Lecture 07: Distribution of Random Variable Function

设 X 是一个随机变量，函数 $g(x):\mathbb{R}\to\mathbb{R}$。构造随机变量 Y，当 $X=x$ 时，$Y=g(x)$，称 Y 是 X 的函数，记为 $Y=g(X)$。在已知 X 的分布时，我们希望求出 Y 的分布。

Discrete Situation

设 X 是一个离散型随机变量，其分布律为

x_1	x_2	$\cdots$	x_k	$\cdots$
p_1	p_2	$\cdots$	p_k	$\cdots$

那么 $Y\in\{g(x_k)\}_{k=1}^n$，去重后可以写成 $\{y_1,y_2,\cdots,y_k,\cdots\}$，显然 Y 也是离散型随机变量。

考虑 $P(Y=y)=P(g(X)=y)=P(x\in\{x|g(x)=y\})$，由可列可加性可知 $P(Y=y)=\sum_{x:g(x)=y}P(X=x)$。

【例】X 是离散型随机变量，$P(X=0)=0$，对于任意 $k\in\mathbb{N}$，$P(X=k)=P(X=-k)=p^k$。求 $Y=X^2$ 的分布律。

> 解：首先解出 p：$\sum_{k=1}^{\infty}P(X=\pm k)=\sum_{k=1}^{\infty}2P(X=k)=2\sum_{k=1}^{\infty}p^k=1$，解得 $p=\frac{1}{3}$。
>
> 显然 Y 的取值只能是正整数。对于任意 $n\in\mathbb{N}$，有
>
> $$P(Y=n)=P(X^2=n)=P(X=\pm\sqrt{n})=\begin{cases}2(\frac{1}{3})^{\sqrt{n}} & ,\sqrt{n}\in\mathbb{N}\\ 0 & ,otherwise\end{cases}$$

注：可以看出即使 $P(X=k)\neq P(X=-k)$，只要 $P(X=\pm k)=2p^k$，Y 的分布就是上述结果。也就是说，随机变量的分布和随机变量函数的分布并不是一一映射。

Continuous Situation

设 X 为连续型随机变量，$y=g(x)$ 为连续函数，$Y=g(X)$，一般地，可如下求 Y 的分布函数 $F_Y(y)$ 和密度函数 $P_Y(y)$：

$$F_Y(y)=P(Y\leq y)=P(g(X)\leq y)=P(x\in\{x|g(x)\leq y\})=\int_{x:g(x)\leq y}p_X(x)dx$$

最后的积分式含参数 y，结果是关于 y 的表达式。若 $F_Y(y)$ 可导，则 $p_Y(y)=F'_Y(y)$。

【例】$X\sim N(0,1)$，求 $Y=X^2$ 的分布。

> 解：注意到 $Y\geq 0$，所以对于任意 $y<0$，$F_y(y)=0$。
>
> 对于 $y\geq 0$，有
>
> $$F_Y(y)=P(Y\leq y)=P(X^2\leq y)=P(-\sqrt{y}\leq X\leq\sqrt{y})=F_X(\sqrt{y})-F_X(-\sqrt{y})$$
>
> 欲求 $p_Y(y)$，我们要对 $F_Y(y)$ 求导，注意使用链式法则：

$$\begin{aligned}P_Y(y) &= F_Y'(y) = F_X'(\sqrt{y}) - F_X'(-\sqrt{y}) \\ &= p_X(\sqrt{y})\frac{1}{2\sqrt{y}} - p_X(-\sqrt{y})\frac{-1}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{y}}p_X(\sqrt{y}) = \frac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}}\end{aligned}$$

> 综上，
>
> $$p_Y(y) = \begin{cases}\frac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}} & ,y \geq 0 \\ 0 & ,y < 0\end{cases}$$

注：$Y = X^2$ 称为服从一个自由度的 χ^2 分布。χ^2 分布是统计学中的一个重要分布。

上述做法需要先计算 $F_Y(y)$ 再求导计算 $p_Y(y)$。若 $y = g(x)$ 严格单调，则对 $g(x)$ 加以一些可导条件，可之间诶计算密度函数 $p_Y(y)$。

$\boxed{\text{Theorem 7.1}}$ 设 X 为连续型随机变量，密度函数为 $p_X(x)$。设 $y = g(x)$ 严格单调处处可导且恒有 $g'(x) > 0$ 或 $g'(x) < 0$，则 $Y = g(X)$ 也为连续型随机变量，且

$$p_Y(y) = \begin{cases}p_X(g^{-1}(y)) \cdot \left|(g^{-1}(y))'\right| & ,Y\text{可取到}y \\ 0 & ,Y\text{取不到}y\end{cases}$$

其中 $g^{-1}(y)$ 是 $g(x)$ 的反函数。

注：当 $g(x)$ 单调且 $g'(x_0)$ 存在非零，可以证明 $g^{-1}(y)$ 在 $y_0 = g(x_0)$ 处可导。但即使严格单调，$g'(x)$ 仍可能在某个 x_0 取到 0 (如 $f(x) = x^3$ 的 $x = 0$ 处)，此时 $g^{-1}(y)$ 在 $y_0 = g(x_0)$ 处的可导性存在问题，故要求 $g'(x)$ 不能取 0。

> 证明：这里仅证单调递增的情况，单调递减大同小异 (最终使结论添上绝对值)：
>
> $$F_Y(y_0) = P(Y \leq y_0) = P(g(X) \leq y_0) = \int_{-\infty}^{g^{-1}(y_0)} p_X(x)dx$$
>
> 作变量替换 $x = g^{-1}(y)$：
>
> $$F_Y(y_0) = \int_{-\infty}^{y_0} p_X(g^{-1}(y))dg^{-1}(y) = \int_{-\infty}^{y_0} p_X(g^{-1}(y))(g^{-1}(y))'dy$$
>
> 根据微积分基本定理，
>
> $$p_Y(y_0) = F_Y'(y_0) = p_X(g^{-1}(y_0))(g^{-1}(y_0))'$$

【例】$X \sim N(\mu, \sigma^2)$，$Z = \frac{X-\mu}{\sigma}$，即 $g(x) = \frac{x-\mu}{\sigma}$，$Z = g(X)$。求 Z 的分布。

> 解：显然 $g(x)$ 单调递增，且 $g'(x) = \frac{1}{\sigma} > 0$。$g^{-1}(z) = \sigma z + \mu$。
>
> $$\begin{aligned}p_Z(z) &= p_X(g^{-1}(z))\left|(g^{-1}(z))'\right| = p_X(\sigma z + \mu) \cdot \sigma \\ &= \sigma \cdot \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\sigma z+\mu-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\end{aligned}$$
>
> 即 $Z \sim N(0,1)$。

【例】若 X 服从 $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ 上的均匀分布，$Y = \tan X$，求 Y 的分布。

> 解：$y = g(x) = \tan x$，$x = g^{-1}(y) = \arctan y$，$(\arctan y)' = \frac{1}{1+y^2}$。于是

> $$\begin{aligned} p_Y(y) &= p_X(\arctan y) \cdot \left| \frac{1}{1+y^2} \right| \\ &= \frac{1}{\pi} \cdot \frac{1}{1+y^2} \end{aligned}$$

注：Y 的分布称为柯西分布。