期望值 II

1 选择数字游戏

这里是一个游戏，您和我都可以发挥这揭示了一个期望的一个奇怪的属性。

首先，您要想出一个在自然数上概率密度函数。您的分配可以绝对是您想的任何数。举例来说，您可能会选择一个在 1,2 ， … ， 6 上的均匀分布，像公平骰子的滚动有一样的结果。或者您可能会选择一个在 0,1,…,n 上的二项式分布 。您甚至可以给每一个自然数非零的概率，只要这些所有的概率的和是 1。

接 下来，根据您的分布，我挑选一个随机数 z。然后，根据同一分布您拿起一个随机数字 y_1。如果您的数字是大于我的（ $y_1 > Z$ ），则游戏结束。否则，如果我们的数字等于或大于您的($z \geq y_1$ ） ，然后您选择一个新的有同一分布数字与 y_2，并继续选择值，y_3， y_4 等，直到您得到的价值，是严格大于我的数，z,什么是选择的预期您必须选择的人数？

当然，您总是需要至少有一个取数字，因此预期数大于 1。如第 2 或第 3 的答案听起来是合理的，虽然可能有人会怀疑答案取决于分配。让我们看看有没有这样的直觉是正确的。

1.1 分析游戏

您取的数，一定要做出一个自然数数值的随机变量。而且，正如我们已经看到，有一个很好的自然数值的随机变量的公式为期望的公式：

$$\text{Ex}\left(\#\text{ times you pick}\right) = \sum_{k=0}^{\infty} \Pr\left(\#\text{ times you pick} > k\right) \tag{1}$$

假设我已经选择我的数字 z ，和您选择了 k 个数字 y_1 ， y_2，…，y_k。有两种可能性：

• 如果在我们的选择中有一个唯一的大数，那么我的数字是可能就是这个数字，像您的任何一个数字。因此，我的数字是大于的您的所有的数字的概率是 1/（k+1 ），您必须再次选择。

否则，有几个数字并列为最大。我的数字，是有可能称为其中的一个，像您的任何数字一样，所以有概率大于 1 / (k+1） ，您必须选择。

在这两种情况下，用概率至少 1 / （k+1 ） ，您需要比 k 更多的选择来击败我。或者换句话说：

$$\Pr\left(\#\text{ times you pick} > k\right) \geq \frac{1}{k+1} \tag{2}$$

这表明，为了尽量减少您的滚动，您应该选择了是非常罕见的分布。例如，

您可能会选择在{ 1,2 ,..., 10^{100} }上的均匀分布。在这种情况下，概率，您需要更多的比 k 选择次数更多来击败我，是非常接近 1 / （k+1 ），对于对 k 的中度值而言。举例来说，概率，您需要更多的超过 99 次选择的概率，是几乎是 1%。这听起来对您而言是非常有前途的；直觉地，您可能会想在一个合理的平均数目的选择上获胜！

不幸的直觉，有一个最简单的证据，证明您需要的选择数目的期望是无限的，无论其分布如何！让我们把（ 2 ）带入到（ 1 ）：

$$\begin{aligned} \text{Ex}(\#\text{ times you pick}) &= \sum_{k=0}^{\infty} \frac{1}{k+1} \\ &= \infty \end{aligned}$$

这一现象可引起各种混乱！举例来说，假设您有一个通信网络的每个数据包有一个 1/k 的机会用 k 或更多的步骤延迟。这听起来很好，有只有百分之一的机会，而被推迟，有 100 步或更多的步骤。但包延迟的期望其实是无限的！

有一个较大点这里说得好：不是每个随机变量有良定义的期望。这个想法可能首先是令人不安的，但不要忘了预期值只是一个加权平均值。甚至有很多的数字的集合没有传统意义上的平均值，例如：

{ 1 ， -2,3 ， -4,5 ， -6 ,...}

严格来说，我们应该限定几乎所有的涉及的期望的定理，用短语，如" ... ...提供了所有的预期存在。" 但我们正在全力以赴含蓄地离开这个假设。幸运的是，随机变量的期望在实践中不会经常出现。

