第二节 · 随机变量

司继春

上海对外经贸大学统计与信息学院

1 一元随机变量

上面介绍了一般的概率空间构建所需要的步骤,即一个概率空间 $(\Omega,\mathscr{F},\mathscr{P})$,我们需要一个样本空间 Ω,一个性质良好的集合族 $\mathscr{F}$ 以及定义在这个集合族上的概率函数 P。特别地,当我们选取样本空间 $\Omega=\mathbb{R}$ 时,我们有 Borelσ-代数 $\mathscr{B}$ 以及由分布函数 F 定义的概率函数 P 组成的概率空间 $(\mathbb{R},\mathscr{B},P)$。

尽管对于一般的样本空间 Ω,我们都可以使用以上的方法构建概率空间,然而很多时候,直接对原始的样本空间 Ω 和集合族 $\mathscr{F}$ 进行分析并不是非常的方便。比如,Ω 作为样本空间,我们并没有限制 Ω 具有代数结构,因而一般我们不能对样本点进行加减等运算。再比如如果我们的研究对象为抛 1000 次硬币,那么我们的样本空间有 2^{1000} 个元素,而这些元素不能相加相减(比如没有正面 + 正面 这样的运算)。为了方便分析,我们一般会把原始的概率空间 Ω 映射到实轴 $\mathbb{R}$ 上进行分析,于是就有了随机变量的概念。

定义 1. (**随机变量**)对于概率空间 $(\Omega,\mathscr{F},\mathscr{P})$,映射 $X:\Omega\to\mathbb{R}^*$ 满足:对于任意的 $B\in\mathscr{B}$,有:

$$X^{-1}\left(B\right)\triangleq\left\{\omega:X\left(\omega\right)\in B\right\}\in\mathscr{F}$$

那么我们称 X 为随机变量(random variable, r.v.)。

例 1. 对于抛硬币的实验,$\Omega=\{H,T\}$,我们可以定义一个随机变量 X 如下:

$$\begin{cases}X\left(H\right)=0\\X\left(T\right)=1\end{cases}$$

对于 $\mathscr{F}=\{\emptyset,\{H\},\{T\},\Omega\}$,我们有 $X^{-1}\left(0\right)=\{H\},X^{-1}\left(1\right)=\{T\}$,对于其他任何 Borel 集 B,如果 $1\in B$ 则 $T\in X^{-1}\left(B\right)$,如果 $0\in B$ 则 $H\in X^{-1}\left(B\right)$。如此我们便定义了一个 $\Omega\to\mathbb{R}^*$ 的随机变量 X。

例 2. 对于泊松分布,其 $\Omega=\{0,1,2\ldots\}=\mathbb{Z}$,定义 $X\left(\omega\right)=\omega,\omega\in\mathbb{Z}$,同上,我们定义了从自然数集合 $\mathbb{Z}\to\mathbb{R}$ 的随机变量 X。

例 3. 电灯泡的寿命的样本空间为 $\Omega=(0,+\infty)$，我们可以定义 $X(\omega)=\omega,\omega\in\Omega$，如此我们定义了从正实数集 $\mathbb{R}^{+}\to\mathbb{R}$ 的随机变量 X。

在此之前我们定义了离散型的样本空间和离散型的分布函数，以上例 (1) 和例 (2) 都属于这种情况。下面我们来定义离散型的随机变量：

定义 2. （**离散型随机变量**）如果存在一个可数集 $B\in\mathscr{B}$，满足 $P(X\in B)=1$，则随机变量 X 成为离散型随机变量。

在得到随机变量的定义之后，我们还需要定义在 $(\mathbb{R}^{*},\mathscr{B})$ 上的概率函数才能完成随机变量的概率空间的定义。由于随机变量是定义在一个一般的样本空间 Ω 及其对应的概率空间 $(\Omega,\mathscr{F},\mathscr{P})$ 上，因而自然的想法是使用原概率空间中的概率函数 $\mathscr{P}$ 来定义 $(\mathbb{R}^{*},\mathscr{B})$ 上的新的概率函数 P。

定理 1. 对于一个随机变量 $X:\Omega\to\mathbb{R}$，定义

$$P_X(B)=\mathscr{P}\left(X^{-1}(B)\right)=\mathscr{P}(\{\omega:X(\omega)\in B\})$$

则 P_X 为概率函数，$(\mathbb{R},\mathscr{B},P_X)$ 为概率空间，我们称 $(\mathbb{R},\mathscr{B},P_X)$ 为 $(\Omega,\mathscr{F},\mathscr{P})$ 导出的概率空间。

