apts.ac.uk
Academy for PhD Training in Statistics

Flexible Regression

Session 4 - Quantile Regression Extensions & Alternative Approaches (Ch. 5)

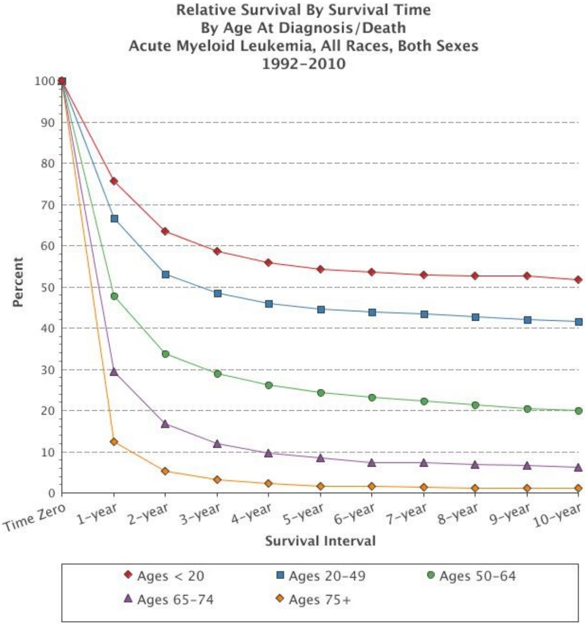
Notes: `https://warwick.ac.uk/fac/sci/statistics/apts/students/resources/`
Slides: `https://www.stats.gla.ac.uk/~claire/APTS_FR_session_4.pdf`

Claire Miller & Tereza Neocleous

Session outline

1. Part I: Censored Regression Quantiles
2. Part II: Alternative Approaches
   - Bayesian Quantile Regression
   - Genelarised Additive Models for Location, Scale and Shape

Part I (extension): quantile regression for censored data/survival analysis applications



Cancer sites include invasive cases only unless otherwise noted.
The annual survival estimates are calculated using monthly intervals.
Source: http://seer.cancer.gov/faststats/selections.php?series=age

Survival data analysis

Censoring

- Characteristic of survival (*duration*/*time-to-event*) data
- Data are censored if observation on the subject/experimental unit had ceased before the event of interest had occurred.
- Censoring could occur for instance if:
  - a clinical trial of a new lung cancer therapy terminates before all the patients in the study are dead due to lung cancer;
  - the subject dies from a reason completely unconnected with the disease;
  - the researchers may simply have lost contact with a subject.

Assumptions

Right censoring

- A censored observation involves a subject whose time to event is unknown (except that it is at least greater than the time for which the subject was observed).
- *E.g.* some subjects in the study may have not experienced the event of interest at the end of the study or time of analysis.
- As the incomplete nature of this observation occurs in the upper tail of the distribution of the time to the event of interest, the observation is termed right-censored.

Other types of censoring

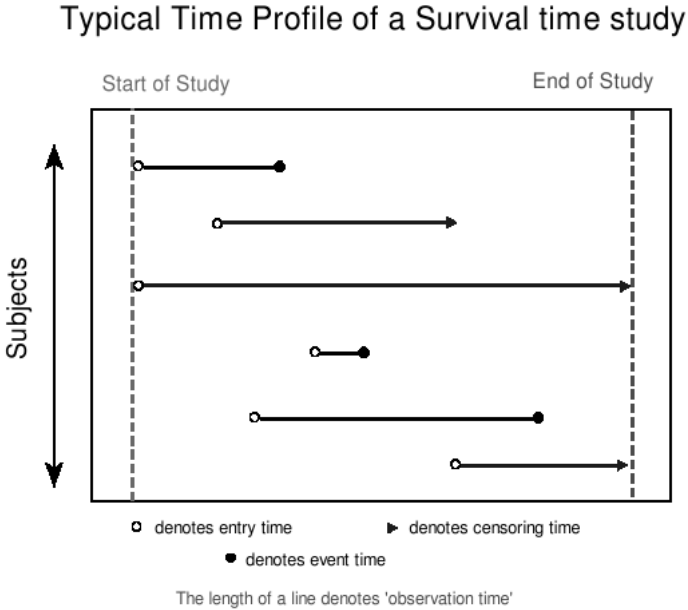
- Left censoring, interval censoring

Assumptions

Data

- Observations are either (complete) event times (*e.g.* a death time) or censored (*e.g.* left, right or interval censored).
- For the purposes of this course, only right-censored data will be considered.
- Let t_i denote the observation time for the ith individual ($i = 1, \ldots, n$) in a sample of n individuals. Further define an indicator function $\delta_i = 1$ if the ith observation time is an event time, $\delta_i = 0$ if the ith observation time is right-censored.
- Each individual's data consist of an observation time and an "event type" indicator function, (t_i, δ_i).

Time horizons



Survival function

Definition and properties

- Denote the time to event as T, a positive random variable, with distribution function $F(t) = P(T < t)$ and density function $f(t)$.
- The survival function, $S(t)$, is defined as

$$S(t) = P(T \geq t) = 1 - F(t) \text{ for any } t > 0.$$

- $S(t)$ is a strictly non-increasing function with a value of 1 at the origin $(t = 0)$ and decreasing to 0 at $t = \infty$.
- Event time distributions are usually skewed and hence the most appropriate 'central' summary of the distribution is provided by the median survival time, i.e. the value t^* such that $S(t^*) = 0.5$.