2 礼券收藏家问题

我每次购买我一次 Taco Bell 的儿童饭菜，我很慷慨地展示”Racin 的火箭“汽车的同一个缩影和我的发射设备，加上一个发射装置，使我以很高的速度可弹射我的新车辆跨越任何桌面或平滑楼。说真的，我的高兴是没有疆界的。

有 n 个不同类型的 racin 火箭车（蓝，绿，红，灰等）。该类型的汽车，由在 Taco Bell 注册，并看起来平均分布和随机独立选择的和蔼的女子每一天颁发给我。为了获取每类 racin '火箭车至少有一个汽车，我必须购买次孩子的饭的期望值是什么？

同一数学问题，在许多名目中出现，例如：为了找出至少一人有每一个可能的生日，您一定要调查的人的期望是什么？在这里，而不是收集 racin 火箭车，您是在收集生日。一般性问题，是所谓的礼券收藏家问题后，又一次诠释。

2.1 采用线性的期望的解决方案

期望的线性是有点递归法和鸽巢原理；一种简单的想法，可用于各种巧妙的方法。举例来说，我们可以以一个聪明的方法来用线性期望来解决代金券收藏家的问题。假设有 5 个不同类型的 racin 火箭，而我接受这个顺序：

蓝色 绿色 绿色 红色 蓝色 桔子色 蓝色 桔子色 灰色

让的划分序列分为 5 部分：

$$\underbrace{\text{blue}}_{X_0}\quad\underbrace{\text{green}}_{X_1}\quad\underbrace{\text{green}\quad\text{red}}_{X_2}\quad\underbrace{\text{blue}\quad\text{orange}}_{X_3}\quad\underbrace{\text{blue}\quad\text{orange}\quad\text{gray}}_{X_4}$$

规则是一个部分结束时，不论何时我得到了一种新型的车。举例来说，中间段结束时，我第一次看到一条红色的赛车。这样，我们就可以打破收集每一个类型的车的问题为阶段。那么我们可以单独分析每一阶段和用线性的期望组装结果。

让我们回到一般的情况下，我收集 n 个 racin 火箭。让空间 X_K 得到第 k 部分的长度。我要购买的孩子的吃饭的总数，以得到所有 n 个 racin 火箭是所有所有这些环节的长度：

$$T = X_0 + X_1 + \ldots + X_{n-1}$$

现在，让我们把注意力集中在 X_K ，第 k 部分的长度。在段 k 的开始，我有 k 不同类型的汽车，以及段的尾部，当我获得一个新的类型。当我拥有的 k 类型，每个孩子的饭含有一个类型，我已经有了与概率的 K /n，因此，每一餐含有一种新型的车用概率 1 -k / n = （ n-k）/n。因此，预计有多少餐，直到我得到了一种新型的汽车是 n / （n-k） ，由"平均失败事件"的公式，我们最后一次算出了。因此，我们必须：

$$\mathrm{Ex}\left(X_k\right) = \frac{n}{n-k}$$

线性期望，加上这种观察，解决了礼券收藏家的问题：

$$
\begin{aligned}
\operatorname{Ex}(T) &= \operatorname{Ex}(X_0 + X_1 + \ldots + X_{n-1}) \\
&= \operatorname{Ex}(X_0) + \operatorname{Ex}(X_1) + \ldots + \operatorname{Ex}(X_{n-1}) \\
&= \frac{n}{n-0} + \frac{n}{n-1} + \ldots + \frac{n}{3} + \frac{n}{2} + \frac{n}{1} \\
&= n\left(\frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{3} + \frac{1}{2} + \frac{1}{1}\right) \\
&= nH_n
\end{aligned}
$$

