Numerical random variables

If \( X \sim P \) is RV with values in \( \mathbb{R} \):

**Expectation value (mean):** \( E[X] = E[X] = \sum_{x} p(x) \cdot x \)

\[ E[f(X)] = \sum_{x} p(x) \cdot f(x) \]

"law of the unconscious statistician"

\[ E[cX] = c \cdot E[X] \quad \& \quad E[X+Y] = E[X] + E[Y] \] \( \text{□} \)

\( X \sim \text{Uniform}(\{ \pm 1 \}), \ Y = -X \Rightarrow E[XY] = -1, \ E[X] = E[Y] = 0 \)

**Variance:** \( \text{Var}(X) = E[(X - E[X])^2] \)

\[ = \sum_{x} p(x)(x - E[X])^2 = E[X^2] - E[X]^2 \]

\( \text{Var}(cX) = c^2 \text{Var}(X) \)

If \( X, Y \) independent:

\[ \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) \] \( \text{□} \)

Examples:

<table>
<thead>
<tr>
<th>P</th>
<th>Bernoulli (f)</th>
<th>Binomial (n,f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>f</td>
<td>n \cdot f</td>
</tr>
<tr>
<td>Var</td>
<td>f(1-f)</td>
<td>n \cdot f \cdot (1-f)</td>
</tr>
</tbody>
</table>

\[ E[(X - E[X])^2] = E[(X - f)^2] = f (1-f)^2 + (1-f) (0-f)^2 = f (1-f) \]

Three results that give these meaning:

**Markov inequality:** If \( X \geq 0 \):

\[ \Pr(X \geq t) \leq \frac{E[X]}{t} \quad (\forall t \geq 0) \]

Pf: \( \Pr(X \geq t) = \sum_{x \geq t} p(x) \leq \sum_{x \geq t} p(x) \cdot \frac{x}{t} \leq \frac{E[X]}{t} \) \( \square \)

**Chebyshev inequality:**

\[ \Pr(|X - E[X]| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2} \]

Pf: Apply Markov to \( Y = (X - E[X])^2 \). \( \square \)
Law of large numbers: Suppose $X_1, \ldots, X_n \sim P$ with mean $\mu$ and variance $\sigma^2$.

Let $\bar{X} := \frac{1}{n} (X_1 + \ldots + X_n)$. Then:

$$\Pr \left( |\bar{X} - \mu| \geq \varepsilon \right) \leq \frac{1}{n} \frac{\sigma^2}{\varepsilon^2}$$

WHP: empirical average $\approx$ expectation value

**Proof:**

- $E\bar{X} = \mu$ & $\text{Var}(\bar{X}) = \frac{1}{n^2} \text{Var}(X_1 + \ldots + X_n) = \frac{\sigma^2}{n}$. (Chebyshev)

Convex and concave functions (§2.7)

Suppose $f : I \rightarrow \mathbb{R}$ is function on interval $I = (a,b)$.

We say $f$ is **convex** if $f'' \geq 0$

- **Concave** if $f'' \leq 0$

Jensen's inequality: Let $Z$ be a RV.

- If $f$ is convex: $E[f(Z)] \geq f(EZ)$
- If $f$ is concave: $E[f(Z)] \leq f(EZ)$

If $f'' > 0$ or $f'' < 0$: " = " holds only if $Z$ is constant

Entropy (§2.4)

Entropy of a random variable (RV) $X$ with distribution $P$:

$$H(X) = H(P) = \sum_x P(x) \cdot \log \frac{1}{P(x)} = E \left[ \log \frac{1}{P(x)} \right]$$

Unit "bit"

E.g. $X \sim \text{Bernoulli}(p)$: binary entropy

$$H(X) = p \cdot \log \frac{1}{p} + (1-p) \cdot \log \frac{1}{1-p}$$

Entropy of Bernoulli variables

$$H(P) = - \sum_p P(p) \cdot \log P(p)$$

E.g.

$$P = \frac{1}{2}$$

$$H(P) = \frac{1}{2} \log \frac{1}{\frac{1}{2}} + \frac{1}{2} \log \frac{1}{1 - \frac{1}{2}} = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1 = 1$$

$$P = 0.5$$

$$H(P) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0 = 0$$
Properties:

* \( H(X) \geq 0 \), iff constant \( p \log_2 p \geq 0 \) \( \forall p \in [0,1] \), iff \( p=0 \) or \( p=1 \)

* \( H(X) = \log \# \{ x : P(x) > 0 \} = \log \# \mathcal{X} \)

\[ H(X) = \log \# \mathcal{X} \implies X \text{ uniformly random} \]

\[ \begin{align*}
\text{Proof: } & \text{Apply Jensen with } f = \log \text{ and } Z = \frac{1}{P(x)}; \\
& E \left[ \log \frac{1}{P(x)} \right] = \log \left( E \left[ \frac{1}{P(x)} \right] \right) \\
& \text{with equality iff } P(x) \text{ constant, i.e. } \\
& P(x) > 0, P(y) > 0 \implies P(x) = P(y) \\
\end{align*} \]

* **NOTATION:** \( H(X,Y) = H(X|Y) = \text{entropy of joint distribution } P(x,y) \)

1. If \( X, Y \) independent: \( H(X,Y) = H(X) + H(Y) \)

**Proof:** Since \( P(x,y) = P(x)P(y) \) we have \( \log \frac{1}{P(x,y)} = \log \frac{1}{P(x)} + \log \frac{1}{P(y)} \)

\[ \text{no take expectation values.} \]

**Interpretation?** Let us call \( h(x) = h(x=x) = \log_2 \frac{1}{P(x)} \) the information content (or "surprise") of an outcome \( x \in \mathcal{X} \).

\[ \implies H(X) = E[h(X)] \text{ is average information content.} \]

Why is this a good definition? Three suggestive examples:

1. **Uniformly random number in \( \{0,1,2,3,4,5\} \):** \( H(X) = \log_2 26 \approx 8 \text{ bit} \)

2. **Poor man's submarine game:** Single submarine hidden; other player asks if submarine in some square \( \rightarrow \) hit/miss

   1st move: \( P(\text{hit}) = \frac{1}{64} \implies h(\text{hit}) = 6 \text{ bit} \)
   \( P(\text{miss}) = \frac{63}{64} \implies h(\text{miss}) \approx 0.0227 \text{ bit} \)

   Leaned precise location (8 bit) \( \rightarrow \) leaned little (63 remaining)

3. **"English" has 215 words in \( \{A,...,Z\}^5 \) s.t. frequency of single letters matches English. Let \( w \) be uniformly random word in this list.

   \( H(w) = 15 \text{ bit} \), i.e. on average 3 bit/letter

   \[ \text{but e.g. } p(w_1 = Z) = 0.19 \implies h(w_1 = Z) \approx 10 \text{ bit} \]

   \[ \leq \frac{1}{26} \]
Consider data source modeled by RV $X$. Assume we know distribution $P_X$.

E.g. $X$ could be a letter and we assume $P(X) = \text{P(English}(x)$

How well can we compress?

Today on Thursday we consider symbol codes, which compress one symbol (letter, source message, ...) at a time:

$$X \xrightarrow{\text{Compression}} C \xrightarrow{\text{One or more bits}} \hat{X} \xrightarrow{\text{Decompression}} \hat{X} = X$$

**Goal:** Show that lossless compression one symbol at a time can achieve $H(X) \leq L < H(X) + 1$, where $L =$ average length of codeword.

At least one more bit than entropy

**Notation:**

- $S^+ = \bigcup_{N \geq 1} S^N$ = nonempty strings over $S$
- $l(w) =$ length of string $w \in S^+$

**Symbol Code:** $C_i: A \rightarrow \{0,1\}^+$ for alphabet $A$

* average length: $L(C_i(P)) = \mathbb{E}[l(C_i(X))] = \sum_{X \in \text{alphabet}} P(X) \cdot l(C_i(X)) = \mathbb{E}[l(C_i(X))]$

* extended code: $C_i^+: A^+ \rightarrow \{0,1\}^+$. $C_i^+(x_1, ..., x_n) := C_i(x_1) \cdots C_i(x_n)$

Two important classes of codes:

* $C_i$ is called uniquely decodable (UD) if $C_i^+(w) = C_i^+(w') \Rightarrow w = w'$ for any $w, w' \in \bigcup_{n \geq 1} S^+$

* $C_i$ is called a prefix code if no codeword $C_i(x)$ is prefix of any other.

Any prefix code is UD.