Hazard function

Definition and properties

- The hazard rate or hazard function, $\lambda(t)$, is expressed as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(T \leq t + \Delta t \mid T \geq t)}{\Delta t}$$

- Rate at which an individual is likely to experience the event of interest in the next small time interval, Δt, given that the individual has survived up to that point (up to time t).
- $\lambda(t) \geq 0$
- Cumulative hazard (total hazard up to time t):

$$\Lambda(t) = \int_0^t \lambda(u) du$$

Relationships between functions

$$S(t) = 1 - F(t)$$

$$F(t) = \int_0^t f(u)du;$$

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}$$

$$\lambda(t) = \frac{f(t)}{S(t)}$$

$$S(t) = \exp\left\{-\int_0^t \lambda(u)du\right\}$$

$$f(t) = \lambda(t)\exp\left\{-\int_0^t \lambda(u)du\right\}$$

Any one of these defines all the others.

Empirical distribution function

Definition: empirical distribution function (EDF)

Given a sample entirely composed of complete data, the EDF is defined as

$$\hat{S}_{EDF}(t) = \frac{\text{number surviving beyond } t}{\text{number in the sample}}$$

Problem

In the presence of censoring, the EDF estimator is biased and a modification is needed.

The Kaplan-Meier estimator

Definitions and assumptions

- Sample consisting of n observation times $t_1, \ldots, t_n$
- m event times and $n - m$ censored observations
- Ordered event times: $t_{(1)} < t_{(2)} < \ldots < t_{(m)}$ where $m \leq n$
- Assume $t_{(0)} = 0$
- r_j: number of individuals who are alive (and hence at risk) just before $t_{(j)}$, for $j = 1, \ldots, m$
- d_j: number of individuals who die at time $t_{(j)}$.

The Kaplan-Meier estimator

Probability estimates

- Since there are r_j individuals at risk just before $t_{(j)}$ and d_j deaths at $t_{(j)}$, the probability that an event of interest occurs during the interval $[t_{(j)}, t_{(j+1)})$ is estimated by

$$d_j/r_j.$$

- Probability of surviving through this interval:

$$1 - d_j/r_j = s_j/r_j$$

  where s_j is the number of individuals in the sample who survive at least beyond time $t_{(j)}$.

The Kaplan-Meier estimator

Survival function:

$$
\begin{aligned}
S(t_{(i)}) &= P(T > t_{(i)}) = P(T > t_{(i)} \mid T > t_{i-1}) P(T > t_{i-1}) \\
&= P(T > t_i \mid T > t_{(i-1)}) P(T > t_{(i-1)} \mid T > t_{(i-2)}) P(T > t_{(i-2)}) \\
&= \dots \\
&= \prod_{j=1}^{i} P(T > t_{(j)} \mid T > t_{(j-1)})
\end{aligned}
$$

The Kaplan-Meier estimator

Sample-based estimate: the KM (*product-limit*) estimator

$$\hat{S}_{KM}(t) = \prod_{j=1}^{i} \frac{s_j}{r_j} \text{ for } t_{(i)} \leq t < t_{(i+1)}, i = 1, \ldots, m$$

since the estimate of $P(T > t_{(j)} \mid T > t_{(j-1)})$ will be s_j/r_j.

The Kaplan-Meier estimator

Redistributing to the right (Efron 1967)

- Ordered observation times with mass $1/n$ associated with each observation initially.
- Distribute the mass $1/n$ of the first censored observation encountered equally to all times to its right.
- Continue until the mass of all the censored observations has been distributed.
- This resulting distribution of masses, or weights, is precisely the KM estimator.
- The KM estimator is similar to the EDF but with different weights.

Illustrative example of the Kaplan-Meier estimator

- 10 observations: 3,4,5+,6,6+,8+,11,14,15,16+.
- The Kaplan-Meier estimator is calculated as follows:

t	d_i	r_i	$S(t)$
3	1	10	0.9
4	1	9	$(1-1/9)*0.9=0.8$
6	1	7	$(1-1/7)*0.8=0.686$
11	1	4	$(1-1/4)*0.686=0.514$
14	1	3	$(1-1/3)*0.514=0.343$
15	1	2	$(1-1/2)*0.343=0.171$
16	1	1	0

- Since the last observation is censored, the survival function could either stay at the same level or be set to 0.