在总结对 nexttolast

下一行到最后一行是第 n 个有计算反向顺序的调和和。您们可能记得，这个和表示为 H_n，大约是 ln。

让我们用这个一般的解决办法，以回答一些具体的问题。举例来说，预期骰子滚动要求看到每个每从 1 至 6 的数字是：

$$6H_6 = 14.7 \ldots$$

以及您一定要调查的期望的人数，以找出至少一人有每一个可能的生日是：

$$365H_{365} = 2364.6 \ldots$$

3 乘积的期望值

对求和来说够了！随机变量的一个乘积的预期值怎么羊？如果 R_1 和 R_2 是独立的，那么期望值的乘积是该乘积它们期望值的乘积。

定理 1 随机变量 R1 和 R2 ：

$\operatorname{Ex}(R_1 \cdot R_2) = \operatorname{Ex}(R_1) \cdot \operatorname{Ex}(R_2)$

证明：我们将变换右侧到左侧：

$$
\begin{aligned}
\operatorname{Ex}(R_1) \cdot \operatorname{Ex}(R_2) &= \left(\sum_{x \in \operatorname{Range}(R_1)} x \cdot \Pr(R_1 = x)\right) \cdot \left(\sum_{x \in \operatorname{Range}(R_1)} y \cdot \Pr(R_2 = y)\right) \\
&= \sum_{x \in \operatorname{Range}(R_1)} \sum_{y \in \operatorname{Range}(R_2)} xy \Pr(R_1 = x) \Pr(R_1 = y) \\
&= \sum_{x \in \operatorname{Range}(R_1)} \sum_{y \in \operatorname{Range}(R_2)} xy \Pr(R_1 = x \cap R_1 = y)
\end{aligned}
$$

第二行是来自乘以出和的乘积。然后我们用事实，即 R_1 和 R_2 都是独立的。现在我们的组合乘积 xy 是相同的项：

$$
\begin{aligned}
&= \sum_{z\in \mathrm{Range}(R_1\cdot R_2)} \sum_{x,y:\ xy=z} xy \Pr\left(R_1 = x \cap R_1 = y\right) \\
&= \sum_{z\in \mathrm{Range}(R_1\cdot R_2)} \left(z \sum_{x,y:\ xy=z} \Pr\left(R_1 = x \cap R_1 = y\right)\right) \\
&= \sum_{z\in \mathrm{Range}(R_1\cdot R_2)} z\cdot \Pr\left(R_1\cdot R_2 = z\right) \\
&= \mathrm{Ex}\left(R_1\cdot R_2\right)
\end{aligned}
$$

二

3.1 两个独立的骰子的乘积

假设我们扔的两个独立的，公正的骰子，并乘以碰到的数字。这种乘积的期望值是什么？

设随机变量 R_1 和 R_2 的两个骰子上被显示的数字。我们可以计算出预期乘积价值如下：

$$
\begin{aligned}
\mathrm{Ex}\left(R_1\cdot R_2\right) &= \mathrm{Ex}\left(R_1\right)\cdot \mathrm{Ex}\left(R_2\right) \\
&= 3\frac{1}{2}\cdot 3\frac{1}{2} \\
&= 12\frac{1}{4}
\end{aligned}
$$

对第一行，我们使用定理 1 。那么，我们使用从上次演讲的结果，从上演讲表示，预期值是 $3\frac{1}{2}$。

3.2 两独立骰子的乘积

假设这两个骰子是不是独立的;事实上，假设说，第二骰子始终是和第一个一样。这是否改变预期乘积价值？在定理 1 中的独立的条件真的必要吗？

一如以往，令随机变量 R_1 和 R_2 是在两个骰子上所显示的数字。我们可以直接计算出乘积的期望值，如下：

$$\begin{aligned}\mathrm{Ex}(R_1 \cdot R_2) &= \mathrm{Ex}\left(R_1^2\right) \\ &= \sum_{i=1}^{6} i^2 \cdot \Pr(R_1 = i) \\ &= \frac{1^2}{6} + \frac{2^2}{6} + \frac{3^2}{6} + \frac{4^2}{6} + \frac{5^2}{6} + \frac{6^2}{6} \\ &= 15\frac{1}{6}\end{aligned}$$

