Lecture 4

条件分布

ruanyl@buaa.edu.cn
September 2016, Beihang University

随机变量的独立性

定义 随机变量 X, Y 称为是独立的, 如果:

$$P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B)$$

对任意 $A, B \subset \mathbb{R}$. 严格地说 A, B 应该是由简单区间（开区间, 闭区间, 或半开闭区间）生成的"可测集", 这部分内容已超出本课程要求, 在此不做深入讨论. 但我们有如下定理

定理 随机变量 X, Y 是独立的当且仅当对 $\forall x, y \in \mathbb{R}$

$$P(X \leqslant x, Y \leqslant y) = P(X \leqslant x)\, P(Y \leqslant y)\,.$$

定理中 $\leqslant$ 替换为 $<$ 时仍然成立. 这一定理是告诉我们随机变量 X, Y 独立的充分必要条件是联合分布函数等于边际分布函数乘积. 从而借助于分布函数与密度函数的关系, 我们还有

定理 假设 X, Y 其具有密度函数. 随机变量 X, Y 是独立的当且仅当其联合密度函数 $f(x,y)$ 满足

$$f(x,y)=f_X(x)f_Y(y),\ a.e.\ x,y\in\mathbb{R}.$$

如果 $f(x,y), f_X(x), f_Y(y)$ 都是连续的, 那么独立性等价于

$$f(x,y)=f_X(x)f_Y(y),\ \forall\, x,y\in\mathbb{R}.$$

注记 对于离散型随机变量, 类似的充分必要条件同样成立, 只需将上面的密度函数替换为离散点概率函数.

例题 投掷两个均匀硬币, 用 X, Y 分别表示正面是否出现, 正面时取值为 1, 否则取值为 0. 那么 X, Y 是两个相互独立的随机变量.

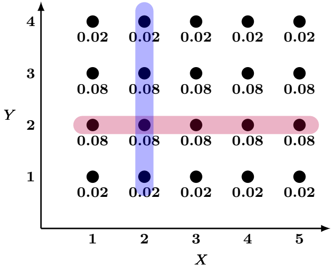
- 由题 $X, Y\in\{0,1\}$. 容易写出 X, Y 各自的分布函数以及联合

分布函数, 通过定义直接验证独立性.

例题 随机变量 X, Y 的联合分布由如下坐标格点所示, 说明 X, Y 是相互独立的.

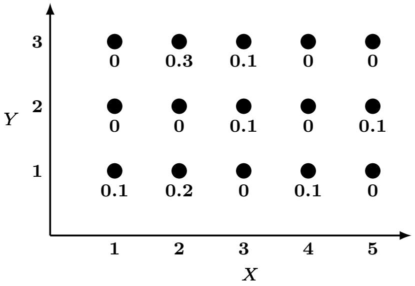
■ 由图 $X \in \{1, 2, 3, 4, 5\}$, $Y \in \{1, 2, 3, 4\}$. 从图中看出, 观测到 X 并不会改变 Y 的分布, 事实上对任意 $i \in \{1, 2, 3, 4, 5\}$

$$P\left(Y = j | X = i\right) = \begin{cases} 0.1, & j = 1, 4 \\ 0.4, & j = 2, 3 \end{cases} = P\left(Y = j\right),$$



反过来观测到 Y 也不会改变 X 的分布. 由此知 X, Y 是相互独立的.

例题 随机变量 X, Y 的联合分布由如下坐标格点所示, X, Y 不是相互独立的.



- 容易计算

$$P(X=2)=\sum_{j} P(X=2, Y=j)=0.5$$

$$P(Y=3)=\sum_i P(X=i,Y=3)=0.4$$

但 $P(X=2,Y=3)=0.3\neq P(X=2)P(Y=3)$.

随机变量的独立性是我们之前遇到的事件独立性的一个延伸概念, 本质上是一样的. 当我们说到随机变量 X,Y 的独立性时, 我们说的是：与 X 相关的任何事件独立于与 Y 相关的任何事件.

注记 事实上如果随机变量 X,Y 是相互独立的, 它们各自的函数也是相互独立的, 即，如果 $f(x)$ 和 $g(y)$ 分别是定义在 X 和 Y 的取值空间上的适当的函数, 那么随机变量 $f(X)$ 和 $g(Y)$ 是相互独立的.

