随机变量

我们已经使用概率建模各种各样的实验、游戏和测试。 在中，我们设法计算事件的概率。 我们问，例如，您赢得 Monty Hall 比赛事件的概率是什么？天下雨事件的概率是什么，给出那天气预报人员今天拿了他的伞？您得一种罕见的疾病的事件的概率是什么，假设您检查呈阳性？

但是您可能询问更加一般的问题实验。雨能下多大？ 这病症将持续多久？整天玩 6.042 游戏我将丢失多少？ 这些问题是功能上不同和不容易根据事件表述。 问题是事件或发生或不发生： 您赢了或输了，您病了或者不病。 但是这些新的问题是关于程度问题： 多少，多么大，多久？ 要求解这些问题，我们需要一个新的数学工具。

1 随机变量

让我们从例子开始。 考虑扔三独立公正硬币的实验。令 C 是出现正面的数量。 令 M＝1，如果三枚硬币都是正面或都是反面，否则 M =0。 现在三次硬币翻转的每个结果都唯一地确定 C 和 M 的值，例如，如果我们翻转正面，反面，正面，那么 C =2 且 M =0。 如果我们翻转反面，反面，反面，则 C =0 且 M=1。实际上， C 计数正面的数量，并且 M 表明所有硬币都匹配。

因为每个结果单独地确定 C 和 M，我们可以认为是结果的它们的数字的映射。 对于这个实验，样本空间是：

$$S = \{HHH,\ HHT,\ HTH,\ HTT,\ THH,\ THT,\ TTH,\ TTT\}$$

现在 C 是如下的映射在样本空间的每个结果映射为数字的函数：

$$\begin{array}{rclrcl} C(HHH) & = & 3 & C(THH) & = & 2 \\ C(HHT) & = & 2 & C(THT) & = & 1 \\ C(HTH) & = & 2 & C(TTH) & = & 1 \\ C(HTT) & = & 1 & C(TTT) & = & 0 \end{array}$$

同样， M 是映射每个结果的函数的另一个方式：

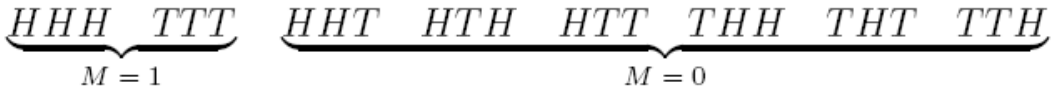
$$\begin{array}{rclrcl} M(HHH) & = & 1 & M(THH) & = & 0 \\ M(HHT) & = & 0 & M(THT) & = & 0 \\ M(HTH) & = & 0 & M(TTH) & = & 0 \\ M(HTT) & = & 0 & M(TTT) & = & 1 \end{array}$$

函数 C 和 M 是随机变量的例子。 一般来说，一个随机变量是定义域为样本空间的函数。 (值域通常可以是任意的，但是我们通常使用实数的一个子集。) 注意命名“随机变量”是用词不当的，随机变量实际上是一个函数!

1.1 指示器随机变量

指示器随机变量(或完全指示或者 Bernoulli 随机变量)是映射每个结果到 0 或 1 的一个随机变量。 随机变量 M 是例子。 如果全部三枚硬币匹配，那么 M =1; 否则， M =0。

指示器随机变量是密切相关对事件。 特别是，指示器把样本空间分成结果映射到 1 部分和那些结果映射到 0 的部分。 例如，指示器 把样本空间 M 分为如下的两个块：



同样地，事件划分样本空间为这些在事件中的结果和那些不在事件里结果。所以，每个事件自然地同某一指示随机变量联系在一起反之亦然： 事件 E 的显示器是一个指示器随机变量，对于在 E 中的结果是 1，不在 E 中的结果是 0。因此。因此， M 是全部三枚硬币匹配的事件的指示随机变量。

1.2 随机变量和事件

事件和更加一般的随机变量之间的一个很强的关系。随机变量把样本空间分成几个块。 例如， C 分割样本空间如下：

因此每个块是样本空间的一个子集并且是事件。 因此，我们可以把涉及到一个随机变量的等式或不平等作为事件。 例如，事件 C =2 包括结果 THH、HTH 和 HHT。 事件 C ≤ 1 包括结果 TTT、TTH、THT 和 HTT。

自然足够的，我们可以谈论等式定义的事件的概率和涉及到随机变量的不平等的概率。 例如：

$$\begin{aligned}\Pr(M=1) &= \Pr(TTT) + \Pr(HHH) \\ &= \frac{1}{8} + \frac{1}{8} \\ &= \frac{1}{4}\end{aligned}$$

另一个例子：

$$\begin{aligned}\Pr(C \geq 2) &= \Pr(THH) + \Pr(HTH) + \Pr(HHT) + \Pr(HHH) \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\ &= \frac{1}{2}\end{aligned}$$

这是相当狂放的; 一个人通常认为等式和不平等或者为真实，或者为假。 但是，当变量被随机变量替换，就有了关系成立的概率!

