Summary:
Quantile Regression

github.com/s-saisw/HandbookOfLaborEconomics

September 6, 2019

- If dependent variable is a dummy, the mean describes the entire distribution. But many variables have continuous distribution, the mean cannot capture all the change. For instance, after program participation, wages can become more compressed.
  - Mean earning after job training may be the same but variance may have decreased.

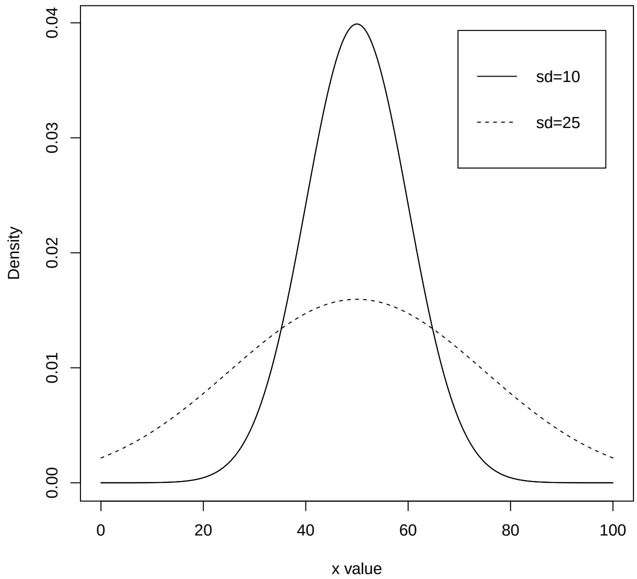

Figure 1: These two distributions have the same mean of 50 but different standard deviations.

```
setwd('/Summary/')

x <- seq(0, 100, 0.1)

hx <- dnorm(x, mean = 50, sd = 10)

plot(x, hx, type="l", lty=1, xlab="x value",
     ylab="Density")

lineType <- 1:2
s <- c(10, 25)

for (i in 1:2){
```

```
  lines(x, dnorm(x, mean = 50, sd = s[i]), lty = lineType[i])
}

labels <- c("sd=10", "sd=25")

legend("topright", inset=.05,
        labels, lwd=1, lty=c(1, 2))

dev.copy(pdf, "normalDist.pdf")
dev.off()
```

- Quantile regression can be used to see whether job training affects earning inequality as well as average earnings.

1 Quantile regression model

- Define conditional quantile function (CQF), $Q_\tau(Y_i|X_i)$.

$$Q_\tau(Y_i|X_i) = F_Y^{-1}(\tau|X_i) \tag{1}$$

Explanation: Suppose $F_Y(y|X_i) = \tau$. Then it follows that $F_Y^{-1}(\tau|X_i) = y = Q_\tau(Y_i|X_i)$. That is $Q_\tau(Y_i|X_i)$ is the value of Y that gives area τ on the cumulative distribution function conditional on X.

- CQF is the conditional-quantile function of the conditional expectation function (CEF). Recall that $\mathbb{E}(Y_i|X_i)$ is the best predictor of Y_i, given any function $m(\cdot)$, where $m : X \to Y$.

$$\mathbb{E}(Y_i|X_i) = \underset{m(X_i)}{\arg\min}\, \mathbb{E}[(Y_i - m(X_i))^2] \tag{2}$$

Explanation: We want to predict Y using X. But Y can be any function of X, $m(X_i)$. But we are lucky because we can use $\mathbb{E}(Y_i|X_i)$ as the minimizer, since $\mathbb{E}(Y_i|X_i)$ minimizes the mean-squared error.

- Similarly, CQF solves the following minimization problem,

$$Q_\tau(Y_i|X_i) = \underset{q(X)}{\arg\min}\, \mathbb{E}[\rho_\tau(Y_i - q(X_i))], \tag{3}$$

where the check loss function $\rho_\tau(u) = (\tau - 1(u \leq 0))u$, i.e.

$$\rho_\tau = \begin{cases} (\tau - 1)u & \text{when } u < 0 \\ 0 & \text{when } u = 0 \\ \tau u & \text{when } u > 0 \end{cases} \tag{4}$$

