# 概率论:Week 14

## 1 积分

测度空间 $(X, \mathscr{F}, \mu)$ 上可测函数积分的定义是用典型方法经过三个步骤来完成的.

### 1.1 非负简单函数的积分

设 $f$ 是简单函数. 那么存在 $(X, \mathscr{F}, \mu)$ 的有限可测分割 $\{A_i, i=1, \cdots, n\}$ 和实数 $\{a_i, i=1, \cdots, n\}$ 使 $f=\sum_{i=1}^{n} a_i I_{A_i}$. 如上述表达式中有 $a_i \geqslant 0$, 则称 $f$ 是非负的. 显然,简单函数的表达方式不惟一, 但它是不是非负与其表达方式无关. 设 $f$ 是一个非负简单函数而 $f=\sum_{i=1}^{n} a_i I_{A_i}$ 是它的一个表达式. 称

$$\int_X f \, \mathrm{d}\mu \stackrel{\text{def}}{=} \sum_{i=1}^{n} a_i \mu\left(A_i\right)$$

为 $f$ 的积分. 可以证明: 非负简单函数的积分具有一意性.

### 1.2 非负可测函数的积分

对于测度空间 $(X, \mathscr{F}, \mu)$ 上的非负可测函数 $f$, 称

$$\int_X f \, \mathrm{d}\mu \stackrel{\text{def}}{=} \sup \left\{\int_X g \, \mathrm{d}\mu : g \text{ 非负简单且 } g \leqslant f\right\}$$

为它的积分.

### 1.3 一般可测函数的积分

测度空间(X,$\mathscr{F}, \mu$)上的可测函数 $f$ 如果满足

$$\min \left\{\int_x f^{+} \, \mathrm{d}\mu, \int_x f^{-} \, \mathrm{d}\mu\right\}<\infty,$$

则称其积分存在或积分有意义；如果满足

$$\max \left\{\int_x f^{+} \, \mathrm{d}\mu, \int_X f^{-} \, \mathrm{d}\mu\right\}<\infty,$$

则称它是可积的.在上述两种情况下，把

$$\int_X f \mathrm{d}\mu \stackrel{\text{def}}{=} \int_X f^+ \,\mathrm{d}\mu - \int_X f^- \,\mathrm{d}\mu$$

叫做 $f$ 的积分或积分值.

**注记.** $f$ 可积当且仅当$|f|$可积.只有当级数 $\sum f(x_i)a_i$ 绝对收敛时，$f$ 才是可积的.

# 2　数学期望

## 2.1　例子

**例 1** (数学期望定义的要求)**.** 随机变量$X$取值

$$X = (-1)^k \frac{2^k}{k}, \quad k = 1, 2, \cdots$$

对应的概率是$p_k = \dfrac{1}{2^k}$, 根据$p_k \geq 0, \sum p_k = 1$, 因此是概率分布, 而且

$$\sum (-1)^k \frac{2^k}{k} p_k = \sum (-1)^k \frac{1}{k} < \infty$$

但是

$$\sum |x_k| p_k = \sum \frac{1}{k} < \infty$$

**例 2** (混管检测)**.** 在一个人数很多的单位中普查某种疾病, $N$ 个人去验血. 对这些人的血的化验可以用两种办法进行. (1) 每个人的血分别化验, 这时需要化验 $N$ 次; (2) 把 $k$ 个人的血混在一起进行化验. 如果结果是阴性的, 那么对这 $k$ 个人只作一次检验就够了; 如果结果是阳性的, 那么必须对这 $k$ 个人再逐个分别化验, 这时对这 $k$ 个人共需作 $k+1$ 次化验. 假定对所有的人来说, 化验是阳性反应的概率都是 $p$, 而且这些人的反应是独立的.我们来说明在 $p$ 相当小的场合, 采用办法 (2) 能减少化验的次数.

若记 $q = 1 - p$, 则 $k$ 个人的混血呈阳性反应的概率为 $1 - q^k$, 用 (2) 的方法验血时, 每个人的血需要化验的次数 $\xi$ 是随机变量, 其分布列为

| $\xi$ | $\frac{1}{k}$ | $1+\frac{1}{k}$ |
|---|---|---|
| $P$ | $q^k$ | $1-q^k$ |

因此

$$E\xi = \frac{1}{k} \cdot q^k + \left(1 + \frac{1}{k}\right)\left(1 - q^k\right) = 1 - q^k + \frac{1}{k}$$

**例 3** (Cauchy分布)**.** 另一个期望不存在的例子是*Cauchy*分布, 它的*pdf*是

$$f_X(x) = \frac{1}{\pi} \frac{1}{1 + x^2}, \quad -\infty < x < +\infty$$

易证 $\displaystyle\int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = 1$. 但是 $E|X| = +\infty$. 事实上,

$$\mathrm{E}|X| = \int_{-\infty}^{+\infty} \frac{|x|}{\pi} \frac{1}{1 + x^2} \,\mathrm{d}x = \frac{2}{\pi} \int_0^{+\infty} \frac{x}{1 + x^2} \,\mathrm{d}x$$

而对任意正数 $M$,

$$\int_0^M \frac{x}{1+x^2}\,\mathrm{d}x = \left.\frac{\log\left(1+x^2\right)}{2}\right|_0^M = \frac{\log\left(1+M^2\right)}{2}$$

于是,

$$\mathrm{E}|X| = \lim_{M\to+\infty}\frac{2}{\pi}\int_0^M \frac{x}{1+x^2}\,\mathrm{d}x = \frac{1}{\pi}\lim_{M\to+\infty}\log\left(1+M^2\right) = \infty,$$

故 EX 不存在.

**定理** (佚名统计学家公式)**.** 若 $g(x)$是一元博雷尔函数，而 $\eta = g(\xi)$ ,则

$$\int_{-\infty}^{\infty} y\mathrm{d}F_\eta(y) = \int_{-\infty}^{\infty} g(x)\,\mathrm{d}F_\xi(x) \quad (4.1.17)$$

即这两个积分中，若有一个存在，则另一个也存在，而且两者相等.