The reweighting-to-the-right algorithm

Data	Step 0	Step 1	Step 2
3	1/10	0.1	0.1
4	1/10	0.1	0.1
5+	1/10	0.0	0.0
6	1/10	1/10+(1/7)1/10=0.114	0.114
6+	1/10	1/10+(1/7)1/10=0.114	0.0
8+	1/10	1/10+(1/7)1/10=0.114	0.114+(1/5)0.114=0.137
11	1/10	1/10+(1/7)1/10=0.114	0.114+(1/5)0.114=0.137
14	1/10	1/10+(1/7)1/10=0.114	0.114+(1/5)0.114=0.137
15	1/10	1/10+(1/7)1/10=0.114	0.114+(1/5)0.114=0.137
16+	1/10	1/10+(1/7)1/10=0.114	0.114+(1/5)0.114=0.137
Data	Step 3	S(t)	
3	0.1	0.9	
4	0.1	0.8	
5+	0.0	0.8	
6	0.114	0.686	
6+	0.0	0.686	
8+	0.0	0.686	
11	0.137+(1/4)0.137=0.171	0.515	
14	0.137+(1/4)0.137=0.171	0.343	
15	0.137+(1/4)0.137=0.171	0.171	
16+	0.137+(1/4)0.137=0.171	0.0	

Motivation for Censored Regression Quantiles (CRQ)

- Portnoy (2003): generalisation of reweighting-to-the-right algorithm to allow covariates
- Main idea: split censored observations by assigning weights to them as they get crossed by the quantile hyperplane.

Censored Regression Quantiles

Consider the linear censored quantile regression model:

Random variables: $\{(\mathbf{x}_i,\ T_i) :\ i = 1,\ ...,\ n\}$ with $x_i \in \mathbb{R}^p$ and $\boldsymbol{\beta}(\tau) \in \mathbb{R}^p$ satisfying

$$P_{\mathbf{x}_i}\left\{T_i \le \mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}(\tau)\right\} = \tau \qquad i = 1,\ ...,\ n$$

Also assume censoring points: $\{C_i :\ i = 1,\ ...,\ n\}$ such that the observables are $Y_i = \min\{T_i,\ C_i\}$

Censoring indicator: $\delta_i \equiv I\{T_i \le C_i\}$

Generally: $\{(T_i,\ \mathbf{x}_i,\ C_i) :\ i = 1,\ ...,\ n\}$ are assumed to be i.i.d. and $(T_i,\ C_i) \mid \mathbf{x}_i$ independent.

The grid algorithm for CRQ

Let $\varepsilon > 0$ be given and define a grid of τ-values:

$$\varepsilon \leq t_1 < t_2 < \cdots < t_M \leq 1 - \varepsilon$$

$$\boldsymbol{\beta}_k = \boldsymbol{\beta}(t_k), \quad k = 1, \cdots, M$$

$$\boldsymbol{\beta} = (\beta_1, \cdots, \beta_M) \in \mathbb{R}^{Mp}$$

As in Portnoy (2003), assume the usual regression quantile at $\tau = t_1$ (using all the data) lies below all censored points.

Define the initial $\hat{\boldsymbol{\beta}}_1$ to be this regression quantile solution, and define weights $\hat{w}_i(t_1) \equiv 1$ $(i = 1, \cdots, n)$

We obtain $\hat{\boldsymbol{\beta}}_k$ inductively as follows:
Suppose we have all $\hat{\boldsymbol{\beta}}_l$ and weights $\hat{w}_i(t_l)$ for $l \leq k$

The regression quantile, $\hat{\boldsymbol{\beta}}_{k+1}$ at t_{k+1} is obtained by minimising a weighted regression quantile objective function.
Specifically define $\hat{\boldsymbol{\beta}}_{k+1}$ to minimise over $\mathbf{b}$

$$\sum_{i=1}^{n} \Big\{ \delta_i \rho_{t_{k+1}}(Y_i - \mathbf{x}_i^\mathsf{T} \mathbf{b})$$

$$+(1-\delta_i)\big[\hat{w}_i(t_{k+1}, \boldsymbol{\beta})\rho_{t_{k+1}}(C_i - \mathbf{x}_i^\mathsf{T} \mathbf{b})$$

$$+(1-\hat{w}_i(t_{k+1}, \boldsymbol{\beta}))\rho_{t_{k+1}}(Y^* - \mathbf{x}_i^\mathsf{T} \mathbf{b})\big] \Big\}$$

where Y^* is sufficiently large.

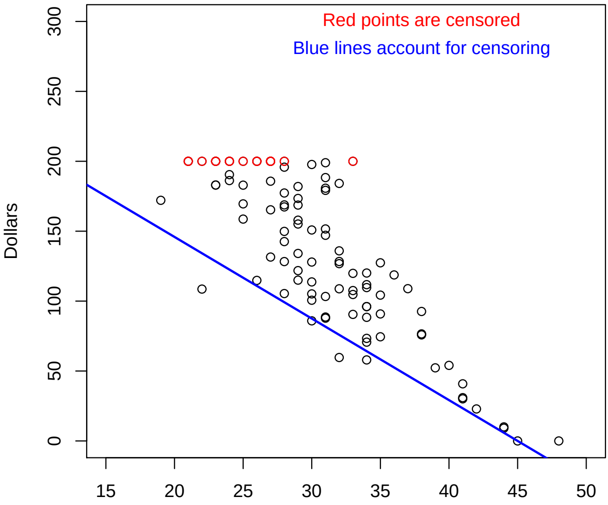
- Before progressing to next grid point, reconsider censored observations with weight 1 (not crossed) before the grid point t_k. When moving from t_k to t_{k+1} some censored observations that were not yet crossed can be crossed. In that case these observations (at C_i) are reweighted with

$$\hat{w}_i(\tau) \equiv (\tau - \tau_i)/(1 - \tau_i)$$

  where $\tau_i(\hat{\beta}) = t_{k+1}$; and an extra contribution to Y^* is added with weight $(1 - \hat{w}_i)$.
- This algorithm stops at the last grid point t_M, or it ends at t_e when only censored observations remain above $\mathbf{x}^\mathsf{T}\hat{\beta}(t_e)$.

