浅谈切比雪夫不等式

苏 淳　江 涛

说到概率论中的切比雪夫不等式, 不少人就会联想到马尔可夫不等式. 其实, 无论是马尔可夫表达式还是切比雪夫不等式, 都是同一种类型的用以估计概率的不等式. 它们都可以由下述的不等式推出:

定理 1. 设 $g(x)$ 是定义在 $[0,\infty)$ 上的非降的非负值函数, 如果对随机变量 Y, 有 $Eg(|Y|)<\infty$, 则对任何使得 $g(a)>0$ 的 $a>0$, 都有

$$P(|Y|\geq a)\leq \frac{Eg(|Y|)}{g(a)}. \tag{1}$$

如果熟悉示性函数的话, 这个不等式证明起来是很方便的. 先把 $Eg(|Y|)$ 分成两段, 只留下 $|Y|\geq a$ 的一段, 利用 $g(x)$ 的非降性, 就能用不等号把 $Eg(|Y|)I(|Y|\geq a)$ 与 $g(a)P(|Y|\geq a)$ 连接起来, 具体做法如下:

$$\begin{aligned}
Eg(|Y|) &= Eg(|Y|)I(|Y|\geq a)+Eg(|Y|)I(|Y|<a) \quad (\text{分成两段})\\
&\geq Eg(|Y|)I(|Y|\geq a) \quad (\text{只留下一段})\\
&\geq g(a)EI(|Y|\geq a) \quad (\text{利用 } g(x) \text{ 的非降性})\\
&= g(a)P(|Y|\geq a) \quad (\text{因为 } EI(|Y|\geq a)=P(|Y|\geq a)).
\end{aligned}$$

亦即

$$Eg(|Y|)\geq g(a)P(|Y|\geq a),$$

整理后即得 (1) 式.

一般来说, 不等式 (1) 就称为切比雪夫不等式. 只要恰当选择其中的随机变量 Y 和函数 $g(x)$, 就可以得到一系列不同形式的概率不等式.

- 如果对随机变量 Y, 有 $E|Y|^r<\infty$ $(r>0)$, 就可以在 (1) 中取 $g(x)=x^r$, 得到

$$P(|Y|\geq x)\leq \frac{E|Y|^r}{x^r}, \quad \forall\, x>0. \tag{2}$$

- 特别地, 在 $r=2$ 时, 可以通过在 (1) 中选取 $g(x)=x^2$ 和 $Y=X-EX$, 得到

$$P(|X-EX|\geq x)\leq \frac{E(X-EX)^2}{x^2}=\frac{DX}{x^2}, \quad \forall\, x>0. \tag{3}$$

在一些书上把 (3) 式叫做切比雪夫不等式, 而把 (2) 式称为一般情形下的推广的切比雪夫不等式.

- 至于有人所称的关于非负随机变量的如下的马尔可夫不等式:

$$P(X \geq a) \leq \frac{EX}{a},$$

其中 $a > 0$, 更是 (1) 式的极为特殊的形式. 事实上, 只需在 (1) 式中取 $g(x) = x$ 和 $Y = X$ 即可.

有鉴于此, 我们把 (1) 式以及由它所派生出来的各种不等式统统称为切比雪夫不等式.

需要指出, 切比雪夫不等式并不是在任何情况下都能给出概率的精确估计的. 看如下的例子:

例 1. **(1)** 设 $X \sim \mathcal{N}(0,1)$, 试估计概率 $P(|X-0.5| > 0.4)$;
(2) 设 $X \sim \mathcal{N}(\mu,\sigma^2)$, 试估计概率 $P(|X-\mu| > 2\sigma)$.

解: **(1)**如果采用切比雪夫不等式, 取 $g(x) = x^2$, 有

$$P(|X-0.5| > 0.4) \leq \frac{E(X-0.5)^2}{0.4^2} = \frac{EX^2 - EX + 0.5^2}{0.16} = \frac{1.25}{0.16} \approx 8.$$

这个估计不能说它不对, 但确实是个笑话, 不用估计, 任何概率都不大于 1, 何至于大到将近 8 的地步? 这里根本用不上切比雪夫不等式, 因为我们知道随机变量 X 的分布, 所以正确做法是查正态分布表:

$$\begin{aligned} P(|X-0.5| > 0.4) &= P(X-0.5 > 0.4) + P(X-0.5 < -0.4) \\ &= 1 - \Phi(0.9) + \Phi(0.1) = 1 - 0.8159 + 0.5398 = 0.7239. \end{aligned}$$

