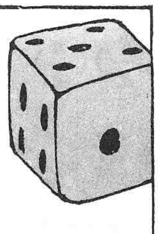

機率專題

謝比雪夫不等式及簡生不等式

許世雄

在機率論中比較常見的不等式有謝比雪夫不等式(Chebyshev's inequality)及簡生不等式(Jensen's inequality)我們在這裏做個簡單地介紹。

考慮隨機變數X,其機率函數(或密度函數)為$p(x)$。則對任意$t>0$,

$$P(|X|\ge t)=\sum_{|x|\ge t} p(x) \quad \left(\text{或} \int_{|x|\ge t} p(x)\,dx\right)$$

$$\le \frac{1}{t}\sum_{|x|\ge t}|x|\,p(x)$$

$$\le \frac{1}{t}\sum_{x}|x|\,p(x)=\frac{E|X|}{t}。$$

即

(1) $$P(|X|\ge t)\le \frac{1}{t}E|X|,$$

若ϕ為一對稱實函數,$\phi(x)=\phi(-x)$;$\phi(x)$在$x\ge 0$上是遞增正函數,則對$t>0$

$$P(|X|\ge t)=P(\phi(X)\ge\phi(t))$$

$$\le \frac{1}{\phi(t)}E(\phi(X)),$$

即

(2) $$P(|X|\ge t)\le \frac{1}{\phi(t)}E(\phi(X))。$$

通常我們稱(2)式為**謝比雪夫不等式**。

特別地,若$\phi(x)=x^2$,則由(2)式得

(3) $$P(|X|\ge t)\le \frac{1}{t^2}E(X^2)。$$

若X之平均為μ_X,標準差為σ_X,則由(3)式可得

$$P(|X-\mu_X|\ge t\sigma_X)$$

$$\le \frac{1}{t^2\sigma_X^2}E(X-\mu_X)^2=\frac{1}{t^2},$$

即得一般常見的謝比雪夫不得式:

(4) $$P(|X-\mu_X|\ge t\,\sigma_X)\le \frac{1}{t^2},$$

(4)式告訴我們:觀察值X離開平均μ_x 超過t個標準差σ_X的機率不超過$\frac{1}{t^2}$。因此若某班某次考試成績平均是60分,標準差是10分,則90分以上的人數絕對不超過全班人數的$\frac{1}{9}$。

顯然,(4)式可改寫為

(5) $$P(|X-\mu_X|\ge t)\le \frac{\sigma_X^2}{t^2}。$$

現在考慮一系列的獨立且具有相同分布的隨機變數$\{X_n,\ n\ge 1\}$。設$E(X_n)=\mu$,$Var(X_n)=\sigma^2<\infty$。令

$\bar{X}_n=\frac{1}{n}\left(\sum_{i=1}^{n}X_i\right)$,則$E(\bar{X}_n)=\mu$,

$Var(\bar{X}_n)=\frac{\sigma^2}{n}$。因此,

$$P(|\bar{X}_n-\mu|\geq\varepsilon)\leq\frac{Var(X_n)}{\varepsilon^2}$$

$$=\frac{\sigma^2}{n\varepsilon^2}\to 0\text{，當 }n\to\infty$$

或

(6) $$\lim_{n\to\infty}P(|\bar{X}_n-\mu|\geq\varepsilon)=0,$$

$\forall\varepsilon>0$。

(6)式即是所謂的**弱大數法則**（weak law of large numbers）。如果 f 是均匀連續且有界的函數，$|f(x)|\leq M<\infty$，$\forall\varepsilon>0$，$\exists\delta=\delta(\varepsilon)$，使得

$|x-y|<\delta\Rightarrow|f(x)-f(y)|<\varepsilon$。因此，若 $p_n(x)$ 表 $\bar{X}_n$ 的機率函數（或密度函數），則

$$|Ef(\bar{X}_n)-f(\mu)|$$
$$=|\sum_x f(\lambda)p_n(x)-f(\mu)|$$
$$\leq\sum_x|f(x)-f(\mu)|p_n(x)$$
$$=\sum_{|x-\mu|<\delta}|f(x)-f(\mu)|p_n(x)$$
$$+\sum_{|x-\mu|\geq\delta}|f(x)-f(\mu)|p_n(x)$$
$$\leq\varepsilon+2M\cdot P(|X_n-\mu|\geq\delta)$$
$$\leq\varepsilon+2M(\frac{\sigma^2}{n\delta^2})$$
$$\leq 2\varepsilon\text{，若 }n\geq\frac{2M\sigma^2}{\varepsilon\delta^2}\text{。}$$