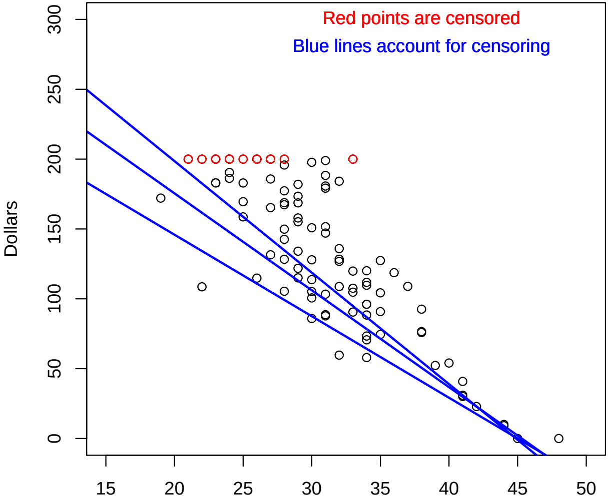
CRQ (Portnoy) algorithm illustration

Start at $\tau = 0.05$. No censored observations crossed at this point.



CRQ (Portnoy) algorithm illustration

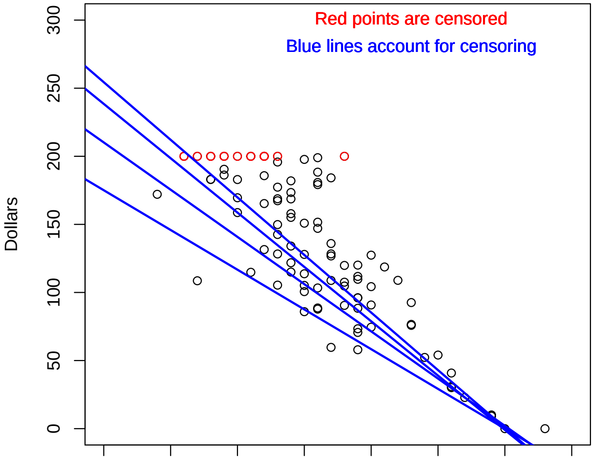
$\tau = 0.05, 0.15, 0.25$: still no censored observations crossed.



CRQ (Portnoy) algorithm illustration

$\tau = 0.05, 0.15, 0.25, 0.35$. Weights are calculated as

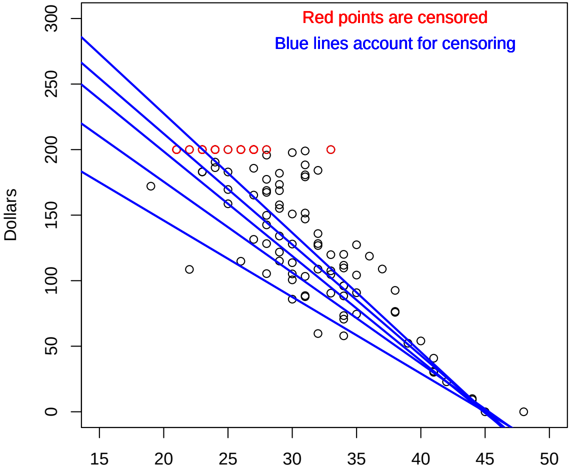
$$\tilde{\tau}_i = 0.35, \quad w_i = \frac{\tau - 0.35}{1 - 0.35} \text{ for } \tau > 0.35.$$



CRQ (Portnoy) algorithm illustration

$\tau = 0.05, 0.15, 0.25, 0.35, 0.45$. Weight is calculated as

$$\tilde{\tau}_i = 0.45, \quad w_i = \frac{\tau - 0.45}{1 - 0.45} \text{ for } \tau > 0.45.$$



Relationship with the Cox Proportional Hazards (PH) model

- CRQ vs Cox Proportional Hazards model: agreement under Accelerated Failure Time (AFT) model if log time is used as the response in CRQ.
- Time T, random censoring to the right
- Survival function $S_T(t \mid \mathbf{x}) = P(T > t \mid \mathbf{x})$
- Hazard $\lambda(t) = \frac{f(t)}{S(t)}$
- Cox PH model: $\lambda(t \mid \mathbf{x}) = \lambda_0(t) \exp(\mathbf{x}^\mathsf{T} \boldsymbol{\beta})$
- AFT model: $\log(T_i) = \mathbf{x}_i^\mathsf{T} \boldsymbol{\beta} + u_i$, u_i i.i.d. with $F(u) = 1 - e^{-e^u}$ gives a Cox proportional hazards model with Weibull baseline hazard.

