as $h \to 0$, $\hat{F}_h(y) \to \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(Y_i \le y)$ ← this is the empirical dsn f'n! ⇒ MLE!

1.2.7 Censored Data

Censored data occur when the value is only partially known. This is different from *truncation*, in which the data does not include any values below (or above) a certain limit.

truncation:

For example, we might sample only households that have an income above a limit, L_0. If all incomes have distribution $F(x; \boldsymbol{\theta})$, then for $y > L_0$,

$$P(Y_1 \le y | Y_1 > L_0) = \frac{P(Y_1 \le y, Y_1 > L_0)}{P(Y_1 > L_0)} = \frac{F(y;\theta) - F(L_0;\theta)}{1 - F(L_0;\theta)}$$

The likelihood is then

$$L(\theta | y) = \prod_{i=1}^{n} \left\{ \frac{f(Y_i;\theta)}{1-F(L_0;\theta)} \right\}$$

this is just an iid likelihood w/ densities adjusted to take into account that $Y_i > L_0$.

1.2.7.1 Type I Censoring

Suppose a random variable X is normally distributed with mean μ and variance σ^2, but whenever $X \le 0$, all we observe is that it is less than or equal to 0. If the sample is set to 0 in the censored cases, then define

censoring ≠ truncation.

$$Y = \begin{cases} 0 & \text{if } X \le 0 \\ X & \text{if } X > 0. \end{cases}$$

Y: observed

X: latent/hidden

$X \sim N(\mu, \sigma^2)$.

The distribution function of Y is

at $y = 0$: $F_Y(0) = \underbrace{P(Y \le 0)}_{P(Y=0)} = P(X \le 0)$

$= P(\sigma Z + \mu \le 0)$

$= P(Z \le -\frac{\mu}{\sigma}) = \Phi(-\frac{\mu}{\sigma})$. where $Z \sim N(0,1)$, Φ st. normal cdf.

for $y > 0$: $F_Y(y) = P(Y \le y) = \Phi(\frac{y-\mu}{\sigma})$

for $y < 0$: $F_Y(y) : P(Y \le y) = 0$

Suppose we have a sample $Y_1, \dots, Y_n$ and let n_0 be the number of sample values that are 0. Then $m = n - n_0$ and

$$L_h(\underline{\theta} \mid \underline{y}) = \left(\frac{1}{2h}\right)^m \prod_{i=1}^{n} \left\{ F_y(y_i + h; \underline{\theta}) - F_y(y_i - h; \underline{\theta}) \right\}$$

$$= \left\{ \Phi\left(\frac{h-\mu}{\sigma}\right) - 0 \right\}^{n_0} \times \prod_{y_i > 0} \left\{ \frac{\Phi((y_i + h - \mu)/\sigma) - \Phi((y_i - h - \mu)/\sigma)}{2h} \right\}$$

$$L(\underline{\theta} \mid \underline{y}) = \lim_{h \to 0^+} L_h(\underline{\theta} \mid \underline{y})$$

$$= \underbrace{\left\{ \Phi\left(-\frac{\mu}{\sigma}\right) \right\}^{n_0}}_{P_\theta(Y \text{ censored})^{n_0}} \underbrace{\prod_{y_i > 0} \left\{ \frac{1}{\sigma} \phi\left(\frac{y_i - \mu}{\sigma}\right) \right\}}_{\prod_{y_i \text{ not censored}} f(y_i \mid \underline{\theta})}$$

essentially same as our mixture!

More generally, for left censoring @ L_0

$$L(\underline{\theta} \mid \underline{y}) = \{F(L_0)\}^{n_0} \prod_{y_i > L_0} f_y(y_i \mid \underline{\theta})$$

L_0 fixed = "type I censoring". Type II censoring = "Let's observe first r failures"

We might have censoring on the left at L_0 and censoring on the right at R_0, but observe all values of X between L_0 and R_0. Suppose X has density $f(x; \boldsymbol{\theta})$ and distribution function $F(x; \boldsymbol{\theta})$ and

$$Y_i = \begin{cases} L_0 & \text{if } X_i \le L_0 \\ X_i & \text{if } L_0 < X_i < R_0 \\ R_0 & \text{if } X_i \ge R_0 \end{cases}$$

If we let n_L and n_R be the number of X_i values $\le L_0$ and $\ge R_0$ then the likelihood of the observed data $Y_1, \dots, Y_n$ is

$$L(\underline{\theta} \mid \underline{y}) = \underbrace{\{F(L_0; \underline{\theta})\}^{n_L}}_{\text{left censored part}} \underbrace{\left\{ \prod_{L_0 < y_i < R_0} f_y(y_i; \underline{\theta}) \right\}}_{\text{observed}} \underbrace{\{1 - F(R_0; \underline{\theta})\}^{n_R}}_{\text{right censored part}}$$

not surprising.