**例 4** (均匀分布与指数分布的联系)**.** 设随机变量$X$服从区间$(0,1)$上的均匀分布, 即$X$的$pdf$是:

$$f_X(x) = \begin{cases} 1, x\in(0,1) \\ 0, x\notin(0,1) \end{cases}$$

定义新的随机变量 $g(X) = -\log X$, 则

$$\mathrm{E}g(X) = \mathrm{E}(-\log X) = \int_0^1 -\log x\,\mathrm{d}x = x - x\log x|_0^1 = 1$$

事实上, 可知 $Y = -\log X$ 的累积分布函数为 $1-\mathrm{e}^{-y}$, 从而其概率密度函数为 $f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}(1-\mathrm{e}^{-y}) = \mathrm{e}^{y}, 0 < y < \infty$, 显然是指数型概率密度函数当参数 $\lambda = 1$时的特例. 于是根据指数分布也可得 $\mathrm{EY} = 1$.

## 2.2 距离最小化

现在讨论随机变量期望的另一个性质, 此时我们将 E$X$ 看成对 $X$ 取值的一个估计.

以 $(X-b)^2$ 度量随机变量 $X$ 与常数 $b$ 之间的距离, 则 $b$ 越接近 $X$, 这个量越小. 我们可以求出使 $\mathrm{E}(X-b)^2$ 最小的 $b$, $b$ 的这个取值就是 $X$ 的一个很好的估计 (注意并非求使 $(X-b)^2$ 最小的 $b$, 因此这样求得的 $b$ 依赖于 $X$, 对估计 $X$ 毫无帮助).

我们可以利用微积分知识求解 $\mathrm{E}(X-b)^2$ 的最小值问题, 不过这里给出一种更简单的方法. 由于 E$X$ 具有很好的性质, 我们做如下变换:

$$\begin{aligned}
& \mathrm{E}(X-b)^2 \\
= \;& \mathrm{E}(X-\mathrm{E}X+\mathrm{E}X-b)^2 && (\text{添加 } \pm\mathrm{E}X \text{ 两项, 不改变原式的值}) \\
= \;& \mathrm{E}((X-\mathrm{E}X)+(\mathrm{E}X-b))^2 && (\text{对项分组}) \\
= \;& \mathrm{E}(X-\mathrm{E}X)^2+(\mathrm{E}X-b)^2+2\mathrm{E}((X-\mathrm{E}X)(\mathrm{E}X-b)) && (\text{展开平方})
\end{aligned}$$

注意, 我们有

$$\mathrm{E}((X-\mathrm{E}X)(\mathrm{E}X-b)) = (\mathrm{E}X-b)\mathrm{E}(X-\mathrm{E}X) = 0$$

这是因为 $(\mathrm{E}X-b)$ 是常数, 从而可以提至期望运算之外, 此外还有 $\mathrm{E}(X-\mathrm{E}X)=\mathrm{E}X-\mathrm{E}X=0$, 于是, 有

$$\mathrm{E}(X-b)^2=\mathrm{E}(X-\mathrm{E}X)^2+(\mathrm{E}X-b)^2$$

显然, $b$ 不影响式右端第一项的取值; 而其第二项的取值恒大于等于 0 ,并且当 $b=\mathrm{E}X$ 时取值为 0 . 故

$$\min_b \mathrm{E}(X-b)^2=\mathrm{E}(X-\mathrm{E}X)^2$$

关于中位数也有一个类似的结论.

## 2.3 矩和矩母函数

概率分布的矩是一类重要的期望.

**定义.** *对任意整数 $n$, $X$ (或 $F_X(x)$ ) 的 $n$ 阶矩 ( nth moment), 记作 $\mu_n'$,定义为:*

$$\mu_n'=\mathrm{E}X^n$$

*$X$ 的 $n$ 阶中心矩 ( nth central moment), 记作 $\mu_n$, 定义为:*

$$\mu_n=\mathrm{E}(X-\mu)^n$$

*其中 $\mu=\mu_1'=\mathrm{E}X$.*

**例 5.** *高阶矩存在, 则低阶矩存在.*

对于随机变量 $X$ 来说, 除期望 $\mathrm{E}X$ 以外最重要的矩莫过于二次中心矩, 这个矩又常称作方差.

**定义.** *随机变量 $X$ 的二阶中心矩称作 $X$ 的方差 (variance), 记作 $\operatorname{Var}X=\mathrm{E}(X-\mathrm{E}X)^2$. $\operatorname{Var}X$ 的正平方根称作 $X$ 的标准差 (standard deviation).*

方差以其期望为基点度量了随机变量的分散度. 在例 2.2.6 中我们已知: 当 $b=\mathrm{E}X$ 时 $\mathrm{E}(X-b)^2$ 取到最小值, 现在考察该最小值的大小. 事实上根据方差的定义,方差越大意味着 $X$ 的取值浮动性越大; 特别地, 若 $\operatorname{Var}X=\mathrm{E}(X-\mathrm{E}X)^2=0$, 则 $X$等于 $\mathrm{E}X$ 的概率为 1 , 故 $X$ 的取值恒定不变. 标准差的大小也有类似的解释: 标准差越小表明 $X$ 越接近 $\mathrm{E}X$, 标准差越大表明 $X$ 的取值浮动性越大. 由于标准差的单位与随机变量的单位相同, 而方差的单位是其平方, 所以我们通常使用标准差度量 $X$ 以 $\mathrm{E}X$ 为基点的分散度.

