

## 复习练习 20 的笔记

**问题 1** 以下二个部分不是相关的。 尝试它们，确保您了解随机变量、分布、概率密度函数等等的专科术语。询问您的 TA，如果您不了解或记得什么某一短语的意思。

(a) 假设  $X_1$ 、 $X_2$  和  $X_3$  是三个相互独立随机变量，每一个都是平均的分布。

$$\Pr(X_i = k) \text{ 等于 } 1/3, \text{ 对每一个 } k=1, 2, 3.$$

令  $M$  是给这三个随机变量的最大值另一个随机变量。密度函数  $M$  是什么？

解： 这可以通过计数可能的结果容易地解决：

$M$  是 1 的概率是  $1/27$

是 2 的概率是  $7/27$

是 3 的概率是  $19/27$

(b) 假设  $X$ 、 $Y$  分别是有参数  $(n, p)$  和  $(m, p)$  的二个独立二项式随机变量。  $\Pr(X + Y = k)$  是什么？

解：  $X$  的概率密度函数是把  $n$  个独立硬币翻转中以偏好  $p$  抛掷  $k$  个正面的概率。同样对  $Y$  和  $m$  翻转。 因为，  $X$  和  $Y$  是独立的，  $X + Y$  的概率密度函数和  $n + m$  独立翻转对应，即，  $X + Y$  是有参数  $(n + m, p)$  的二项式变量。 因此，

$$\binom{m+n}{k} p^k (1-p)^{m+n-k}.$$

**问题 2:** 我是上帝。 很严肃的。 因此，我知道大家想的一切。 特别是，我知道谁每一个 250,000,000 个美国人要投票支持在即将来临的竞选。 我知道他们中  $p = 0.52$  部分想要投票支持当前总统。

您必然要死的。 在时空的一个无意义小点。但是在小点之中的相当重大小点。 您工作接近总统，并且，在一个星期之内，您必须回答他痛苦的问题：“我获胜了吗？”或，由于算术专科术语(但是一样痛苦)：“ $p > 1/2$ ”吗？ 1

您的第一想法是要求我(我是上帝)。 但是您已经很长期没有谈了话，因此您知道我不想告诉您。 您的第二个想法是告诉每个美国人，询问他们，然后用 250 百万除回答是的人数。但是您很快体会没有足够时间(有代表性民主的一个原因)。 您的第三个想法。 您没有第三个

想法! 在您的恐慌中, 星期几乎结束, 您随机开始选取美国人, 告诉他们, 并且询问!

令人惊奇的, 那是正确方法。 但是您应该小心您对总统说的东西! 让我们看见。

(a) 在您第一个电话中, 您一致地随机选 1 个美国人, 打电话, 并且问他或她是否将投票支持总统。答复是“是”的概率多大... (i)从我的方面? (ii)从您的方面? 根据硬币翻转您怎么会塑造此?

解: 从我的方面, 它是 0.52。从您的方面, 它也是 0.52。唯一的问题是您可能不知道那个, 因此您称它  $p$ 。清楚地, 从您的方面, 第一个电话是相同的像翻转与未知的偏好的一枚硬币称  $p$  (和我知道 0.52)。

(b) 在您的第二个电话中, 您一致地随机再选 1 个美国人, 打电话, 并且问他或她是否将投票支持总统。但是等等! 当选择第二个选民的时候, 您难道不应该排除您在第一个打电话的人? 如果您排除他或她, 有多坏?

解: 如果您做此, 您修改您翻转的硬币。偏好将增加或减少, 依赖第一个人是否说“是”或“不是”。分析将是杂乱, 因此您不想要这么做。

(c) 因此, 在每一个  $n$  电话, 您一致地随机选 1 个美国人并且询问。您的计划是最后是由  $n$  除正面答复的数量  $M$  得到  $P = M/n$ 。一个 MIT 朋友告诉您那, 作为随机试验的数字结果, 这  $P$  是一个随机变量; 并且那, 根据他的演算,

$$\Pr(|P - p| \leq 0.03) \geq 0.95. \quad (1)$$

当您做打电话的时候, 您做除法得到  $P$  和它是 0.53。您打电话给总统。您说什么?

- 总统先生  $p = 0.53$ !

- 总统先生,  $p$  是在 0.53 的 0.03 之内的的概率至少是 95%!

总统先生, 任一  $p$  是在 0.53 的 0.03 之间或的非常奇怪的事发生(比 100 个发生 5 个少)

对于每个声明答复: (i)您辩解来声称它? (ii)这是否是真实的?

解: 第一个称述明显地不相关。

(i) 因为您可能没有询问所有美国人, 能只发表关于  $p$  的概率性称述。

(ii) 陈述也是错误的, 因为  $p = 0.52$ 。然而, 与选民一个不同的选择, 可能是真实的。当然, 甚而在那种情况中, 您被认为是不恰当的来声称它。

第二称述也是错误的。

(i) 未知数常数的  $p$  在 0.53 的 0.03 之内或远离 0.53 超过 0.03。在第一种情况中, 它和您声称的一致可能性是 1; 在第二种情况中, 它是 0。关键的点是您可能不知道哪个情况成立。您可能提出上述声明, 只有当您知道您是在第一种情况中。悲哀地, 您可能不知道。

(ii) 声称在这种情况下实际上是真的。因为  $p = 0.52$ , 未知的常数的确在 0.53 的 0.03 之内。因此您谈论的概率是 1, 并且至少 95%。但是, 正如同我们说, 它可能是 0 然后声明是错误的。

第三声明是正确一个。

- (i) 你声明这个陈述是正当的。为了说明为什么，从称述开始：  
或者  $|0.53 - p| \leq 0.03$  或  $|0.53 - p| > 0.03$ 。