We could also let each X_i be subject to its own censoring values L_i and R_i. For the special case of right censoring, define $Y_i = \min(X_i, R_i)$. In addition, define $\delta_i = \mathbb{I}(X_i \le R_i)$. Then the likelihood can be written as

$$= \begin{cases} 1 & \text{if observe } X_i \\ 0 & \text{if } X_i \text{ censored} \end{cases}$$

$$L(\underline{\theta}|\underline{y}) = \prod_{i=1}^{n} f(y_i;\underline{\theta})^{\delta_i} \underbrace{\left[1 - F(R_i;\underline{\theta})\right]^{1-\delta_i}}_{\text{right censored}}$$

Example (Equipment failure times): Pieces of equipment are regularly checked for failure (but started at different times). By a fixed date (when the study ended), three of the items had not failed and therefore were censored.

$R_2 = 72 = Y_2$ $R_4 = 50 = Y_4$ $R_{10} = 21 = Y_{10}$

y	2	72	51	50	33	27	14	24	4	21
delta	1	0	1	0	1	1	1	1	1	0

censored!

Suppose failure times follow an exponential distribution $F(x;\sigma) = 1 - \exp(-x/\sigma), x \ge 0$. Then

Let n_R be # obs. were right censored.

$$L(\sigma|\boldsymbol{Y}) = \prod_{i=1}^{n} \left[\frac{1}{\sigma}\exp\left(-\frac{y_i}{\sigma}\right)\right]^{\delta_i} \exp\left(-y_i/\sigma\right)^{1-\delta_i}$$

$$= \left(\frac{1}{\sigma}\right)^{n-n_R} \exp\left(-n\bar{Y}/\sigma\right)$$

$$\ell(\sigma) = -(n-n_R)\log\sigma - n\bar{Y}/\sigma$$

$$\frac{d\ell}{d\sigma} = -\frac{(n-n_R)}{\sigma} + \frac{n\bar{Y}}{\sigma^2} \overset{\text{set}}{=} 0$$

$$n\bar{y} = \sigma(n-n_R) \Rightarrow \hat{\sigma}_{MLE} = \frac{n\bar{Y}}{n-n_R}$$

$$= \frac{10 \cdot 30.8}{7} = 44.0$$

1.2.7.2 Random Censoring

So far we have considered censoring times to be fixed. This is not required.

e.g. in medical studies patients enter at different times, modeled as random variables (w/ fixed end date)

How is this different than previous example?

↳ don't know when patients will enter!

This leads to random censoring times, e.g. R_i, where we assume that the censoring times are independent of $X_1, \ldots, X_n$ and iid with distribution function $G(t)$ and density $g(t)$.

Again $Y_i = \min(X_i, R_i)$ and $\delta_i = \mathbb{I}(X_i \le R_i)$

Let's consider the contributions to the likelihood:

due to $(\underbrace{Y_i, \delta_i = 1}_{Y_i \text{ observed}})$:

$$\frac{P(Y_i \in (y-h, y+h], \delta_i = 1)}{2h} = P(X_i \in (y-h, y+h], X_i \le R_i)/2h$$
$$= \frac{1}{2h}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{I}(y-h < t \le y+h, t \le r) f(t;\underline{\theta}) g(r)\,dt\,dr$$
$$= \frac{1}{2h}\int_{y-h}^{y+h}\left[\int_{-\infty}^{\infty} \mathbb{I}(t \le r) g(r)\,dr\right] f(t;\underline{\theta})\,dt$$
$$\text{FTC} \quad = \frac{1}{2h}\int_{y-h}^{y+h}\left[1 - G(t)\right] f(t;\theta)\,dt$$
$$\xrightarrow{h \to 0^+} [1 - G(y)]\, f(y;\theta)$$

due to $(\underbrace{Y_i, \delta_i = 0}_{Y_i \text{ censored}})$:

$$\frac{P(Y_i \in (y-h, y+h], \delta_i = 0)}{2h} = P(R_i \in (y-h, y+h], X_i > R_i)/2h$$
$$= \frac{1}{2h}\int_{y-h}^{y+h}\left[\int_{-\infty}^{\infty} \mathbb{I}(t > r) f(t;\underline{\theta})\,dt\right] g(r)\,dr$$
$$\text{FTC} \quad = \frac{1}{2h}\int_{y-h}^{y+h}\left[1 - F(r;\underline{\theta})\right] g(r)\,dr$$
$$\xrightarrow{h \to 0^+} [1 - F(y;\underline{\theta})]\, g(y)$$

which results in

$$L(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{\delta}) = \left\{\prod_{i=1}^n f(y_i;\underline{\theta})^{\delta_i}[1 - G(y_i)]^{\delta_i}\right\}\left\{\prod_{i=1}^n g(y_i)^{1-\delta_i}\left[1 - F(y_i;\underline{\theta})\right]^{1-\delta_i}\right\}$$
$$= \prod_{i=1}^n \underbrace{f(y_i;\underline{\theta})^{\delta_i}[1 - F(y_i;\underline{\theta})]^{1-\delta_i}}_{\text{Same as for fixed!}}\ \underbrace{g(y_i)^{1-\delta_i}[1 - G(y_i)]^{\delta_i}}_{\text{not necessary for MLE}}$$

1.3 Likelihoods for Regression Models

We will start with linear regression and then talk about more general models.