(2) 若用切比雪夫不等式, 取 $g(x) = x^2$, 有

$$P(|X-\mu| > 2\sigma) \leq \frac{DX}{4\sigma^2} = \frac{1}{4}.$$

但是事实上, 由 $X \sim \mathcal{N}(\mu,\sigma^2)$, 可知

$$\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1),$$

所以

$$P(|X-\mu| > 2\sigma) = P\left(\left|\frac{X-\mu}{\sigma}\right| > 2\right) = 2(1-\Phi(2)) = 0.0456$$

查表得出来的结果较之用切比雪夫不等式要精确得多.

但是在随机变量 X 的分布未知, 只知道它的期望与方差时, 切比雪夫不等式却可以帮我们得到一些有用的概率估计; 看几个例子.

例 2. 已知某校某门课程每年平均有 70 人选修.

(1) 估计下次该课程至少有 80 人选修的概率;

(2) 如果已知该课程各年选修人数的方差是 8, 再估计上述概率.

解:以 X 表示选课人数, 则 X 是非负整数值随机变量.

(1) 由于仅知道平均人数 $EX = 70$, 所以在 (2) 式中取 $r = 1$, 得到

$$P(X \geq 80) \leq \frac{EX}{80} = \frac{7}{8}.$$

这是一个很可观的概率, 表明下一次选修人数超过 80 人的概率小于 $\frac{7}{8}$, 但到底小到何种程度, 却无从估计.

(2) 由于除了均值之外, 现在还知道了方差 $DX = 8$, 所以在使用切比雪夫不等式时就可以采用 (3) 式:

$$P(X \geq 80) = P(X - 70 \geq 10) \leq P(|X - 70| \geq 10) \leq \frac{DX}{100} = 0.08.$$

这样一来, 我们就心中有数了, 即下一次选修人数超过 80 人的概率不大, 不会超过 8%. 这就是切比雪夫不等式带来的一个好处, 它在分布未知的情况下, 仅凭期望和方差, 有时也可以给出概率上界的一个较好的估计.

人们更多地是以切比雪夫不等式作为工具, 讨论随机变量的性质以及随机变量序列的依概率收敛性, 包括弱大数律. 我们先来看运用切比雪夫不等式讨论随机变量性质的一个例子.

例 3. 方差为 0 的随机变量是退化的随机变量, 即有 $DX = 0$, 则对某个常数 c, 有 $P(X = c) = 1$.

证明: 由于 X 的方差存在, 所以它的数学期望存在, 记 $c = EX$, 则由切比雪夫不等式, 对任何 $\varepsilon > 0$, 都有

$$P(|X - c| > \varepsilon) = P(|X - EX| > \varepsilon) \leq \frac{DX}{\varepsilon^2} = 0,$$

这表明, 对任何 $\varepsilon > 0$, 都有

$$P(|X - c| \leq \varepsilon) = 1.$$

在上式中令 $\varepsilon \to 0$, 由概率的上连续性即得 $P(|X - c| = 0) = 1$, 即 $P(X = c) = 1$.

下面是运用切比雪夫不等式估计概率上界的有一个例子.

例 4. 若 X 是非负整数值随机变量, 则有

$$1-EX \leq P(X=0) \leq \frac{DX}{(EX)^2}.$$

证明: 由于 X 是非负整数值随机变量, 故有 $EX>0$ (如果 $EX=0$, 则易证 X 是退化于 0 的随机变量), 从而由切比雪夫不等式知

$$P(X=0)=P(X-EX=-EX) \leq P(|X-EX| \geq EX) \leq \frac{DX}{(EX)^2}.$$

此即所证不等式的右半部. 用这个不等式估计概率的方法称为二阶矩方法, 尽管很粗糙, 但在 $DX<(EX)^2$ 的情况下, 对证明某些极限定理还是相当有用的. 另一方面, 若在 (2) 式中令 $r=1$, 则可得

$$P(X \geq 1) \leq EX,$$

故知

$$P(X=0)=1-P(X \geq 1) \geq 1-EX.$$

此即所证不等式的左半部. 用这个不等式估计概率的方法称为一阶矩方法.