例 4. 对于例 (1) 中的随机变量 X，如果原概率空间 $(\Omega,\mathscr{F},\mathscr{P})$ 中定义 $\mathscr{P}(H)=\mathscr{P}(T)=0.5$，则 $P_X(\{1\})=\mathscr{P}\left(X^{-1}(1)\right)=\mathscr{P}(T)=0.5$，同理可定义 $P(\{0\})$，从而完成概率空间 $(\mathbb{R},\mathscr{B},P_X)$ 的定义。

在此之前，我们定义了分布函数，而对于每一个随机变量，由于其值域均为 $\mathbb{R}$，因而其总对应一个分布函数。

定义 3. （累积分布函数）对于一个随机变量 X，函数

$$F_X(x)=P_X((-\infty,x])=\mathscr{P}\left(X^{-1}((-\infty,x])\right),\forall x\in\mathbb{R}$$

为一个分布函数（满足分布函数定义的要求），我们称其为**累积分布函数**（**cumulative distribution function, c.d.f.**）。

对于随机变量来说，累积分布函数包含了所有概率函数 P_X 的信息，因而使用 P_X 和使用累积分布函数 F_X 是等价的。因而我们通常使用标记 $X\sim F_X(x)$ 表示随机变量 X **服从** F_X **分布**。此外，如果随机变量 X 和 Y 具有同样的分布，则记为 $X\sim Y$。

例 5. （累积分布函数）泊松分布和 Logistic 分布函数如图 (1.1) 和 (1.2) 所示：

定义 4. 如果两个随机变量的累积分布函数 $F_X(x)=F_Y(x)$，则我们称两个随机变量**同分布**。

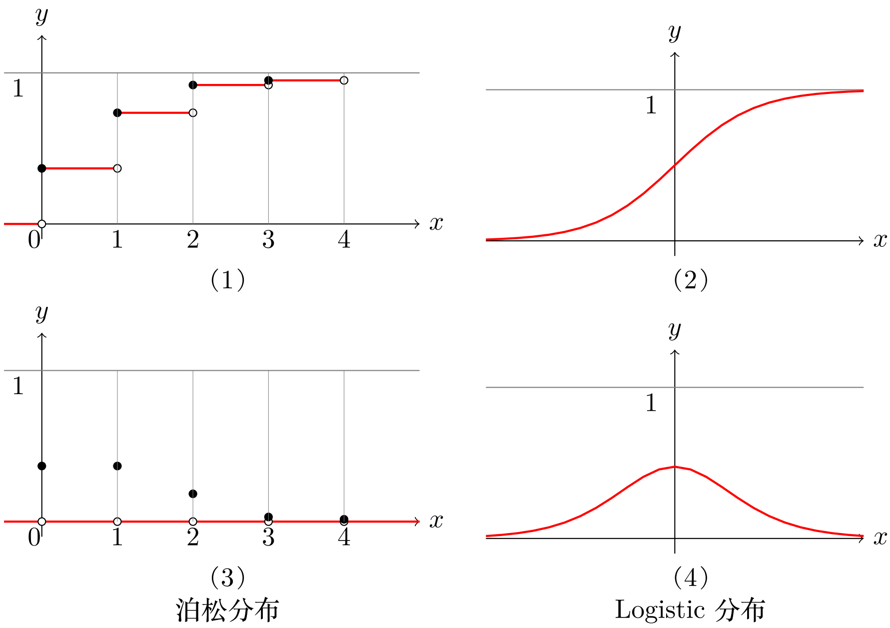


图 1: 累积分布函数与概率密度函数

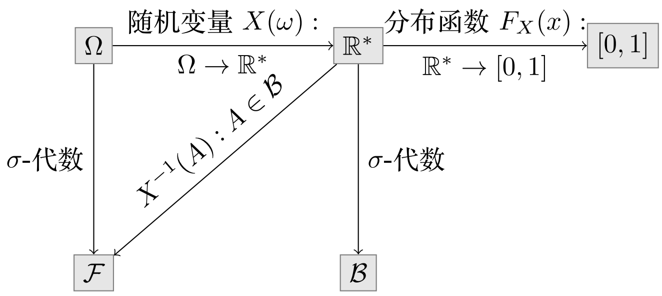


图 2: 随机变量

总结一下，图 (2) 回顾了从原概率空间 $(\Omega, \mathscr{F}, \mathscr{P})$ 定义随机变量 X 并定义新的概率空间 $(\mathbb{R}, \mathscr{B}, P_X)$，并在此基础上定义分布函数的整个过程。由于随机变量极大的简化了分析，因此下面的分析中主要以随机变量为主要研究手段研究概率与统计问题。