1.3.1 Linear Model

Consider the familiar linear model

$$Y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \qquad i = 1, \ldots, n,$$ independent observations.

where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are known nonrandom vectors.

$E[\varepsilon_i] = 0$

$Var[\varepsilon_i] = \sigma^2$

often we estimate β by $\hat{\beta}_{OLS}$, which does not require a distribution for ε_i.

For likelihood-based estimation, we need a distribution for ε_i! Namely start w/ $\varepsilon_i \sim N(0, \sigma^2)$.

$$\Rightarrow \quad L(\boldsymbol{\beta}, \sigma | \{Y_i, \boldsymbol{x}_i\}_{i=1}^n) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) \exp\left(-\frac{(Y_i - x_i^\top \beta)^2}{2\sigma^2}\right)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - x_i^\top \beta)^2\right)$$

take logs,
take derivatives,
set $= 0$, solve $\rightsquigarrow$

$$\hat{\beta}_{MLE} = (X^\top X)^{-1} X^\top Y$$

same as $\hat{\beta}_{OLS}$

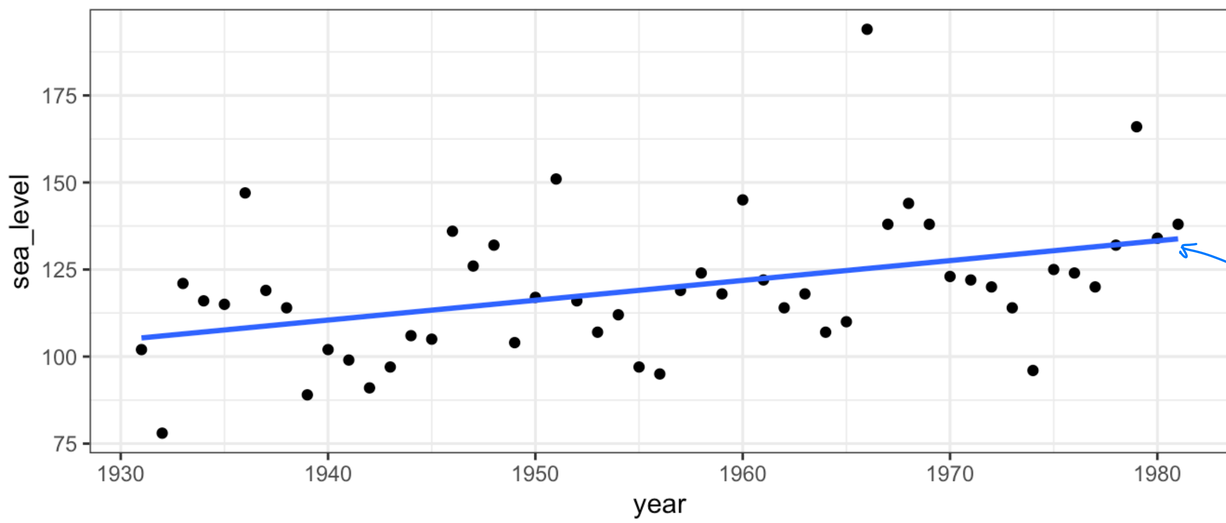
$$\hat{\sigma}^2_{MLE} = \frac{1}{n} \sum_{i=1}^n (Y_i - X_i^\top \hat{\beta}_{MLE})^2$$

only asymptotically unbiased.

What do you do when ϵ_i are not Gaussian?

- transform data so $\hat{\varepsilon}_i$ look gaussian
- Use a different distribution for ε_i!

Example (Venice sea levels): The annual maximum sea levels in Venice for 1931−1981 are :



least squares line.

We know maxima not Gaussian!

Approach 1: OLS $E[\varepsilon_i]=0$, $Var[\varepsilon_i]=\sigma^2$ no distributional assumption.

extreme value distribution

Approach 2: Assume $\varepsilon_i \sim$ Gumbel, use ML

$$f_\varepsilon(y) = \frac{1}{\sigma}\exp\left(-\frac{y}{\sigma}\right)\exp\left(-\exp\left(-\frac{y}{\sigma}\right)\right)$$

$$\Rightarrow L(\beta,\sigma \mid \{y_i, x_i\}_{i=1}^n) = \prod_{i=1}^n f_\varepsilon\left(y_i - x_i^T\beta\right) = \prod_{i=1}^n \frac{1}{\sigma}\exp\left(-\frac{y_i - x_i^T\beta}{\sigma}\right)\exp\left(-\exp\left(-\frac{y_i - x_i^T\beta}{\sigma}\right)\right)$$

YOUR TURN: Fit both approaches to Venice data!

OLS:

$\hat\beta_0 = 104.8$, $\hat\beta_1 = .567$ (se .177)

ML, GUMBEL

$\hat\beta_0 = 96.8$, $\hat\beta_1 = .564$ (s.e. .136)

Difference in $\hat\beta_0$? $E[\varepsilon_i] = 0.577\sigma = 0.577\,\overset{14.5}{\hat\sigma_{MLE}} \neq 0$ $(96.8 + .577(14.5) = 105.1)$

OLS or ML? If EV model is correct, more efficient (note: st. errors).