今天我们来讨论一下（弱）大数定理。内容很简单很好记，很好”理解”：

”如果随机变量 X 的期望 μ 存在，并且 variance 有限（大部分我们用到的随机变量都能满足这个条件），那么当我们对此随机变量进行 n 次独立采样得到样本 $(X_1, X_2, \cdots, X_n)$ 之后算出来的平均值 $\bar{X}_n(=\frac{X_1+X_2+\cdots+X_n}{n})$ 在样本数量 n 足够大的时候将会是对随机变量 X 的期望 μ 的一个很好的近似。”

基于我们的生活经验，我们觉得这个定理很是自然而然。但是怎么证明它？现在我们来”自然而然”地证明一下。

首先，对于任何非负的随机变量 Y，我们看一下它的期望：$E(Y) = \int yf(y)dy$, 对于任意正数 v, 我们可以得到 $E(y) = \int_0^v yf(y)dy + \int_v^\infty yf(y)dy$, 这个式子的右边两项都是正的，所以我们可以看出 $E(y) \geq \int_v^\infty yf(y)dy \geq v\int_v^\infty f(y)dy$, 所以我们得到

$\int_v^\infty f(y)dy <= \frac{E(y)}{v}$，也就是 $P(y > v) <= \frac{E(y)}{v}$（这个叫 Markov 不等式）

把这个不等式使用到随机变量 $(X-\mu)^2$, 我们得到 $\frac{E((X-\mu)^2)}{v} >= P((X-\mu)^2 > v) = P(|X-\mu| > \sqrt{v})$, 也就是 $P(|X-\mu| > \sqrt{v}) <= \frac{var(X)}{v}$（这个叫 Chebyshev 不等式）

现在把这个不等式使用到随机变量 $\bar{X}_n$ 上，则得到：

$P(|\bar{X}_n - \mu| > \sqrt{v}) <= \frac{var(\bar{X}_n)}{v} = \frac{\frac{\sigma^2}{n}}{v}$

也就是

$P(|\bar{X}_n - \mu| > \sqrt{v}) <= \frac{\sigma^2/v}{n}$

对于任意正数 v, 我们看到当 n 足够大的时候，上式右边趋于零 — 这就是 $\bar{X}_n$ 趋于 μ 的意思（严格地说就是依概率收敛的意思），就是说 $\bar{X}_n$ 是对 μ 的足够好的近似 — 这就是（弱）大数定理）。