微元法与随机变量的变换

微元法

假设 X 是一个（连续型）随机变量，其概率密度函数为 f_X 。我们知道，对于事件 $\{x < X < \Delta x + x\}$，其概率

$$P(x < X < \Delta x + x) = \int_x^{\Delta x + x} f_X(u)\,\mathrm{d}u = f_X(\xi) \cdot |\Delta x|\,, \quad \xi \in (x, \Delta x + x) \tag{1}$$

令$\Delta x \ \to 0$，我们可以近似认为

$$P(X = x) = f_X(x)\,|\mathrm{d}x| \tag{2}$$

即 X 在点 x 的概率等于该点的密度函数乘x的微元。这个式子的几何意义是容易理解的。而如果考虑 x 在某一个范围内的概率，相当于对左侧求和，也即相当于对右侧做积分，回到了公式（1）.

微元法可以便捷的用于分析随机变量与求解密度函数。随机变量的变换这一知识点就可以用微元法进行推导得出。

随机变量的变换

问题

假设 X, Y 是两个随机变量（向量），其概率密度函数分别为 f_X 和 f_Y 。其中随机变量 X 及其概率密度函数 f_X 是已知的，$Y = g(X)$ 是通过X进行变换 g（已知）得到的随机变量。一个自然的问题是：Y 的概率密度函数 f_Y 如何计算?

如：已知 X 的概率密度函数，计算 $X^2, X+1, 1/X, (X-\bar{X})^2$ 的概率密度函数。

方法

由于**变换前后，同样的事件发生的概率是不变的，只是表示方式不同。**故对于事件 $A = \{x < X < \Delta x + x\}$，我们假设 g 是连续函数，则其变换之后用 Y 表示应为 $A = \{y < Y < \Delta y + y\}$, 其中 $y = g(x)$.

于是就有

$$P(A) = \int_x^{\Delta x + x} f_X(u)\,\mathrm{d}u = \int_y^{\Delta y + y} f_Y(v)\,\mathrm{d}v$$

令 $\Delta x \ \to 0$，则 $\Delta y \to 0$ ......发现是似曾相识的过程对不对？再继续做下去就是复杂版的微元法的推导了。但这没必要，我们直接用微元法考虑问题：

$$P(X = x) = f_X(x)\,|\mathrm{d}x| = f_Y(y)\,|\mathrm{d}y| = P(Y = y)\,, \quad y = g(x)$$

我们再假设 g 是可逆函数[1]，其逆函数为 h，则 $x = h(y)$，于是

$$\begin{aligned} f_Y(y)\,|\mathrm{d}y| &= f_X(x)\,|\mathrm{d}x| \\ &= f_X(h(y))\,|\mathrm{d}(h(y))| \\ &= f_X(h(y))\,|h'(y)|\,|\mathrm{d}y| \end{aligned} \tag{3}$$

推出

$$f_Y(y) = f_X(h(y))\,|h'(y)| \tag{4}$$

对于多元函数/随机向量，$h'(y)$ 即 Jacobi行列式 $\frac{\partial \vec{x}}{\partial \vec{y}}$，故公式（4）也可写作

$$f_Y(y) = f_X(h(y))\,|\frac{\partial \vec{x}}{\partial \vec{y}}| \tag{5}$$

可以简单记为公式（3）两边除掉 $|\partial y|$

例子

1. 已知 $X(>0)$ 的概率密度函数，计算 X^2 的概率密度函数

> 因为 $Y = X^2$ ，故 $X = \sqrt{Y}$ ，因此
> $$\begin{aligned} f_{X^2}(y) &= f_X(\sqrt{y})(\sqrt{y}') \\ &= f_X(\sqrt{y})/(2 * \sqrt{y}') \qquad ,y > 0 \end{aligned}$$

2. 已知 X 的概率密度函数，计算 $X+1$ 的概率密度函数

> 因为 $Y = X + 1$ ，故 $X = Y - 1$ ，因此
> $$\begin{aligned} f_{X+1}(y) &= f_X(y-1)((y-1)') \\ &= f_X(y-1) \end{aligned}$$