条件分布

离散随机变量的条件分布

定义 假设离散联合分布 (X, Y) 的概率函数为 $f(x, y)$. $f_X(x)$ 和 $f_Y(y)$ 分别是边际概率函数. 如果 $f_Y(y) > 0$, 那么

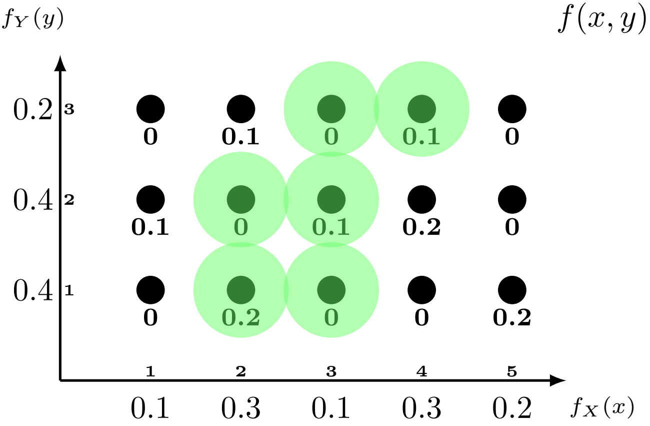
$$f_X(x|y) = \frac{f(x, y)}{f_Y(y)}$$

称为 X 在**给定** $Y = y$ **时的条件概率函数**.

正如事件的条件概率是一个真正的概率, 条件概率函数 $f_X(\cdot|y)$ 也是一个真正的概率函数, 它是非负的, 并满足

$$\sum_x f_X(x|y) = \sum_x \frac{f(x, y)}{f_Y(y)} = \frac{\sum_x f(x, y)}{f_Y(y)} = 1.$$

例题 下图给出 X 和 Y 的联合分布, 试写出关于 X 和 Y 的条件分布. 着色区域的概率是否可以用关于 X 或 Y 的条件分布与边际分布表示出来?



连续随机变量的条件分布

定义 给定集合 A, 以及以 $f_X(x)$ 为密度函数的随机变量 X, A 关于 X 的**条件概率**是指满足下面条件的**可积分函数** $h_A(x)$,

$$P(A, X \leqslant u) = \int_{-\infty}^{u} h_A(x) f_X(x)\, dx, \forall u \in \mathbb{R} \tag{4.1}$$

条件概率 $h_A(x)$ 通常用记号 $P(A|X = x)$ 来表示.

注记 在定义 (4.1) 中, 可要求对任意 x, $P(\cdot|X = x)$ 是一个概率.

在这个定义中如果 $A = \{Y \leqslant v\}$, Y 是以 $f_Y(y)$ 为密度函数的随机变量. 那么

$$P(X \leqslant u, Y \leqslant v) = \int_{-\infty}^{u} P(Y \leqslant v|X = x) f_X(x)\, dx, \forall u \in \mathbb{R} \tag{4.2}$$

如果随机变量 X, Y 具有联合分布 $F(x, y)$, 以及联合密度 $f(x, y)$,

那么上式可以写作

$$
\begin{aligned}
F(u,v) &= \int_{-\infty}^{v}\int_{-\infty}^{u} f(x,y)\,dxdy \\
&= \int_{-\infty}^{u} P(Y \leqslant v|X=x) f_X(x)\,dx
\end{aligned}
$$

等式两边对 u 求导数得到[1]

$$
\frac{\partial F(u,v)}{\partial u} = \int_{-\infty}^{v} f(u,y)\,dy = P(Y \leqslant v|X=u) \cdot f_X(u) \tag{4.3}
$$

(4.3) 两边继续对 v 求导数得到

$$
f(u,v) = \frac{\partial^2 F(u,v)}{\partial v \partial u} = f_X(u) \cdot \frac{\partial}{\partial v} P(Y \leqslant v|X=u) \tag{4.4}
$$

[1]事实上, 这里的导数在除去一个概率为 0 的集合后处处存在.