设 $X$ 是一随机变量且其方差有限, 则对任意常数 $a,b$, 有:

$$\operatorname{Var}(aX+b)=a^2\operatorname{Var}X$$

事实上, 根据定义, 有

$$\begin{aligned}\operatorname{Var}(aX+b)&=\mathrm{E}((aX+b)-\mathrm{E}(aX+b))^2\\&=\mathrm{E}(aX-a\mathrm{E}X)^2\quad(\mathrm{E}(aX+b)=a\mathrm{E}X+b)\\&=a^2\mathrm{E}(X-\mathrm{E}X)^2\\&=a^2\operatorname{Var}X\end{aligned}$$

有时利用方差的下列替代公式可以简化计算:

$$\mathrm{Var}\,X = \mathrm{E}X^2 - (\mathrm{E}X)^2$$

该式可由下述过程推导得出:

$$\begin{aligned}\mathrm{Var}\,X &= \mathrm{E}(X - \mathrm{E}X)^2 = \mathrm{E}\left[X^2 - 2X\mathrm{E}X + (\mathrm{E}X)^2\right] \\ &= \mathrm{E}X^2 - 2(\mathrm{E}X)^2 + (\mathrm{E}X)^2 \\ &= \mathrm{E}X^2 - (\mathrm{E}X)^2\end{aligned}$$

其中利用了 $\mathrm{E}(X\mathrm{EX}) = (\mathrm{E}X)(\mathrm{EX}) = (\mathrm{EX})^2$ 的事实 (因为 $\mathrm{E}X$ 是常数).

下面我们介绍一类与概率分布有关的新的函数——矩母函数。顾名思义，矩母函数可以用于求矩，不过在大多数情形下直接计算矩比用矩母函数计算更为简单. 事实上，矩母函数的主要用处不在于求矩，而是它能够唯一地确定概率分布——如果使用得当，这个性质将非常有用.

**定义** (矩母函数)**.** *设随机变量 $X$ 的累积分布函数为$F_x$. $X$(或$F_X$) 的矩母函数 (moment generating function, 简记为 mgf), 记作 $M_X(t)$ , 定义为:*

$$M_X(t) = \mathrm{E}e^{\iota X}$$

$M_X(t) = \mathrm{E}e^{\iota X}$ *这里假定当$t$ 在0 的某邻域内时上式中的期望存在，即，存在$h > 0$ 使得对任意$-h < t < h, \mathrm{E}e^{\iota X}$都存在，如果在0的任意邻域内该期望都不存在，则称矩母函数不存在.*

我们可以将 X 的矩母函数更准确地表示为:

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx} f_X(x)\mathrm{d}x, \text{如果 } X$$

或

$$M_X(t) = \sum_x e^x P\left(X = x\right),$$

利用矩母函数计算矩的方法非常简单, 我们可以总结为: $n$阶矩等于对矩母函数求$n$次导数再在0点取值.

**定理.** *设 $X$ 的矩母函数为 $M_X(t)$, 则*

$$\mathrm{E}X^n = M_X^{(n)}(0)$$

*其中,*

$$M_X^{(n)}(0) = \left.\frac{\mathrm{d}^n}{\mathrm{d}t^t} M_X(t)\right|_{t=0}$$

*即, $X$ 的 $n$ 阶矩等于 $M_X(t)$ 在 $t = 0$ 处的 $n$ 阶导数.*

证明. 假设可以将求导运算提至积分符号内 , 则有

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} M_X(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{-\infty}^{+\infty} e^{tx} f_X(x)\mathrm{d}x \\ &= \int_{-\infty}^{+\infty}\left(\frac{\mathrm{d}}{\mathrm{d}t}e^{tx}\right) f_X(x)\mathrm{d}x \\ &= \int_{-\infty}^{+\infty}\left(xe^{tx}\right) f_X(x)\mathrm{d}x \\ &= \mathrm{EX}^{tX}\end{aligned}$$

故

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}M_X(t)\right|_{t=0} = \left.\mathrm{E}X\mathrm{e}^{tX}\right|_{t=0} = \mathrm{E}X$$

同理可证

$$\left.\frac{\mathrm{d}^n}{\mathrm{d}t^n}M_X(t)\right|_{t=0} = \left.\mathrm{E}^n\mathrm{e}^{tX}\right|_{t=0} = \mathrm{E}X^n$$

□

**例 6.** *求二项分布的三阶矩.*

**例 7.** *二项分布. 参数为 $(n, p)$ 的二项概率pmf, 故*

$$M_X(t) = \sum_{x=0}^{n} \mathrm{e}^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} \left(p\mathrm{e}^t\right)^x (1-p)^{n-x}$$

*根据二项式定理:*

$$\sum_{x=0}^{n} \binom{n}{x} u^x v^{n-x} = (u+v)^n$$

*令其中 $u = p\mathrm{e}^t, v = 1-p$, 则有*

$$M_X(t) = \left[p\mathrm{e}^t + (1-p)\right]^n$$

我们不加证明的指出: 当随机变量的支集有界时, 矩能唯一确定概率分布.

**定理.** *设 $F_X(x), F_Y(y)$ 是两个累积分布函数, 且其对应的全部矩都存在.*

1. *如果 $X$ 和 $Y$ 的支集有界, 则对任意 $u$ 有 $F_X(u) = F_Y(u)$ 当且仅当对任意整数 $r = 0, 1, 2, \cdots$ 都有 $\mathrm{EX}^r = \mathrm{EY}^r$;*

2. *如果 $X$ 和 $Y$ 的矩母函数都存在, 并且对 0 的某邻域内的任意 $t$, 都有 $M_X(t) = M_Y(t)$, 则对任意 $u$ 有 $F_X(u) = F_Y(u)$.*

由于$M_X(t)$ 的定义式

$$M_X(t) = \int_{-\infty}^{+\infty} \mathrm{e}^{tx} f_X(x)\mathrm{d}x$$

就定义了一个 Laplace 变换 ($M_X(t)$ 是 $f_X(x)$ 的一个 Laplace 变换. Laplace 变换的一个重要性质是唯一性, 即: 若对任意 $|t| < h(h > 0)$ 成立, 则对于给定的 $M_X(t)$ 只存在唯一一个函数 $f_X(x)$ 满足式上式. 基于这一事实, 唯一性就十分直观.