Example: when $\tau = 0.5$, $\rho_\tau = \frac{1}{2}|u|$. Then $\arg\min_{q(X)} \mathbb{E}[\rho_\tau(Y_i - q(X_i))] = \arg\min_{q(X)} \mathbb{E}[\frac{1}{2}|u|(Y_i - q(X_i))] = \arg\min_{q(X)} \mathbb{E}[|u|(Y_i - q(X_i))]$. In this case, $Q_\tau(Y_i|X_i)$ is conditional median.

```
setwd('/Summary/')

fun <- function(u,t){(u>0)*t*u + (u<=0)*(t-1)*u}

u <- seq(-1,1,0.01)
t <- c(0.25, 0.5, 0.75)
res <- mapply(fun, list(u), t)
lineType <- 1:3
```

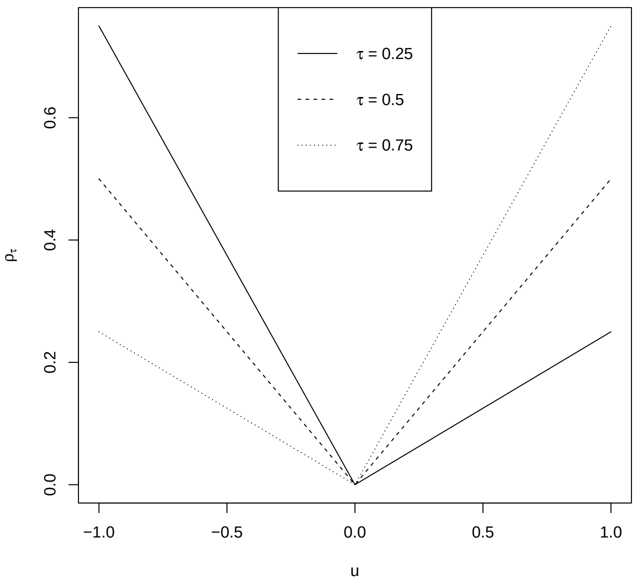

Figure 2: Check loss function looks like a check-mark.

```
matplot(u, res, type="l", lty=lineType, lwd=1,
        xlab="u", ylab=expression(rho[tau]), col = "black")

legend("top",
       legend = c(expression(paste(tau, " = ", 0.25)),
                                 expression(paste(tau, " = ", 0.5)),
                                 expression(paste(tau, " = ", 0.75))),
       lwd=1, lty = 1:3)

dev.copy(pdf, "checkLoss.pdf")
dev.off()
```

- With high dimensional X_i, it may be hard to estimate $Q_\tau(Y_i|X_i)$. We can assume $q(X_i)$ is linear and boil down the function to

$$\beta_\tau = \arg\min_{b \in \mathbb{R}^d} \mathbb{E}[\rho_\tau(Y_i - X_i'b)] \tag{5}$$

- If quantile coefficients are constant across quantiles, it is called a "location shift". This is related to the concept of homoscedasticity, i.e. constant variance across groups. In this case, the dependent variable Y_i can be described using

$$Y_i \sim N(X_i'\beta, \sigma_\epsilon^2). \tag{6}$$

  This implies

$$P[Y_i - X_i'\beta < \sigma_\epsilon \Phi^{-1}(\tau)|X_i] = \tau, \tag{7}$$

  where $\Phi^{-1}(\tau)$ is the inverse of standard normal CDF. Then we have

$$Q_\tau(Y_i|X_i) = X_i'\beta + \sigma_\epsilon \Phi^{-1}(\tau). \tag{8}$$

- If variance changes across groups, then σ_ϵ^2 is replaced by a function of X_i. It follows that

$$Y_i \sim N(X_i'\beta, \sigma^2(X_i)), \tag{9}$$

where $\sigma^2(X_i) = (\lambda' X_i)^2$. This implies

$$P[Y_i - X_i'\beta < (\lambda' X_i)\Phi^{-1}(\tau)|X_i] = \tau, \tag{10}$$

and

$$Q_\tau(Y_i|X_i) = X_i'\beta + (\lambda' X_i)\Phi^{-1}(\tau) = X_i'[\beta + \lambda\Phi^{-1}(\tau)]. \tag{11}$$

- Suppose our dependent variable is topcoded at c, i.e., $Y_{i,obs} = Y_i \cdot 1[Y_i < c]$. We use censored quantile regression estimator

$$\beta_\tau^c = \underset{b\in\mathbb{R}^d}{\arg\min}\,\mathbb{E}[1[X_i'\beta^c < c] \cdot \rho_\tau(Y_i - X_i'b)]. \tag{12}$$

2 Quantile regression approximation theorem

- Define the quantile regression specification error

$$\Delta_\tau(X_i, \beta_\tau) \equiv X_i'\beta_\tau - Q_\tau(Y_i|X_i) \tag{13}$$

- **Quantile regression approximation theorem:**
  Suppose that

  1. $f_Y(y|X_i)$ exists almost surely
  2. $\mathbb{E}[Y_i], E[Q_\tau(Y_I|X_i)]$, and $\mathbb{E}||X_i||$ are finite
  3. β_τ uniquely solves $\arg\min_{b\in\mathbb{R}^d}\mathbb{E}[\rho_\tau(Y_i - X_i'b)]$.

