Distribution functions

Remark 1 *$P(a \le X < b) = F(b) - F(a)$, $P(a < X < b) = F(b) - F(a+0)$, $P(a \le X \le b) = F(b+0) - F(a)$, $P(X = a) = F(a+0) - F(a)$ (last equality means, if F is continuous, each point has probability 0).*

Definition 1 *X is an **absolutely continuous random variable**, if the distribution function F can be written as $F(z) = \int\limits_{-\infty}^{z} f(t)dt$, where $f(t)$ is the **density function** of X.*

The density function is non-negative, and $\int\limits_{-\infty}^{\infty} f(t)dt = 1$. Moreover, the integral function of f, which meets the previously mentioned 2 properties, is a distribution function.

Example 1 *If $X \sim U(a,b)$, then $f(x) = \begin{cases} 0 & x \le a \\ \frac{1}{b-a} & a < x \le b \\ 0 & x > b \end{cases}$.*

Example 2 *If $X \sim Exp(\lambda)$, then $f(x) = \begin{cases} 0 & x \le 0 \\ \lambda e^{-\lambda x} & 0 < x \end{cases}$.*

Properties of density function:

1. if F is absolutely continuous, then the derivative of F is f, i.e., $F' = f$, where F is differentiable
2. $P(a < X < b) = \int_a^b f(t)dt \approx f(a)(b-a)$

Proposition 1 *Assume that $g : \mathbb{R} \to \mathbb{R}$ is a function, and X is an absolutely continuous random variable. Then $g(X)$ is also a random variable, but not necessarily absolutely continuous ($g(x) = 1$ function gives a counterexample, because it gives a degenerate random variable for any X).*

Example 3 *If g is a linear function - let us say $g(x) = ax + b$ -, and F_X is the distribution function of random variable X, then $F_{aX+b}(z) = F_X\left(\frac{z-b}{a}\right)$ if $a > 0$, and $F_{aX+b}(z) = 1 - F_X\left(\frac{z-b}{a}\right)$ if $a < 0$. In the absolutely continuous case, the density of random variable $aX + b$ is $f_{aX+b}(z) = \frac{1}{|a|} f_X\left(\frac{z-b}{a}\right)$.*

Proposition 2 *More generally, if g is strictly monotonic, continuously differentiable and $g' \ne 0$, then*

$$f_{g(X)}(z) = \frac{f_X(g^{-1}(z))}{|g'(g^{-1}(z))|}.$$

Definition 2 *A random variable X has a **standard normal distribution**, if the density function of X is $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$. Notation: $X \sim N(0,1)$.*

Proposition 3 *The above defined function is a density function.*

Proof: For all $x \in \mathbb{R}$ $f(x) > 0$, thus we have to count the integral on $\mathbb{R}$ (it has to be 1).

$$\left(\int\limits_{-\infty}^{\infty} f(x)dx\right)^2 = \int\limits_{-\infty}^{\infty} f(x)dx \int\limits_{-\infty}^{\infty} f(y)dy = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\left(\frac{x^2}{2}+\frac{y^2}{2}\right)} dxdy =$$

$$= \int\limits_{0}^{\infty}\int\limits_{0}^{2\pi} \frac{1}{2\pi} r e^{-\frac{r^2}{2}} d\varphi dr = \int\limits_{0}^{\infty} r e^{-\frac{r^2}{2}} dr = \left[-e^{-\frac{r^2}{2}}\right]_{r=0}^{\infty} = 1.$$

□

Definition 3 *The expected value of absolutely continuous random variables is $E(X) = \int_{-\infty}^{\infty} z f(z)dz$. Our intuition is that we take the partition of $\mathbb{R}$, where the partitions are small intervals of length $\delta > 0$. Thus - as in the discrete case - $E(X) = \sum\limits_{z} zP(z < X < z+\delta) \approx \sum z\delta f(z) \approx \int zf(z)dz$.*

Properties as $E(aX+b) = aE(X)+b$ and $E(X+Y) = E(X)+E(Y)$ remain.

Example 4 *If* $X \sim U(a,b)$*, then* $E(X) = \int_a^b \frac{y}{b-a} dy = \left[\frac{y^2}{2(b-a)}\right]_{y=a}^{b} = \frac{a+b}{2}$.

Example 5 *If* $X \sim Exp(\lambda)$*, then (using integration by parts)* $E(X) = \int_0^\infty \lambda y e^{-\lambda y} dy = \left[-ye^{-\lambda y}\right]_{y=0}^{\infty} + \int_0^\infty e^{-\lambda y} dy = \frac{1}{\lambda}$.

Example 6 *If* $X \sim N(0,1)$*, then* $E(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \left[-\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\right]_{x=-\infty}^{\infty} = 0$.

Proposition 4 *For a random variable* X*, the expected value can be calculated using the distribution function as follows.* $E(X) = \int\limits_0^\infty (1-F_X(y))dy - \int\limits_{-\infty}^0 F_X(y)dy$*. Thus, if* $X \geq 0$ *is a nonnegative random variable, then* $E(X) = \int\limits_0^\infty (1-F_X(y))dy$.

Definition 4 *Suppose that* m *is an arbitrary and* σ *is a positive real number. If* X *is a standard normal variable, then the random variable* $Y = \sigma X + m$ *has* ***normal distribution*** *with parameters* (m,σ)*. The density function can be calculated using the formula in 3:* $f_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$.

Proposition 5 *Let* $f(x)$ *be the density function of absolutely continuous random variable* X*,* $g(x): \mathbb{R} \to \mathbb{R}$ *an arbitrary function, and* $Y = g(X)$*. Then* $E(Y) = \int\limits_{-\infty}^{\infty} g(y) f_X(y) dy$.

Remark 2 *The definition of variance and properties remain in the absolutely continuous case, i.e.,* $Var(X) = E\left((X-E(X))^2\right)$ *etc. Also remark that for an arbitrary* $k > 0$ *integer,* $E(X^k) = \int_{-\infty}^{\infty} z^k f(z) dz$.