特別地，當 $\{X_n\}$ 表獨立柏努利試驗（Bernoulli trials）時，$P(X_n=1)=\mu$，$P(X_n=0)=1-\mu$，$0\leq\mu\leq1$，μ 表成功之機率，則 $E(X_n)=\mu$，$Var(X_n)=\mu(1-\mu)\leq\frac{1}{4}$，$Ef(\bar{X}_n)=\sum_x f(x)p_n(x)$

$=\sum_{k=0}^{n}f(\frac{k}{n})\binom{n}{k}\mu^k(1-\mu)^{n-k}$。所以

當 $n\geq\frac{M}{2\varepsilon\delta^2}\geq\frac{2M\sigma^2}{\varepsilon\delta^2}$ 時

(7) $$\sup_{0\leq\mu\leq1}|Ef(\bar{X}_n)-f(\mu)|$$

$$=\sup_{0\leq\mu\leq1}|f(\mu)-\sum_{k=0}^{n}f(\frac{k}{n})\binom{n}{k}\cdot\mu^k(1-\mu)^{n-k}|\leq2\varepsilon\text{。}$$

注意：$\sum_{k=0}^{n}f(\frac{k}{n})\binom{n}{k}\mu^k(1-\mu)^{n-k}$ 是 μ 之多項式，若 f 是定義在〔0，1〕上之連續函數（因此是均勻連續且有界的函數）。(7)式告訴我們：任何定義在〔0，1〕上的連續函數都可用多項式（polynomial）均勻地逼近。此即有名的Weierstrass 定理。以上的證明是S. Berstein給的。

底下我們談簡生不等式。考慮凸函數 $f(x)$。由凸函數的性質我們知道：對任意 x，y，λ，$0\leq\lambda\leq1$，$f(x)$ 滿足 $f(\lambda x+(1-\lambda)y)\leq\lambda f(x)+(1-\lambda)f(y)$。利用歸納法可證：對任意 $x_1\cdots$，x_n，λ_1，…，λ_n，$\lambda_i\geq0$，$1\leq i\leq n$，$\sum_{i=1}^{n}\lambda_i=1$，$f(x)$ 滿足

(8) $$f(\sum_{i=1}^{n}\lambda_i x_i)\leq\sum_{i=1}^{n}\lambda_i f(x_i),$$

如果 X 表隨機變數其機率函數 $p(x)$ 滿足 $p(x_i)=\lambda_i$，$1\leq i\leq n$，則(8)式說：$f(E(X))\leq Ef(X)$。由於 $f(X)$ 是凸函數，基本上 $f(X)$ 是連續的。由連續的性質，前面所說的不等式對任意隨機變數 X 都成立（只要 $E(X)$，$Ef(X)$ 有意思）。總之，若 $f(X)$ 是凸函數，X 是隨機變數，則有

(9) $$f(E(X))\leq E(f(X)),$$

(9)式就是**簡生不等式**。

很多常見的不等式都是簡生不等式的結果。例如：若 X 的機率函數 $p(x)$ 滿足 $p(a_i)=\frac{1}{n}$，$a_i>0$，$1\leq i\leq n$。取 $f(x)=x^p$，$x>0$，$p\geq1$，由(9)式得

(10) $$(\frac{a_1+a_2+\cdots+a_n}{n})^p\leq\frac{a_1{}^p+a_2{}^p+\cdots+a_n{}^p}{n}\text{。}$$

若 X 的機率函數 $p(x)$ 滿足 $p(\log a_i)=\lambda_i$，

$a_i>0$，$1\le i\le n$；取 $f(x)=e^x$，$-\infty<x<\infty$，由(9)式得

(11) $$\prod_{i=1}^{n} a_i^{\lambda_i}\le \sum_{i=1}^{n}\lambda_i a_i 。$$

特別地，當 $\lambda_i=\dfrac{1}{n}$，$1\le i\le n$ 時，得

(12) $$\left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}}\le \frac{1}{n}\left(\sum_{i=1}^{n} a_i\right)。$$