AFT and CRQ

- Conditional quantile $Q_T(\tau \mid \mathbf{x}) = \inf\{t : P(T \leq t \mid \mathbf{x}) \geq \tau\}$
- Log transformation: $Q_T(\tau \mid \mathbf{x}) = \exp(Q_{\log(T)}(\tau \mid \mathbf{x}))$
- AFT correspondence: use $\log(T)$ as the response in CRQ
- $Q_{\log(T)}(\tau \mid \mathbf{x}) = \mathbf{x}^\mathsf{T}\beta + F_u^{-1}(\tau)$: $\mathbf{x}^\mathsf{T}\boldsymbol{\beta}$ shifts the location of $\log(T)$
- Introduce heterogeneity in the conditional distribution of $\log(T)$ by allowing $\boldsymbol{\beta}$ to vary with τ: $Q_{\log(T)}(\tau \mid \mathbf{x}) = \mathbf{x}^\mathsf{T}\boldsymbol{\beta}(\tau)$

Some comments on CRQ

- CRQ vs regular quantile regression for uncensored data:
  - CRQ: iterative estimating process (like KM);
  - RQ: computes a single quantile at a time, *e.g.* median
- Confidence intervals for the CRQ regression coefficients are obtained using bootstrap.
- Software implementation: `crq()` function in `quanteg` package (setting `method` argument to `"portnoy"`).
- Another method is `"penghuang"` , which generalises the martingale representation of the Nelson-Aalen estimator.

Peng and Huang's CRQ method

Peng and Huang (2008) extend the martingale representation of the Nelson-Aalen estimator of the cumulative hazard function to produce an "estimating equation" for conditional quantiles.

- $\Lambda_T(t|\mathbf{x}) = -\log\{1 - F_T(t|\mathbf{x})\}$: cumulative hazard function of T conditional on $\mathbf{x}$;
- $N_i(t) = I(Y_i \le t, \delta_i = 1)$;
- $M_i(t) = N_i(t) - \Lambda_T\{t \wedge Y_i|\mathbf{x}_i)$ is a martingale process so that $E\{M_i(t)|\mathbf{x}_i\} = 0$ for all $t \ge 0$.

So

$$E\left[N_i\{\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\tau)\} - \Lambda_T\{\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\tau) \wedge Y_i\}|\mathbf{x}_i\right] = 0.$$

Connection between Λ_T and the quantile functions:

$$\begin{aligned} & \Lambda_T\{\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\tau) \wedge Y_i|\mathbf{x}_i\} = H(\tau) \wedge H\{F_T(Y_i|\mathbf{x}_i)\} \\ = & \int_0^{\tau} I\{Y_i \geq \mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}_0(u)\} dH(u), \end{aligned}$$

where $H(u) = -\log(1-u)$ for $0 \leq u \leq 1$.

The estimating equation becomes

$$n^{-1/2} \sum_{i=1}^{n} \mathbf{x}_i \left[N_i(\mathbf{x}_i^T \boldsymbol{\beta}) - \int_0^{\tau} I\{Y_i \geq \mathbf{x}_i^T \boldsymbol{\beta}(u)\} dH(u) \right] = 0.$$

Approximating the integral on a grid, $0 = \tau_0 < \tau_1 < \cdots < \tau_J < 1$ yields a simple linear programming formulation to be solved at the gridpoints,

$$\alpha_i(\tau_j) = \sum_{k=0}^{j-1} I\{Y_i \geq \mathbf{x}_i^T \hat{\boldsymbol{\beta}}(\tau_k)\}\{H(\tau_{k+1}) - H(\tau_k)\},$$

yielding Peng and Huang's final estimating equation,

$$n^{-1/2} \sum \mathbf{x}_i \left[N_i\{\mathbf{x}_i^T \boldsymbol{\beta}(\tau)\} - \alpha_i(\tau) \right] = 0.$$

Setting $r_i(\mathbf{b}) = Y_i - \mathbf{x}_i^T \mathbf{b}$, this convex function for the Peng and Huang problem takes the form

$$R(\mathbf{b}, \tau_j) = \sum_{i=1}^{n} r_i(\mathbf{b}) \left[\alpha_i(\tau_j) - I\{r_i(\mathbf{b}) < 0\} \delta_i \right] = \min!$$