**定义.** *对任意常数 $a, b$, 随机变量 $aX + b$ 的矩母函数为*

$$M_{aX+b}(t) = \mathrm{e}^{bt} M_X(at)$$

证明. 由定义, 有

$$\begin{aligned} M_{aX+b}(t) &= \mathrm{E}\left(\mathrm{e}^{(aX+b)t}\right) \\ &= \mathrm{E}\left(\mathrm{e}^{(aX)t}\mathrm{e}^{bt}\right) && \text{(根据指数的性质)} \\ &= \mathrm{e}^{bt}\mathrm{E}\left(\mathrm{e}^{(at)X}\right) && \text{(因为 } \mathrm{e}^{bt} \text{ 是常数)} \\ &= \mathrm{e}^{bt} M_X(at) && \text{(根据矩母函数的定义)} \end{aligned}$$

定理得证. □

### 2.4　多维场合

**定义.** 随机向量 $(\xi_1, \xi_2, \cdots, \xi_n)$ 的数学期望为 $(E\xi_1,\ E\xi_2, \cdots, E\xi_n)$, 其中

$$E\xi_i = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i \,\mathrm{d}F\left(x_1, \cdots, x_n\right) = \int_{-\infty}^{\infty} x_i \,\mathrm{d}F_i\left(x_i\right)$$

这里 $F_i\left(x_i\right)$ 是 $\xi_i$ 的分布函数 $(i = 1, 2, \cdots, n)$.

**例 8** (超级和分布的期望)**.** 超几何分布

$$p_m = \frac{\begin{pmatrix} M \\ m \end{pmatrix}\begin{pmatrix} N-M \\ n-m \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}, \quad m = 0, 1, \cdots, n$$

的数学期望. 当然可以用 $\sum\limits_{k=0}^{n} k p_k$ 直接求出, 但也可用下面方法来计算. 设想一个相应的不放回抽样, 令

$$\xi_i = \begin{cases} 1, & \text{第 } i \text{ 次抽得次品} \\ 0, & \text{第 } i \text{ 次抽得好品} \end{cases}$$

则 $P\left\{\xi_i = 1\right\} = \dfrac{M}{N}$, 因此 $E\xi_i = \dfrac{M}{N}$, 而 $\xi = \xi_1 + \cdots + \xi_n$ 表示 $n$ 次不放回抽样中抽出的次品数, 它服从上述超几何分布, 利用可加性得到

$$E\xi = E\xi_1 + \cdots + E\xi_n = \frac{nM}{N}$$

## 3　不等式和恒等式

### 3.1　概率不等式

最著名、或许也是最有用的一个概率不等式是 Chebychev (切比雪夫) 不等式. 尽管其应用非常广泛, 证明却相当容易.

**定理** (Chebyshev)**.** 设 $X$ 为随机变量, $g(x)$ 为非负函数, 则对任意 $r > 0$ 有

$$P(g(X) \geqslant r) \leqslant \frac{\mathrm{E}g(X)}{r}$$

证明.

$$\begin{aligned}
\mathrm{E}g(X) &= \int_{-\infty}^{+\infty} g(x) f_X(x) \mathrm{d}x \\
&\geqslant \int_{\langle x_x:(x) \geqslant r|} g(x) f_X(x) \mathrm{d}x \quad \text{(因为 } g \text{ 是非负函数)} \\
&\geqslant r \int_{(x_2 g(x) \geqslant r\}} f_X(x) \mathrm{d}x \\
&= r P(g(X) \geqslant r) \quad \text{(根据概率密度函数的定义)}
\end{aligned}$$

上式整理后即得 Chebychev 不等式. □

**注记** (与书上的联系)**.** 令 $g(x)=(x-\mu)^2/\sigma^2$, 其中 $\mu=\mathrm{E}X, \sigma^2=\operatorname{Var}X$. 为方便计算将 $r$ 写作 $t^2$, 则

$$P\left(\frac{(X-\mu)^2}{\sigma^2}\geqslant t^2\right)\leqslant\frac{1}{t^2}\mathrm{E}\frac{(X-\mu)^2}{\sigma^2}=\frac{1}{t^2}$$

由上式易知

$$P(|X-\mu|\geqslant t\sigma)\leqslant\frac{1}{t^2}$$

## 3.2 递推关系的恒等式

若 $X$ 服从参数为 $\lambda$ 的 Poisson 分布, 则

$$P(X=x+1)=\frac{\lambda}{x+1}P(X=x)$$

这就使我们能够从 $P(X=0)=\mathrm{e}^{-\lambda}$ 出发递归地计算 Poisson 概率.

## 3.3 分部积分的概率恒等式

**定理** (Stein引理)**.** 设 $X\sim\mathrm{n}\left(\theta,\sigma^2\right)$, $g$ 是满足 $\mathrm{E}\left|g'(X)\right|<+\infty$ 的可导函数, 则

$$\mathrm{E}[g(X)(X-\theta)]=\sigma^2\mathrm{E}g'(X)$$

证明. 上式左端为

$$\mathrm{E}[g(X)(X-\theta)]=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}g(x)(x-\theta)\mathrm{e}^{-(x-\theta)^2/\left(2\sigma^2\right)}\mathrm{d}x$$

对 $u=g(x)$ 和 $\mathrm{d}v=(x-\theta)\mathrm{e}^{-(x-\theta)^2/2\sigma^2}\,\mathrm{d}x$ 运用分部积分法, 可得

$$\mathrm{E}[g(X)(X-\theta)]=\frac{1}{\sqrt{2\pi}\sigma}\left[-\sigma^2g(x)\mathrm{e}^{-(x-\theta)^2/\left(2\sigma^2\right)}\Big|_{-\infty}^{+\infty}+\sigma^2\int_{-\infty}^{+\infty}g'(x)\mathrm{e}^{-(x-\theta)^2/\left(2\sigma^2\right)}\mathrm{d}x\right]$$