1.3 条件概率

混合条件变量和涉及随机变量的事件创造了新的困难。例如， $\Pr(C \geq 2|M=0)$ 是至少二枚硬币是正面的概率($C \geq 2$)，假设不是所有的三枚硬币是相同的($M = 0$)。 使用有条件概率的定义，我们可以计算这个概率：

$$\begin{aligned}\Pr(C \geq 2 \mid M=0) &= \frac{\Pr(C \geq 2 \cap M = 0)}{\Pr(M=0)} \\ &= \frac{\Pr(\{THH, HTH, HHT\})}{\Pr(\{THH, HTH, HHT, HTT, THT, TTH\})} \\ &= \frac{3/8}{6/8} \\ &= \frac{1}{2}\end{aligned}$$

在第一行的表达式 $C \geq 2 \cap M = 0$ 看起来很奇怪;不等式和等式之间的集运算∩做了什么？ 但是回忆，在这上下文， $C \geq 2$ 且 $M = 0$ 是事件，结果的集合。 因此采取它们的交集点是完全有效的!

1.4 独立

独立的概念也可以从事件转入到随机变量。 随机变量 R_1 和 R_2 是独立的，如果

$$\Pr(R_1 = x_1 \cap R_2 = x_2) = \Pr(R_1 = x_1) \cdot \Pr(R_2 = x_2)$$

对于所有的在 R_1 的 x_1 和在 R_2 的 x_2。

正如事件，我们可以公式化随机变量的独立，用一个等价的可能更加直觉的方式： 随机变量 R_1 和 R_2 是独立的，当且仅当：

$$\Pr(R_1 = x_1 \mid R_2 = x_2) = \Pr(R_1 = x_1) \text{ or } \Pr(R_2 = x_2) = 0$$

对于所有的在 R_1 值域的 x_1 和在 R_2 值域的 x_2 的。换句话说， R_1 取一个特殊的值的概率没有收到 R_2 的值的影响。

作为一个例子， C 是否是和 M 独立的？ 直观地，答案应该是“不”。 正面的数量， C，完全地取决于全部三枚硬币是否匹配; 即是否 M =1。 但是为了核实这种直觉我们必须发现某一 x_1， $x_2 \in R$ 这样：

$$\Pr(C = x_1 \cap M = x_2) \neq \Pr(C = x_1) \cdot \Pr(M = x_2)$$

一个适当的值的选择是 x_1 =2 和 x_2 =1。 在这种情况中，我们有：

$$\Pr(C = 2 \cap M = 1) = 0 \qquad \text{but} \qquad \Pr(C = 2) \cdot \Pr(M = 1) = \frac{3}{8} \cdot \frac{1}{4} \neq 0$$

独立的概念可以泛化到随机变量的集合。 随机变量 R_1， R_2，…， R_n 是相互独立的，如果

$$\begin{aligned}\Pr(R_1 = x_1 \cap R_2 = x_2 \cap \cdots \cap R_n = x_n) \\ = \Pr(R_1 = x_1) \cdot \Pr(R_2 = x_2) \cdots \Pr(R_n = x_n)\end{aligned}$$

为所有在 R_1，…，R_n 值域的 x_1，…。

相互独立的这个定义的后果是变量子集的一个赋值的概率等于单独赋值的概率的乘积。因此，例如，如果 R_1， R_2，…， R_{100} 是相互的有值域 N 的随机变量，那么它跟随：

$$\Pr(R_1 = 9 \cap R_7 = 84 \cap R_{23} = 13) = \Pr(R_1 = 9) \cdot \Pr(R_7 = 84) \cdot \Pr(R_{23} = 13)$$

(这通过求和其他随机变量的所有可能的价值而成立; 我们省去细节。)

1.5 骰子的一个例子

假设我们滚动二个公平，独立的骰子。 这个实验的样本空间包括所有对(r_1，r_2) 这里 r_1， r_2 ∈{1,2,3,4,5,6}。 例如，因此结果(3,5)对应于第一个骰子滚出 3 和第二个骰子滚出 5。 在样本空间每个结果的概率的是 1/6 • 1/6 =1/36，因为骰子是公平和独立的。

我们能把每个骰子出现的数字作为随机变量 D_1 和 D_2。 因此 D_1 (3,5) =3 和 D_2 (3,5) =5。 那么表达式 $D_1 + D_2$ 是另一个随机变量;让我们用 T 代表 “全”。 精确地，我们定义：

T (w)= D_1 (w)+ D_2 (w)，为每个结果 w

因此 T (3,5) = D_1 (3,5) + D_2 (3,5) =3+ 5 =8。一般来说，随机变量的所有函数本身是一个随机变量。 例如 $\sqrt{D_1} + \cos(D_2)$ 是一个奇怪的，但是明确定义的随机变量。

让我们也定义了指示随机变量 S，为两个骰子的总的数和是 7 的事件：

$$S(w) = \begin{cases} 1 & \text{if } T(w) = 7 \\ 0 & \text{if } T(w) \neq 7 \end{cases}$$

