1) $Y_t = g(X_t) + \varepsilon_t \qquad \varepsilon_t \sim N(0,\sigma^2) \quad X_t$ has density f_X

we (falsely) fit the model $Y_t = \beta X_t + \varepsilon_t \quad [\varepsilon_t \sim N(0,\sigma^2)]$

to the data.

(i) $\hat{\beta} = \dfrac{\sum Y_t X_t}{\sum X_t^2}$

(ii) By the continuous mapping theorem we have

$$\hat{\beta} = \frac{\frac{1}{T}\sum Y_t X_t}{\frac{1}{T}\sum X_t^2} \longrightarrow \frac{E(Y_t X_t)}{E(X_t^2)} = \frac{\int g(x)x f(x)dx}{\int x^2 f(x)\,dx} := \beta_0$$

(iii) There are 2 ways to approach this problem.

Either consider the marginal distribution of Y_t for both the correct model and the best fitting linear model or consider the joint distribution of (Y_t, X_t) for both the best fitting and correct model.

In the case of the marginal distributions the true density is:

$$h(y) = \int \varphi[y - g(x)]\, f_X(x)\, dx.$$ standard normal density

the best fitting is:

$$f_\beta(y) = \int \varphi[y - \beta_0 x]\, f_X(x)\, dx.$$

The KL-criterion is then $\mathbb{E}_h\left[\log \frac{f_\beta(y)}{h(y)}\right]$, however ②

it does not give an interesting explicit expression.

Let us, instead, consider the joint distributions.

The true joint distribution is

- $f_{Y|X}(y)\, f_X(x) = \varphi_\varepsilon(y-g(x))\, f_X(x)$

The best fitting distribution is.

- $f_{Y|X}(y)\, f_X(x) = \varphi_\varepsilon(y-\beta_0 x)\, f_X(x)$.

Thus the KL-criterion is

$$\mathbb{E}_g\left\{\log \frac{\varphi_\varepsilon(Y-\beta X)\,\cancel{f_X(X)}}{\varphi_\varepsilon(Y-g(X))\,\cancel{f_X(X)}}\right\}$$

$$= \mathbb{E}\left\{-\frac{1}{2\sigma^2}(Y-\beta X)^2+\frac{1}{2\sigma^2}(Y-g(X))^2\right\}$$

$$= -\frac{1}{2\sigma^2}\int(\beta x-g(x))^2 f(x)\,dx$$

[after some calculation and using that $\mathbb{E}(YX)=\mathbb{E}(g(X)X)$].

This has a rather nice form. It will always be negative (as theory tells us), but the better the approximation the closer it will be to zero.

2) let us suppose that Y_t is a discrete r.v. and X_t are regressors which are known to influence Y_t. If we treat X_t as random, then conditional on X_t we have

$\mathbb{E}(Y_t | X_t) = e^{\beta X_t}$ and $\mathrm{var}(Y_t | X_t) = e^{\beta X_t}[1 + \xi e^{\beta X_t}]$, where $\xi \geq 0$. The estimating equation.

$$G(\beta) = \sum_t' (Y_t - e^{\beta X_t}) X_t = 0$$ is used to estimate β.

We denote this estimate as $\hat{\beta}$.

By the delta method we have

$$\sqrt{T}(\hat{\beta} - \beta) \longmapsto N(0, V_T).$$

(i) $V_T = A_T^{-1} B_T A_T^{-1}$, where

$$A_T = \mathbb{E}\left\{\frac{1}{T} \sum_t' e^{\beta X_t} X_t^2\right\} = \mathbb{E}\{e^{\beta X_t} X_t^2\}$$

$$B_T = \mathrm{var}\left\{\frac{1}{\sqrt{T}} \sum_t (Y_t - e^{\beta X_t}) X_t\right\}$$

$$= \frac{1}{T} \sum_t' \mathbb{E}\{\mathrm{var}(Y_t - e^{\beta X_t} | X_t) X_t^2\}$$

Since $\cancel{\mathrm{var}}\,\mathbb{E}[(Y_t - e^{\beta X_t}) X_t] = 0$

$$= \frac{1}{T} \sum_t' \mathbb{E}\{X_t^2 e^{\beta X_t}[1 + \xi e^{\beta X_t}]\}$$

$$= \mathbb{E}[X_t^2 e^{\beta X_t}\} + \xi \mathbb{E}\{X_t^2 e^{2\beta X_t}\}$$

(ii) To derive explicit expressions for A_T and B_T when X_t are standard normal distributions, we use the identity:

$$\frac{1}{\sqrt{2\pi}}\int x\, e^{\gamma x} e^{-\frac{1}{2}x^2}\,dx = \frac{e^{\gamma^2/2}}{\sqrt{2\pi}}\int x\, e^{-\frac{1}{2}(x-\gamma)^2}\,dx$$

$$= e^{\gamma^2/2}\left\{\frac{1}{\sqrt{2\pi}}\int (x-\gamma)\, e^{-\frac{1}{2}(x-\gamma)^2}\,dx + \frac{\gamma}{\sqrt{2\pi}}\int e^{-\frac{1}{2}(x-\gamma)^2}\,dx\right\}$$