此即所謂的"算術平均數大於幾何平均數"的不等式。

我們稱 $E(|X|^p)$，$0<p<\infty$，爲隨機變數 X 的 **p 次絕對動差**。很多有關動差的不等式都是簡生不等式的結果。首先，在(9)式中以 $|X|$ 代替 X 並且取 $f(x)=x^p$，$x>0$，$p\ge 1$，則得 $(E|X|)^p\le E(|X|^p)$

或

(13) $$E|X|\le (E(|X|^p))^{\frac{1}{p}},\ p\ge 1。$$

若 $1\le r\le r'<\infty$，則在(13)式中以 $|X|^r$ 代替 $|X|$ 並且令 $p=\dfrac{r'}{r}>1$，可得

$$E|X|^r\le (E(|X|^{r'}))^{\frac{r}{r'}}；$$

或

(14) $$(E(|X|^r))^{\frac{1}{r}}\le (E(|X|^{r'}))^{\frac{1}{r'}},$$
$$\infty>r'\ge r\ge 1$$

(14)式稱爲**利亞普諾夫不等式**（Liapounov's inequality）。

現在考慮兩個隨機變數 X 和 Y。在(11)式中，令 $n=2$，$\lambda_1=\dfrac{1}{p}$，$\lambda_2=\dfrac{1}{q}$，$1<p<\infty$，$1<q<\infty$，$\dfrac{1}{p}+\dfrac{1}{q}=1$，可得對任意正數 a_1，a_2，我們有 $a_1^{\frac{1}{p}}a_2^{\frac{1}{q}}\le\dfrac{a_1}{p}+\dfrac{a_2}{q}$。令 $a_1=|X|^p/E(|X|^p)$，$a_2=|Y|^q/E(|Y|^q)$，則得

$$|XY|/(E(|X|^p))^{\frac{1}{p}}(E(|Y|^q))^{\frac{1}{q}}$$
$$\le\frac{|X|^p}{pE(|X|^p)}+\frac{|Y|^q}{qE(|Y|^q)},$$

兩邊取期望值，則有

(15) $$E|XY|\le (E(|X|^p))^{\frac{1}{p}}(E(|Y|^q))^{\frac{1}{q}}。$$

(15)式稱爲**賀德不等式**（Hölder's inequality）。特別地，當 $p=q=2$ 時，由(15)式得

(16) $$E|XY|\le (E(X^2))^{\frac{1}{2}}(E(Y^2))^{\frac{1}{2}}$$

此即大家所熟悉的**哥西—史瓦茲不等式**（Cauchy-Schwarz's inequality）。

現在，令 $p>1$，由(15)式知

$$\begin{aligned}
&E(|X+Y|^p)\\
&=E(|X+Y|^{p-1}|X+Y|)\\
&\le E(|X+Y|^{p-1}|X|)\\
&+E(|X+Y|^{p-1}|Y|)\\
&\le (E(|X+Y|^{(p-1)q}))^{\frac{1}{q}}(E(|X|^p))^{\frac{1}{p}}\\
&+(E(|X+Y|^{(p-1)q}))^{\frac{1}{q}}(E(|Y|^p))^{\frac{1}{p}}\\
&,\frac{1}{p}+\frac{1}{q}=1\\
&=(E(|X+Y|^p))^{\frac{1}{q}}[(E(|X|^p))^{\frac{1}{p}}\\
&+(E(|Y|^p))^{\frac{1}{p}}],
\end{aligned}$$

兩邊除以 $(E(|X+Y|^p))^{\frac{1}{q}}$ 得

(17) $$(E(|X+Y|^p))^{\frac{1}{p}}$$
$$\le (E(|X|^p))^{\frac{1}{p}}+(E(|Y|^p))^{\frac{1}{p}},$$

顯然，若 $p=1$ 時，(17)式仍成立，因此(17)式對 $p\ge 1$ 都成立。此不等式稱爲**明可斯基不等式**（Minkowski's inequality）。若 L_p 表所有滿足 $E|X|^p<\infty$ 的隨機變數（我們把兩隨機變數 X 與 Y 看成一樣，若 $P(X=Y)=1$），則(17)式告訴我們 L_p 是個向量空間，而且在 L_p 中可引進距離 $d(X,Y)=(E(|X-Y|^p))^{\frac{1}{p}}$。事實上，可證 L_p 在距離 d 下是完備的，這些牽涉到泛函數分析（functional analysis）的東西，非本文所要討論，我們就此打住。

（本文作者任教於清華大學應用數學所）