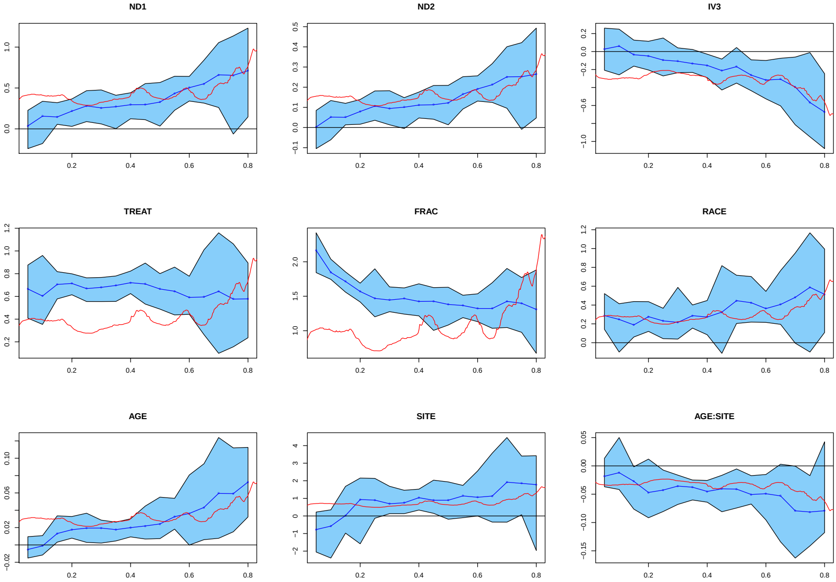
Example: UIS data

In R:

```
> library(quantreg)
> library(survival)
> data(uis)
> fit <- crq(Surv(log(TIME), CENSOR) ~  ND1 + ND2 + IV3 +
             TREAT + FRAC + RACE + AGE * SITE,
             method = "Portnoy", data = uis)
> Sfit <- summary(fit,1:19/20)
> PHfit <- coxph(Surv(TIME, CENSOR) ~  ND1 + ND2 + IV3 +
              TREAT + FRAC + RACE + AGE * SITE, data = uis)
> plot(Sfit, CoxPHit = PHfit)
```

Reference: Koenker, 2008

Estimated quantile coefficients



Part II: alternative approaches to conditional quantile estimation

Bayesian quantile regression

- Let us revisit the linear quantile regression equation

$$Q_\tau(Y|\mathbf{x}) = \mathbf{x}^\mathsf{T}\boldsymbol{\beta}(\tau).$$

- Given the observations $\mathbf{y} = (y_1, \ldots, y_n)$, the posterior distribution of $\boldsymbol{\beta}(\tau)$, $\pi(\boldsymbol{\beta}|\mathbf{y})$ is given by

$$\pi(\boldsymbol{\beta}|\mathbf{y}) \propto L(\mathbf{y}|\boldsymbol{\beta})\pi(\boldsymbol{\beta})$$

  where $\pi(\boldsymbol{\beta})$ is the prior distribution of $\boldsymbol{\beta}$ and $L(\mathbf{y}|\boldsymbol{\beta})$ is the likelihood function.

What should the likelihood be?

- Semiparametric and nonparametric Bayesian methods can be used
- Usually involve mixtures of Dirichlet processes

+ Flexible

- Computation is hard

- Yu and Moyeed (2001): asymmetric Laplace distribution

Asymmetric Laplace distribution

- Suppose that the random variable Z follows the asymmetric Laplace distribution.
- Density:
$$f_\tau(z) = \tau(1-\tau)\exp[-\rho_\tau(z)]$$
for $0 < \tau < 1$ where $\rho_\tau(z) = z(\tau - I(z < 0))$.
- If $\tau = 0.5$, this reduces to the (symmetric) Laplace distribution.
- Mean:
$$E(Z) = \frac{1-2\tau}{\tau(1-\tau)}$$
- Variance:
$$\mathrm{Var}(Z) = \frac{1-2\tau+2\tau^2}{\tau^2(1-\tau)^2}$$

- Incorporate location and scale parameters μ and σ to obtain

$$f_\tau(z; \mu, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau\left(\frac{z-\mu}{\sigma}\right)\right\}$$

- Minimising the quantile regression objective function is equivalent to maximising the likelihood

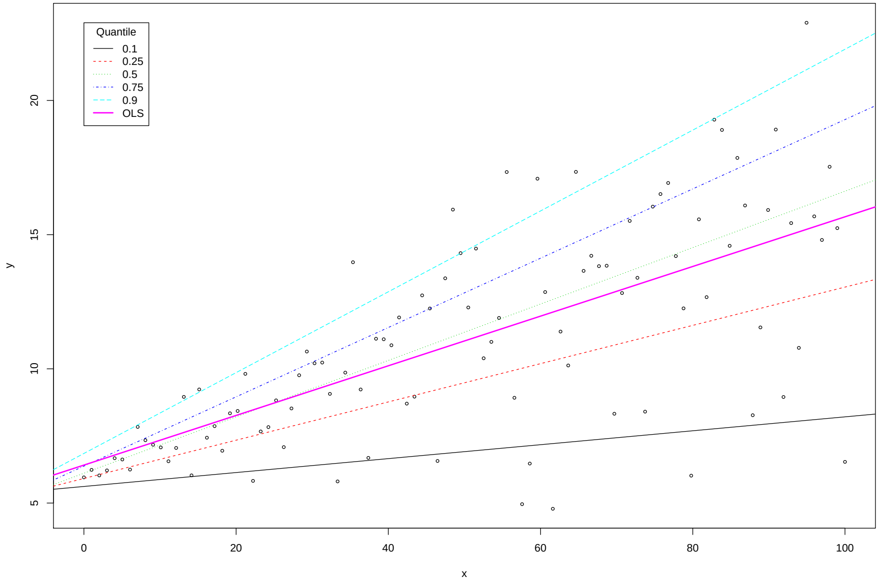
$$L(\boldsymbol{y}|\boldsymbol{\beta}) = \{\tau(1-\tau)\}^n \exp\left\{-\sum_{i=1}^{n} \rho_\tau\{y_i - \mathbf{x}_i^\mathsf{T}\boldsymbol{\beta})\right\},$$