因此，当总和是 7 时候，S 等于 1；否则 S 等于 0。例如， S (4,3) =1，但是 S (5,3) =0。

让我们考虑关于独立的两个问题。 首先， D_1 和 T 是否是独立的？知觉的，答案将似乎是“不”，因为在第一出来的数字强烈影响这两个骰子的值。 但是为了证明这个，我们必须发现这样的整数 x_1 和 x_2：

$$\Pr(D_1 = x_1 \cap T = x_2) \neq \Pr(D_1 = x_1) \cdot \Pr(T = x_2)$$

例如，我们也许选择 x_1 =2，并且 x_2 =3，我们有

$$\Pr(T = 2 \mid D_1 = 3) = 0$$

因为共和不可能只是 2 当一个骰子是 3 的时候。另一方面，我们有：

$$\Pr(T=2)\cdot\Pr(D_1=3)=\Pr(\{1,1\})\cdot\Pr(\{(3,1),\ (3,2),\ \ldots,\ (3,6)\})$$
$$=\frac{1}{36}\cdot\frac{6}{36}\neq 0$$

因此，如同我们怀疑，这些随机变量不是独立的。

S 和 D_1 是否是独立的？ 再次，直觉建议答案是“no”。 在在第一个骰子上的数字应该影响总和是否是等于 7。但是这次直觉结果是错误的! 这两个随机变量实际上是独立的。

证明二个随机变量是独立需要花费一些工作。 (幸运地，这是一项不凡的任务; 通常独立是一个建模的假定。随机变量很少意想不到地变成是独立的。) 在这种情况下，我们必须说明：

$$\Pr(S=x_1\cap D_1=x_2)=\Pr(S=x_1)\cdot\Pr(D_1=x_2) \qquad (1)$$

对于所有 $x_1\ \in\ \{0,1\}$和所有 $x_2\ \in\ \{1,2,3,4,5,6\}$。 我们可以通过两个部分来求出这些概率：

•假设 $x_1=1$。 那么对于 x_2 的每值我们有：

$$\Pr(S=1)=\Pr((1,6),\ (2,5),\ \ldots,\ (6,1))=\frac{1}{6}$$
$$\Pr(D_1=x_2)=\Pr((x_2,1),\ (x_2,2),\ \ldots,\ (x_2,6))=\frac{1}{6}$$
$$\Pr(S=1\cap D_1=x_2)=\Pr((x_2,7-x_2))=\frac{1}{36}$$

因为$1/6\cdot 1/6=1/36$，独立条件是满足的。

·否则，假设 $x_1=0$。 那么我们有 Pr (S=0) =1 − Pr (S=1) =5/6 和 Pr (D_1 = x_2) =1/6 正如以前。 现在事件

$$S=0\cap D_1=x_2$$

包括 5 个结果： 所有(x_2,1)， (x_2,2)，…， (x_2,6)除了(x_2,7− x_2)。 所以，这个事件的概率是 5/36。 从 5/6 • 1/6 =5/36，独立条件满足。

因此， 第一个骰子滚动出的结果是和总和是 7 的事实是独立的。 这是一个奇怪，被隔绝的结果; 例如，第一次滚动不是总和是 6 或 8 除 7 之外的事实独立

的。 但是这个例子说明，独立随机变量的数学概念－－当紧密和“无关量”的直觉概念关联是—不是同一件事。

2 概率分布

随机变量被定义为是试验的样本空间的函数。 通常，然而，与同样属性的随机变量显示出完全不同的试验。例如，和投票有关的随机变量，在主要测试中和硬币翻转中，所有都共享的某一属性。 如果我们能以抽象、从任何试验的细节中分离出来，那么我们的结论可能应用到所有的这种随机变量出现的试验中。这种一般的结论是非常有用的。有两个工具捕获随机变量的重要属性，但是把相关试验的细节放到了后面。

有值域 V 的随机变量 R 的密度函数是一个函数：

PDF_R：V à [0,1]，定义为：

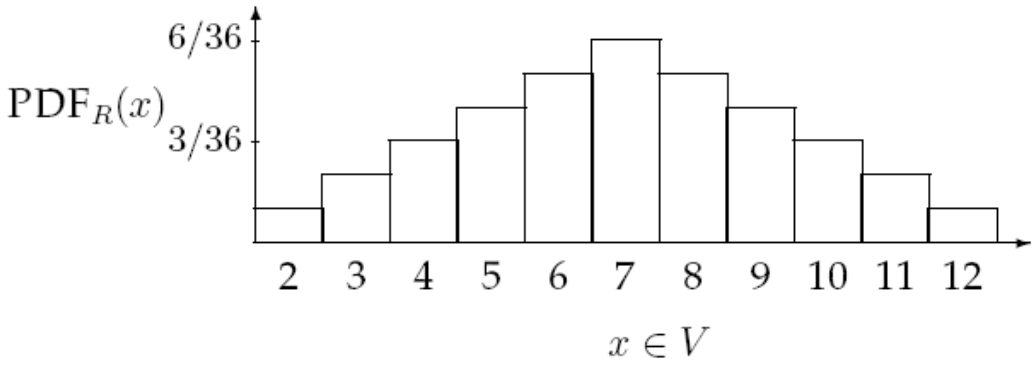
$$\mathrm{PDF}_R(x) = \Pr(R = x)$$

这个定义的结果是：

$$\sum_{x \in V} \mathrm{PDF}_R(x) = 1$$

因为随机变量总是取集合 V 中的一个值。

作为为例，让我们回归到滚动两个公平，独立骰子的实验。 正如以前，令 T 是两个转动的共计。 这个随机变量从集合 v= {2,3，…， 12}取值。 概率密度函数的分布图如下所示：



