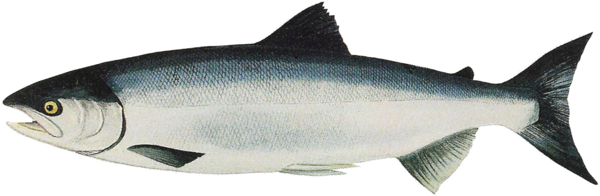
Information Theory
TUESDAY

Mathias Winther Madsen
mathias@gmail.com
github.com/mathias-madsen/nasslli2025/



NASSLLI, June 2025

The Variability of Sums

Problem

Which of the following is more probable?

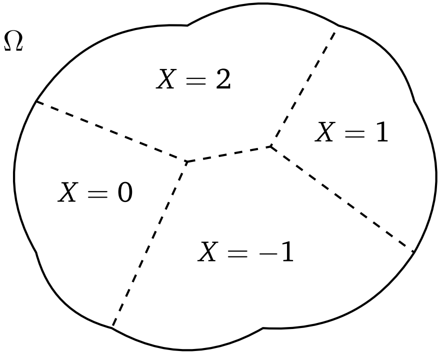
1. a sum greater than 40 in 10 dice rolls
2. a sum greater than 400 in 100 dice rolls

$$
\begin{aligned}
5+1+5+3+5+6+4+6+1+6 &= 42 \\
2+3+3+4+2+4+5+2+2+6 &= 33 \\
6+6+2+3+6+2+6+6+5+6 &= 48 \\
2+3+1+2+2+4+2+3+5+1 &= 25 \\
&\vdots
\end{aligned}
$$

Random Variables

Definition: Random Variable

A **random variable** is a variable whose numeric value varies between different (randomly sampled) states of the world.



Expected Value

Definition: Expected Value

$$EX = \sum_{x \in \mathbb{A}} p(x)\, x$$

Theorem: Linearity of Expectations

For any any constant $r \in \mathbb{R}$ and any random variable X,

$$E(rX) = r(EX).$$

For any random variables X and Y,

$$E(X+Y) = EX + EY.$$

Variance

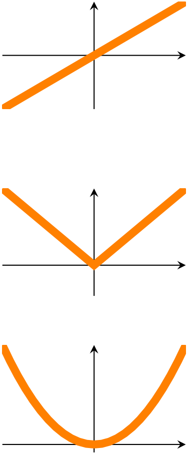
Problem

Suppose X follows the distribution

x	1	2	4
$p(x)$	1/2	1/4	1/4

Compute the expected values of

- X
- $X - EX$
- $|X - EX|$
- $(X - EX)^2$



Variance

Definition: Variance

$$Var(X) \;=\; E[(X - EX)^2] \;=\; E(X^2) - (EX)^2.$$

Theorem: Variance of Linear Transformations

For any random variable X and any constant $r \in \mathbb{R}$ and any random variable X,

$$Var(rX) \quad = \quad r^2\, Var(X)$$

For *independent* random variables X and Y,

$$\begin{aligned} E(XY) &= E(X)E(Y) \\ Var(X+Y) &= Var(X) + Var(Y) \end{aligned}$$

(These formulas are usually false for dependent X and Y.)

The Convergence of Averages

Problem

Suppose that $X_1, X_2, \ldots, X_n$ are independent random variables with a common mean m and a common variance v. Let further

$$A_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

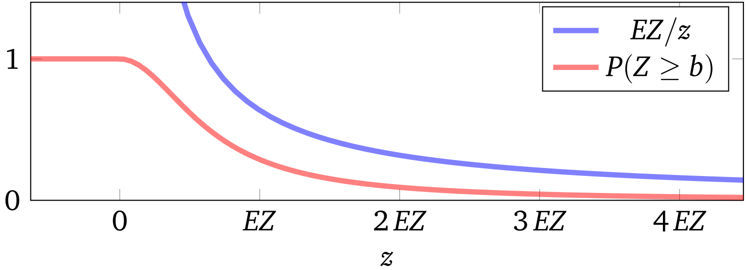
What is $E(A_n)$ and $Var(A_n)$?

The Law of Large Numbers

Markov's Inequality

Suppose $P(Z \geq 0) = 1$; then

$$P(Z > z) \;\leq\; \frac{EZ}{z}.$$

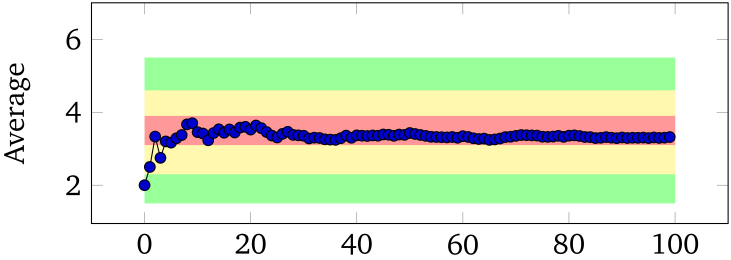


The Law of Large Numbers

Chebyshev's Inequality

Suppose $E(X_i) = m$ and $Var(X_i) = v$ for $i = 1, 2, \ldots, n$; then

$$P\left[(m - A_n)^2 > b\right] \;\leq\; \frac{v}{nb}.$$



Chebyshev: "Des Valeurs Moyennes,"
Journal de Mathématiques Pures et Appliquées (1867)

The Law of Large Numbers

The Law of Large Numbers

Averages converge to expecations: for any $\varepsilon > 0$,

$$P\left(|A_n - EX| \leq \varepsilon\right) \quad \rightarrow \quad 1$$

for $n \rightarrow \infty$ provided that $Var(X) < \infty$.