由 $g'$ 所满足的条件易知, 上式右端第一项为 $0$ , 而余下的项恰好是 $\sigma^2\mathrm{E}g'(X)$. □

**例 9** (应用: 求高阶矩)**.** *Stein* 引理可以很大程度上简化高阶矩的计算. 例如, 如果 $X\sim\mathrm{N}\left(\theta,\sigma^2\right)$, 则有

$$\begin{aligned}
\mathrm{E}X^3&=\mathrm{E}X^2(X-\theta+\theta)\\
&=\mathrm{E}X^2(X-\theta)+\theta\mathrm{E}X^2\\
&=2\sigma^2\mathrm{E}X+\theta\mathrm{E}X^2\quad\left(g(x)=x^2,g'(x)=2x\right)\\
&=2\sigma^2\theta+\theta\left(\sigma^2+\theta^2\right)\\
&=3\theta\sigma^2+\theta^3
\end{aligned}$$

## 3.4 Cauchy-Schwarz不等式

**定理** (柯西-施瓦茨 (Cauchy-Schwarz) 不等式)**.** 对任意随机变量 $\xi$ 与 $\eta$ 都有

$$|E\xi\eta|^2\leqslant E\xi^2\cdot E\eta^2$$

等式成立当且仅当

$$P\{\eta=t_0\xi\}=1$$

这里 $t_0$ 是某一个常数.

证明. 对任意实数 $t$, 定义 $u(t)=E(t\xi-\eta)^2=t^2E\xi^2-2tE\xi\eta+E\eta^2$, 显然对一切 $t, u(t)\geqslant 0$, 因此二次方程 $u(t)=0$ 或者没有实根或者有一个重根. 所以判别式

$$[E\xi\eta]^2-E\xi^2\cdot E\eta^2\leqslant 0$$

此外, 方程 $u(t)=0$ 有一个重根 $t_0$ 存在的充要条件是

$$[E\xi\eta]^2-E\xi^2E\eta^2=0$$

这时 $E\left(t_0\xi-\eta\right)^2=0$, 因此

$$D\left(t_0\xi-\eta\right)=0,\quad E\left(t_0\xi-\eta\right)=0$$

从而

$$P\left\{t_0\xi-\eta=0\right\}=1$$

定理证毕. □

由柯西-施瓦茨 (Cauchy-Schwarz) 不等式立即可以推出, 若两随机变量的方差存在, 则它们的协方差也存在.

# 4 相关系数

**定义** (协方差)**.** *称*

$$\sigma_{ij}=\operatorname{cov}\left(\xi_i,\xi_j\right)=E\left[\left(\xi_i-E\xi_i\right)\left(\xi_j-E\xi_j\right)\right]$$
$$i,j=1,2,\cdots,n$$

*为 $\xi_i$ 与 $\xi_j$ 的协方差 (covariance). 不难验算*

$$\operatorname{cov}\left(\xi_i,\xi_j\right)=E\xi_i\xi_j-E\xi_i\cdot E\xi_j$$
$$D\left(\sum_{i=1}^n\xi_i\right)=\sum_{i=1}^nD\xi_i+2\sum_{1\in i<j\in n}\operatorname{cov}\left(\xi_i,\xi_j\right)$$

*特别地*

$$D\left(\xi_i\pm\xi_j\right)=D\xi_i+D\xi_j\pm 2\operatorname{cov}\left(\xi_i,\xi_j\right)$$

*方差是协方差的特例, 显然 $\sigma_{ii}=D\xi_i$. 矩阵*

$$\boldsymbol{\Sigma}=\begin{pmatrix}\sigma_{11}&\sigma_{12}&\cdots&\sigma_{1n}\\\sigma_{21}&\sigma_{22}&\cdots&\sigma_{2n}\\\vdots&\vdots&&\vdots\\\sigma_{n1}&\sigma_{n2}&\cdots&\sigma_{nn}\end{pmatrix}$$

*称为 $\xi$ 的协方差矩阵, 简记作 $D\xi$, 显然这是一个对称矩阵. 此外, 对任何实数 $t_j(j=1,2,\cdots,n)$ 有*

$$\sum_{j,k}\sigma_{jk}t_jt_k=E\left[\sum_{j=1}^nt_j\left(\xi_j-E\xi_j\right)\right]^2\geqslant 0$$

*因此 $\boldsymbol{\Sigma}$ 是一个非负定矩阵, 所以若以 $\det\boldsymbol{\Sigma}$ 记 $\boldsymbol{\Sigma}$ 的行列式, 则有*

$$\det\boldsymbol{\Sigma}\geqslant 0$$

更常用的是如下“标准化”了的协方差.

**定义** (相关系数)**.** *称*

$$\rho_{ij} = \frac{\operatorname{cov}(\xi_i, \xi_j)}{\sqrt{D\xi_i}\sqrt{D\xi_j}}$$

*为 $\xi_i$ 与 $\xi_j$ 的相关系数 (correlation coefficient), 这里当然要求 $D\xi_i$与 $D\xi_j$ 不为零.*

$$\operatorname{cov}(a\xi + b, c\eta + d) = ac\operatorname{cov}(\xi, \eta)$$

因此当 $ac > 0$ 时,

$$\rho_{a\xi+b, c\pi+d} = \frac{ac\operatorname{cov}(\xi, \eta)}{|ac|\sqrt{D\xi}\sqrt{D\eta}} = \rho_{\xi\eta}$$

当 $ac < 0$ 时, $\rho_{a\xi+b,c\eta+d} = -\rho_{\xi\eta}$, 但总有 $\rho^2_{a\xi+b,c\eta+d} = \rho^2_{\xi\eta}$.

## 4.1 应用: 抽样理论

袋中有 $N$ 张卡片, 各记以数字 $Y_1, Y_2, \cdots, Y_N$, 不放回地从中抽出 $n$ 张, 求其和的数学期望与方差.