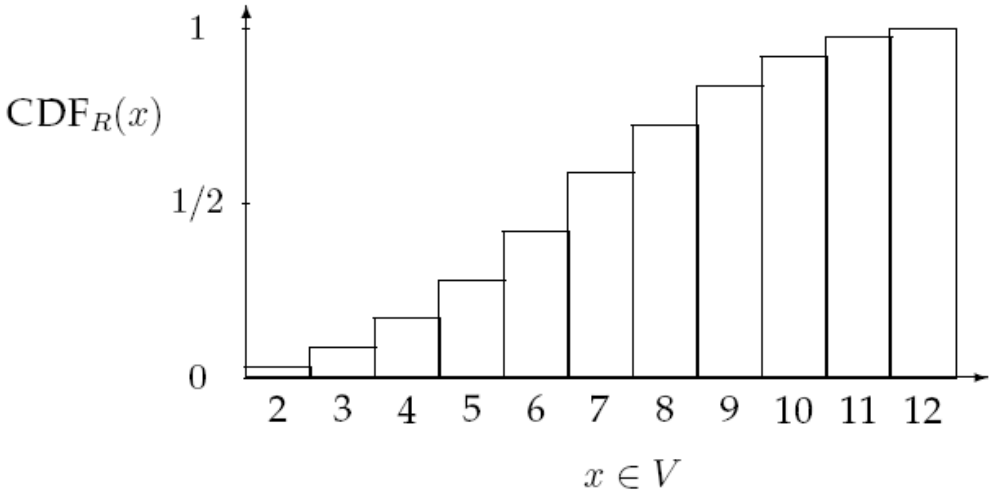
在中部的隆起表明总和接近 7 是很可能。 所有长方形总面积是 1，因为骰子必须恰好从 V = {2,3，…， 12}取一个值为总和。

密切相关的想法是一个随机变量的累积分布函数(cdf)

这个函数$\mathrm{CDF}_R : V \to [0,1]$定义为：

$$\mathrm{CDF}_R(x) = \Pr(R \leq x)$$

作为例子，随机变量 T 的累积分布函数如下显示：



在累积分布函数的第 i 个长条的高度的等于在概率密度函数左边的 i 个条的高度的总和。这和 pdf 和 cdf 的定义是一致的。

$$\begin{aligned}\mathrm{CDF}_R(x) &= \Pr(R \leq x) \\ &= \sum_{y \leq x} \Pr(R = y) \\ &= \sum_{y \leq x} \mathrm{PDF}_R(y)\end{aligned}$$

总之， PDF_R (x)测量 R = x 的概率和 CDF_R(x)测量 R ≤ X 的概率。 PDF_R 和 CDF_R 捕获关于随机变量 R 的同样信息－－您能从另一获得另一个---但是有时一个是更加方便的。 这里的关键不是概率密度函数，也不是涉及到实验的样本空间的累积分布函数。 因此，通过这些函数，我们可以不参考特殊试验来学习随机变量。

于今天剩下的部分，我们看看三重要分布和它们的一些应用。

2.1Bernoulli 分布

由于指示随机变量也许最普通的类型，因为它们和事件的紧密关联。随机变量的指示随机函数 B 的概率密度是

$$\mathrm{PDF}_B(0) = p$$
$$\mathrm{PDF}_B(1) = 1 - p$$

这里 0≤ p≤ 1。 对应的累积分部函数是：

$$\mathrm{CDF}_B(0) = p$$
$$\mathrm{CDF}_B(1) = 1$$

这个被称为 Bernoulli 分布。 在一枚硬币上的翻转为正面的数目（可能是有偏差的）是一个 Bernoulli 分部。

2.2 平均分布

随机变量以相同的概率取每个可能的值称为平均。 例如， 在集合{1,2，…，N}上的随机变量 U 的概率密度函数是：

$$\mathrm{PDF}_U(k) = \frac{1}{N}$$

并且累积分布函数是：

$$\mathrm{CDF}_U(k) = \frac{k}{N}$$

平均分部总是碰到。 例如，在公平滚动的骰子的数字是在集合{1,2，…， 6}上的一致分部。

2.3 数字游戏

让我们玩一个游戏！我有二个信封。 其中每一包含一个在范围 0,1 整数，…100，并且数字是不同的。 要赢得游戏，您必须确定哪个信封包含一个更大的数。 要给您机会，我会让您偷看在随机选择的一个信封的数字。 您