第一步，利用这一事实的结果，第二个骰子始终是和第一个骰子一样。那么我们使用我们期望的公式扩大 $\mathrm{Ex}\left(R_1^2\right)$。现在骰子不再是独立的，但预期的乘积值发生了变化为 $15\frac{1}{6}$ 。所以依赖随机变量的乘积的期望不需要等于期望的乘积。

3.3 推论

定理 1 延伸到相互独立的变量的聚集。

推论 2 ：如果随机变量 R_1 和 R_2 ,...,R_n 是相互独立，那么

$$\mathrm{Ex}(R_1 \cdot R_2 \cdots R_n) = \mathrm{Ex}(R_1) \cdot \mathrm{Ex}(R_2) \cdots \mathrm{Ex}(R_n)$$

使用归纳法证明，定理 1 ，和相互独立的定义。我们将略去细节。 我们现在已经知道的随机变量的求和和乘积的期望值。不幸的是，期望值，是不那么容易的找到特征

。这里是一个有缺陷的尝试。

虚假推论 3 如果 R 是的一个随机变量，则

$$\mathrm{Ex}\left(\frac{1}{R}\right)=\frac{1}{\mathrm{Ex}\,(R)}$$

作为一个反例，假设随机变量 R 是有概率 1/2 且是 2 的概率是 1/2。那么，我们有：

$$\begin{aligned}\frac{1}{\mathrm{Ex}\,(R)} &= \frac{1}{1\cdot\frac{1}{2}+2\cdot\frac{1}{2}}\\ &= \frac{2}{3}\\ \mathrm{Ex}\left(\frac{1}{R}\right) &= \frac{1}{1}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}\\ &= \frac{3}{4}\end{aligned}$$

两个数量不能划等号的，因此推论必然要假的。但在这里，是另一种虚假的推论，即我们实际上可以"证明" ！

虚假推论 4：如果 $\mathrm{Ex}\,(R/T)>1$ > 1 ，则 $\mathrm{Ex}\,(R)>\mathrm{Ex}\,(T)$。

"证明" ：我们开始如果部分 ，双方都乘以 Ex（ t ），然后再运用定理 1 ：

$$\begin{aligned}\mathrm{Ex}\,(R/T) &> 1\\ \mathrm{Ex}\,(R/T)\cdot\mathrm{Ex}\,(T) &> \mathrm{Ex}\,(T)\\ \mathrm{Ex}\,(R) &> \mathrm{Ex}\,(T)\end{aligned}$$

□

这个"证明”是假的！第一步就是有效，如果只有 Ex(T) > 0 。更重要的是，我们不能应用定理 1 在第二个步骤，因为 R/ T 和 T 是不一定独立。不幸的是，事实推论 4 是虚假的，并不意味着它是从未用过！

3.3.1 的 RISC 悖论

下列数据是从由一些著名的教授写的论文中。他们想说明在 RISC 处理器上的程序一般都短于在 CISC 处理器上的程序。为此，他们作出了一个一些基准程序问题的长度的表，看上去是这样的：

Benchmark	RISC	CISC	CISC / RISC
E-string search	150	120	0.8
F-bit test	120	180	1.5
Ackerman	150	300	2.0
Rec 2-sort	2800	1400	0.5
Average			1.2

每一行都包含一个基准程序的数据。在前面两列的数字为每个类型的处理器程序长度。第三列包含 CISC 程序长度和 RISC 程序长度的比例。在所有的基准上给出值 1.2 在右下。作者总结说"的 CISC 程序平均是 20 ％长的" 。

但有一个相当严重的问题。假设我们重做最后列，使用反比， RISC/CISC 不是的 CISC / RISC 的。

Benchmark	RISC	CISC	RISC / CISC
E-string search	150	120	1.25
F-bit test	120	180	0.67
Ackerman	150	300	0.5
Rec 2-sort	2800	1400	2.0
Average			1.1

