从矩生成函数到特征函数

黄建祺

2022 年 12 月 8 日

1 矩生成函数

随机变量 X 的高阶距可能很难算，因此需要用定义一个所谓的矩生成函数来求各阶距.

Definition 1.1 (矩母生成函数 (Moment Generating Function),MGF)**.** *随机变量 X 的 MGF 可定义为*

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_{x\in\Omega_X} e^{tx} f_X(x), X \text{ 为离散} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)dx, X \text{ 为连续} \end{cases}$$

MGF 的存在性：若对于 t 在 0 的某个领域 $(-\epsilon, \epsilon)$ 上是存在的，则称 $M_X(t)$ 是存在的，若不存在就对于 X 分布来说是不存在的.

因此从 MGF 可以推导到一个各阶距是否是存在的，若对于所有的 t 都存在 MGF，则可推出在所有阶的导数存在.

证明. $e^{tx} = \sum_{k=0}^{\infty} \frac{(tx)^k}{k!}$，$M_X(t)$ 的麦克劳林展开为

$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k)$$

反过来对于某个阶的期望为 ∞，则对于 $t > 0$ 所有的矩母函数都不存在. □

Theorem 1.1. *若 MGF 的 t 在某个领域 $(-\epsilon, \epsilon)$ 内存在，则 $M_X(0) = 1$*

证明. $M_X(t) = \int_{\infty}^{\infty} e^{tx} dF_X(x)$，有 $M_X(0) = \int_{-\infty}^{\infty} dF_X(x) = 1$ □

Theorem 1.2. *若 MGF $M_X(t)$ 对于 t 在 0 的某个邻域内存在,则对所有正整数 $k=1,2,\cdots$ 有*

$$M_X^{(k)}(0)=E(X^k)$$

其中的 k 表示的 k 阶导.

证明.

$$\begin{aligned}M_X^{(k)}(t)&=\frac{d^k}{dt^k}\int_{-\infty}^{\infty}e^{tx}dF_X(x)\\&=\int_{-\infty}^{\infty}\frac{d^k}{dt^k}(e^{tx})dF_X(x)\\&=\int_{\infty}^{\infty}x^ke^{tx}dF_X(x)\end{aligned}$$

令 t=0，则

$$M_X^{(k)}(0)=\int_{-\infty}^{\infty}x^kdF_X(x)=E(X^k)$$

□

因此若矩母函数存在，可以在原点处对其进行求导 ($k=1,2,\cdots$)，从而可以得到 X 的各阶矩. 因此 $M_X(t)$ 也称为矩生成函数. 利用这一性质，我们可以很快的推导出一个分布的任意阶矩.

Example 1.1.

$$\begin{aligned}&M_X^{(1)}(0)=\mu_X\\&M_X^{(2)}(0)=E(X^2)=\sigma_X^2+\mu_X^2\end{aligned}$$

Example 1.2. 令 $X\sim exp(1)$，$x\subset[0,\infty]$ 则

$$M_X(t)=E[e^{tX}]=\int_0^{\infty}e^{tx}e^{-x}dx=\int_0^{\infty}e^{(t-1)x}=\frac{1}{1-t}$$

因此就有 $E[X]=M_X`(0)=1,Var(X)=E[X^2]-E[X]^2=2-1=1$

Theorem 1.3. *若 $Y=a+bX$，其中 a 和 b 为两个常数，并且对于在 0 处的某个小邻域内的所有 t，X 的 $MGFM_X(t)$ 存在，则对于 0 在某个小邻域内的所有 t，Y 的 MGF 也存在，且为*

$$M_Y(t)=e^{at}M_X(bt)$$

Theorem 1.4. *假设两个随机变量 X 和 Y 的 MGF $M_X(t)$ 和 $M_Y(t)$ 在 0 的某个邻域内的 $N_\epsilon(0) = \{t \in \mathbb{R} : -\epsilon < t < \epsilon\}$ 存在，则对任意的 $z \in R$ 都有 $F_X(z) = F_Y(z)$，当且仅当所有的 $t \in N_\epsilon(0)$，X 和 Y 都有相同的 MGF $M_X(t)$ 和 $M_Y(t)$.*

因此对于一个猜想分布来说，若能够满足 MGF 的定义，最终的分布就是所猜想的分布.

Example 1.3.

$$M_X(t) = \frac{1}{2} + \frac{1}{4}e^{-t} + \frac{1}{4}e^{t}, -\infty < t < \infty$$

其中

$$M_X(t) = E(e^{tX}) \qquad = \sum_x e^{tx} f_X(x) = \frac{1}{2}e^{0\cdot t} + \frac{1}{4}e^{(-1)\cdot t} + \frac{1}{4}e^{1\cdot t}$$

猜想 X 的概率密度为

$$f_X(x) = \begin{cases} \frac{1}{4} & x = -1 \\ \frac{1}{2} & x = 0 \\ \frac{1}{4} & x = 1 \end{cases}$$