能否给您构想出一个赢取的比 50%更好的机会的战略？

例如，您可能随机取一个信封和猜测它包含更大的数。 但是这个战略只赢取 50%的时间。 您的挑战做得将更好的。

因此您也许设法是更加聪明的。 假设您偷看在左信封和看到数字 12。 因为 12 是一个小数字，您也许猜测另一个数字是更大。 或许，我在两个信封都放了小数字。那么您的猜测也许不是好的！

这里重点是在信封的数字可能不是任意的。 我要我想能打败您的猜测战略的方法选择数字。 我将使用随机数来选择满足我这边的数字： 让您输了！

2.3.1 在赢的战略之后的直觉

令人惊奇的是，有赢取超过 50%时间，不管我在信封投入什么数字的战略！

假设您莫名其妙地知道我的编号下限和高数量范围之间的数字 x。现在您在信封偷看并且看见一或看到另一个数字。 如果它大于 x，则您知道您正在看一个较大的数字。 如果它小于 x，然后偷看较小的数。 换句话说，如果您知道我在较低和较高的数字之间，那么你一定可以赢得游戏。

这个聪明的战略的仅有的缺是你不知道 x。哦，好吧。

但是若您要尝试猜测 x 呢？ 您恰好有某以概率猜测的某一概率。 在这种情况下，您赢取 100%时间。另一方面，如果您不正确地猜测，那么您不会比前面更糟糕; 您的赢取的机会仍然是 50%。 结合这两个情况，您整体的获胜的机会比 50%好！

关于概率的非正式的论据，像这一个，经常合理振振有词，但是不阻止受到严密细查。 相反，这个论据完全地听起来完全不合理—，但是实际上是正确！

2.3.2 分析获胜策略

对于一般性的，假设我能从集合{0,1 ， … ， n}选择数字。
名为较小的数和较大的数。

你的目标是猜出了在 L 和 H 之间的数字，以避免混淆相等的情况，您选择随机从半个整数中选择 x：

$$\left\{\frac{1}{2},\ 1\frac{1}{2},\ 2\frac{1}{2},\ \ldots,\ n-\frac{1}{2}\right\}$$

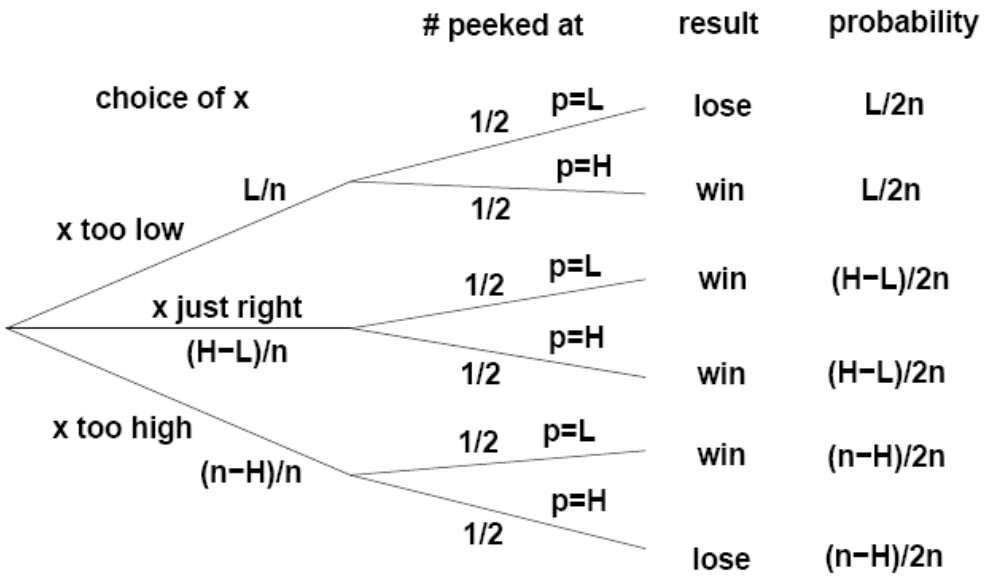
但你应该用的概率分布是什么吗？

均匀分布，变为您最好的选择。一个非正式的理由是， 如果我猜出您不可能选择一些号码比如说$50\frac{1}{2}$，那么我将总是把 50 和 51 放到信封中。那么您不大可能从 L 和 H 之间选择一个 x，当然有较少的机会获胜了。

当您选择了数字 x 后，您偷看了一个信封，且看到一些数字 p。如果 p > x，那么您猜测您看到的是更大的数字。如果 p < x，那么您猜测另外的数字是较大的。

所有的剩余的要判断的是这个策略成功的概率。我们能用通常的 4 步方法和数图来完成这个。

第 1 步：找到样本空间。您或者选择 x 太小（<L），或者太高(>H)，或者正好（L<x<H）。那么您或者偷看较小的数字(p=L)或者更高的数字(p=H)。这个给出六个可能的输出。）



