Probability I
Law of Large Numbers

Daniel Yu

October 8, 2024

Contents

1 Interpretations of Probability

1.1 Frequentist Perspective

Adherents believe that probability is simply a measurement of the outcomes over repeated experiments. For example,

Note. Let $\{X_1, \ldots, X_n\}$ be iid discrete random variables. Denote $range(X_i) = \{x_1, x_2, \ldots, x_m\}$ and let $k_i(n) = \{\text{number of times } x_i \text{ appeared in the sequence}\}$. Then let $\frac{k_i(n)}{n}$ be the relative frequency of x_i. However $\frac{k_i(n)}{n}$ is itself a random variable!

Remark. Example
Toss a fair 4-sided die n times. Say $n = 2$. Let $\frac{k_1(2)}{2}$ be a random variable.

$$\frac{k_1(2)}{2} \in \{0, \frac{1}{2}, 2\}.$$

Then,

$$\begin{aligned} P[\frac{k_1(2)}{2}] &= (\frac{3}{4})^2 = \frac{9}{16} \\ P[\frac{k_1(2)}{2}] &= 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8} \\ P[\frac{k_1(2)}{2}] &= \frac{1}{16}. \end{aligned}$$

This is scaled binomial distribution (ex consider $bin(2, \frac{1}{4}), P(X = k) = \binom{2}{k} \frac{1}{4}^k (\frac{3}{4})^{2-k}$). A binomial distribution is random variable.

1.1.1 Law of Large Numbers

Can be thought of as a consequence of the frequentist perspective of probability. Not unique to it though.

Theorem 1. Let X_1 be a discrete random variable with outcomes $\{y_1, y_2, \ldots, y_m\}$, and let $k_i(n) = \{\text{number of times that } y_i \text{ appears in } n \text{ trials }\}$. Then, as $n \to \infty$

$$\{\frac{k_1(n)}{n}\}_n \to P[X_i = y_i] \text{ in probability.}$$

Corollary 1. Law of Large Numbers
If X_1 is a dsicrete random variables with outcomes $\{y_1, \ldots, y_m\}$ with X_i are iid (so we are sampling from the same random distribution). And let

$$E_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n).$$

Then, $E_n \to E[X_1]$ as $n \to \infty$ (X_i are iid). The time average (average of experiment throughout time, E_n) = space average (space of possiblities, multiverse perspective, $E[X_1]$)

Proof.

$$\begin{aligned}
E_n &= \frac{1}{n} \cdot \sum_{i=1}^{m} y_i \cdot \text{ number of times } y_i \text{ appears} \\
&= \frac{1}{n} \sum_{i=1}^{m} y_i \cdot \frac{k_i(n)}{n} \\
&\text{take n} \to \infty\text{, use the theorem} \\
&= \frac{1}{n} \sum_{i=1}^{m} y_i \cdot P[X_i = y_i] \\
&= E[X].
\end{aligned}$$

□

1.1.2 Probability Ineqaulities

Note true in general but we are only going to deal with the discrete case

Theorem 2. Markov's Theorem
Let X be a non-negative random variable. Then, $\forall a > 0$,

$$P[X \geq a] \leq \frac{E[X]}{a}.$$

In other words the probability that a random variable is greater than or equal to some value is always less than or equal to the expected value (mean) of the random variable divided by the value.

$$\begin{aligned}
E[X] &= \sum_{y \in range(x)} y \cdot P[X = y] \\
&\geq \sum_{y \in range(x) \cap y \geq a} y \cdot P[X = y] \\
&\geq \sum_{y \in range(x) \cap y \geq a} a \cdot P[X = y] \\
&= a \cdot P[X \geq a] \\
\frac{E[X]}{a} &\geq P[X \geq a].
\end{aligned}$$

□

Note. Example
Consider tossing fair 4 sided dice. $K_1(n)$ are the numbers of 1s. What is probability $P[\frac{k_1(n)}{n} \geq \frac{1}{3}]$

Proof.