R implementation

- Function `bayesQR()` from `library(bayesQR)`
- Usual arguments `formula`, `quantile`, plus number of MCMC draws

```
fit.b <- bayesQR(y~x, quantile=c(.1,.25,.5,.75,.9), ndraw=5000)

plot(x, y, main="", cex=.6, xlab="x")

sum.b <- summary(fit.b, burnin=500)
for (i in 1:length(sum.b)){
  abline(a=sum.b[[i]]$betadraw[1,1],
         b=sum.b[[i]]$betadraw[2,1],lty=i,col=i)}

fit.OLS <- lm(y~x)
abline(fit.OLS,lty=1,lwd=2,col=6)

legend(x=0,y=max(y),legend=c(.1,.25,.50,.75,.9,"OLS"),
       lty=c(1,2,3,4,5,1),lwd=c(1,1,1,1,1,2),
       col=c(1:6),title="Quantile")
```

Bayes QR fit



Properties of the AL-based Bayesian QR

- If a flat prior $\pi(\boldsymbol{\beta}) \propto 1$ is used, then
  - the posterior distribution of $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta}|\mathbf{y})$ is proper;
  - the posterior mode is the frequentist estimator $\hat{\boldsymbol{\beta}}(\tau)$.
- However, when the AL likelihood is misspecified,
  - the posterior chain from the Bayesian AL quantile regression does not lead to valid posterior inference;
  - correction to the covariance matrix of the posterior chain is possible to enable an asymptotically valid posterior inference (Yang, Wang and He, 2016).

Other related methods

Geraci and Bottai (2007, 2013) use the asymmetric Laplace approach to fit linear quantile mixed models (`lqmm` package in R).

- Quantile regression with a random intercept effect:

$$Q_\tau(Y_{ij}|\mathbf{x}_{ij}, b_i) = \mathbf{x}_{ij}^\mathsf{T}\boldsymbol{\beta} + b_i.$$

- Assume $(Y_{ij}|\mathbf{x}_{ij}, \boldsymbol{\eta}, b_i) \sim AL(\mathbf{x}_{ij}^\mathsf{T}\boldsymbol{\beta} + b_i, \sigma, \tau)$ and $b_i \sim N(0, \varphi^2)$, where $\boldsymbol{\eta} = (\boldsymbol{\beta}, \sigma, \varphi)$.
- Estimate $\boldsymbol{\eta}$ using an EM algorithm by integrating out b_i from $f(\mathbf{y}, \mathbf{b}|\boldsymbol{\eta}) = f(y|\boldsymbol{\eta}, \mathbf{b})f(\mathbf{b}|\boldsymbol{\eta})$.

Other related methods

- Tsionas (2003), Kozumi and Kobayashi (2011): Gibbs sampling procedures for Bayesian quantile regression assuming AL likelihood using a conditional Gaussian representation
- Li, Xi and Lin (2010): Bayesian regularised quantile regression
- Lum and Gelfand, 2012: Bayesian spatial quantile regression assuming asymmetric Laplace process
- Yang and He (2012): Bayesian empirical likelihood
- Kottas (2009): Mixtures with Dirichlet process priors
- Reich et al. (2010), Reich et al. (2011): Bayesian QR for clustered/spatial data

$\vdots$

Part II: alternative approaches to conditional quantile estimation

Generalised Additive Models for Location, Scale and Shape

- GAMLSS, Ribgy and Stasinopoulos (2005): generalisation of GLMs and GAMs.
- Generalised: large number of response distributions/link functions
- Additive: allow for non-parametric smooth terms as well as the usual linear regression terms
- Location, Scale and Shape: focus not only on the mean but also on how the spread and shape of the distribution of the response depend on explanatory variables.
- Parameters:
  - μ: location
  - σ: spread
  - τ: skewness (no relationship to the quantile level!)
  - ν: kurtosis

Example: Box-Cox, Cole and Green distribution

$$f(y|\mu,\sigma,\nu) = \frac{1}{\sqrt{2\pi}\sigma}\frac{y^{\nu-1}}{\mu^{\nu}}\exp\left(-\frac{z^2}{2}\right)$$

where

$$z = \begin{cases} \dfrac{(y/\mu)^{\nu}-1}{\nu\sigma} & \text{if } \nu \neq 0 \\[2ex] \dfrac{\log(y/\mu)}{\sigma} & \text{if } \nu = 0. \end{cases}$$

Implementation in R: `gamlss` package

Additive terms	function in R
P-splines	`pb()`, `pbm()`, `cy()`
Varying coefficient	`pvc()`
Cubic splines	`cs()`
Loess/ neural networks	`lo()`, `nn()`
Fractional/piecewise polynomials	`fp()`, `fk()`
Non-linear fit	`nl()`
Random effects	`random()`, `re()`
Ridge regression	`ri()`
Simon Wood's GAM	`ga()`
Decision trees	`tr()`
Random walk and AR	`rw()`, `ar()`

Lots of functionality! See Stasinopoulos et al. (2018) for a tutorial.