Theorem 1.5. *假设 $\{X_n, n = 1, 2 \cdots\}$ 为随机变量序列，每一个随机变量 X_n 有 MGF 和 CDF，进一步的假设，对于 t 在 0 点某个小领域内的任意值. 有*

$$\lim_{n\to\infty} M_n(T) = M_X(t)$$

同时 $F_X(x)$ 也是弱收敛的，即

$$\lim_{n\to\infty} F_n(X) = F_X(x)$$

Theorem 1.6. *假设 $F_X(x)$ 和 $F_Y(y)$ 是有界支撑的 CDE，对于所有的整数 $E(X^k) = E(Y^k)$，则任意的 $z \in (-\infty, +\infty)$ 都有 $F_X(x) = F_Y(y)$*

证明. 因为 X 和 Y 所属于的是一个有界支撑的集合，因此对于一个充分大的数 M，$P(|X| \leq M) = 1$ 及 $P(|Y| \leq M) = 1$，又因 $E\,|X| \leq M^k$ 和 $E\,|Y|^k \leq M^k$ 有

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k) \\ &\leq \sum_{0}^{\infty} \frac{t^k}{k!} E\,|X|^k \\ &\leq \sum_{k=0}^{\infty} \frac{(tM)^k}{k!} = e^{tM} < \infty \end{aligned}$$

同样的有 $M_Y(t) \le e^{tM} < \infty$ 根据 MGF 的公式可得到 $M_X(t) = M_Y(t)$，其中 MGF 的唯一性定理则 $F_X(z) = F_Y(z)$ 对于 $\forall z$ □

1.1 联合矩生成函数

从一维到多维，定义同样没有发生较大的变化

$$M_{XY}(t_1, t_2) = E(e^{t_1X+t_2Y}), -\infty < t_1, t_2 < \infty$$

其中的期望对于在 $(0, 0)$ 范围内的所有 (t_1, t_2) 均成立．假设所有的 $(0, 0)$ 的某个领域内存在，则对于所有的非负整数 $r, s \ge 0$ 有

$$E(X^rY^s) = M_{XY}^{(r,s)}(0, 0)$$

同时

$$cov(X^r, Y^s) = M_{XY}^{(r,s)}(0, 0) - M_X^{(r)}(0)M_Y^{(s)}(0)$$

特别的

$$cov(X, Y) = M_{XY}^{(1,1)}(0, 0) - M_X^{(1)}(0)M_Y^{(1)}(0)$$

Theorem 1.7. *假设 $U = g_1(X)$ 和 $V = g_2(Y)$ 是连续可导一一映射的可测函数，则 X 和 Y 相互独立，当且仅当 U 和 V 相互独立．*

1.2 独立性和期望

假设 (X, Y) 相互独立，则对于任意的可测函数 $h(X)$ 和 $q(Y)$ 有

$$E[h(X)q(Y)] = E[h(X)]E[q(Y)]$$

或者等等价于协方差为 0．

证明.

$$E[h(X)q(Y)] = E[h(X)]E[q(Y)]$$

□

Lemma 1.1. 假设 X 和 Y 相互独立，并且二者的边际 $MGFM_X(t)$ 和 $M_Y(t)$ 对于 t 在 *0* 的某个邻域内存在，则 $M_{X+Y}(t)$ 对于 t 在 *0* 的某个邻域内的所有 t 都存在．且

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

证明.

$$g(X,Y) = e^{t(X+Y)} = e^{tX}e^{tY} = h(X)q(Y)$$

由上独立性定理可得

$$E[e^{t(X+Y)}] = E(e^{tX})E(e^{tY})$$

即

$$M_{X+Y} = M_X(t)M_Y(t)$$

□

这个的定理可以帮助我们在一些重要分布中得到使用，对于一对 $X \sim N(\mu_1, \sigma_1)$ 和 $Y \sim N(\mu_2, \sigma_2)$ 正态分布来说，若相互独立则

$$X \pm Y \sim N(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$$

同样，这个可推及其他分布，柏松分布、卡方分布也有相同的性质．

2 特征函数

对于一些重要分布来说，MGF 并不存在，因此就需要引入一个对任何概率分布都存在的特征函数．

Example 2.1. 假设有这样的随机变量 X 有如下的函数：

$$f(x) = \frac{1}{\pi(1+x^2)}$$

称 X 服从柯西分布．而其矩母函数:

$$\begin{aligned} E(e^{tX}) &= \int_{-\infty}^{+\infty} e^{tx}\frac{1}{\pi(1+x^2)}dx \\ &\geq \int_{0}^{+\infty} e^{|t|x}\frac{1}{\pi(1+x^2)}dx \\ &\geq \int_{0}^{+\infty} \frac{1+|t|x}{\pi(1+x^2)} = +\infty \end{aligned}$$