$$\begin{aligned}
E[\frac{k_1(n)}{n}]n &= E[\frac{1}{n}\sum_{i=1}^{n} 1_{x_i=1}] \\
&= \frac{1}{n}\sum_{i=1}^{n} E[1_{x_i=1}] \\
&= \frac{1}{n} \cdot n \cdot \frac{1}{4} \\
&= \frac{1}{4}.
\end{aligned}$$

By markov's inequality:

$$P[\frac{k_1(n)}{n} \geq \frac{1}{3}] \leq \frac{E[\frac{k_1(n)}{n}]}{\frac{1}{3}} = \frac{3}{4}.$$

Obvious tbh, so kinda useluess and doesn't tell us anything. We want a tighter bound! □

Proposition 1. Let $\{Y_n\}$ be a sequence of non-negative R.V. If,

$$E[Y_n] \to 0.$$

then,

$$Y_n \to 0.$$

in probability.

Proof. Rough sketch
Shiow $\forall \varepsilon$,

$$\lim_{n\to\infty} P[|Y_n| > \varepsilon] = 0.$$

Since $Y_n \geq 0$, $|Y_n| = Y_n$.

$$\lim_{n\to\infty} P[Y_n > \varepsilon] \leq \lim_{n\to\infty} \frac{E[Y_n]}{\varepsilon} = 0.$$

□

Note. Clique Problem (NP-hard)
n people in a room, each pair of people are friends with probability $\frac{1}{n}$ independent of every other pair. Let Q_n = number of cliques of size 4 (i.e 4 participants that are mutually friends) i.e $\binom{4}{2} = 6$ pairs are friends.

Proof. Understanding this is very hard because as n increases, the number of cliques increases but the probability of each pair being friends decreases ($\frac{1}{n}$). However, using prop, even thought $P[Q_n = 0]$ is difficult to compute, $E[Q_n]$ is not.

$$\begin{aligned}
E[Q_n] &= E[\sum_{\text{quadruplets}} 1_{\text{quadruplet are mutual friends}}] \\
&= \sum_{\text{quadruplets}} E[1_{\text{quadruplet are mutual friends}}] \\
&= \frac{n!}{(n-4)!4!}\left(\frac{1}{n}\right)^6 \\
&= \frac{n(n-1)(n-2)(n-3)}{4!} \cdot (\frac{1}{n})^6 \\
&\leq \frac{1}{24n^2} \\
&\to 0.
\end{aligned}$$

Therefore, $P[Q_n \geq 1] \leq \frac{E[Q_n]}{1} \leq \frac{1}{24n^2} \to 0.$ $\square$

Remark. Note that if the clique was size 3, then this would converge... (let K_n = number of size 3 cliques)

$$\begin{aligned} E[K_n] &= \binom{n}{3} \frac{1}{n}^3 \\ &= \frac{n(n-1)(n-2)}{3!} \frac{1}{n}^3 \\ &= \frac{n^3 - 3n^2 + 2n}{6} \frac{1}{n}^3 \\ &\to \frac{1}{6}. \end{aligned}$$

Theorem 3. Chebyslev's Inequality
Let X be a random variable with finite variance. Then,

$$P[|X - E[X]| \geq a] \leq \frac{Var(x)}{a^2}.$$

If we set $\sigma = \sqrt{Var(x)}$ (standard deviation). Then,

$$P[|X - E[X]| > r\sigma] \leq \frac{1}{r^2}.$$

Proof.

$$\begin{aligned} P[|X - E[X]| \geq a] &= P[(|X - E[X]|)^2 \geq a^2] \\ &\leq \frac{1}{a^2} E[(X - E[X])^2] \\ &= \frac{Var(x)}{a}. \end{aligned}$$

$\square$

Note. Example
Back to $\frac{k_1(n)}{n}$. $E[\frac{k_1(n)}{n} = \frac{1}{4}]$. What is $var(\frac{k_1(n)}{n})$?