虽然累计分布函数描述了随机变量的所有特征，然而很多时候，使用概率密度函数可以使分析和计算更为便利。

定义 5. 对于离散随机变量，**概率质量函数**（**probability mass function**, p.m.f）定义为：

$$f_X(x) = P(X = x), \text{for all } x$$

定义 6. （概率密度函数）对于连续型随机变量，**概率密度函数**（**probability density function**, p.d.f），$f_X(x)$ 定义为：

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt, \text{for all } x$$

例 6. （概率密度函数）泊松分布的 p.m.f 如图 (1.3) 所示，Logistic 分布的 p.d.f 如图 (1.2) 所示。

2 期望及其性质

2.1 期望的定义与性质

数学期望（**mathematical expectation**）的概念在概率论和统计学中有着最广泛的应用。在一个一般的概率空间 $(\Omega, \mathscr{F}, \mathscr{P})$ 中，数学期望实际上就是在这个概率空间中的积分。下面我们将给出一个在一般的概率空间中期望（积分）的定义。

定义 7. （离散型随机变量的期望）在概率空间 $(\Omega, \mathscr{F}, \mathscr{P})$ 中，对于正的离散型随机变量 X：

$$X(\omega) = b_j \; if \; \omega \in \Lambda_j$$

其中 Λ_j 为样本空间 Ω 的划分，$b_j \geq 0$。期望定义为：

$$\mathbb{E}(X) = \sum_j b_j \mathscr{P}(\Lambda_j)$$

注意在上面的定义中，我们要求随机变量为离散型随机变量，而对于样本空间是否离散并没有做假定。

例 7. 对于例 (2) 中定义的随机变量 $X(\omega) = \omega, \omega \in \mathbb{Z}$，其期望为：

$$\mathbb{E}(X) = \sum_{j=0}^{\infty} j \cdot \mathscr{P}(\{j\}) = \sum_{j=0}^{\infty} j \cdot \frac{\lambda^j}{j!} e^{-\lambda} = \lambda \sum_{j=1}^{\infty} \frac{\lambda^{j-1}}{(j-1)!} e^{-\lambda} = \lambda$$

特别的，如果令 $X(\omega)=1_A(\omega)$，其中 1_A 为指示函数，即：

$$1_A(\omega)=\begin{cases}1 & if\ \omega\in A\\ 0 & otherwise\end{cases}$$

那么其期望：

$$\mathbb{E}(1_A)=1\cdot\mathscr{P}(A)+0=\mathscr{P}(A)$$

下面我们使用离散型随机变量的数学期望继续定义连续型随机变量的数学期望。令 X 为定义在 $(\Omega,\mathscr{F},\mathscr{P})$ 上任意的正的随机变量（$X(\omega)\geq 0$），定义集合：

$$\Lambda_{mn}=\left\{\omega:\frac{n}{2^m}\leq X(\omega)<\frac{n+1}{2^m}\right\}\subset\mathscr{F}$$

如此，对于任意一个 m，我们可以定义一个离散随机变量 X_m：

$$X_m(\omega)=\frac{n}{2^m}\,if\ \omega\in\Lambda_{mn}$$

注意对于任意 m，有：

$$X_m(\omega)\leq X_{m+1}(\omega);0\leq X(\omega)-X_m(\omega)<\frac{1}{2^m},\forall\omega\in\Omega$$

即 X_m 为单调递增的随机变量，从而

$$\lim_{m\to\infty}X_m(\omega)=X(\omega),\forall\omega\in\Omega$$

根据以上离散随机变量的定义，X_m 的期望可以定义为：

$$\mathbb{E}(X_m)=\sum_{n=0}^{\infty}\frac{n}{2^m}\mathscr{P}\left(\left\{\frac{n}{2^m}\leq X(\omega)<\frac{n+1}{2^m}\right\}\right)$$

如果存在一个 m 使得 $E(X_m)=+\infty$，那么定义 $E(X)=+\infty$，否则，定义：

$$\mathbb{E}(X)=\lim_{m\to\infty}\mathbb{E}(X_m)$$

如果上述极限存在且小于 $+\infty$，我们称随机变量 X 是**可积**（integrable）的。

至此我们定义了任意正随机变量的期望。对于任意的随机变量 X，定义 $X^+(\omega)=\max\{X(\omega),0\}$，$X^-(\omega)=\max\{-X(\omega),0\}$，则 $X=X^+-X^-$，$|X|=X^++X^-$。X^+ 和 X^- 都是正的随机变量，如果 $|X|$ 是可积的，那么我们称随机变量 X 是**可积**的，此时可以定义随机变量 X 的期望为：

$$\mathbb{E}(X)=\mathbb{E}\left(X^+\right)-\mathbb{E}\left(X^-\right)$$

至此，任意随机变量的期望即定义完毕。由于该期望是由概率函数 $\mathscr{P}$ 定义的，我们一般记期望为：

$$\mathbb{E}(X)=\int_{\Omega}X(\omega)\mathscr{P}(d\omega)$$

更一般的，对于 $A\in\mathscr{F}$，定义随机变量 X 在集合 A 上的积分为：

$$\int_{A}X(\omega)\mathscr{P}(d\omega)=\mathbb{E}(X\cdot 1_A)=\int_{\Omega}X(\omega)\cdot 1_A(\omega)\mathscr{P}(d\omega)$$