$$= e^{\gamma^2/2}\left[0+\gamma\right] = \gamma\, e^{\gamma^2/2} \qquad (1)$$

$$\frac{1}{\sqrt{2\pi}}\int x^2 e^{\gamma x} e^{-\frac{1}{2}x^2}\,dx =$$

$$e^{\gamma^2/2}\left[\frac{1}{\sqrt{2\pi}}\int (x-\gamma)^2 e^{-\frac{1}{2}(x-\gamma)^2}\,dx + \frac{2\gamma}{\sqrt{2\pi}}\int x\, e^{-\frac{1}{2}(x-\gamma)^2}\,dx - \frac{\gamma^2}{\sqrt{2\pi}}\int e^{-\frac{1}{2}(x-\gamma)^2}\,dx\right]$$

$$= e^{\gamma^2/2}\left[1+\gamma^2\right] \qquad (2).$$

Thus

$$A_T = \int e^{\beta x_t} x^2 \varphi(x)\,dx \overset{(\text{by }(1))}{=} e^{\beta^2/2}(\beta^2+1)$$

$$B_T = E\left[X_t^2 e^{\beta X_t}\right] + \xi\, E\left(X_t^2 e^{2\beta X_t}\right)$$

$$= e^{\beta^2/2}(1+\beta)^2 + \xi\, e^{(2\beta)^2/2}\left[1+(2\beta)\right]^2. \qquad (3)$$

(iii) Estimate B_T with $\frac{1}{T}\sum_t \left\{(y_t - e^{\beta x_t})x_t\right\}^2 = \hat{B}$

Given the estimate of $\hat{\beta}$

using

$$\hat{B} = e^{\hat{\beta}^2/2}\left(1+\hat{\beta}^2\right) + \xi\, e^{(2\hat{\beta})^2/2}\left[1+(2\hat{\beta})^2\right]$$

nowe we can solve for ξ, to obtain an estimate of ξ.

There also exists graphical methods for estimating ξ too.

3) we have the survival likelihood; (6)

$$\mathcal{L}_T(\theta) = \sum_i \delta_i \log f(T_i;\theta) + \sum_i (1-\delta_i) \log \mathcal{F}(Y_i;\theta).$$

(ii) Since the observations are iid taking expectations give

$$\mathrm{E}\{\mathcal{L}_T(\theta)\} = \pi\, \mathrm{E}\{\log f(T_i;\theta)\} + (1-\pi)\, \mathrm{E}\{\log \mathcal{F}(Y_i;\theta)\}$$

since δ_i and (T_i, C_i) are independent.

Differentiating $\mathrm{E}\{\log f(T_i;\theta)\}$ wrt θ it is easily seen that $\dfrac{\partial \mathrm{E}\{\log f(T_i;\theta)\}}{\partial\theta} = 0$ at $\theta = \theta_0$.

We now consider the term $\mathrm{E}\{\log \mathcal{F}(Y_i;\theta)\}$ and show that (usually) the derivative will not equal zero at $\theta=\theta_0$.

First we need the density of Y_i. Since $Y_i = \min(T_i, C_i)$ we see that

$$P\{\min(T_i, C_i) \le y\} = 1 - P\{\min(T_i, C_i) > y\}$$
$$= 1 - \mathcal{F}(y;\theta_0)\, G(y)$$

Thus the density of Y_i is

$$f(y;\theta_0)\, G(y) + \mathcal{F}(y;\theta_0)\, \overbrace{g(y)}^{\text{density of } C}.$$

Using the above we have

$$\mathbb{E}\{\log F(Y_i;\theta)\} = \int [\log F(y;\theta)]\{f(y;\theta_0)G(y) + F(y;\theta_0)g(y)\}\,dy$$

To check whether the above is maximum at θ_0 we diff the above wrt θ to give

$$\frac{\partial}{\partial\theta}\mathbb{E}\{\log F(Y_i;\theta)\} = \int \frac{1}{F(y;\theta)}\frac{\partial F(y;\theta)}{\partial\theta}\{f(y;\theta_0)G(y) + F(y;\theta_0)g(y)\}\,dy$$

At the true parameter θ this gives

$$= \underbrace{\frac{\partial}{\partial\theta}\int \overset{F(y;\theta)}{\cancel{F(y;\theta)}}\, g(y)\,dy}_{=0 \text{ since } \int g(y)dy = 1} + \underbrace{\int \frac{\partial F(y;\theta)}{\partial\theta}\frac{G(y)}{F(y;\theta_0)} f(y;\theta_0)\,dy}.$$

Actually $\neq 0$! Yours truely got it wrong! – we all make mistakes ☹

This term will not $= 0$ (in general).