3. 已知随机向量 $X = (X_1, X_2)$ 的概率密度函数，计算 $Y = X_1 + X_2$ 的概率密度函数

> 由于 Y 只有1维，而 X 却有2维，因此 X 到 Y 的映射显然是不可逆的，无法直接使用公式（5）。故我们可以构造随机向量 $\tilde{Y} = (Y, X_1)$ ，对 X 和 $\tilde{Y}$ 使用公式（5）可得：
> $$f_{\tilde{Y}}(y, x_1) = f_X(h(y, x_1))\,\frac{\partial(x_1, x_2)}{\partial(y, x_1)}$$
> 其中
> $$\begin{aligned} h(y, x_1) &= (x_1, x_2) = (x_1, y - x_1) \\ \frac{\partial(x_1, x_2)}{\partial(y, x_1)} &= \frac{\partial(x_1, y - x_1)}{\partial(y, x_1)} \\ &= |\begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix}| \\ &= 1 \end{aligned}$$
> 故
> $$\begin{aligned} f_{\tilde{Y}}(y, x_1) &= f_X(h(y, x_1))\,\frac{\partial(x_1, x_2)}{\partial(y, x_1)} \\ &= f_X(x_1, y - x_1) \end{aligned}$$
> 那么 Y 的分布就是 $\tilde{Y}$ 的边缘分布，所以
> $$f_Y(y) = \int_{dom(X_1)} f_X(x_1, y - x_1)\,\mathrm{d}x_1$$

4. 已知随机向量 $X = (X_1, X_2)$ 的概率密度函数，计算 $Y = X_1 - X_2$ 的概率密度函数

> 类似上题，同样方法计算可得
> $$h(y, x_1) = (x_1, y + x_1)$$

> $$\frac{\partial(x_1,x_2)}{\partial(y,x_1)}=1$$
> $$f_{\tilde{Y}}(y,x_1)=f_X(x_1,y+x_1)$$
> $$f_Y(y)=\int_{dom(X_1)}f_X(x_1,y+x_1)\,\mathrm{d}x_1$$

5. 已知随机向量 $X=(X_1,X_2)$ 的概率密度函数，计算 $Y=X_1X_2$ 的概率密度函数

> 类似上题，同样方法计算可得
>
> $$h(y,x_1)=(x_1,y/x_1)$$
> $$\frac{\partial(x_1,x_2)}{\partial(y,x_1)}=1/x_1$$
> $$f_{\tilde{Y}}(y,x_1)=f_X(x_1,y/x_1)/|x_1|$$
> $$f_Y(y)=\int_{dom(X_1)}f_X(x_1,y/x_1)/|x_1|\,\mathrm{d}x_1$$

总结

微元法——公式（2）

随机变量的变换——公式（3）（4）（5）

拓展

如果 g 不可逆怎么办?

回到我们思维的起点：**变换前后，同样的事件发生的概率是不变的，只是表示方式不同。**那么如果 g 在点 x 处不可逆，设 $g^{-1}(x)=\{y_i\}_{i\in I}$ ，则

$$P(X=x)=f_X(x)\,|\mathrm{d}x|$$
$$P(Y=y_i)=f_Y(y_i)\,|\mathrm{d}y_i|$$
$$P(Y=y_1\ or\ y_2\ or\ \dots)=\sum_{i\in I}f_Y(y_i)\,|\mathrm{d}y_i|$$

设在 y_i 处的逆变换为 h_i，即 $x=h_i(y_i)$ ，则应有

$$f_Y(y)=\sum_{i\in I}f_X(h_i(y_i))\,|h_i'(y_i)| \tag{6}$$

不过实际中如果真的出现了不可逆的情况，除了直接应用公式（6），**更自然的想法是将X的定义域限制为若干个可逆的部分，利用可逆的情况分别解决后再合并。**这两种方法其实是一回事。

我们尝试用新的想法求解之前我们已经解决的问题：

> 已知独立同分布的随机变量 X_i 的概率密度函数，计算 $Y=X_1+X_2$ 的概率密度函数。

> 令 $y=g(x_1,x_2)$， 本身 g 不可逆。但我们考虑在 $X_1=x_1$ 时，$y=g_{x_1}(x_2)$，有反函数 $x_2=h_{x_1}(y)=y-x_1$ ，
>
> 那么根据公式（6），我们有

> $$
> \begin{aligned}
> f_{X_1+X_2}(y) &= \sum_{x_1 \in dom(X_1)} f_{X_2|X_1=x_1}(h_{x_1}(y))\,|h'_{x_1}(y)| \\
> &= \int_{dom(X_1)} f_{X_1,X_2}(x_1, y-x_1)\,\mathrm{d}x_1
> \end{aligned} \tag{7}
> $$

不过这个方法用于记忆公式（7）还行，实际如果用于计算还是太不严谨了，仅作为不严谨的拓展思考......

[1]注意到对于一元连续函数，可逆与严格单调是等价的。因此在一维的情况下也可以直接假设 g 严格单调。