特别的，令 $X(\omega)=1$ 为退化的随机变量，根据以上积分的定义，同样有：

$$\mathscr{P}(A)=\mathbb{E}(1_A)=\int_{A}\mathscr{P}(d\omega)$$

积分有以下性质：

定理 2. （积分的性质）

1. (线性性) $\int_A(aX(\omega)+bY(\omega))\mathscr{P}(d\omega)=a\int_A X(\omega)\mathscr{P}(d\omega)+b\int_A Y(\omega)\mathscr{P}(d\omega)$
2. (可加性）如果 A_n 不相交，则 $\int_{\bigcup_n A_n}X\mathscr{P}(d\omega)=\sum_n\int_{A_n}X\mathscr{P}(d\omega)$
3. 如果在 A 上，$X\geq 0\ a.s.$，则 $\int_A X\mathscr{P}(d\omega)\geq 0$
4. (单调性) 如果在 A 上, $X_1\leq X\leq X_2 a.s.$, 则 $\int_A X_1\mathscr{P}(d\omega)\leq\int_A X\mathscr{P}(d\omega)\leq\int_A X_2\mathscr{P}(d\omega)$
5. (均值定理) 如果在 A 上有 $a\leq X\leq b a.s.$, 那么 $a\mathscr{P}(A)\leq\int_A X\mathscr{P}(d\omega)\leq b\mathscr{P}(A)$
6. $\left|\int_A X\mathscr{P}(d\omega)\right|\leq\int_A|X|\mathscr{P}(d\omega)$
7. (有界收敛定理）如果 $\lim_{n\to\infty}X_n=X\ a.s.$，且存在一个常数 M 使得 $\forall n:|X_n|\leq M\ a.s.$，那么
$$\lim_{n\to\infty}\int_A X_n\mathscr{P}(d\omega)=\int_A X\mathscr{P}(d\omega)=\int_A\left(\lim_{n\to\infty}X_n\right)\mathscr{P}(d\omega)\qquad(1)$$
8. (单调收敛定理）如果 $X_n\geq 0$ 且 $X_n\uparrow X\ a.s.$，那么 (1) 式成立
9. (控制收敛定理）如果 $\lim_{n\to\infty}X_n=X\ a.s.$，且存在一个随机变量 Y 使得 $|X_n|\leq Y\ a.s.$，且 $\int_A Y\mathscr{P}(d\omega)<\infty$，则 (1) 式成立。
10. 如果 $\sum_n\int_A|X_n|\mathscr{P}(d\omega)<\infty$，那么 $\sum_n|X_n|<\infty\ a.s.\ on\ A$，从而
$$\int_A\sum_n X_n\mathscr{P}(d\omega)=\sum_n\int_A X_n\mathscr{P}(d\omega)$$

由于期望即定义为 Ω 上的积分，因而以上性质对于期望都成立，比如由积分的线性性可以得到期望的线性性：

$$\mathbb{E}\left(aX+bY\right)=a\mathbb{E}\left(X\right)+b\mathbb{E}\left(Y\right)$$

而定理 (2.7-2.9) 则解决了极限符号和积分符号互换的问题。注意如果不满足以上定理的条件，积分和极限符号不必然可以互换。

例 8. （积分与极限互换）如果令 $\Omega=\mathbb{R}$，分布函数为：

$$F\left(\omega\right)=\begin{cases}0 & \omega<-1\\ \frac{1}{2}\omega+\frac{1}{2} & -1\leq\omega\leq1\\ 1 & \omega>1\end{cases}$$

即在 $[-1,1]$ 上的均匀分布，进而使用此分布函数构建 $\mathbb{R}$ 上的概率测度 P。随机变量

$$X_n\left(\omega\right)=\begin{cases}0 & if\ \ |\omega|>\frac{1}{n}\\ n & if\ \ |\omega|\leq\frac{1}{n}\end{cases}$$

因而除了在一个点 $\omega=0$ 处之外，$\lim_{n\to\infty}X_n\left(\omega\right)=0$，或者 $\lim_{n\to\infty}X_n\left(\omega\right)=0\ a.s.$，因而[1]

$$\int_\Omega\lim_{n\to\infty}X_nP\left(d\omega\right)=0$$

然而由于 $\int_\Omega X_nP\left(d\omega\right)=1$，因而 $\lim_{n\to\infty}\int_\Omega X_nP\left(d\omega\right)=1$。因而在这个例子里 $\lim_{n\to\infty}\int_\Omega X_nP\left(d\omega\right)\neq\int_\Omega\lim_{n\to\infty}X_nP\left(d\omega\right)$。