第 2 步：定义感兴趣的事件。在您获胜的事件中的 4 个结果是在树图中标记。

第 3 步：分配输出概率。首先，我们分配边概率。您猜测 x 是太小的概率 L/n，太高的概率(n-H)/n，正好的概率(H-L)/n。下一步，您以相同的概率看了或者小的或者大的数字。把从根到叶的路径上的数值给出结果概率。

第 4 步：计算事件概率。您获胜的事件的概率是 4 个结果的概率的和。

$$\begin{aligned}\Pr(\text{win}) &= \frac{L}{2n} + \frac{H-L}{2n} + \frac{H-L}{2n} + \frac{n-H}{2n} \\ &= \frac{1}{2} + \frac{H-L}{2n} \\ &\geq \frac{1}{2} + \frac{1}{2n}\end{aligned}$$

最后不等式依赖于一个事实，即数目较高的 H 是至少大于数目较低的数字 L 1，因为他们必须有所区别。

果然，你用这个策略赢有一半以上的时间，不管数量多少，在信封！举例来说，如果我在 0 ， 1 ， 。 。 。 ， 100 范围选择号码， 那么您获胜的概率至少 $\frac{1}{2} + \frac{1}{200} = 50.5\%$。甚至更好，如果我只允许在 0…10 范围内的数字，那么您的获胜的机率上升到 55％ ！由拉斯维加斯的标准，这些都是伟大的赔率！

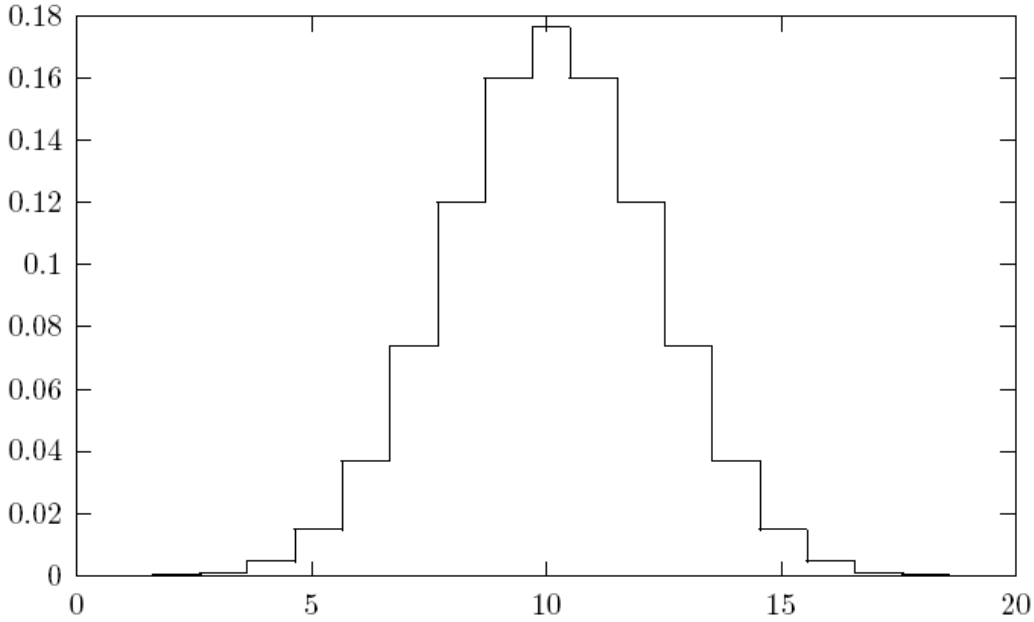
2.4 二项分布(binomial distribution)

对较复杂的分布，在计算机科学中二项分布肯定是最重要的。标准的例子，一个有二项分布随机变量是在 n 个独立的翻转的硬币正面数的分布;称这个为 H。如果硬币是公平的，那么 H 是一个公平的二项式密度函数：

$$\text{PDF}_H(k) = \binom{n}{k} 2^{-n}$$

这个是成立的，因为存在 n 个硬币抛掷的 $\binom{n}{k}$ 序列，恰好有 k 个正面和每个这样的序列有概率 2^{-n}。

这是公平概率密度函数 $\text{PDF}_H(k)$ 的分布图，根据 n = 20 硬币翻转。最有可能的结果是 K = 10 正面，为了一个 k 的较大和较小的值。这些下滑到峰的左边和右边的地区域，通常称为分布的尾部。



在计算机科学中的大量的分析来证明这二项式的尾部和类似的分布非常小的。问题的上下文中，这通常意味着有非常小的概率发生“坏”的事情，这可能是一个服务器或通信链路超载或随机算法运行异常长的时间或生产了错误的结果。