Thus we have that

$$\left.\frac{d\,\mathbb{E}\{\mathcal{L}_T(\theta)\}}{d\theta}\right|_{\theta=\theta_0} = \pi \underbrace{\left.\frac{d\,\mathbb{E}\{\log f(Y_i;\theta)\}}{\partial\theta}\right|_{\theta=\theta_0}}_{=0} + (1-\pi)\underbrace{\frac{d\,\mathbb{E}\{\log F(Y_i;\theta)\}}{\partial\theta}}_{\text{not necessarily } = 0.}$$

Since the derivative of expectation of the criterion is not zero at the true parameter θ_0, the estimator $\hat{\theta}_T$ is unlikely to consistently estimate θ_0. It will be a biased (in probability) estimator.

Given my huge mistake, thinking that

$$\frac{\partial}{\partial \theta} \int f(y;\theta) g(y)\, dy = 0 \quad (*)$$

[not quite sure what came over me].

the derivation of the information

$$E\left\{ \frac{-\partial^2 \mathcal{L}_T}{\partial \theta^2} \right\} = E\left\{ \frac{\partial \mathcal{L}_T}{\partial \theta} \right\}^2$$

is under the assumption.

You should try and see where this derivation falls apart, when (*) is not true!

So I made a huge mistake but it is good practice for you to understand where things work and do not work.

(ii) Now we consider the special case that $G(x) = F(x;\theta_0)$. In this case the density of $\min(T_i, C_i)$ is $2f(x;\theta_0)F(x;\theta_0)$. Now by substituting this into the calculations in (i) it can be seen that

$$\frac{\partial \mathbb{E}\{\log F(Y_i;\theta)\}}{\partial\theta}\Big\rfloor_{\theta=\theta_0} = 0.$$

Thus in this special case the estimator will be consistent.

(iii). Taking second derivatives of

$$\pi\,\mathbb{E}\{\log f(T_i;\theta)\} + (1-\pi)\,\mathbb{E}\{\log f(Y_i;\theta)\}$$

gives

$$\pi\,\mathbb{E}\left\{-\frac{1}{f(T,\theta)^2}\left(\frac{\partial f}{\partial\theta}\right)^2 + \overbrace{\frac{1}{f(T;\theta)}\frac{\partial^2 f}{\partial\theta^2}}^{=0}\right\}$$

$$+(1-\pi)\,\mathbb{E}\left\{-\frac{1}{F(Y_i;\theta)}\left(\frac{\partial F(Y_i;\theta)}{\partial\theta}\right)^2 + \underbrace{\frac{1}{F(Y_i;\theta)}\frac{\partial^2 F(Y_i;\theta)}{\partial\theta^2}}\right\}$$

Under the assumptions on the distribut given in (ii)

$$\mathbb{E}\left\{\frac{1}{F(Y_i;\theta)}\left(\frac{\partial^2 F(Y_i;\theta)}{\partial\theta^2}\right)\right\} = 0.$$

Thus, if $G(x) = \mathcal{F}(x;\theta_0)$ we have

$$\frac{\partial^2}{\partial\theta^2}\left[\pi\,\mathbb{E}\left[\log f(T;\theta)\right] + (1-\pi)\,\mathbb{E}\left\{\log \mathcal{F}(Y_i;\theta)\right\}\right]$$

$$= \pi\,\mathbb{E}\left\{\frac{-1}{f(T;\theta)^2}\left(\frac{\partial f}{\partial\theta}\right)^2\right\} + (1-\pi)\,\mathbb{E}\left\{\frac{-1}{\mathcal{F}(Y_i;\theta)}\left(\frac{\partial \mathcal{F}(Y;\theta)}{\partial\theta}\right)^2\right\}$$

$$= -\,\mathbb{E}\left[\frac{\partial}{\partial\theta}\left[\delta_i \log f(T;\theta) + (1-\delta_i)\log \mathcal{F}(Y_i;\theta)\right]\right]^2.$$

Thus the limiting distribution of the estimator (based on the survival likelihood) is

$$\sqrt{T}\,(\hat{\theta} - \theta_0) \rightarrow N(0, I_1),$$

where

$$I_1 = \pi\,\mathbb{E}\left\{\frac{1}{f(T;\theta)^2}\left(\frac{\partial f}{\partial\theta}\right)^2\right\} + (1-\pi)\,\mathbb{E}\left\{\frac{1}{\mathcal{F}(Y_i;\theta)}\left(\frac{\partial \mathcal{F}(Y_i;\theta)}{\partial\theta}\right)^2\right\}.$$

(iv) An alternative method is to maximise the 'non-censored' part of the density, which is

$$\mathcal{L}_2(\theta) = \sum_i{}' \delta_i \log f(T_i;\theta).$$

The limiting distribution of this estimator is

$$\sqrt{T}\,(\hat{\theta}_2 - \theta_0) \rightarrow N(0, I_2),$$ where

$$I_2 = \pi \, E\left\{ \frac{1}{f(T;\theta)^2} \left(\frac{\partial f}{\partial \theta} \right)^2 \right\}.$$

We can easily see that $I_1 > I_2$, thus the limiting variance of $\hat{\theta}_1$ is smaller than $\hat{\theta}_2$. Hence in the special case that $G(x) = F(x;\theta)$, it makes sense to use the survival likelihood to estimate θ_0.