# [番外]蒙特卡罗MC的更新公式怎么来的？

[](https://www.zhihu.com/people/zhang-si-jun-52)

[张斯俊](https://www.zhihu.com/people/zhang-si-jun-52" \t "/home/admin/文档\\x/_blank)

愿成为一把梯子，助你跨过无数的坑。

在上一篇说过，蒙特卡罗的V，其实就是多次G的平均值。如果不明白，请先看上篇内容。

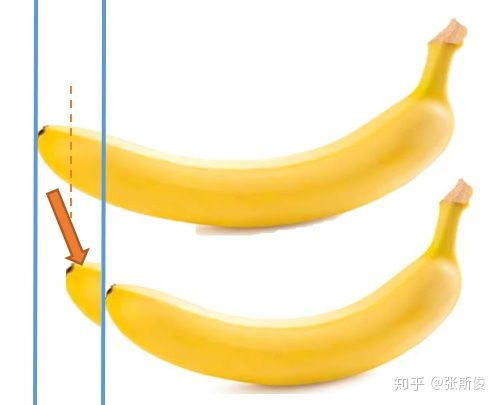
[张斯俊：如何用蒙特卡罗方法（Monte-Carlo）估算V值？​](https://zhuanlan.zhihu.com/p/109755443" \t "/home/admin/文档\\x/_blank)

这样计算其实相当麻烦，因为每一个状态都需要建立一个空间，记录每个轨迹下的G值。

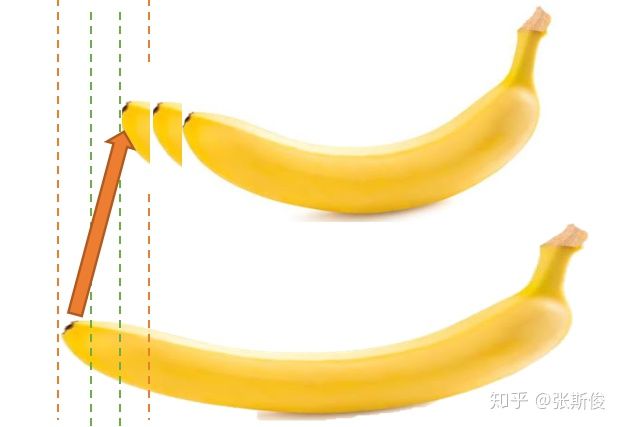
那有没有一种方法，可以用更少的空间计算V值呢？当然有.

### 平均香蕉

有两条长度分别为A，B的香蕉(并假设:A>B)。 如果我要知道他们平均有多长。我只需要把A切成和B长；然后把多出来的那一节，再切一半，接到B就可以了。这时候，称那条加长了的B香蕉为平均香蕉。



如果这时候有第三条C呢？其实原理也一样，比较一下C和平均香蕉，然后把差切出来。但这个时候因为有3条香蕉要平均，所以要分3份，再接到平均香蕉上。再来一根香蕉也一样。



明白了吗？

V是G的平均值，但可以用**增量更新**的方式： 现在只需要记录之前的平均值V，新加进来的G，和次数N。把V和G的差，除以N，然后再加到原来的平均值V上，就能计算到新的V值。

V\_new = (V\_old - G) \* (1 / N) + V\_old

V\_old：原来的V值，也就是平均香蕉 G：这一次回溯后，计算出来的G值，也就是新加进来的香蕉 N: 这个状态被经过多少次了。 V\_new：新计算出来的V值，也就是平均香蕉。

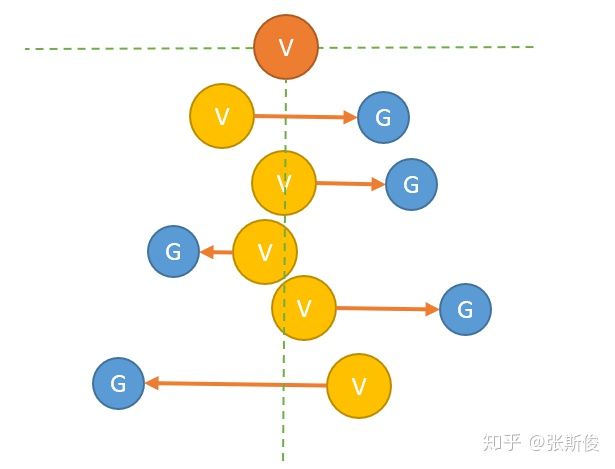
### 更进一步

但其实这样计算还是比较麻烦，甚至可以不用记录N，把(1/N)设置成为一个固定的数，例如0.1、0.2(还记得超参数吗？)。把这个值称为**学习率**。

这就相当于，新来的香蕉G和平均香蕉V的差的十分之一，会被加到平均香蕉上！也就是说，每一次G都会引导V增加一些或者减少一些，而这个V值慢慢就会接近真正的V值。

但这数值不就是不对了吗？但为什么可以这样做呢？ 这是因为V值是一个"客观存在"的值，可以用任何其他方式去求它。而G值是围绕V值的值，G每次出现会都往自己身边拉一下。 就像很多个G在拔河，有的G比V大，有的比G小，有的G力气大，有的G力气小。但没问题，当有足够多的G在的时候，他们一起拉，V就会接近于那个客观存在的V值。

图中：假设橙色的V是客观存在的那个V值，G值会一次又一次把现在的V值(黄色)往左右拉，当次数足够多的时候。V值就和真实的V值很接近了。

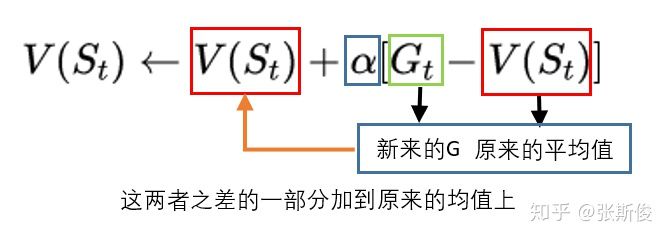


所以，把这里的G，也称为V的**更新目标**。

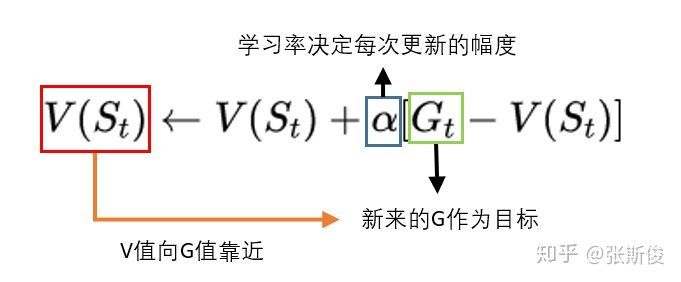
而学习率就可以理解为，每次V向目标靠近的幅度；学习率越大，表示向G靠近的幅度越大，反之则越小。但是不是越大越好呢？大家可以仔细想想。

因此，可以用两种角度理解MC的更新公式：

第一种，用“平均香蕉”来理解。



第二种，用“G的拔河”来理解。



所以大家不用记公式，理解更重要。下一篇TD算法再见！

===========你的支持，就是在下努力的原动力===========

如果专栏对你有用，请点赞并关注在下喔。如果发现有问题，也可以在文章下留言。

你的每一点关注，都是在下的继续努力的动力来源！感激！