以上我们介绍了数学期望的定义，然而给定一个随机变量，使用概率空间 $(\Omega,\mathscr{F},\mathscr{P})$ 计算数学期望非常不方便，我们通常希望使用导出的概率空间 $(\mathbb{R},\mathscr{B},P)$ 计算数学期望，因此我们有以下定理：

定理 3. 如果概率空间 $(\Omega,\mathscr{F},\mathscr{P})$ 根据定理 *(1)* 导出了概率空间 $(\mathbb{R},\mathscr{B},P)$，令 g 为一个可测函数，则：

$$\mathbb{E}\left(g\left(X\right)\right)=\int_\Omega g\left(X\left(\omega\right)\right)\mathscr{P}\left(d\omega\right)=\int_{\mathbb{R}}g\left(x\right)P\left(dx\right)$$

如果等式两边积分都存在。

如果 F 为对应于概率函数 P 的分布函数，则在 $(a,b]$ 上的积分也可以写为：

$$\int_{(a,b]}g\left(x\right)P\left(dx\right)=\int_{(a,b]}g\left(x\right)dF\left(x\right)$$

[1]在定义积分时，当遇到 $0\cdot\infty$ 时，定义 $0\cdot\infty=0$。

因而随机变量 X 的数学期望可以写为:

$$\mathbb{E}(X) = \int_{\mathbb{R}} x dF(x)$$

此外，根据积分以及分布函数的定义:

$$\int_{(a,b]} dF(x) = P((a,b]) = F(b) - F(a)$$

例 9. （Cauchy 分布的期望）Cauchy 的密度函数为:

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

如果一个随机变量服从 Cauchy 分布，则:

$$\mathbb{E}(|X|) = \int_{\mathbb{R}} \frac{|x|}{\pi} \frac{1}{1+x^2} dx = \frac{2}{\pi} \int_{[0,\infty)} \frac{x}{1+x^2} dx$$

对于任意正数 M:

$$\int_{[0,M]} \frac{x}{1+x^2} dx = \left.\frac{\log(1+x^2)}{2}\right|_0^M = \frac{\log(1+M^2)}{2}$$

因此:

$$\mathbb{E}(|X|) = \frac{1}{\pi} \lim_{M\to\infty} \log(1+M^2) = \infty$$

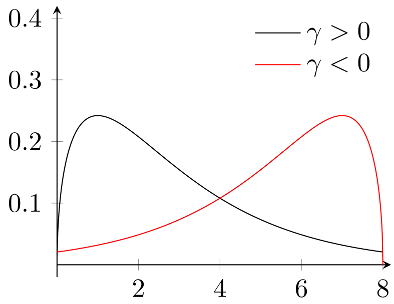
因而该随机变量是不可积的。

对于任意的正数 p，如果 $\mathbb{E}(|X|^p) < \infty$，则记 $X \in L^p = L^p(\Omega, \mathscr{F}, \mathscr{P})$。对于整数 r，随机变量 X 的 r **阶矩**被定义为 $\mathbb{E}(X^r)$。一阶矩即为随机变量 X 的期望。此外，随机变量 X 的 r **阶中心矩**被定义为 $\mathbb{E}([X - E(X)]^r)$。特别的，当 $r = 2$ 时，2 阶中心矩即为随机变量的**方差**（**variance**），记为 $\mathrm{Var}(X)$ 或者 $\sigma^2(X)$，**标准差**（**standard deviation**）定义为 $\sigma(X) = \sqrt{\mathrm{Var}(X)}$。$X$ 的方差可以使用一阶和二阶矩计算得到:

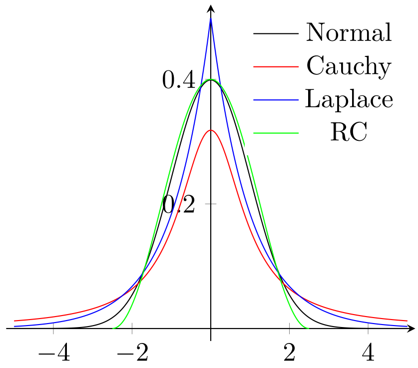
$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}\left([X - \mathbb{E}(X)]^2\right) = \mathbb{E}\left(X^2 - 2\mathbb{E}(X)\cdot X + \mathbb{E}(X)^2\right) \\
&= \mathbb{E}\left(X^2\right) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2
\end{aligned}$$