完全使用的作者相同的推理所，我们可以断定的 RISC 程序 10 ％长，平均比的 CISC 节目！ 怎么回事？

3.3.2 概率解释

为了阐明这个悖论，我们可以用机械概率论建模的 RISC 和 CISC 之间的辩论与。

令样本空间是一套基准程序。让随机变量 R 是的 RISC 的程序的长度，令随

机变量 C 是 CISC 程序的长度。我们想比较 RISC 程序的平均长度，Ex（ R ），和 CISC 程序的长度。

为了比较平均程序的长度，我们必须为每个样本点指定一个概率；实际上，这会为每一个基准程序分配一个"权重" 。有人可能会想衡量基准，在实际中有多少相似的程序出现。但是让我们按照原来作者马首是瞻。他们发以它们的平均分配每个比例相等的权重，因此这含蓄地假设类似程序以相等的概率出现。让我们这样做的同时，使样本空间一致。现在，我们可以按照如下计算出 Ex（R）和 Ex（C） ，内容如下：

$$\begin{aligned} \mathrm{Ex}(R) &= \frac{150}{4} + \frac{120}{4} + \frac{150}{4} + \frac{2800}{4} \\ &= 805 \\ \mathrm{Ex}(C) &= \frac{120}{4} + \frac{180}{4} + \frac{300}{4} + \frac{1400}{4} \\ &= 500 \end{aligned}$$

所以 RISC 程序的平均长度的的其实就是 $\mathrm{Ex}(R) / \mathrm{Ex}(C) = 1.61$ 倍大于一个 CISC 程序平均程度。 RISC 比任何以前的两份答案暗示的更差！

更具我们的概率模型，作者对每一个采样点计算 C/R，然后平均每场获得特的 Ex(C/R)是 1.2 。这多少是正确的。不过，他们解释这是指的 CISC 程序时间超过 RISC 的程序中介绍了平均水平。因此，关键的结论，这一具有里程碑意义的论文，就基于推论 4 的 ，我们知道它是假的！

3.3.3 一个比较简单的例子

问题的根源是在下面更明确，简单的例子。假设数据如下。

Benchmark	Processor A	Processor B	B/A	A/B
Problem 1	2	1	1/2	2
Problem 2	1	2	2	1/2
Average			1.25	1.25

现在的处理器 A 和 B 的统计数字是完全对称。然而，从第三个列我们的结论是，处理器 B 的程序平均长 25 ％，并从第 4 个栏，我们会断定处理器 A 的程序平

均长 25 %。这两个结论显然是错误的。平均数的比率的模型，可以是很具误导性。更普遍的，如果您计算一个商数的期望，三思；您将会为滥用和错误解释获得成熟的值。

4 全期望定理

早些时候我们谈到条件概率。例如，您可能想知道有人处理至少有两个 A 的概率，因为他们处理的方片 A。同样地，一个可以谈条件期望。例如，您可以决定公平模具骰子碰到期望数，个定滚动是偶数。

有 几种方法来计算条件期望，正如有几种方法来计算一个普通的期望。事实上，在条件期望公式是和一般的期望公式是一样的，除非所有的概率成为有条件的概率。如果 R 是的一个随机变量且 E 是一个事件，那么，给出事件 E 的 R 的预期值是表示为 Ex(R|E)，定义为：

$$\begin{aligned}\operatorname{Ex}(R \mid E) &= \sum_{w \in S} R(w) \Pr(w \mid E) \\ &= \sum_{x \in \text{range}(R)} x \cdot \Pr(R = x \mid E)\end{aligned}$$

举例来说，令 R 是公平骰子上碰到的数字，令 E 是数字是偶数的事件。让我们计算 Ex(R|E) ，骰子滚动的预期值，给出其结果是偶数。

$$\begin{aligned}\operatorname{Ex}(R \mid E) &= \sum_{w \in \{1,\ldots,6\}} R(w) \cdot \Pr(w \mid E) \\ &= 1 \cdot 0 + 2 \cdot \frac{1}{3} + 3 \cdot 0 + 4 \cdot \frac{1}{3} + 5 \cdot 0 + 6 \cdot \frac{1}{3} \\ &= 4\end{aligned}$$