下面我们来讨论随机变量序列的弱大数定律, 在这些讨论中, 切比雪夫不等式是一个重要工具.

定理 1. 设 $\{X_n,\ n \in \mathbb{N}\}$ 为随机变量序列, $S_n=\sum\limits_{k=1}^{n} X_k$. 如果存在实数序列 $\{a_n,\ n \in \mathbb{N}\}$ 和正数序列 $\{b_n,\ n \in \mathbb{N}\}$, 使得

$$\frac{S_n-a_n}{b_n} \xrightarrow{p} 0, \tag{4}$$

亦即

$$\lim_{n \to \infty} P\left(\left|\frac{S_n-a_n}{b_n}\right| \geq \varepsilon\right)=0, \quad \forall\, \varepsilon>0, \tag{5}$$

就说 $\{X_n,\ n \in \mathbb{N}\}$ 服从弱大数律. 其中 $\{a_n,\ n \in \mathbb{N}\}$ 称为中心化数列, $\{b_n,\ n \in \mathbb{N}\}$ 称为正则化数列.

研究弱大数律, 就是对随机变量序列 $\{X_n,\ n \in \mathbb{N}\}$ 寻找存在中心化数列 $\{a_n,\ n \in \mathbb{N}\}$ 和正则化数列 $\{b_n,\ n \in \mathbb{N}\}$, 使得 (4) 式成立的条件. 如果 $E|X_n|<\infty,\ n \in \mathbb{N}$, 那么一般就会取 $a_n=ES_n,\ b_n=n,\ n \in \mathbb{N}$, 讨论使得

$$\frac{S_n-ES_n}{n} \xrightarrow{p} 0 \tag{6}$$

成立的条件.

下面是一些弱大数律的例子.

例 5. (马尔可夫弱大数律) 若对随机变量序列 $\{X_n,\ n\in\mathbb{N}\}$, 有

$$\lim_{n\to\infty}\frac{DS_n}{n^2}=0, \tag{7}$$

则有如 (6) 式的弱大数律成立.

证明: 在切比雪夫不等式中令 $g(x)=x^2$, 可知, 对任何 $\varepsilon>0$, 当 $n\to\infty$ 时, 都有

$$\begin{aligned}P\left(\left|\frac{S_n-ES_n}{n}\right|\geq\varepsilon\right)&=P\left(|S_n-ES_n|\geq n\varepsilon\right)\\&\leq\frac{E(S_n-ES_n)^2}{n^2\varepsilon^2}=\frac{1}{\varepsilon^2}\frac{DS_n}{n^2}\to 0,\end{aligned}$$

所以有如 (6) 式的弱大数律成立.

在马尔可夫弱大数律中没有对序列 $\{X_n,\ n\in\mathbb{N}\}$ 中的随机变量之间的相互关系作任何假定, 所以是一个较为广泛的结论.

例 6. (切比雪夫弱大数律)如果序列 $\{X_n,\ n\in\mathbb{N}\}$ 中的随机变量两两不相关且存在常数 $C>0$, 使得 $DX_n\leq C$, $\forall\, n\in\mathbb{N}$, 那么就有如 (6) 式的弱大数律成立.

证明: 由于 $\{X_n,\ n\in\mathbb{N}\}$ 中的随机变量两两不相关, 所以

$$DS_n=\sum_{k=1}^{n}DX_k\leq nC,$$

即有条件 (7) 成立. 故由 Markov 弱大数律得知切比雪夫弱大数律成立.

例 7. (贝努里弱大数律)如果在独立重复随机试验中, 每次成功的概率是 $p\,(0<p<1)$. 以 Z_n 表示 n 次独立重复试验中的成功次数, 则有

$$\frac{Z_n}{n}\xrightarrow{p}p. \tag{8}$$

证明: 可写 $Z_n=\sum\limits_{k=1}^{n}X_k:=S_n$, 其中 $\{X_k\}$ 是一列相互独立的同分布的随机变量, 有 $P(X_k=1)=p,\ P(X_k=0)=1-p$. 从而 $EX_k=p,\ DX_k=pq\leq 1$. 所以由切比雪夫弱大数律得知有如 (6) 式的弱大数律成立, 亦即有 (8) 成立.