注意 $\mathbb{E}X^2 \geq (\mathbb{E}X)^2$。此外，根据方差定义，有 $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$。

除了前两阶矩之外，经常我们还会关心更高阶的矩。其中，**偏度**（**skewness**）和**峰度**（**kurtosis**）是最经常被关心的高阶矩。其中偏度为随机变量的三阶中心



(a) 左偏与右偏



(b) 尖峰厚尾

图 3: 偏度与峰度

矩，即定义：

$$\gamma = \mathbb{E}\left(\left[\frac{X-\mathbb{E}(X)}{\sigma(X)}\right]^3\right)$$

如果 X 为对称分布，那么必然有 $\gamma = 0$。顾名思义，偏度与分布的不对称性有关，如图 (3.a) 所示，当 $\gamma > 0$ 时，分布函数右边的尾巴比较厚，我们称其分布为右偏，反之为左偏。

而峰度则是随机变量的四阶中心矩，即：

$$\mathrm{Kurt}(X) = \mathbb{E}\left(\left[\frac{X-\mathbb{E}(X)}{\sigma(X)}\right]^4\right) = \frac{\mathbb{E}\left([X-\mathbb{E}(X)]^4\right)}{[\mathrm{Var}(X)]^2}$$

尽管我们一般将其称为「峰度」，然而这一称呼并不准确，更准确的称呼应该为「尾厚度」。如图 (3.a) 所示，其中 RC 代表 Raised cosine 分布。如果比较正态分布，会发现正态分布的尾巴比 Raised cosine 分布要厚，而相应的正态分布的峰要「尖」一点，所以我们经常说「尖峰厚尾」。实际上，正态分布的峰度为 3，而图中 Raised cosine 分布的峰度约为 $2.40623 < 3$。

然而注意到，峰度大的并不一定代表峰更「尖」，而仅仅是尾巴更「厚」。如图中 Laplace 分布和 Cauthy 分布相比，Laplace 分布的峰更尖，但是其峰度小于 Cauthy 分布的峰度。实际上，Cauthy 分布的峰度为 $+\infty$。因而判断峰度时，不能顾名思义只看其峰的尖锐程度，而应该看尾巴的厚度。

我们一般会把峰度与正态分布的峰度相比，定义超额峰度（excess kurtosis）为峰度减 3，因而如果超额峰度 > 0，那么其尾巴比正态分布的尾巴要厚，而如果超额峰度 < 0，那么其尾巴要比正态分布的尾巴要薄。

2.2 积分与导数互换

在实际应用中，我们经常需要对积分进行求导，比如：

$$\frac{d}{d\theta}\int_{\mathbb{R}} g(x,\theta)\, dF(x)$$

然而积分常常难以计算，经常我们希望使用：

$$\int_{\mathbb{R}} \frac{dg(x,\theta)}{d\theta} dF(x)$$

来计算。然而这一计算方法是有条件的。

定理 4. （积分与微分互换）如果函数 $g(x,\theta)$ 在 $\theta=\theta_0$ 处可微，即对于任意 x，极限

$$\lim_{\delta\to 0}\frac{g(x,\theta_0+\delta)-g(x,\theta_0)}{\delta} = \left.\frac{\partial}{\partial\theta} g(x,\theta)\right|_{\theta=\theta_0}$$

存在，并且存在一个函数 $h(x,\theta_0)$ 以及一个常数 δ_0，有

1. 对于任意的 x 和 $|\delta|\le\delta_0$，有：$\left|\frac{g(x,\theta_0+\delta)-g(x,\theta_0)}{\delta}\right| \le h(x,\theta_0)$
2. $\int_{\mathbb{R}} h(x,\theta_0)\, dF(x) < \infty$

那么：

$$\left.\frac{d}{d\theta}\int_{\mathbb{R}} g(x,\theta)\, dF(x)\right|_{\theta=\theta_0} = \int_{\mathbb{R}}\left[\left.\frac{dg(x,\theta)}{d\theta}\right|_{\theta=\theta_0}\right] dF(x)$$

该定理即控制收敛定理的应用。该定理表明，在一定的条件下，积分与微分操作可以交换顺序。在该条件成立下，进一步我们有莱布尼茨法则：

定理 5. （***Leibnitz* 法则**）如果 $g(x,\theta), a(\theta), b(\theta)$ 对 θ 可微，那么：

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} g(x,\theta)\, dx = g(b(\theta),\theta)\frac{d}{d\theta}b(\theta) - g(a(\theta),\theta)\frac{d}{d\theta}a(\theta) + \int_{a(\theta)}^{b(\theta)}\frac{\partial}{\partial\theta} g(x,\theta)\, dx$$