条件期望对于分解期望成部分的计算是有用的。分解通过和全概法则一致来判定。

定理 5 （全期望） 令 E_1，…，E_N 是分割样本空间的事件，并有非零概率。如果 R 是的一个随机变量，那么：

$$\mathrm{Ex}(R) = \mathrm{Ex}(R \mid E_1) \cdot \Pr(E_1) + \cdots + \mathrm{Ex}(R \mid E_n) \cdot \Pr(E_n)$$

举例来说，令 R 一个公平骰子上出现的数字和 E 是结果是偶数的事件，正如以前。那么，$\overline{E}$ 是结果是奇数的事件。因此，全期望定理的内容为：

$$\underbrace{\mathrm{Ex}(R)}_{=7/2} = \underbrace{\mathrm{Ex}(R \mid E)}_{=4} \cdot \underbrace{\Pr(E)}_{=1/2} + \underbrace{\mathrm{Ex}\left(R \mid \overline{E}\right)}_{=?} \cdot \underbrace{\Pr(E)}_{=1/2}$$

唯一的数量在这里，我们不知道的是 $\mathrm{Ex}\left(R \mid \overline{E}\right)$，这是预期的骰子的滚动，由于其结果是奇数的。为这个未知数求解这个方程，我们的结论是 $\mathrm{Ex}\left(R \mid \overline{E}\right) = 3$。

为了证明全期望定理，我们一开始就有一个引理。

引理：设 R 是一个随机变量， E 是一个有正面概率的事件，且 Ie 是对 E 的指示变量。那么

$$\mathrm{Ex}(R \mid E) = \frac{\mathrm{Ex}(R \cdot I_E)}{\Pr(E)} \tag{3}$$

证明：注意，对任何结果，s，在样本空间中，

$$\Pr(\{s\} \cap E) = \begin{cases} 0 & \text{if } I_E(s) = 0, \\ \Pr(s) & \text{if } I_E(s) = 1, \end{cases}$$

同样：

$$\Pr(\{s\} \cap E) = I_E(s) \cdot \Pr(s). \tag{4}$$

现在：

$$
\begin{aligned}
\mathrm{Ex}(R \mid E) &= \sum_{s \in S} R(s) \cdot \Pr(\{s\} \mid E) && (\text{Def of } \mathrm{Ex}(\cdot \mid E)) \\
&= \sum_{s \in S} R(s) \cdot \frac{\Pr(\{s\} \cap E)}{\Pr(E)} && (\text{Def of } \Pr(\cdot \mid E)) \\
&\sum_{s \in S} R(s) \cdot \frac{I_E(s) \cdot \Pr(s)}{\Pr(E)} && (\text{by } (4)) \\
&= \frac{\sum_{s \in S} (R(s) \cdot I_E(s)) \cdot \Pr(s)}{\Pr(E)} \\
&= \frac{\mathrm{Ex}(R \cdot I_E)}{\Pr(E)} && (\text{Def of } \mathrm{Ex}(R \cdot I_E))
\end{aligned}
$$

现在我们证明全期望定理：

证明：由于 E_i 的分割样本空间，

$$
R = \sum_i R \cdot I_{E_i} \tag{5}
$$

任何随机变量 R，有：

$$
\begin{aligned}
\mathrm{Ex}(R) &= \mathrm{Ex}\left(\sum_i R \cdot I_{E_i}\right) && (\text{by } (5)) \\
&= \sum_i \mathrm{Ex}(R \cdot I_{E_i}) && (\text{linearity of } \mathrm{Ex}()) \\
&= \sum_i \mathrm{Ex}(R \mid E_i) \cdot \Pr(E_i) && (\text{by } (3))
\end{aligned}
$$

$\llcorner$