

方差的定义

作者：姜鱼

日期：2025 年 2 月 28 日

对于离散型随机变量

$$E[X] = \sum_x xP\{X = x\}$$

我们定义它的期望，类似于均值，是这组数据中能得到的平均结果。但是不同于直接取均值的理念，我们从期望的角度，用性质为其定义：如果一处变量在整体中的占比（或者说概率）越大，那么期望就更有可能落在这个数上；此外，概率相同的多个变量在，我们再考虑从变量本身的大小出发，取其均值，定义期望落在这个均值上。

基于这样的思路，我们最终定义了以 $x \times P$ 为单元再对其求和的计算方法。实际上可以理解为加权，以变量的数值加权变量的占比。

离散变量的概率，是对每处的变量值考察其概率。概率学依赖于集合，所以我们永远可以视其为某一部分集合 A 在另一部分集合 B 中的占比，个体之于整体。所以过渡到连续变量，则是对一段变量出现的区间考察其概率，相当于范围之于整体。

所以概率需要积分：

$$P\{a \leq x \leq b\} = \int_a^b p(x) dx$$

所以需要一段区间中每处变量为其概率的加权：

$$E[x] = \int_a^b xp(x) dx$$

当然，乘法是有交换律的，我们也可以反向思考由概率为变量加权进行求和或者积分，殊途同归。