2.3 常用不等式

下面我们介绍几个常用的不等式。

定理 6. （***Chebyshev* 不等式**）如果函数 ψ 满足：$\psi(u)=\psi(-u)\ge 0$，且在 $(0,\infty)$ 上单调递增，X 为随机变量，且 $\psi(X)<\infty$，则对于 $u>0$，有：

$$\mathscr{P}(|X|\ge u) \le \frac{\mathbb{E}[(\psi(X))]}{\psi(u)}$$

Proof. 根据均值定理：

$$\mathbb{E}[\psi(X)] = \int_{\Omega}\psi(X)\,\mathscr{P}(d\omega) \ge \int_{\{|X|\ge u\}}\psi(X)\,\mathscr{P}(d\omega) \ge \psi(u)\,\mathscr{P}(|X|\ge u)$$

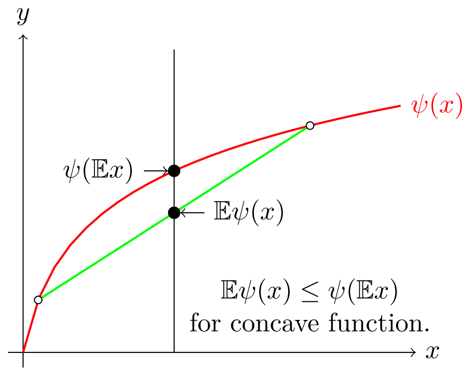


图 4: Jensen 不等式

□

特别的，令 $Y = \frac{[X-\mathbb{E}(X)]}{\sigma(X)}$，则 $\mathbb{E}(Y)=0$，$\mathbb{E}\left(Y^2\right)=1$。令 $\psi(x)=x^2$，有：

$$\mathscr{P}(|Y| \geq u) \leq \frac{1}{u^2}$$

定理 7. （***Jensen*** **不等式**）如果 ψ 为凹函数，且随机变量 X 和 $\psi(X)$ 可积，则：

$$\psi(\mathbb{E}(X)) \geq \mathbb{E}[\psi(X)]$$

Jensen 不等式表明，对于一般的非线性函数，期望的函数与函数的期望并不相等，如图 (4) 所示。

作为 Jensen 不等式的一个应用，考虑 $0<r<s$，并令 $p=\frac{s}{r}>1$，注意 $\psi(x)=|x|^p$ 为凸函数，因而根据 Jensen 不等式，有：

$$\mathbb{E}\left(|X|^{rp}\right)=\mathbb{E}\left(|X|^{s}\right) \geq \left[\mathbb{E}\left(|X|^{r}\right)\right]^{p}=\left[\mathbb{E}\left(|X|^{r}\right)\right]^{\frac{s}{r}}$$

整理之后可以得到：

$$\mathbb{E}\left(|X|^{r}\right) \leq \left[\mathbb{E}\left(|X|^{s}\right)\right]^{\frac{r}{s}}$$

以上不等式被称为 **Liapounov 不等式**。

定理 8. （***Cauchy-Schwarz*** **不等式**）对于两个随机变量 X, Y，若满足可积性，有：

$$|\mathbb{E}(XY)| \leq \mathbb{E}(|XY|) \leq \left[\mathbb{E}\left(|X|^2\right)\right]^{\frac{1}{2}}\left[\mathbb{E}\left(|Y|^2\right)\right]^{\frac{1}{2}}$$

3 随机变量的变换

对于一个可测函数 $g(\cdot)\mathbb{R} \to \mathbb{R}$，$Y=g(X)$ 也是一个随机变量。如果我们已知 X 的分布，如何获得新的随机变量 Y 的分布呢？

首先仿照随机变量的定义，我们定义函数 $g(\cdot)$ 对于一个集合的逆为：

$$g^{-1}(A) = \{x \in \mathbb{R} : g(x) \in A\}$$

因而对于单点集 $g^{-1}(\{y\}) = \{x \in \mathbb{R} : g(x) = y\}$。

对于离散型的随机变量 X，$g(X)$ 也是离散型的，因而随机变量 Y 的 p.m.f. 为：

$$f_Y(y) = P(Y = y) = \sum_{x \in g^{-1}(\{y\})} P(X = x) = \sum_{x \in g^{-1}(\{y\})} f_X(x)$$

由此可以确定随机变量 $Y = g(X)$ 的概率质量函数。

对于一个一般的随机变量 X，随机变量 $Y = g(X)$ 的累积分布函数可以使用如下定义计算：

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(g(X) \le y) \\ &= P(\{x : g(x) \le y\}) \\ &= \int_{\{x : g(x) \le y\}} dF_X \end{aligned}$$