特别地, 如果 $f_X(u) \neq 0$, 那么 (4.3) 变为

$$P(Y \leqslant v | X = u) = \int_{-\infty}^{v} \frac{f(u, y)}{f_X(u)} dy \tag{4.5}$$

而 (4.4) 变为

$$\frac{\partial}{\partial v} P(Y \leqslant v | X = u) = \frac{f(u, v)}{f_X(u)} \tag{4.6}$$

(4.6) 被称为条件密度函数, 具体地我们有

定义 假设连续型随机变量 X, Y 具有联合分布 $F(x, y)$, $f_X(x)$ 和 $f_Y(y)$ 分别是边际密度函数. 如果 $f_X(x) > 0$, 那么

$$f_Y(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{\partial}{\partial y} P(Y \leqslant y | X = x)$$

称为 Y 在给定 $X = x$ 时的**条件密度函数**. $f_X(x) = 0$ 时, $f(x|y)$ 可自由定义为任何密度函数.

同样地可定义 X 给定 $Y = x$ 时的条件密度函数 $f_X(x|y)$.

定理 $f_Y(y|x)$ 是一个关于 y 的密度函数. $f_X(x|y)$ 是一个关于 x 的密度函数.

连续型随机变量的条件密度 $f_Y(y|x)$ 不能像离散情形一样定义, 因为对任意点 x, 总有 $P(X=x)=0$. 但可以通过事件条件概率的极限来理解, 当 $\Delta_x\to 0$,

$$\begin{aligned}
&P\left(y\leqslant Y\leqslant y+\Delta_y\middle| x\leqslant X\leqslant x+\Delta_x\right)\\
&=\frac{P\left(y\leqslant Y\leqslant y+\Delta_y, x\leqslant X\leqslant x+\Delta_x\right)}{P\left(x\leqslant X\leqslant x+\Delta_x\right)}\\
&=\frac{\int_x^{x+\Delta_x}\left(\int_y^{y+\Delta_y} f(u,v)\,dv\right)du}{\int_x^{x+\Delta_x} f_X(u)\,du}\\
&\approx\frac{f(x,y)\,\Delta_x\Delta_y}{f_X(x)\,\Delta_x}=\frac{f(x,y)}{f_X(x)}\Delta_y
\end{aligned}$$

由于右侧并不依赖于 Δ_x, 因此令 $\Delta_x \to 0$ 我们可以得到

$$P\left(y \leqslant Y \leqslant y+\Delta_y \middle| X=x\right) \approx \frac{f(x,y)}{f_X(x)} \Delta_y = f_Y(y|x)\,\Delta_y$$

这个式子正是在给定 $X=x$ 的条件下 Y 取值于小区间 $[y, y+\Delta_y]$ 的概率, 因此我们可以有

$$P\left(Y \in B \middle| X=x\right) = \int_B f_Y(y|x)\,dy$$

注记 上面给出的仅仅是一个直观解释, 将条件概率理解为事件条件概率的极限有一定的局限性, 极限存在需要许多假设条件.

例题 假设随机变量 X, Y 的联合分布具有密度函数

$$f(x,y) = \begin{cases} \frac{21}{4}x^2 y, & x^2 \leqslant y \leqslant 1 \\ 0, & \text{其他} \end{cases}$$

试求出条件密度函数 $f_Y(y|x)$, 并确定 $P\left(Y \geqslant \frac{3}{4} \middle| X=\frac{1}{2}\right)$.

■ 我们已经计算过

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{x^2}^{1} \frac{21}{4} x^2 y dy = \frac{21}{8}\left(x^2 - x^6\right)$$

因此

$$f_Y(y|x) = \begin{cases} \frac{2y}{1-x^4}, & x^2 \leqslant y \leqslant 1 \\ 0, & \text{其他} \end{cases}$$

从而

$$P\left(Y \geqslant \frac{3}{4} \middle| X = \frac{1}{2}\right)$$

$$= \int_{3/4}^{1} f_Y\left(y \middle| \frac{1}{2}\right) dy$$

$$= \int_{3/4}^{1} \frac{2y}{1-\left(\frac{1}{4}\right)^4} dy$$

$$= \frac{16}{15} y^2 \Big|_{y=3/4}^{1} = \frac{7}{15}$$

乘法公式

按照条件概率定义直接有

$$f(x, y) = f_Y(y|x) \cdot f_X(x)$$

这正是乘法公式.