Proof.

$$\begin{aligned} Var(\frac{k_1(n)}{n}) &= \frac{1}{n}^2 Var(k_1(n)) \\ &= \frac{1}{n}^2 Var(\sum_{i=1}^{n} 1_{x_i=1}) \\ &\text{remember each roll } X_i \text{ is iid} \\ &= \frac{1}{n}^2 \sum_{i=1}^{n} Var(1_{x_i=1}) \\ &= \frac{1}{n}^2 \sum_{i=1}^{n} [E[1_{x_i=1}^2] - E[1_{x_i=1}]^2] \\ &= \frac{1}{n}^2 \sum_{i=1}^{n} [\frac{1}{4} - \frac{1}{16}] \\ &= \frac{3}{16n}. \end{aligned}$$

$\square$

1.1.3 Proof of law of large numbers

Recall that the law of large number states: Let X_1 be a discrete random variable with outcomes $\{y_1, y_2, \ldots, y_m\}$, and let $k_i(n) = \{$number of times that y_i appears in n trials $\}$. Then, as $n \to \infty$

$$\{\frac{k_1(n)}{n}\}_n \to P[X_i = y_i] \text{ in probability.}$$

Proof. Let $\frac{k_i(n)}{n} = \frac{1}{n} \times$ number of times y_i shows up. Then, we can think of it as being equivalent to saying the following

$$\begin{aligned} E[\frac{k_i(n)}{n}] &= \frac{1}{n}E[\sum_{i=1}^{n} 1_{x_j=y_j}] \\ &= P[X_i = Y_i]. \end{aligned}$$

So if the probability goes to 0, then the expected value goes to 0 as well. So all we need to show as $n \to \infty$

$$P[|\frac{k_i(n)}{n} - E[\frac{k_i(n)}{n}]| < \varepsilon] \to 0.$$

By chebyslev's inequality,

$$\begin{aligned} P[|\frac{k_i(n)}{n} - E[\frac{k_i(n)}{n}]| > \varepsilon] &\leq \frac{1}{\varepsilon^2}Var(\frac{k_i(n)}{n}) \\ &\leq \frac{1}{\varepsilon^2 n^2}Var(\frac{k_i(n)}{n}) \\ var(k_i(n)) &= n \cdot Var(1_{x_i=y_i}) \\ &\leq \frac{n}{4} \\ P[|\frac{k_i(n)}{n} - E[\frac{k_i(n)}{n}]| > \varepsilon] &\leq \frac{1}{4\varepsilon^2 n}. \end{aligned}$$

This goes to 0 as $n \to \infty$ □

Note. Random Walk Mistake
Consider a random walk $s_1 = 1, -1$ with probability $\frac{1}{2}$ and $\{s_i\}_{i=1}^{n}$ be iid sequence. Clearly, $E[S_i] = 0$. Now, denote $W_n = \sum_{i=1}^{n} S_i$. Does the weak law of large numbers imply that $W_n \to 0$?

Proof. Lets compute $P[W_n = 0]$. realize that if n odd, then $W_n \neq 0$. So consider $P[W_{2n} = 0]$.

$$\begin{aligned} \{W_{2n}\} &\Leftrightarrow \text{steps with } +1 = \text{steps with -1} \\ P[W_{2n} = 0] &= \binom{2n}{n} \cdot \frac{1}{2^{2n}} \\ &= \frac{(2n)!}{(n!)^2 4^n} \\ &\text{using sterling's approximation } n! \approx \left(\frac{n}{\varepsilon}\right)^n \cdot \sqrt{2\pi n} \\ &\approx \frac{\left(\frac{2n}{\varepsilon}\right)^{2n}\sqrt{4\pi n}}{\left((\frac{n}{\varepsilon})^n\sqrt{2\pi n}\right)^2 4^n} \\ &= \frac{\sqrt{4\pi n}}{2\pi n} \\ &= \frac{1}{\sqrt{\pi n}} \\ &\to 0 (n \to \infty). \end{aligned}$$

However, $W_n \not\to 0$ in probability (consider the random walks of stocks, they don't converge to 0)!!!! This is not a contradiction because the weak law of large numbers states that $\frac{W_n}{n} \to 0$ in probability not W_n!. In truth,

$$P[|\frac{W_n}{n}| \geq \varepsilon] \to 0$$
$$P[|W_n| \geq \varepsilon n] \to 0.$$

□