特别的，如果 $g(\cdot)$ 严格单调递增，则：

$$F_Y(y) = F_X\left(g^{-1}(y)\right)$$

如果 $g(\cdot)$ 严格单调递减，则：

$$F_Y(y) = 1 - F_X\left(g^{-1}(y)\right)$$

现在，假设我们有一个随机变量 $U \sim Uniform(0,1)$，即 $(0,1)$ 区间上的均匀分布，而 $F(\cdot)$ 为一个分布函数，我们可以定义一个新的随机变量 $X = F^{-1}(U)$。注意如果分布函数 $F(\cdot)$ 存在「平台」，即 $F(x) = c,\ for\ a \le x < b$，那么定义 $F^{-1}(u) = \inf\{x : F(x) \ge u\}$。注意对于一个存在「平台」的分布函数 F，$F(x) = c,\ for\ a \le x < b$，$P(X \le x) = P(X \le a),\ for\ a \le x < b$，也就是说如果某个随机变量的分布函数为 F，那么其取值在 $[a,b)$ 范围内的概率为 0。根据以上推理，随机变量 $X = F^{-1}(U)$ 的分布函数为：

$$\begin{aligned} F_X(x) &= \int_{F^{-1}(u) \le x} dG(u) \\ &= \int_{F^{-1}(u) \le x} du \\ &= \int_{u \le F(x)} du \\ &= F(x) \end{aligned}$$

其中第二个等号由于均匀分布的分布函数 $G(u)=u, 0\leq u\leq 1$。这也就意味着，如果我们有分布函数 $F(\cdot)$，可以生成一个 $(0,1)$ 区间内的随机变量 U，进而生成 $X=F^{-1}(U)$，那么我们就生成了一个新的随机变量，其分布函数为 F。

例 10. （指数分布）若 $U\sim Uniform(0,1)$，令 $F(x)=1-e^{-x}, x>0$，即指数分布的分布函数。令 $X=F^{-1}(U)$，则随机变量 X 的分布函数为：

$$\begin{aligned} F_X(x) &= \int_{F^{-1}(u)\leq x} dG(u) \\ &= \int_{F^{-1}(u)\leq x} du \\ &= \int_{u\leq(1-e^{-x})} du \\ &= 1-e^{-x} \end{aligned}$$

因而为了生成指数分布的随机变量，只要生成均匀分布 U，并令 $X=-\ln u$ 即可。

在绝大多数计算机语言中，都有生成均匀分布的指令。比如在 C 语言中，可以使用标准库 <stdlib.h> 中的 rand() 函数生成均匀分布：

```
double x=(double)rand()/RAND_MAX;
```

在 Python 中，可以使用 Python 标准库的 random 包，或者 NumPy 的 random 包，比如：

```
import random as rd
x=rd.random()
```

或者：

```
import numpy.random as nprd
x=nprd.random()
```

即可以生成均匀分布。在 Excel 中，也可以使用 RAND() 生成均匀分布的随机数。

习题

练习 1. 在研究中，对收入等变量取对数是非常常见的处理手段。如果 $X>0$ 代表总体收入，那么 $\mathbb{E}(X)$ 和 $\exp[\mathbb{E}(\ln X)]$ 哪一个更大？

练习 2. 如果 r.v.$X\sim Binomial(n,p)$，求随机变量 $Y=n-X$ 的概率质量函数。

练习 3. 求对数正态分布（$Y=e^X, X\sim N(0,1)$）的概率密度函数。

练习 4. 证明:对于一个随机变量 $X\sim F_X$,随机变量 $Y=F_X(X)\sim Uniform(0,1)$。

练习 5. 使用任何编程语言，通过均匀分布生成 100 个 Logistic 分布（分布函数 $\frac{e^x}{1+e^x}$）的随机数，并将理论的分布函数及其经验分布函数画在一张图中。其中经验分布函数的定义为:

$$\hat{F}(x)=\frac{1}{N}\sum_{i=1}^{N}1(X_i\leq x)$$

即样本中小于 x 的样本的 s 比例。观察两者是否贴近? 生成 1000 个数据，重复以上步骤，并比较两张图的差异。

练习 6. 使用任何编程语言，通过均匀分布生成 100 个服从泊松分布（$\lambda=1$）的随机数，并计算均值。

参考文献

[1] Ash, R.B., Doleans-Dade, C., 2000. Probability and measure theory. Academic Press.

[2] Athreya, K.B., Lahiri, S.N., 2006. Measure Theory and Probability Theory. Springer, New York.

[3] Casella, G., Berger, R.L., 2002. Statistical inference. Duxbury Pacific Grove, CA.

[4] Chung, K.L., 2001. A Course in Probability Theory, 3rd editio. ed. Elsevier Ltd., Singapore.

[5] Shao, J., 2007. Mathematical Statistics, 2nd ed. Springer, New York.