The Convergence of Frequencies

Problem

A source produces a random word

$$X^{10} = (X_1, X_2, \ldots, X_{10}),$$

by sampling its letters independently from the distribution

x	T	S	E
$p(x)$	$1/4$	$1/2$	$1/4$

- What is the probability of the word $X^{10} = \texttt{STETSESSES}$?
- What is the most probable 10-letter word?
- What is the most probable number of E's?

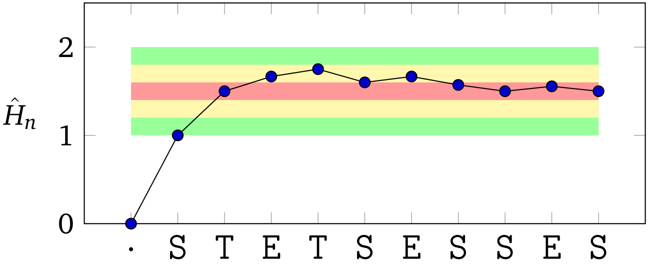
Typical Sequences

Theorem: Convergence of Sample Entropy

Let $X_1, X_2, \ldots, X_n$ be drawn independently from a distribution with entropy H, and let

$$\hat{H}_n = -\frac{\log_2 p(X_1) + \log_2 p(X_2) + \cdots + \log_2 p(X_n)}{n}.$$

Then $\hat{H}_n \to H$ in the sense that $P(|\hat{H}_n - H| \leq \varepsilon) \to 1$ for $n \to \infty$.

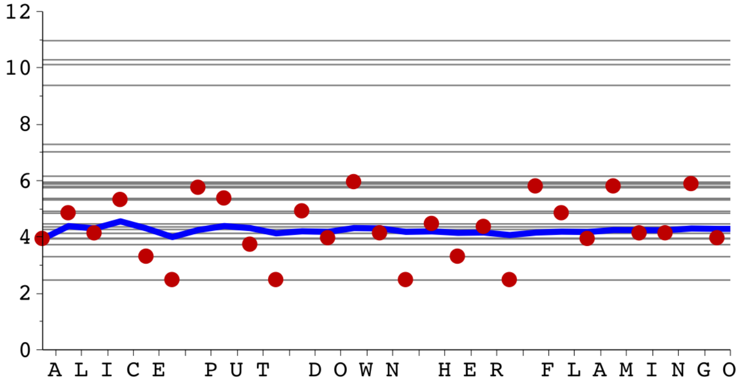


Typical Sequences

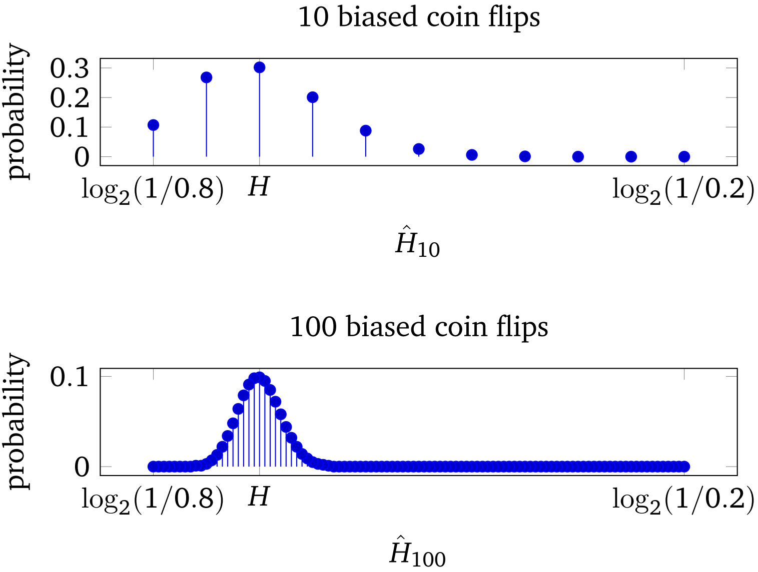
$X_1, X_2, \ldots, X_{40}$	$\hat{H}_{40}$
SO_SHE_BEGAN_VERY_CAUTIOUSLY:_'BUT_I_DON	4.489
U'VE_HAD!'_'OH,_I'VE_HAD_SUCH_A_CURIOUS_	4.532
F_YOU_PROVIDE_ACCESS_TO_OR_DISTRIBUTE_CO	4.325
'I_DIDN'T_KNOW_THAT_CHESHIRE_CATS_ALWAYS	4.286
LY_ONE_A-PIECE_ALL_ROUND._'BUT_SHE_MUST_	4.369
TS_EYES_WERE_GETTING_EXTREMELY_SMALL_FOR	4.321
W_LARGER,_I_CAN_REACH_THE_KEY;_AND_IF_IT	4.269

$$H = 4.315$$

Typical Sequences



Typical Sequences



Typical Sequences

The Law of Large Numbers

$$P\left(-\frac{1}{n}\log p(X_1, X_2, \ldots, X_n) \;>\; H + \varepsilon\right) \;\to\; 0.$$

The Asymptotic Equipartition Property

$$P\left(p(X_1, X_2, \ldots, X_n) \;<\; 2^{-n(H+\varepsilon)}\right) \;\to\; 0.$$

In other words: there are about 2^{Hn} reasonably probable texts (and they are about equally probable).