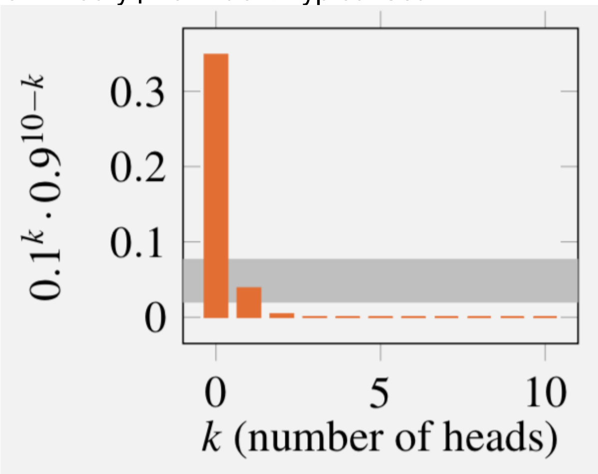
Definition: Typical set

The typical set $A_{\epsilon}^{(n)}$ with respect to P_X is the set of strings $(x_1, \ldots, x_n) \in {\cal X}^n$ such that $$ 2^{-n (H(X) + \epsilon)} \leq P_{X^n}(x_1, \ldots, x_n) \leq 2^{-n(H(X) - \epsilon)} \, , $$ where $P_{X^n}(x_1, \ldots, x_n) = \prod_{i=1}^n P_X(x_i)$.

The typical set is relatively small, but contains almost all of the probability mass.

Example

Consider again the experiment of flipping n biased coins, where the probability of observing "heads" is p, giving $H(X) = h(p)$. The graph below lets you investigate the typical sets for different values of n, p, and ϵ. [Unfortunately, the tool is not quite ready yet, we will supply it as soon as it works...]



On the horizontal axis, the different possible numbers of heads (0 to n) for a sequence x are shown; this is the 'weight' of the sequence x. The vertical axis shows the probability of observing that specific sequence x. (Note that that is *not* the probability of the event "observe $\text{weight}(x)$ heads", because such an event contains many different sequences). The gray area represents the range of probabilities that cause a sequence to fall into the typical set.

For example, set $p = 0.1$, $n = 10$ and $\epsilon = 0.1$. The typical set contains those elements that have probability between $2^{-10 (h(0.1) + 0.1)} \approx 0.0194$ and $2^{-10 (h(0.1) - 0.1)} \approx 0.0775$. Using the fact that a sequence with k heads occurs with probability $p^k \cdot (1-p)^{10-k}$, we can compute that the typical set consists of all sequences with exactly 1 heads and 9 tails (those sequences all have probability approximately 0.0387). We see that indeed, the bar with $\text{weight}(x) = 1$ falls into the gray area. Note that in particular, the all-tails outcome ($\text{weight}(x) = 0$) is *not* included in the typical set, even though is by far the most likely outcome. This is because there is only one sequence with $\text{weight}(x) = 0$, and already 10 sequences with $\text{weight}(x) = 1$: together, those 10 sequences have a higher probability mass (0.387) than the single all-tails outcome (0.349). Hence, the typical set still contains a lot of the probability mass.

This effect intensifies as we increase n: for $n = 1000$, the probability of the typical set (containing elements with 69--131 heads) is $P[A^{(1000)}_{\epsilon}] \approx 0.999$. The `typical' sequences, where roughly one in ten tosses results in a heads, make up almost all of the probability. The high-probability sequences, those with fewer heads, are so scarce that their total probability mass is tiny.