Lecture 11: Examples of 2-Dimensional Random Variable Function

【例】(顺序统计量, order statistics) 设 X, Y 相互独立，分布函数分别为 $F_X(x)$ 和 $F_Y(y)$。令 $M=\max\{X,Y\}$，$N=\min\{X,Y\}$，现考虑 M, N 的分布。

$$
\begin{aligned}
F_M(z) &= P(M\le z)=P(\max\{X,Y\}\le z)=P(X\le z, Y\le z)\overset{\text{独立性}}{=}P(X\le z)P(Y\le z)=F_X(z)F_Y(z)\\
F_N(z) &= P(N\le z)=P(\min\{X,Y\}\le z)=1-P(X>z,Y>z)=1-(1-F_X(z))(1-F_Y(z))
\end{aligned}
$$

注：(1) 设 X, Y 只取整数，则 $P(M=n)=P(M\le n)-P(M\le n-1)=F_M(n)-F_M(n-1)$
。

(2) 上述结果可以推广到 n 个独立随机变量 $X_1,\cdots,X_n$ 的情形，此时有

$$
\begin{aligned}
F_M(z) &= \prod_{k=1}^{n} F_{X_i}(z)\\
F_N(z) &= 1-\prod_{k=1}^{n}(1-F_{X_i}(z))
\end{aligned}
$$

【例】(和的分布) 令 $Z=X+Y$，考虑下列情形中 Z 的分布：

- X, Y 相互独立，且取值均为非负整数。此时 X, Y 显然均为离散型随机变量，记 $P(X=k)=p_k$，$P(Y-k)=q_k$，$k=0,1,2,\cdots$。
- (X,Y) 的密度函数为 $p(x,y)$。

对于离散情形：

$$
P(Z=n)=P(X+Y=n)=\sum_{k=0}^{n}P(X=k,Y=n-k)=\sum_{k=0}^{n}P(X=k)P(Y=n-k)=\sum_{k=0}^{n}p_kq_{n-k}
$$

上式称为 (离散) 卷积 (convolution) 公式。

对于连续情形：

$$
\begin{aligned}
F_Z(z)=P(X+Y\le z) &= \iint_{\{(x,y):x+y\le z\}} p(x,y)dxdy\\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{z-x} p(x,y)dydx\\
&\overset{y=v-x}{=} \int_{-\infty}^{+\infty}\int_{-\infty}^{z} p(x,v-x)dvdx\\
&\overset{\text{换序}}{=} \int_{-\infty}^{z}\left(\int_{-\infty}^{+\infty} p(x,v-x)dx\right)dv
\end{aligned}
$$

因为 $F_Z(z)=\int_{-\infty}^{z}p_Z(u)du$，和上式对比，我们发现括号内的部分正好是 Z 的密度函数。

$$
p_Z(v)=\int_{-\infty}^{+\infty}p(x,v-x)dx
$$

上式称为 (连续) 卷积公式。

进一步地，若 X, Y 独立，则 $p(x,y)=p_X(x)p_Y(y)$，因此 Z 的密度为

$$
p_Z(v)=\int_{-\infty}^{+\infty}p_X(x)p_Y(v-x)dx
$$

【例题】设 $X\sim B(n_1,p),Y\sim B(n_2,p)$，且 X,Y 独立，求 $X+Y$ 的分布。

> 解：$P(X=k)=\binom{n_1}{k}p^k(1-p)^{n_1-k},k=0,\cdots,n_1$，
> $P(Y=k)=\binom{n_2}{k}p^k(1-p)^{n_2-k},k=0,\cdots n_2$。
>
> 令 $Z=X+Y$，则
>
> $$\begin{aligned}P(Z=n)&=\sum_{k=0}^{n}P(X=k,Y=n-k)=\sum_{k=0}^{n}P(X=k)P(Y=n-k)\\&=\sum_{k=\max\{0,n-n_2\}}^{\min\{n,n_1\}}\binom{n_1}{k}p^k(1-p)^{n_1-k}\binom{n_2}{n-k}p^{n-k}(1-p)^{n_2-n+k}\\&=p^n(1-p)^{n_1+n_2-n}\sum_{k=\max\{0,n-n_2\}}^{\min\{n,n_1\}}\binom{n_1}{k}\binom{n_2}{n-k}\\&=p^n(1-p)^{n_1+n_2-n}\binom{n_1+n_2}{n}\end{aligned}$$
>
> 因此 $Z\sim B(n_1+n_2,p)$。
>
> 在多次实验 p 相同的情况下，这个结论是容易理解的：先做 n_1 次再做 n_2 次和一共做 n_1+n_2 次是一样的。