Definition 2.1. 假设随机变量的 *CDF* 为 $F_X(x)$，则其特征函数为

$$\varphi_X(t) = E(e^{itX}) = \int_{\infty}^{\infty} e^{itx}dF_X(x)$$

其中 $i = \sqrt{-1}$，同时

$$e^{itx} = \cos(tx) + i\sin(tx) \tag{1}$$

2.1 特征函数的性质

- 对于任意的概率分布，其特征函数总是存在且有下界的
- $\varphi_X(0) = 1$
- $\varphi_X(0)$ 在 $(-\infty, +\infty)$ 上连续;
- $\varphi_X(-t) = \varphi_X(t)^*$ 其中 $\varphi_X(t)^*$ 表示为复共轭 (complex conjugate)
- 假设 $Y = a + bX$ 其中 a 和 b 为任意实常数，则

$$\varphi_Y(t) = e^{iat}\varphi_X(bt)$$

- 若 MGF$M_X(t)$ 在 0 的某个邻域内的所有 t 都存在，则对于所有的 t，都有

$$\varphi_X(t) = M_X(it)$$

Example 2.2. 假设随机变量 X 服从柯西 $(0,1)$ 分布，其 *PDF* 为

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

则其特征函数存在且为

$$\varphi_X(t) = e^{-|t|}$$

Theorem 2.1. 假设 X 的 k 阶矩存在，则 $\varphi_X(t)$ 对 $t \in (-\infty, +\infty)$ 是 k 阶可导的，且

$$\varphi_X(0)^{(k)} = i^k E(X^k)$$

Example 2.3. 求标准正态分布 $N(0,1)$ 的特征函数:

$$\psi(t) = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} e^{itx} e^{-x^2/2} dx$$

$$\begin{aligned}
\psi'(t) &= \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} ixe^{itx} e^{-x^2/2} dx \\
&= \frac{i}{\sqrt{2\pi}} \int_{\infty}^{\infty} e^{itx} d\left(-e^{x^2/2}\right) \\
&= -\frac{i}{\sqrt{2\pi}} \left(e^{itx - x^2/2} |_{\infty}^{\infty} + t \int_{-\infty}^{+\infty} e^{itx} e^{-x^2/2} dx \right) \\
&= -t\psi(t)
\end{aligned}$$

于是

$$\begin{aligned}
&\frac{d}{dt} \log \psi(t) = -t, \\
&\psi(t) = ce^{-\frac{1}{2}t^2},
\end{aligned}$$

用 $\psi(0) = 1$ 带入求解微分方程得:

$$\psi(t) = e^{-\frac{1}{2}t^2}$$

Theorem 2.2 (唯一性定理)**.** 假设两个随机变量 X 和 Y 的特征函数为 $\varphi_X(t)$ 和 $\varphi_Y(t)$，则 X 和 Y 具有同分布，当且仅当对所有的 $t \in (-\infty, +\infty)$ 有 $\varphi_X(t) = \varphi_Y(t)$

证明. 根据定义，特征函数为 CDF 的傅立叶变换

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} dF_X(x)$$

因此可以看到对于每一个给定的分布 $F_X(x)$ 都有唯一的特征函数. □

对于一个正态分布来说，若随机变量 $X = (X_1, \cdots, X_n)$ 的分布函数为 $F(x_1, \cdots, x_n)$，与随机变量相仿. 类似的可得到其特征函数

$$\psi(t_1, \cdots, t_n) = E(e^{it^T X}) = \int_{\infty}^{\infty} \cdots \int_{\infty}^{\infty} e^{i(t_1x_1 + \cdots + t_nx_n)} dF(x_1, \cdots, x_n)$$

Lemma 2.1. 若 $X \sim N(0, I_n)$，则 X 的任意线性函数 $Y = A_{n\times m}X + \mu$ 服从正态分布 $N(\mu, AA^T)$.

Lemma 2.2. 若 $Y \sim N(\mu, \Sigma)$，则

$$A\boldsymbol{Y} + \boldsymbol{b} \sim N(A\boldsymbol{\mu} + \boldsymbol{b}, A\Sigma A^T)$$

Theorem 2.3. 对于随机向量序列 $\{X_n\}$ 的 *CDF* 和特征函数分别为 $F_n(x)$ 和 $\phi_n(t)$，又假设随机变量 X 的 *CDF* 和特征函数分别为 $F_X(x)$ 和 $\phi_X(t)$，令 $n \to 0$ 若对 $F_X(x)$ 对所有连续点 x，有 $F_n(X) \to F_X(x)$，则对任意的 $t \in (-\infty, \infty)$ 都有 $\phi_n(t) \to \varphi_X(t)$ 若对于任意的，则对于 $F_X(x)$ 的连续点 x，有 $F_n(x) \to F_X(x)$

3 总结

- 在实际应用中，逐个测量事件空间中的各事件发生的概率（或者分布函数）是极端困难的，相反，对大多数分布而言，矩（平均值、方差以及各种高阶矩）往往是容易被测量的；
- 在问题变得复杂之后，再来计算矩（例如均值、方差等等）的时候，如果我们知道分布函数，那么我们要做的是求和与积分，而如果我们知道特征函数，在计算矩的时候，我们要做的只是微分，而通常，求导会比直接积分更容易，而且可以针对各阶矩有更统一的形式.
- 特征函数相当于转换了一个坐标系，能够对两个分布的特征进行进一步更为细致的刻画.