取一张时, 其数字的均值及方差分别为

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

及

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2$$

若以 $\eta_n$ 记 $n$ 张卡片的数字之和, 以 $\xi_i, i = 1, 2\cdots, n$ 记第 $i$ 次抽得的卡片上的数字, 则

$$\eta_n = \xi_1 + \xi_2 + \cdots + \xi_n$$

$$P\{\xi_i = Y_l\} = \frac{1}{N}, \quad l = 1, 2, \cdots, N, \quad i = 1, 2, \cdots, n$$

因此

$$E\xi_i = \bar{Y}, \quad D\xi_i = \sigma^2$$

所以

$$E\eta_n = E\xi_1 + E\xi_2 + \cdots + E\xi_n = n\bar{Y}$$

$$\begin{aligned} D\eta_n &= \sum_{i=1}^{n} D\xi_1 + 2\sum_{1\leqslant i<j\leqslant n}\operatorname{cov}(\xi_i, \xi_j) \\ &= n\sigma^2 + n(n-1)\operatorname{cov}(\xi_1, \xi_2) \end{aligned}$$

这里用到 $\xi_i$ 之间的对称性, 也即抽签与顺序无关. 在 (4.2.33) 中令 $n = N$, 这时 $\eta_N = Y_1 + Y_2 + \cdots + Y_N$ 是一个常数,因此 $D\eta_N = 0$,于是

$$N\sigma^2 + N(N-1)\operatorname{cov}(\xi_1, \xi_2) = 0$$

所以

$$\operatorname{cov}(\xi_1,\xi_2) = -\frac{\sigma^2}{N-1}$$

最后得到

$$\begin{aligned} D\eta_n &= n\sigma^2 - \frac{n(n-1)\sigma^2}{N-1} \\ &= \frac{n(N-n)}{N-1}\sigma^2 \end{aligned}$$

**注记.** 与有放回抽取的方差 $n\sigma^2$ 相比, 多出了一个因子 $\frac{N-n}{N-1}$, 称为有限总体修正因子. 当 $n=1$ 时,它等于 *1*; 而当 $n=N$ 时, 它取值为 *0*.在不放回场合, 添加抽取的信息量较大, 即方差小, 这与直观完全符合.

# 5 变换方法

## 5.1 Z transform(母函数)

若随机变量 $\xi$ 取非负整数值, 且相应的分布列为

| $\xi$ | 0 | 1 | 2 | $\cdots$ |
|---|---|---|---|---|
| $P$ | $p_0$ | $p_1$ | $p_2$ | $\cdots$ |

则称

$$P(s) = \sum_{k=0}^{\infty} p_k s^k$$

为 $\xi$ 的母函数 (generating function). 由佚名统计学家公式知

$$P(s) = Es^{\xi}$$

因为母函数由分布列完全决定, 因此亦称它为该概率分布的母函数. 由于

$$\sum_{k=0}^{\infty} p_k = 1$$

由幂级数的收敛性知道 $P(s)$ 至少在 $|s| \leqslant 1$ 一致收玫且绝对收玫.因此母函数对任何整值随机变量都存在. 对于任一数列 $\{a_n\}$ 也可定义 $\sum_{n=0}^{\infty} a_n s^n$ 为其母函数, 但我们以后只讨论概率分布对应的母函数.

**注记.** 事实上, 有更广泛的情况:

1. *Bilateral Z-transform(双边Z变换)*

$$X(z) = \mathcal{Z}\{x[n]\} = \sum_{n=-\infty}^{\infty} x[n] z^{-n}$$

2. *Unilateral Z-transform(单边Z变换)*

$$X(z) = \mathcal{Z}\{x[n]\} = \sum_{n=0}^{\infty} x[n] z^{-n}.$$

我们不加证明的指出: 母函数与随机变量之间存在一一对应关系,

它的一个应用是求随机变量的期望, 这一点在随机过程中有应用. 当数学期望 $\sum_{k=1}^{\infty} kp_k$ 存在时,

$$P'(1) = \sum_{k=1}^{\infty} kp_k = E\xi$$

当数学期望 $\sum_{k=1}^{x} kp_k = \infty$ 时, $\lim_{s\to 1} P'(s) = \infty$.

卷积定理:

若随机变量 $\xi$ 与 $\eta$ 相互独立, 它们都是整值随机变量, 概率分布分别为 $\{a_k\}$ 及 $\{b_k\}$, 而相应的母函数为 $A(s)$ 及 $B(s)$, 下面计算随机变量 $\zeta = \xi + \eta$ 的概率分布. 显然 $\zeta$ 也是整值随机变量, 若记 $c_r = P\{\zeta = r\}$, 则

$$c_r = a_0 b_r + a_1 b_{r-1} + \cdots + a_r b_0$$

这就是离散卷积公式. 记

$$C(s) = \sum_{r=0}^{\infty} c_r s^r$$

利用母函数在 $|s| \leqslant 1$ 的一致收玫性及绝对收玫性,

$$\begin{aligned} A(s)B(s) &= \sum_{k=0}^{\infty} a_k s^k \cdot \sum_{l=0}^{\infty} b_l s^l = \sum_{k,l} a_k b_l s^{k+l} \\ &= \sum_{r=0}^{\infty} \left( \sum_{k=0}^{r} a_k b_{r-k} \right) s^r = \sum_{r=0}^{\infty} c_r s^r \end{aligned}$$

因此

$$C(s) = A(s)B(s)$$

即两个独立随机变量之和的母函数是这两个随机变量的母函数的乘积. 这是一个相当重要的性质, 由于母函数具有这个性质, 因此在研究独立随机变量和的问题时,母函数很适用.