注：上述结论可以推广到 n 个变量的情形。且类似可证对于独立的泊松分布 $X_k\sim P(\lambda_k),k=1,\cdots n$，有 $X=\sum_{k=1}^{n}X_k\sim P(\sum_{k=1}^{n}\lambda_k)$。

【例题】设 $X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2)$ 且 X,Y 相互独立，求 $Z=X+Y$ 的分布。

> 解：$p_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},p_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$。
>
> 正态分布变量的取值均为 $\mathbb{R}$，因此我们直接使用卷积公式：
>
> $$p_Z(z)=\int_{-\infty}^{+\infty}p_X(x)p_Y(z-x)dx=\frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty}e^{-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{z-x-\mu_2}{\sigma_2}\right)^2\right]}dx$$
>
> 遇到这种 e 上带指数，积分区域是 $\mathbb{R}$ 的积分，常见的处理手法是将其凑成平方的形式，然后套用高斯积分。因此我们现在希望中括号中两个平方式的交叉项消掉。这里给出一个处理技巧：令 $x=y+a$，a 为待定的系数，则
>
> $$p_Z(z)=\frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty}e^{-\frac{1}{2}\left[\left(\frac{y+a-\mu_1}{\sigma_1}\right)^2+\left(\frac{y+a+\mu_2-z}{\sigma_2}\right)^2\right]}dy$$
>
> 令交叉项为零，即
>
> $$\frac{y}{\sigma_1}\cdot\frac{a-\mu_1}{\sigma_1}+\frac{y}{\sigma_2}\cdot\frac{a+\mu_2-z}{\sigma_2}=0$$
>
> 解得
>
> $$a=\frac{\mu_1\sigma_2^2-\mu_2\sigma_1^2+z\sigma_1^2}{\sigma_1^2+\sigma_2^2}$$
>
> 带回原式，有

> $$\begin{aligned} p_Z(z) &= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left[\left(\frac{y}{\sigma_1}\right)^2+\left(\frac{y}{\sigma_2}\right)^2+\frac{(z-\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}\right]}dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2}e^{-\frac{(z-\mu_1-\mu_2)^2}{2\left(\sigma_1^2+\sigma_2^2\right)}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}\cdot\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}y^2}dy \\ &\overset{v=\sqrt{\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}}y}{=} \frac{1}{2\pi\sigma_1\sigma_2}e^{-\frac{(z-\mu_1-\mu_2)^2}{2\left(\sigma_1^2+\sigma_2^2\right)}}\int_{-\infty}^{+\infty} e^{-\frac{v^2}{2}}d\left(\sqrt{\frac{\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2}}y\right) \\ &= \frac{1}{\sqrt{2\pi}\cdot\sqrt{\sigma_1^2+\sigma_2^2}}e^{-\frac{(z-\mu_1-\mu_2)^2}{2\left(\sigma_1^2+\sigma_2^2\right)}} \end{aligned}$$
>
> 因此 $Z\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$。

【例题】(习题 3.25) 设 $X\sim U[0,2],Y\sim U[0,1]$ 且 X,Y 独立，求 $Z=X+Y$ 的密度函数。

> 解：该题随机变量取值不是 $\mathbb{R}$，因此不能套用卷积公式，要从定义出法求解。
>
> $p_X(x)=\begin{cases}\frac{1}{2} & ,0\le x\le 2\\ 0 & ,otherwise\end{cases}$，$p_Y(y)=\begin{cases}1 & ,0\le y\le 1\\ 0 & ,otherwise\end{cases}$。$(X,Y)$ 的有效区域是一个长方形。
>
> $$F_Z(z)=P(X+Y\le z)=\iint_{\{(x,y):x+y\le z\}}p(x,y)dxdy$$
>
> 在有效区域内，有 $p(x,y)=p_X(x)p_Y(y)=\frac{1}{2}$。考虑拿直线 $x+y\le z$ 滑过平面，看直线左侧。
>
> 容易看出 $z\le 0$ 时 $F_Z(z)=0$；$z\ge 3$ 时 $F_Z(z)=1$。剩下的几种情形需要仔细考虑：
>
> - $0<z<1$，此时获得的是一个三角形，$F_Z(z)=\frac{1}{2}\cdot\frac{z^2}{2}=\frac{z^2}{4},p_Z(z)=F_Z'(z)=\frac{z}{2}$。
> - $1\le z<2$，此时获得的是一个梯形，$F_Z(z)=\frac{1}{2}\cdot\frac{z-1+z}{2}=\frac{z}{2}-\frac{1}{4}$，$p_Z(z)=\frac{1}{2}$。
> - $2\le z<3$，此时获得的是矩形减去一个三角形，
> $F_Z(z)=\frac{1}{2}\cdot(2-\frac{1}{2}(3-z)^2)=1-\frac{1}{4}(3-z)^2$，$p_Z(z)=\frac{1}{2}(3-z)$。