5. CENSORING AND TRUNCATION

A common problem in the survival data analysis is censoring and truncation. Roughly speaking, a censored observation contains only partial information about the object of interest. The most common case of censoring is right censoring, which is considered here. We can observe the event under study for a given individual only if we can observe the individual at this moment. In the case of censoring we can only observe that this event is not true yet.

Let the observations consist of $(T_{i1}, T_{i2}, \Delta_{i1}, \Delta_{i2}), i = 1, \cdots, n,$ where $T_{ij} = min\{D_{ij}, Y_{ij}\},\ i = 1, \cdots, n; j = 1, 2$ are the observed times and $\Delta_{ij} = \delta(Y_{ij} - D_{ij})$ are indicators if the event has happened or not, δ is the Heaviside function. Here are $D_{ij}, i = 1, \cdots, n; j = 1, 2$ the moments of death and $Y_{ij}, i = 1, \cdots, n; j = 1, 2$ are independent and identically distributed moments of censoring with cumulative density function $G(x)$.

Therefore, there are four cases for the survival probability:

$$P(T_1 > t_1, T_2 > t_2, d_1 = 1, d_2 = 1) = \int_0^{t_1}\int_0^{t_2}\int_{t_1}^{\infty}\int_{t_2}^{\infty} f(x_1, x_2)g_1(y_1)g_2(y_2)dy_2dy_1dx_2dx_1$$

$$= \int_0^{t_1}\int_0^{t_2} f(x_1, x_2)(1 - G_1(x_1))(1 - G_2(x_2))dx_2dx_1,$$

where $g_i(x)$ and $G_i(x)$ are respectively the density and cumulative density of the censoring times, d_1 and d_2 are indicators if the observations T_1 and T_2 are censored in the meaning mentioned above.

By analogy

$$P(T_1 > t_1, T_2 > t_2, d_1 = 1, d_2 = 0) = \int_0^{t_1}\int_{t_1}^{\infty}\int_0^{t_2}\int_{y_2}^{\infty} f(x_1, x_2)g_1(y_1)g_2(y_2)dx_2dy_2dy_1dx_1$$

$$= \int_0^{t_1}\int_0^{t_2} \frac{\partial}{\partial x_1}(1 - F(x_1, y_2))(1 - G_1(x_1))g_2(y_2)dy_2dx_1$$

and

$$P(T_1 > t_1, T_2 > t_2, d_1 = 0, d_2 = 1) = \int_0^{t_1}\int_0^{t_2} \frac{\partial}{\partial x_2}(1-F(y_1, x_2))(1-G_2(x_1))g_1(y_2)dy_1dx_2.$$

In the last case

$$P(T_1 > t_1, T_2 > t_2, d_1 = 0, d_2 = 0) = \int_0^{t_1}\int_0^{t_2}\int_{t_1}^{\infty}\int_{t_2}^{\infty} f(x_1, x_2)g_1(y_1)g_2(y_2)dx_2dx_1dy_2dy_1$$

$$= \int_0^{t_1}\int_0^{t_2} S(x_1, x_2)g_1(y_1)g_2(y_2)dy_1dy_2$$