这明显地是真的。现在读它如下：或者  $p$  在 0.53 的 0.03 范围之内或它不是并且我的任意随机变量  $P$  从集合中取了值，这个集合是在 100 次仅有 5 次命中。因此，明显地或者  $p$  在 0.53 的 0.03 之内或奇怪的事发生了。

- (ii) 声明是真实的。在特例中，它为真，因为前半它是真实的。

从演讲的事实。假设出现正面的概率为  $p$ ，是硬币翻转  $n$  次。然后为所有  $\alpha < p$

$$\Pr(\# \text{ heads} \leq \alpha n) \leq \frac{1 - \alpha}{1 - \alpha/p} \cdot \frac{2^{nH(\alpha)}}{\sqrt{2\pi\alpha(1-\alpha)n}} \cdot p^{\alpha n}(1-p)^{(1-\alpha)n}$$

这里，

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

**问题 3** 一个硬币以概率  $p$  出现正面是硬币被翻转  $n$  次。发现一个上界：

$$\Pr(\text{正面的个数} \geq \beta n)$$

这里  $\beta > p$ 。考虑反面的个数和把数字带入上面大的公式中。

解：  $\Pr(\text{出现正面的次数} \geq \beta n) = \Pr(\text{出现反面的次数} \leq (1 - \beta)n)$

现在反面出现的概率是  $1 - p$ 。因此答复是和如上所述相同的，用  $1 - \beta$  替换  $\alpha$ ，用  $1 - p$  替换  $p$ ：

$$\Pr(\# \text{ heads} > \beta n) \leq \frac{\beta}{1 - \frac{1-\beta}{1-p}} \cdot \frac{2^{nH(\beta)}}{\sqrt{2\pi\beta(1-\beta)n}} \cdot p^{\beta n}(1-p)^{(1-\beta)n}$$

这里我们使用了事实的， $H(1 - \beta) = H(\beta)$ 。

**问题 4:** 在 11 月 2004 日的电话调查发现 35% 美国的成人人口相信进化论是“有证据充分支持的”。电话调查了 1016 个人并且声称误差幅度只有 3 个百分点。

让我们检查一下电话调查的要求。假设有  $m$  成人美国人， $pm$  相信进化论是充分支持的，且  $(1 - p)m$  不相信。电话调查  $n$  一致地和独立地选择随机的美国人。这些， $qn$  相信进化是足够支持的且  $(1 - q)n$  认为不。电话调查然后估计相信进化论的美国人的分数是

q。

注意到，在这个实验的唯一的随机化是谁电话调查选择投票。因此样本空间是  $n$  成人美国人所有序列。第  $i$  个人被调查为“是”的概率是  $p$ ，为“no”的概率是  $1 - p$ 。此外， $n$  的回答是相互独立的

(a) 给出调查估值将是 0.03 或太低的概率的一个上界。请写表达式；不要求值！

解：我们可以认为每个回答是作为的有概率  $p$  为正面的硬币翻转的响应。用这些术语中， $qn$  是翻转为正面的总数。因此我们有：

$$\begin{aligned} \Pr(qn \leq (p - 0.03)n) \\ \leq \frac{1 - (p - 0.03)}{1 - (p - 0.03)/p} \cdot \frac{2^{nH(p-0.03)}}{\sqrt{2\pi(p-0.03)(1-(p-0.03))n}} \cdot p^{(p-0.03)n}(1-p)^{(1-(p-0.03))n} \end{aligned}$$

b) 给出在调查的估值将是 0.03 或更高的概率的一个上界。再次，请写出表达式。

解：使用前面问题的回答和以前一样推理：

$$\begin{aligned} \Pr(qn > (p + 0.03)n) \\ \leq \frac{p + 0.03}{1 - \frac{1-(p-0.03)}{1-p}} \cdot \frac{2^{nH(p+0.03)}}{\sqrt{2\pi(p+0.03)(1-(p+0.03))n}} \cdot p^{(p-0.03)n}(1-p)^{(1-(p+0.03))n} \end{aligned}$$

(b) 这两个答复的总和无论怎样是，电话民意调查便宜 3 个百分点或更多的概率。不幸地是，这些表达式都依靠  $p$ ——未知的电话调查试图估计的进化论相信者的分数。

然而，这两个表达式的总和最大化，当  $p=0.5$ 。因此用  $p=0.5$  和  $n=1016$  评估总和到上界的电话调查的概率错误是 0.03 或更多。民意测验者通常设法保证有 95% 机会实际百分比  $p$  在调查误差范围之间，在这种情况下是  $q \pm 0.03$ 。电话进化调查适当地被设计？

解：错误是 0.03 或更多的概率约为 0.07，因此在一个投票的分数之间的误差范围意味着  $p$  将存在一个错误的调查分数区间中，有概率 0.93。因此我们的估计建议电话民意测验不是足够大能满足被要求的 0.95 概率。因为电话调查是专业的，我们期望他得到了民意测验的大小是合适的，通过使用一个更加准确的数字估计公式—或他也许认为它合法地让非常大错误边缘下降到 0.03。

(d) 如果我们接受所有电话投票数据和演算，我们能认为，存在一个高概率，认为进化是被支持的事实的成年美国人的数量是  $35 \pm 3\%$  ？

解：不。这是事实问题，是真实或错误的。我们可以说任一上面的声明是真实或者在民意测验期间 1 比 20 的事件发生了；明确地，电话调查选择了一个没有代表性的样品。这也许说服您  $p$  “也许”是在范围  $0.35 \pm 0.03$  的，但是没有办法转换那不拘形式的“也许”成一个数学概率。