## 5.2 随机个随机变量的求和

若 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 是一串相互独立具有相同概率分布的整值随机变量, $P\{\xi_i = j\} = f_j$, 其母函数为

$$F(s) = \sum_{j=0}^{\infty} f_j s^i$$

随机变量 $\nu$ 是取正整数值的, 且 $P\{\nu = n\} = g_n$, 其母函数为

$$G(s) = \sum_{n=1}^{\infty} g_n s^n$$

若 $| \xi_n\}$ 与 $\nu$ 独立, 考虑和 $\eta = \xi_1 + \xi_2 + \cdots + \xi_\nu$, 记

$$P\{\eta = i\} = h_i$$

我们来求 $\eta$ 的母函数

$$H(s)=\sum_{i=0}^{\infty}h_is^i$$

利用全概率公式及 $\{\xi_n\}$ 与 $\nu$ 的独立性

$$\begin{aligned}h_i=P\{\eta=i\}&=\sum_{n=1}^{\infty}P\{\nu=n\}P\{\eta=i\mid\nu=n\}\\&=\sum_{n=1}^{\infty}P\left\{\nu=n\left\{P\left\{\xi_1+\xi_2+\cdots+\xi_n=i\right\}\nu=n\right\}\right.\\&=\sum_{n=1}^{\infty}P\left\{\nu=n\left\{P\left\{\xi_1+\xi_2+\cdots+\xi_n=i\right\}\right.\right.\end{aligned}$$

由于 $\xi_1+\xi_2+\cdots+\xi_n$ 为 $n$ 个相互独立具有相同母函数 $F(s)$ 的随机变量之和, 故其母函数

$$\sum_{i=0}^{\infty}P\left\{\xi_1+\cdots+\xi_n=i\right\}s^i=[F(s)]^n$$

因此

$$\begin{aligned}H(s)=\sum_{i=0}^{\infty}h_is^i&=\sum_{n=1}^{\infty}P\{\nu=n\}\sum_{i=0}^{\infty}P\left\{\xi_1+\cdots+\xi_n=i\right\}s^i\\&=\sum_{n=1}^{\infty}g_n[F(s)]^n=G[F(s)]\end{aligned}$$

可以看到, 随机个相互独立相同分布的随机变量之和的母函数是原来两个母函数的复合. 由于

$$H'(s)=G'[F(s)]\cdot F'(s)$$

因此当 $E\xi_i$ 及 $E\nu$ 存在时, 在上式中令 $s=1$, 得到

$$E\eta=E\nu\cdot E\xi_i$$

## 5.3 特征函数

**定义** (复随机变量)**.** *如果 $\xi$ 与 $\eta$ 都是概率空间 $(\Omega,\mathscr{F},P)$ 上的实值随机变量,则称 $\zeta=\xi+i\eta$ 为复随机变量.*

从定义知道, 对复随机变量的研究本质上是对二维随机向量的研究. 这里举一个例子: 如果二维向量 $(\xi_1,\eta_1)$ 与 $(\xi_2,\eta_2)$ 是独立的,则我们称复随机变量 $\zeta_1=\xi_1+i\eta_1$ 与 $\zeta_2=\xi_2+i\eta_2$ 是独立的. 定义一个复随机变量 $\zeta=\xi+i\eta$ 的数学期望为

$$E\zeta=E\xi+\mathrm{i}E\eta$$

对复随机变量也可以平行于实随机变量建立起一系列结果.例如, 若 $\zeta_1,\zeta_2,\cdots,\zeta_n$ 是相互独立的, 则

$$E\zeta_1\zeta_2\cdots\zeta_n=E\zeta_1E\zeta_2\cdots E\zeta_n$$

**定义.** 若随机变量 $\xi$ 的分布函数为 $F_{\xi}(x)$, 则称

$$f_{\xi}(t)=E\mathrm{e}^{\mathrm{i}t\xi}=\int_{-x}^{x}\mathrm{e}^{\mathrm{i}/x}\,\mathrm{d}F_{\xi}(x)$$

为 $\xi$ 的特征函数 *(characteristic function).*

特征函数是一个实变量的复值函数, 由于 $\left|\mathrm{e}^{\mathrm{i}x}\right|=1$, 所以它对一切实数 $t$ 都有意义. 显然特征函数只与分布函数有关, 因此亦称某一分布函数的特征函数. 对于离散型随机变量,若其分布列为 则

| $\xi$ | $x_1$ | $x_2\cdots$ | $x_n\cdots$ |
|---|---|---|---|
| $P$ | $p_1$ | $p_2\cdots$ | $p_n\cdots$ |

其特征函数为

$$f(t)=\sum_{j=1}^{\infty}p_j\mathrm{e}^{\mathrm{i}x_j}$$

特别地, 对于整值随机变量, 若其母函数为 $P(s)$, 则 $f(t)=P\left(\mathrm{e}^{i}\right)$. 对于连续型随机变量, 若其分布密度函数为 $p(x)$, 则其特征函数为

$$f(t)=\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}tx}p(x)\mathrm{d}x$$

这时, 特征函数是密度函数 $p(x)$ 的傅里叶(Fourier) 变换.

1. $f(0)=1$, 规范性;
2. $|f(t)|\leq 1=f(0)$, 有界性⇒控制收敛;
3. $f(-t)=\overline{f(t)}$, 共轭对称性;
4. 特征函数在$(-\infty,\infty)$上一致连续, 相较于矩母函数的优点是不用考虑收敛半径的问题;
5. 卷积定理:$n$个相互独立随机变量之和的特征函数是特征函数之积;
6. 与数字特征之间的关系, 若$n$阶矩存在, 那么对于$k\leq n$:

   $$f^{(k)}(0)=\mathrm{i}^kE\xi^k$$

7. Taylor展示:

   $$f(t)=1+(\mathrm{i}t)E\xi+\frac{(\mathrm{i}t)^2}{2!}E\xi^2+\cdots+\frac{(\mathrm{i}t)^n}{n!}E\xi^n+o\left(t^n\right)$$

8. 线性组合: 设 $\eta=a\xi+b$, 这里 $a,b$ 为常数, 则

   $$f_{\eta}(t)=\mathrm{e}^{\mathrm{i}bt}f_{\xi}(at)$$

**定理** (逆转公式)**.** 设分布函数 $F(x)$ 的特征函数为 $f(t)$, 又 $x_1,x_2$ 是 $F(x)$ 的连续点, 则

$$F\left(x_2\right)-F\left(x_1\right)=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}x_1}-\mathrm{e}^{-\mathrm{i}ix_2}}{\mathrm{i}t}f(t)\mathrm{d}t$$