Example: GAMLSS for the NHANES BMI data

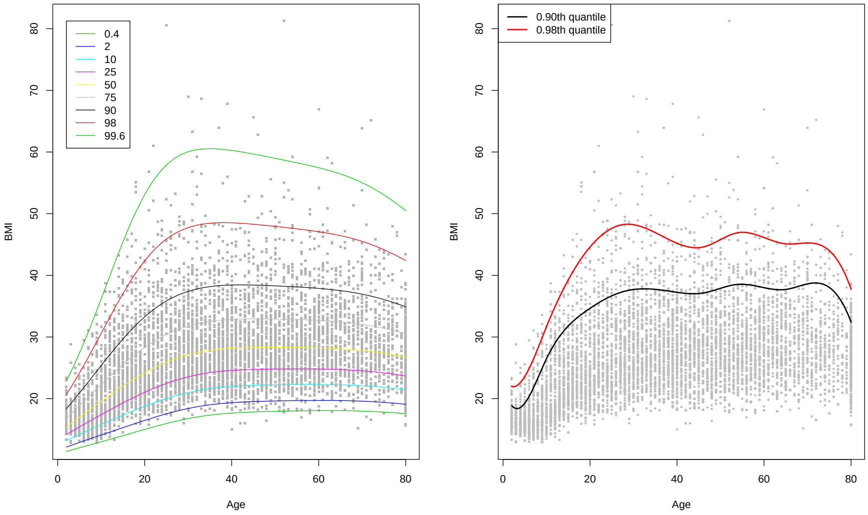
- National Health and Nutrition Examination Survey (US)
- Relationship between body mass index (BMI) and age
- In R:

```
> library(NHANES)
> library(gamlss)
> data(uis)
> model <- gamlss(BMI~ps(Age), sigma.formula=~ps(Age),
                  tau.formula=~ps(Age), data=NHANES,
                  family="BCCG")
> centiles(model.g, xvar=NHANES$Age, xlab="Age",
           ylab="BMI", main="")
```

Example: quantile regression for the NHANES BMI data

```
> library(quantreg)
> plot(BMI~Age, data=NHANES, col="grey", pch=16, cex=0.5)
> newdata <- data.frame(Age=seq(2,80,len=500))
> model <- rq(BMI~bs(Age, df=10), data=NHANES, tau=0.9)
> lines(newdata$Age, predict(model, newdata), col=1, lwd=2)
> model <- rq(BMI~bs(Age, df=10), data=NHANES, tau=0.98)
> lines(newdata$Age, predict(model, newdata), col=2, lwd=2)
> legend("topleft", col=1:2, lwd=2,
        c("0.90th quantile", "0.98th quantile"))
```

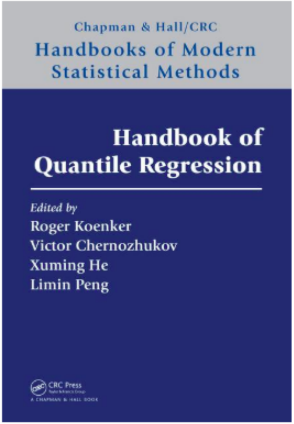
GAMLSS and RQ model for NHANES data



Summary I

Quantile regression: extensions and alternative approaches

Censored regression quantiles: only one of many active research areas in quantile regression



- Survival analysis
- Quantile time series analysis
- Extremal quantile regression
- Methods for longitudinal data
- Methods for measurement error/missing data
- Genetic and genomic applications
- Finance applications
- Applications in ecology and the environmental sciences

Summary II

Quantile regression: extensions and alternative approaches

- Bayesian quantile regression: using the asymmetric Laplace distribution is appealing but several other approaches have been/are being developed – active research area
- GAMLSS methods
  - + are an attractive option for additive models for conditional quantiles
  - + avoid quantile crossing
  - + have a stable R implementation with extensive documentation
  - - make distributional assumptions/involve complex distributions
- R packages/code increasingly available on GitHub, *e.g.* qgam

Aknowledgements

Prof. Judy H. Wang,
GWU

Prof. Xuming He,
U Michigan

Prof. Roger Koenker,
U Illinois

Postscript

A recent less theoretical introductory reference

Statistical Modelling 2018; **18(3–4)**: 203–218

Quantile regression: A short story on how and why

Elisabeth Waldmann[1]
[1]Department of Medical Informatics, Biometry and Epidemiology, Friedrich-Alexander-Universität, Erlangen-Nürnberg,Germany.

Abstract: Quantile regression quantifies the association of explanatory variables with a conditional quantile of a dependent variable without assuming any specific conditional distribution. It hence models the quantiles, instead of the mean as done in standard regression. In cases where either the requirements for mean regression, such as homoscedasticity, are violated or interest lies in the outer regions of the conditional distribution, quantile regression can explain dependencies more accurately than classical methods. However, many quantile regression papers are rather theoretical so the method has still not become a standard tool in applications. In this article, we explain quantile regression from an applied perspective. In particular, we illustrate the concept, advantages and disadvantages of quantile regression using two datasets as examples.

Key words: quantile regression, tutorial article, additive regression models, gradient boosting

Received May 2017; **revised** December 2017; **accepted** December 2017