第 I 部
情報理論

1　自己情報量

・対数の底が 2 の時、単位はビット (bit)
・対数の底が e(ネイピア数) の時、単位はナット (nat)

$$I(x) = -log(P(x)) = log(W(x))$$

自己情報量は確率の log である、という事かな？ log の底は自分の都合で決めて良いと言っている (形が見えればよいのかな?)

珍しい事象 (頻度が小の事象) の情報量を示すので、確率が小さいまたは事象数が多い中の 1 つの事象の、珍しさという (-log(確率) は結局 log(1/確率)) らしい

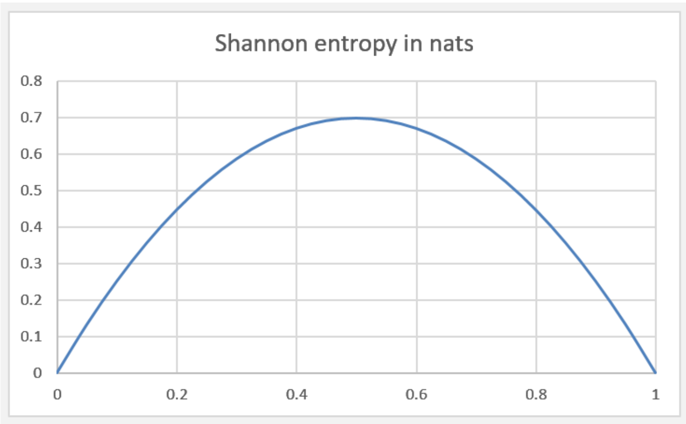
log で示す関数は情報量を示しているという言い方もしている、対数にする意味は大きな値を強調するためだろうか？まぁそのくらいの理解でとりあえずは

2　シャノンエントロピー

・微分エントロピーともいう (differential の誤訳?)
・自己情報量の期待値である

$$H(x) = E(I(x)) = -E(log(P(x))) = -\sum(P(x)log(P(x)))$$

自己情報量の平均を取ればシャノンエントリピーであるという説明 (以下グラフは厳密にはシャノンエントロピーではないけど図示の為でっち上げた)



説明では「コイン投げの例」とされており横軸は表が出る確率と思われる
つまり横軸の 1 は表が必ず出る状態で 0 は必ず裏が出る状態で、0 や 1 の状態は事象数としてはごく稀なのでエントロピーが低い (現実にはイカサマコインじゃないと考え難い)、0.5 付近が一番エントロピーが高い (これは裏表が均一に出現する状態を示す)

シャノンエントロピーは「現実の情報量の予測に使う」期待値で、誤差関数に適用できうるという説明がされている

3　カルバック・ライブラー・ダイバージェンス

・同じ事象・確率変数における異なる確率分布 P、Q の違いを表す

$D_{KL}(P||Q) = \mathbb{E}_{X\sim P}[log\frac{P(x)}{Q(x)}] = \mathbb{E}_{X\sim P}[logP(x) - logQ(x)]$

$\mathbb{E}_{x\sim P}$ は平均を表す

例えば、P が期待値として与えられる確率として、Q が実際に発生した確率、という場合の評価に用いることができる

4　交差エントロピー

・カルバック・ダイバージェンス・エントロピーの一部分を取り出したもの
・Q(=予測情報量) についての自己情報量を P(=現実情報量) の分布で平均している

$D_{KL}(P||Q) = \sum_x P(x)((-logQ(x)) - (-logP(x)))$

$P(x)(-logQ(x))$ の部分を $H(P,Q)$ と置き換えて

$H(P,Q) = H(P) + D_{KL}(P||Q)$　(ただし H はシャノンエントロピー)

$H(P,Q) = -\mathbb{E}_{X\sim P}logQ(x) = -\sum_x P(x)logQ(x)$

説明では現実情報量は予期せぬような事象が起こりうる、予測情報量がきれいな正規分布 (この場合シャノンエントロピーの様な分布) だとしても現実は乖離を持つはず、つまり現実での「Q で予想してなかった事象」はこの方法だと予想でこの事象は確率 0 パーセントの事象という風に見えるので評価しやすい、という事のように解釈できなくもない (多分そういう説明)