**定理** (唯一性)**.** 分布函数由其特征函数唯一决定.

### 5.4 特征函数的应用: 独立随机变量的和用特征函数的积表示

**例 10** (二项分布). 若 $\xi_1$ 服从 $B(m,p)$,$\xi_2$ 服从 $B(n,p)$, 而且 $\xi_1$ 与 $\xi_2$ 独立, 则 $\eta=\xi_1+\xi_2$ 服从 $B(m+n,p)$. 事实上 $f_{\xi_1}(t)=\left(p\mathrm{e}^{\mathrm{i}t}+q\right)^m$, $f_{\xi_2}(t)=\left(p\mathrm{e}^{\mathrm{i}t}+q\right)^n$, 由性质 4 知

$$f_\eta(t)=\left(p\mathrm{e}^{\mathrm{i}t}+q\right)^{m+n}$$

因此由唯一性定理知 $\eta$ 服从 $B(m+n,p)$. 这个事实简记作

$$B(n_1,p)*B(n_2,p)=B(n_1+n_2,p)$$

**例 11** (Poisson分布). 若 $\xi_1$ 服从 $P(\lambda_1)$, $\xi_2$ 服从 $P(\lambda_2)$, 而且 $\xi_1$ 与 $\xi_2$ 独立, 则 $\eta=\xi_1+\xi_2$ 服从 $P(\lambda_1+\lambda_2)$. 事实上

$$f_{\xi_1}(t)=\mathrm{e}^{\lambda_1\left(\mathrm{e}^{it}-1\right)},\quad f_{\xi_2}(t)=\mathrm{e}^{\lambda_2\left(\mathrm{e}^{it}-1\right)}$$
$$f_\eta(t)=\mathrm{e}^{(\lambda_1+\lambda_2)\left(\mathrm{e}^{it}-1\right)}$$

这个事实简记作

$$Poi(\lambda_1)*Poi(\lambda_2)=P(\lambda_1+\lambda_2)$$

**例 12** (Gamma分布). 若 $\xi_1$ 服从 $\Gamma(r_1,\lambda)$, $\xi_2$ 服从 $\Gamma(r_2,\lambda)$, 而且 $\xi_1$ 与 $\xi_2$ 独立, 则 $\eta=\xi_1+\xi_2$ 服从 $\Gamma(r_1+r_2,\lambda)$. 事实上,

$$f_{\xi_1}(t)=\left(1-\frac{\mathrm{i}t}{\lambda}\right)^{-r_1},\quad f_{\xi_2}(t)=\left(1-\frac{\mathrm{i}t}{\lambda}\right)^{-r_2}$$
$$f_\eta(t)=\left(1-\frac{\mathrm{i}t}{\lambda}\right)^{-(r_1+r_2)}$$

这个事实简记作

$$\Gamma(r_1,\lambda)*\Gamma(r_2,\lambda)=\Gamma(r_1+r_2,\lambda)$$

特别地, $\chi_n^2$ 分布即为 $\Gamma\left(\frac{n}{2},\frac{1}{2}\right)$, 也具有再生性:

$$\chi_m^2*\chi_n^2=\chi_{m+n}^2$$

这个性质, 通常称为 $\chi^2$ 分布的可加性, 在数理统计中是最常使用的命题之一.

## 6 条件期望

**定义.** 设 $f$ 为概率空间 $(X,\mathscr{F},P)$ 上积分存在的随机变量. 称 $\mathrm{E}(f\mid\mathscr{G})$ 为 $f$ 关于 $\mathscr{F}$ 的子 $\sigma$ 域 $\mathscr{G}$ 的条件期望, 如果

1. $\mathrm{E}(f\mid\mathscr{G})$ 是 $(X,\mathscr{G},P)$ 上积分存在的可测函数;
2. 对任何 $A\in\mathscr{G}$, 有
$$\int_A \mathrm{E}(f\mid\mathscr{G})\mathrm{d}P=\int_A f\,\mathrm{d}P.$$

$P(A\mid\mathscr{G})\overset{def}{=}\mathrm{E}(I_A\mid\mathscr{G})$ 称为事件 $A\in\mathscr{F}$ 关于子 $\sigma$ 域 $\mathscr{G}$ 的条件概率.如果 $g$ 是 $(X,\mathscr{G},P)$ 到可测空间 $(Y,\mathscr{S})$ 的随机元, 则分别称 $\mathrm{E}(f\mid g)\overset{def}{=}\mathrm{E}(f\mid\sigma(g))$ 和 $P(A\mid g)\overset{def}{=}P(A\mid\sigma(g))$ 为随机变量 $f$ 关于 $g$ 的条件期望和事件 $A\in\mathscr{F}$ 关于 $g$ 的条件概率.

**性质** (Iterative Expectation)**.** 不加证明的给出以下公式, 事实上, 有时是*a.s.*意义下成立, 这里不加区分.

1. $E(A) = E\{E(A \mid B)\}$

2. $E(A \mid C) = E\{E(A \mid B, C) \mid C\}$.

**性质** (EVVE formular)**.** 方差分解公式与条件方差分解公式:

1. $\operatorname{var}(A) = E\{\operatorname{var}(A \mid B)\} + \operatorname{var}\{E(A \mid B)\}$

2. $\operatorname{var}(A \mid C) = E\{\operatorname{var}(A \mid B, C) \mid C\} + \operatorname{var}\{E(A \mid B, C) \mid C\}$

**性质** (Tower Property)**.** 协方差分解公式:

1. $\operatorname{cov}(A_1, A_2) = E\{\operatorname{cov}(A_1, A_2 \mid B)\} + \operatorname{cov}\{E(A_1 \mid B), E(A_2 \mid B)\}$

2. $\operatorname{cov}(A_1, A_2 \mid C) = E\{\operatorname{cov}(A_1, A_2 \mid B, C) \mid C\} + \operatorname{cov}\{E(A_1 \mid B, C), E(A_2 \mid B, C) \mid C\}$