0.1 Law of large numbers

$(X_n)_{n\in\mathbb{N}}$ sequence of iid random variables. Look at the sample average / sample mean

$$\overline{X_n} = \frac{1}{n}(X_1 + \cdots + X_n)$$

It is random! Suppose $mu = \mathbb{E}(X_k)$.

Weak LLN: Suppose also that $\sigma^2 = \mathbb{V}(X_k)$ is finite. Then $\overline{X_n} \to \mu$ in probability:

$$\mathbb{P}[\left|\overline{X_n} - \mu\right| \geq a] \leq \frac{\sigma^2}{na^2}$$

Strong LLN: $\overline{X_n} \to \mu$ almost surely

1 Entropy

1.1 Introduction: measuring information formulas of Hartley, Shannon

"The amount of information contained in a message should be measured by the shortest way to formulate its contents."

We use bits (0s and 1s) (E.g.: 26 letters + symbols, Ä,Ö,Ü, ␣, , , . . .)

32 symbols $\equiv$ binary sequences of length 5: $00000, 00001, \ldots, 11111$)

Unit of information: 1 bit

Amount of information contained in the answer to a yes-no-question disregarding its specific contents.

Example. *Guess a number in* $\{0, \ldots, 2^n - 1\}$.
How many questions?
encode these numbers by binary sequences of length n.
Ask for the digits. This results in needing only n *questions.*

$$n = \log_2(2^n)$$

Definition 1.1. *Set* U_N *of* N *different objects "of equal value". Hartley formula: amount of information to identify one of them is*

$$H(U_N) = \log_2 N$$

Got this chapter from RENYI.

Justification in terms of "natural" axioms

(A) $H(U_2) = 1$ (normalisation)

(B) $H(U_N) \leq H(U_{N+1})$

(C) $H(U_{N \cdot M} = H(U_N) + H(U_M)$

Motivation for (C): Take $M \cdot N$ elements proceed in 2 steps:
group the $M \cdot N$ elements into N groups of M elements each

$$U_{NM} = U_M^{(1)} \uplus U_M^{(2)} \uplus \cdots \uplus U_M^{(N)}$$

1. which group? $H(U_N)$

2. which element of the group found in step 1? $H(U_M)$

Lemma 1.2. $N \mapsto H(U_N) = \log_2(N)$ *is the only function on* $\mathbb{N}$ *which satisfies* (A) - (C).

Proof. Let $N \geq 2$ (fixed).

$$2^{s(k)} \leq N^k \leq 2^{s(k)+1}$$

where $s(k) = \lfloor \log_2(N^k) \rfloor = \lfloor k \log_2 N \rfloor$ and $k \in \mathbb{N}$.

$$\frac{s(k)}{k} \leq \log_2(N) < \frac{s(k)+1}{k}$$

so that $\lim_{k\to\infty} \frac{s(k)}{k} = \log_2(N)$.

$$\begin{aligned}
(B) &\implies H(U_{2^{s(k)}}) \leq H(U_{N^k}) \leq H(U_{2^{s(k)+1}}) \\
(C) &\implies s(k)H(U_2) \leq kH(U_N) \leq (s(k)+1)H(U_2) \\
(A) &\implies \frac{s(k)}{k} \leq H(U_N) \leq \frac{s(k)+1}{k}
\end{aligned}$$

$$k \to \infty : \frac{s(k)}{k} \to \log_2(N) \quad \frac{s(k)+1}{k} \to \log_2(N)$$

□

Variant: one can replace (B) by

(B^*) $H(U_{N+1}) - H(U_N) \to 0$ as $N \to \infty$.

Group into not necessarily equal parts:

$$U_N = U_{N_1} \uplus U_{N_2} \uplus \cdots \uplus U_{N_n}$$

step 1 which group $H_1 =?$

step 2 which element of the group found in step 1?

If we know that it is group k, then we need $\log_2(N_k)$ "questions". The average number of questions needed depends on the sizes of the groups.

$$H_2 = \sum_{k=1}^{n} \frac{N_k}{N} \log_2(N_k)$$

we should have

$$H(U_N) = \underbrace{H_1}_{=?} + H_2$$

$$\begin{aligned}
H_1 &= \log_2(N) - \sum_{k=1}^{n} \frac{N_k}{N} \log_2(N_k) \\
&= -\sum_{k=1}^{n} \frac{N_k}{N} (\log_2(N_k - \log_2(N)) \\
&= -\sum_{k=1}^{n} \frac{N_k}{N} \log_2(\frac{N_k}{N}) \\
&= -\sum_{k=1}^{n} p_k ß log_2 p_k
\end{aligned}$$

where $p_k = \frac{N_k}{N}$ is the probability of k-th group

1.2 Entropy

information associated with / of discrete probability distributions. respectively random variables

Definition 1.3. *Let X be a (discrete) RV taking values in a finite set $\mathcal{X}$.*

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathcal{X}$$

Distribution of X:

$$p_X(x) = \mathbb{P}[X = x]$$

where $x \in \mathcal{X}$. The entropy of X, respectively of the prob. dist p on $\mathcal{X}$ is

$$H(X) = H(p(x)) = -\sum_{x \in \mathcal{X}} p(X) \log_2(p(x))$$

Convention: $0 \log_2 0 = 0$ $f(p) = -p \log p$. add pic

If we enumerate $\mathcal{X} = \{x_1, \dots, x_k\}$, such that $p_k = p(x_k)$, then $(p_1, \dots, p_n)$ is a prob vector

$$H(p_1, \dots, p_k) = -\sum_{k=1}^{n} p_k \log_2(p_k)$$

$$H(X) = \mathbb{E}(-\log_2(p_X(X))$$

Recall:

$$\begin{aligned} g : \mathcal{X} \to \mathbb{R} \\ g(X) = g \circ X : \Omega \to \mathbb{R} \\ g(X)(\omega) = g(X(\omega)) \\ \mathbb{E}(g(X)) = \sum_{x \in \mathcal{X}} g(x) p(x) \end{aligned}$$

Example. $\mathcal{X} = \{0, 1\}$ *and* $p(0) = 1 - \theta$, $p(1) = \theta$

$$p(X) = \begin{cases} 1 - \theta & \textit{if } X = 0 \\ \theta & \textit{if } X = 1 \end{cases}$$

$$H(X) = -\theta \log_2(\theta) - (1 - \theta) \log_2(1 - \theta)$$