PACS Part 2, Lecture 6

Entropy

- entropy = measure of randomness
- for a discrete random variable X:

$$H(X) = -\sum_x \mathbb{P}(X = x) \log_2 \mathbb{P}(X = x)$$

- for a Bernoulli variable with parameter p:

$$H(p) = -p \log_2 p - (1 - p) \log_2 (1 - p)$$

- the maximum is achieved at $H(1/2) = 1$
- continuous extension: $H(0) = H(1) = 0$
- intermediate value: $H(1/4) \approx 0.8113$
- intuition: each Bernoulli try gives $H(p)$ random bits
- uniform distribution on a set of size n: $H(X) = \log_2 n$
- if X and Y are independent, then $H(X, Y) = H(X) + H(Y)$
- we have $\frac{2^{nH(q)}}{n+1} \leq \binom{n}{nq} \leq 2^{nH(q)}$ if nq is an integer

  Proof idea: The term $\binom{n}{nq} q^{qn} (1 - q)^{(1-q)n}$ is dominant in the binomial expansion of $(q + (1 - q))^n$.
- in particular, if $q > 1/2$, then $\binom{n}{\lfloor nq \rfloor} \geq \frac{2^{nH(q)}}{n+1}$

Measure of Randomness

- one formalization of number of random bits extracted from a random variable X: specify an extraction function $\mathrm{Ext} \colon \mathrm{Supp}(X) \to \{0, 1\}^+$
- we would want $\mathbb{P}(\mathrm{Ext}(X) = y \mid |\mathrm{Ext}(X)| = k) = 1/2^{|y|}$ for every $y \in \{0, 1\}^k$ whenever $\mathbb{P}(|\mathrm{Ext}(X)| = k) > 0$
- example: unbiased 8-sided die
  map each of the numbers $0, 1, \ldots, 7$ to its binary representation of length 3
- example: unbiased 12-sided die
  - for numbers $0, 1, \ldots, 7$, map to length-3 binary representations
  - for numbers $8, 9, 10, 11$, map to length-2 binary representations of $0, 1, 2, 3$

- If X is random variable with m equally probable possible values, then there is an extraction function with $\mathbb{E}\ |\mathrm{Ext}(X)| \geq \lfloor H(X) \rfloor - 1$.

  Proof: If m is a power of 2, we use the binary representation. If not, we split m into 2^α and $m - 2^\alpha$ with $\alpha = \lfloor \log_2 m \rfloor$. We show $\mathbb{E}\ |\mathrm{Ext}(X)| \geq \lfloor \log_2 m \rfloor - 1$ by induction. For the induction step, we calculate:

$$\begin{aligned}
\mathbb{E}\ |\mathrm{Ext}(X)| &\geq \frac{2^\alpha}{m}\alpha + \frac{m - 2^\alpha}{m}\left(\log_2(m - 2^\alpha) - 1\right) \\
&= \alpha - \frac{m - 2^\alpha}{m}(\alpha - \lfloor \log_2(m - 2^\alpha) \rfloor + 1) \\
&\geq \alpha - \frac{2^{\beta+1} - 1}{2^\alpha + 2^{\beta+1} - 1}(\alpha - \beta + 1)
\end{aligned}$$

  where $\beta = \lfloor \log_2(m - 2^\alpha) \rfloor$ since $m \leq 2^\alpha + 2^{\beta+1} - 1$. Using $(2^{\beta+1} - 1)(2^{\alpha-\beta-1} + 1) \leq 2^\alpha + 2^{\beta+1} - 1$, we get:

$$\mathbb{E}\ |\mathrm{Ext}(X)| \geq \alpha - \frac{1}{2^{\alpha-\beta-1} + 1}(\alpha - \beta + 1) \geq \alpha - 1$$

- Let $X \sim \mathrm{Bern}(p)$ with $0 < p < 1$ and $\delta > 0$. For sufficiently large n, we have:

  1. there is an extraction function on a sequence of n i.i.d. copies of X that outputs at least $(1 - \delta)nH(p)$ bits in expectation.
  2. the expected number of bits of any extraction function on a sequence of n i.i.d. copies of X is at most $nH(p)$.

  Proof: To prove (1), assume wlog $p > 1/2$ and let Z be the number of successes of the Bernoulli trials. All sequences of n trials with j successes are equally probable. Conditioned on $Z = j$, we thus generate a uniform distribution with $\binom{n}{j}$ possible values. Using the previous result, setting $B = |\mathrm{Ext}(X)|$, we want to show

$$\mathbb{E}B \geq \sum_{j=1}^{n} \mathbb{P}(Z = j)\left(\left\lfloor \log_2 \binom{n}{j} \right\rfloor - 1\right) \geq (1 - \delta)nH(p)$$

  for sufficiently large n.

  We limit the sum to $n(p - \varepsilon) \leq j \leq n(p + \varepsilon)$ and use $\binom{n}{j} \geq \frac{2^{nH(p+\varepsilon)}}{n+1}$. For the sum of the $\mathbb{P}(Z = j)$, we use the Chernoff bound. We thus get:

$$\begin{aligned}
\mathbb{E}B &\geq \left(\log_2 \frac{2^{nH(p+\varepsilon)}}{n+1} - 2\right)\mathbb{P}\left(|Z - np| \leq \frac{\varepsilon}{p}np\right) \\
&\geq (nH(p + \varepsilon) - \log_2(n + 1) - 2)\left(1 - 2e^{-n\varepsilon^2/3p}\right)
\end{aligned}$$

  This is $\geq (1 - \delta)nH(p)$ for sufficiently large n.

To prove (2), we note $2^{|\mathrm{Ext}(x)|}\mathbb{P}(X=x) \leq 1$ and calculate

$$\begin{aligned}\mathbb{E}B &= \sum_x \mathbb{P}(X=x)|\mathrm{Ext}(x)| \\ &\leq \sum_x \mathbb{P}(X=x)\log_2 \frac{1}{\mathbb{P}(X=x)} = H(X) = nH(p)\end{aligned}$$

Compression

- entropy can also be used as a measurement for compressibility
- example: two i.i.d. Bernoulli trials with $p = 3/4$
  - If two successes: code as 0.
  - If first is a success and the second a fail: code as 10.
  - If first is a fail and the second a success: code as 110.
  - If two fails: code as 111.

  The expected number of bits in the code is $27/16 < 2$.
- More generally: Let $X \sim \mathrm{Bern}(p)$ with $0 < p < 1$ and $\delta > 0$. For sufficiently large n, we have:
  1. there is a compression function on a sequence of n i.i.d. copies of X that outputs at most $(1+\delta)nH(p)$ bits in expectation.
  2. the expected number of bits of any compression function on a sequence of n i.i.d. copies of X is at least $(1-\delta)nH(p)$.