2.4.1 一般二项分布

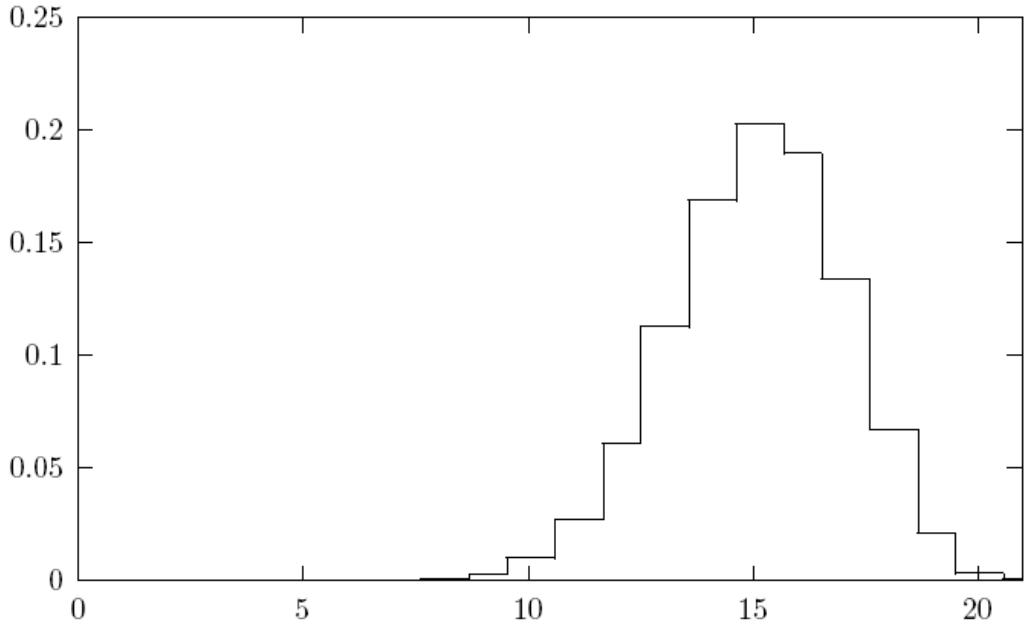
现在令 j 是 n 个独立的硬币出现的正面的个数，每一个正面有概率 p，那么 J 是**一般二项式密度函数**：

$$\mathrm{PDF}_J(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

一如以往，存在 $\binom{n}{k}$ 个有 k 个正面和 n-k 个反面的序列。但是现在每个这样的序列的概率是：$p^k(1-p)^{n-k}$。

因此有 K 序列与 K 元首和 N -k 反面，但现在概率每个这样的序列是的 P K（ 1 - 规划）氮钾。

作为一个例子，下面的分布图说明根据 n=20 的翻转，且出现正面的概率是 p = 0.75 的密度函数$\mathrm{PDF}_J(k)$。该图表明，我们是最有可能获得约当 K = 15 的正面，因为你可能预期的。再次，对 K 的较多或较少的值的概率下落迅速，.



2.4.2 逼近二项式密度函数

存在一个近似的为一般二项密度函数的闭形公式，尽管他有一点使用不便的。首先我们需要一个公式形式的对关键项的逼近，$\binom{n}{k}$。方便的说，让我们用 α n 替换 k，这里 α 是在 0 到 1 之间的数字。那么，由 Stirling 公式，我们发现：

$$\binom{n}{\alpha n} \leq \frac{2^{nH(\alpha)}}{\sqrt{2\pi\alpha(1-\alpha)n}}$$

这里

那里的 H（α），是著名的熵函数：

$$H(\alpha) = \alpha \log_2 \frac{1}{\alpha} + (1-\alpha) \log_2 \frac{1}{1-\alpha}$$

这个在 $\binom{n}{\alpha n}$ 上的上界是非常紧凑并把它作为一个很好的近似。

现在让的把这个公式带入一般二项式密度函数。在 n 次投掷中，翻动 α n 正面的概率得到正面的概率是：

$$\mathrm{PDF}_J(\alpha n) \leq \frac{2^{nH(\alpha)}}{\sqrt{2\pi\alpha(1-\alpha)n}} \cdot p^{\alpha n}(1-p)^{(1-\alpha)n} \qquad (2)$$

这个计算公式是像一个保龄球鞋一样丑陋，但相当有用。举例来说，假设我们翻转一个公平的硬币 n 次。得到恰好 $\frac{1}{2}n$ 正面的概率是什么？带入 α = 1 / 2 和 P = 1 / 2 到这个公式给出：

$$\begin{aligned}\mathrm{PDF}_J(\alpha n) &\leq \frac{2^{nH(1/2)}}{\sqrt{2\pi(1/2)(1-(1/2))n}} \cdot 2^{-n} \\ &= \sqrt{\frac{2}{\pi n}}\end{aligned}$$

因此，例如，如果我们翻转一枚公平的硬币 100 次，确切得到 50 个头的概率是大约 $1/\sqrt{50\pi} \approx 0.079$ 或大约 8%。

2.5 逼近累积二项式分布函数

假设一枚硬币出现正面的概率是 p。 正如前面，让随机变量 J 是在 n 独立翻转出来正面的数量。 然后至多 k 个正面的累积二项式分布函数的概率是：

$$\begin{aligned}\mathrm{CDF}_J(k) &= \Pr\left(J \le k\right) \\ &= \sum_{i=0}^{k} \mathrm{PDF}_J(i) \\ &= \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}\end{aligned}$$