全概率公式

由乘法公式积分得到

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_{-\infty}^{\infty} f_Y(y|x) \cdot f_X(x)\, dx$$

将 $f_Y(y|x)$ 代入 (4.5) (4.2) 我们便得到分布形式的全概率公式

$$P(X \leqslant x, Y \leqslant y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{y} f_Y(v|u)\, dv \right) f_X(u)\, du$$
$$= \int_{-\infty}^{y} \left(\int_{-\infty}^{x} f_Y(v|u) f_X(u)\, du \right) dv$$

因此, 一般地

$$P(X \in A, Y \in B) = \int_{A} \left(\int_{B} f_Y(v|u)\, dv \right) f_X(u)\, du$$
$$= \int_{B} \left(\int_{A} f_Y(v|u) f_X(u)\, du \right) dv$$

将密度函数换为离散概率函数, 积分换为求和, 就得到前面学过的全概率公式.

Bayes 法则

两次运用乘法公式得到

$$f(x,y) = f_Y(y|x) \cdot f_X(x) = f_X(x|y) \cdot f_Y(y)$$

借助于全概率公式我们有下面的 **Bayes 法则**

定理 假设随机变量 X, Y 的联合分布具有密度函数 $f(x,y)$. $f_X(x)$ 和 $f_Y(y)$ 分别是边际密度函数. 如果 $f_Y(y) > 0$, 那么

$$f(x|y) = \frac{f_Y(y|x) \cdot f_X(x)}{f_Y(y)} = \frac{f_Y(y|x) \cdot f_X(x)}{\int_{-\infty}^{\infty} f_Y(y|x) \cdot f_X(x)\, dx}.$$

同样地如果 $f_X(x) > 0$, 那么

$$f(y|x) = \frac{f_X(x|y) \cdot f_Y(y)}{f_X(x)} = \frac{f_X(x|y) \cdot f_Y(y)}{\int_{-\infty}^{\infty} f_X(x|y) \cdot f_Y(y)\, dy}.$$

理论上, 如果已知 X, Y 的联合密度, 那么边际密度以及条件密度都可以求出. 但实际应用当中联合分布往往不是容易得到的, 相反, 关于某一随机变量的条件分布比较容易得到, 在这一情形下 Bayes 法则为估计关于其他变量的条件分布提供了一个途径, 即为 Bayes 推断.

例题 从 $[0,1]$ 按均匀分布取出一点 X, 观测到 $X = x \in (0,1)$ 之后, 再以 $[x,1]$ 上的均匀分布取出另一个点 Y. 确定关于 Y 的边际密度函数, 以及 X 在给定 $Y = y$ 时的条件分布密度.

- 由题设, X 的密度函数为

$$f_X(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$$

当 $X = x$ 时 Y 的条件分布密度为

$$f_Y(y|x) = \begin{cases} \frac{1}{1-x}, & x < y < 1 \\ 0, & \text{其他} \end{cases}$$

由乘法公式 X, Y 的联合密度函数为

$$f(x,y)=\begin{cases}\frac{1}{1-x}, & 0<x<y<1\\ 0, & \text{其他}\end{cases}.$$

因此当 $y\in(0,1)$, 关于 Y 的边际密度函数为

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)\,dx \\ &= \int_0^y \frac{1}{1-x}dx \\ &= -\ln(1-y) \end{aligned}$$

当 $y\notin(0,1)$ 时 $f_Y(y)=0$. 为了求得 $f_X(x|y)$, 我们运用 Bayes 法则得到

$$f_X(x|y)=\frac{f_Y(y|x)\cdot f_X(x)}{f_Y(y)}$$

$$= \begin{cases} \frac{-1}{(1-x)\ln(1-y)}, & 0 < x < y < 1 \\ 0, & \text{其他} \end{cases}$$

作为条件分布的另一应用我们有

定理 随机变量 X, Y 是独立的当且仅当

$$f_X(x|y) = f_X(x)$$

对 $\forall x \in \mathbb{R}$ 以及满足 $f_Y(y) > 0$ 的 y 成立. 对 $f_Y(y|x)$ 也可以叙述同样的结论.

■ 这是因为随机变量 X, Y 是独立的当且仅当其联合密度函数可以写成各个边际分布（密度）乘积的形式.

注记 本课讲述的随机变量的独立性, 条件分布等都可以推广到多元情形, 在此不再累述.