求值这个表达式对于 k 和 n 来说，直接地需要很多工作，并且 n，那么现在逼近是真正地有用的。 再次，我们可以令 k = αn; 即而不是思考正面(k)的绝对数量，我们考虑是头翻转，为正面的比例(α)。 以下逼近是成立的，如果提供 α< p ：

$$\begin{aligned}\mathrm{CDF}_J(\alpha n) &\le \frac{1-\alpha}{1-\alpha/p} \cdot \mathrm{PDF}_J(\alpha n) \\ &\le \frac{1-\alpha}{1-\alpha/p} \cdot \frac{2^{nH(\alpha)}}{\sqrt{2\pi\alpha(1-\alpha)n}} \cdot p^{\alpha n}(1-p)^{(1-\alpha)n}\end{aligned}$$

在第一步，我们界限几何总和对几何级数的总和应用公式。 (细节是愚钝和省去。) 然后我们从前述的部分插入 $\mathrm{PDF}_J(\alpha n)$近似公式(2)。

您必须按在计算器按很多按钮来计算这个公式，对于特定选择的 α， p 和 n。(甚至计算 H(α)也要大量的工作！) 但对于大的 n，求值累积分布函数恰好要求极大的更多的工作! 因此在他们孵化之前，不要以为礼物就已经到手了。 或者其他的事情。

作为例子，在 100 次投掷中，翻转至多 25 正面的概率是通过设置 α =1/4， p =1/2 和 n= 100 获得：

$$\mathrm{CDF}_J(n/4) \le \frac{1-(1/4)}{1-(1/4)/(1/2)} \cdot \mathrm{PDF}_J(n/4) \qquad \le \frac{3}{2} \cdot 1.913 \cdot 10^{-7}$$

这说明，翻 25 或更少的正面是极其不可能的，这是和我们较早时的二项分部的尾部是很小的断言是一致的。事实上，注意到翻动 25 或更少的正面的概率只有比翻转是 25 个正面多 50 %以上的概率。因此，恰好翻动 25 正面两倍于翻转任何数目介于 0 和 24 个正面的概率！

注意：在 $\mathrm{CDF}_J(\alpha n)$ 上的上界的只有当 α <p 的时候成立。如果这不是在您问题中的情况，那么再试图思维补项，也即使，看反面的个数而不是正面的个数。

3 民意调查的哲学

碰到二项分布的地方是在民意调查中。民意调查不仅涉及到技巧数学，而且也涉及到一些哲学问题。

困难是民意调查试图应用概率理论来解决一些事实问题。让我们先考虑一个稍微不同的问题，哪里的问题更严酷。

$$N = 2^{6972607} - 1$$

是一个素数的概率是什么？有人可能会猜 1 / 10 或 1 / 100 。或有人可能会很精明，并指出了素数定理，意味着只有大约 1 到 500 万之间的数字在此范围内是素数。不过，这些答案都是错的。这里没有随机过程，在这里。数 N ，要么是质数，要么是合数。您可以进行很多的"反复较量"如你所愿;答案永远是一样的。因此，似乎概率并不触及这个问题。

然而，根据拉宾和米勒，有一个概率质性测试。如果 n 是合数，
有至少 3 / 4 机会测试就会发现这一点。（在余下的 1 / 4 的时间里，试点是不确定的，它永远不会产生一个错误的答案） 。此外，测试，可运行一次又一次地和结果都是独立的。所以如果 n ，实际上是合数，当 K = 100 重复的 rabin-miller 的概率没有至少发现：

$$\left(\frac{1}{4}\right)^{100}$$

所以 100 个连续的定论的答案，将极有说服力的证据表明， n 是质数！但是，我们还不能说任何关于 N 是质数的概率：这仍然是 0 或 1 ，我们不知道是哪个。

类似的情况出现在民意调查上下文中：我们可以作出一个有说服力的论据，说明了公众舆论是真的，单不能实际说明称述由任何特定的概率为真。假如我们进行一项有 是/不 的就一些问题的调查。那么，我们假定总数的 P 部分会回答"是"和剩下的 1 –p 部分会回答"不" 。 （让我们忘记的人挂断电话就民意测验或开始了关于他们的小狗 Fi-Fi 的长故事－真正的民意测验并没有这样的奢侈！ ） ，现在 P 是一个固定数量，而不是一个随机决定的数量。所以设法确定 P 的一个随机试验，是类似于尝试，判定是否 N 是质数或合数，使用概率质

数测试。

概率滑落到了民意调查，因为民意调查样本的是平均和独立地随机选择。结果是合格的，按照这样说：

"一个可以肯定地说，百分之九十五的信心，最大限度地缘抽样误差是± 3 个百分点" 。

这意味着，在民意调查中的报告的数字是在实际部分 p 的 3%或者其他的不信的 1 在 20 中事件发生，在民意调查过程中；特别的，民意调查的随机样本并不带表大部分人口。这不是同样的事情，正如说的那样 95%的机会民意调查是对的；它或者是或者不是，正如 N 不是质数就是合数，不论 Rabin-Miller 测试的结果。