  Then
  $$\beta_\tau = \underset{b\in\mathbb{R}^d}{\arg\min}\,\mathbb{E}[w_\tau(X_i, b)\cdot\Delta_\tau^2(X_i, \beta_\tau)],$$
  where weighting function is defined as follows,
  $$w_\tau(X_i, b) = \int_0^1 (1-u) f_{\epsilon_\tau}(u\Delta_\tau(X_i, \beta_\tau)|X_i), \text{ and } \epsilon_i(\tau) = Y_i - Q_\tau(Y_I|X_i).$$

- Note that unlike OLS estimator β, quantile estimator β_τ has a subscript τ to indicate that it is estimated at some specific value τ.

- Caveats:

  1. Quantile coefficients tell us about effects on distributions, NOT on individuals.
  2. $Q_\tau(Y_i|X_i) = X_i'\beta_\tau$ does NOT imply $Q_\tau(Y_i) = Q_\tau(X_i)'\beta_\tau$.
     Note that in the case of expectation operator, $\mathbb{E}(Y_i|X_i) = X_i'\beta$ implies $Q_\tau(Y_i) = \mathbb{E}(X_i)'\beta$, by the law of iterated expectation.

3 Quantile treatment effect

- We assume

$$Q_\tau(Y_i|X_i, D_i, D_{1i} > D_{0i}) = \alpha_\tau D_i + X_i'\beta_\tau. \tag{14}$$

- Then it is natural to have

$$Q_\tau(Y_{1i}|X_i, D_{1i} > D_{0i}) - Q_\tau(Y_{0i}|X_i, D_{1i} > D_{0i}) = \alpha_\tau. \tag{15}$$

  α_τ is the effect of the training program "conditional" on X_i and the fact that i is a complier. However, α_τ does not tell us whether treatment changed the quantiles of "unconditional" distribution. Therefore, α_τ is NOT the conditional quantile treatment effects.

- QTE estimator is the sample analog of

$$(\alpha_\tau, \beta_\tau) = \arg\min_{a,b} \mathbb{E}[\rho_\tau(Y_i - aD_i - X_i'b)|D_{1i} > D_{0i}] = \arg\min_{a,b} \mathbb{E}[\kappa_i \rho_\tau(Y_i - aD_i - X_i'b)], \tag{16}$$

  where

$$\kappa_i = 1 - \frac{D_i(1 - z_i)}{1 - P(z_i = 1|X_i)} - \frac{(1 - D_i)z_i}{P(z_i = 1|X_i)}.$$

  Explanation: For i such that $D_i = z_i = 1$ or $D_i = z_i = 0$, $\kappa_i = 1$. That is, for compliers, QTE boils down to quantile estimator. When $D_i \neq z_i$, $\kappa_i < 0$.

- We can replace κ_i with the expected value of κ and simplify the above problem to

$$(\alpha_\tau, \beta_\tau) = \arg\min_{a,b} \mathbb{E}[\mathbb{E}[\kappa_i|Y_i, D_i, X_i]\rho_\tau(Y_i - aD_i - X_i'b)], \tag{17}$$

  where

$$\mathbb{E}[\kappa_i|Y_i, D_i, X_i] = 1 - \frac{D_i(1 - \mathbb{E}[z_i|Y_i, D_i = 1, X_i])}{1 - P(z_i = 1|X_i)} - \frac{(1 - D_i)\mathbb{E}[z_i|Y_i, D_i = 0, X_i]}{P(z_i = 1|X_i)} \tag{18}$$

- Practically,

  1. Probit z_i on Y_i and X_i separately for the $D_i = 0$ and $D_i = 1$ subsamples. Save the fitted values $\hat{z}_{i,D=1}$ and $\hat{z}_{i,D=0}$.
  2. Probit z_i on X_i for the whole sample. Save the fitted values $\hat{P}(z_i = 1|X_i)$.
  3. Use the fitted values to construct $\mathbb{E}[\kappa_i|Y_i, D_i, X_i]$ using (18). Note that D_i is observed.
  4. Use these κ_i's to weight quantile regressions and estimate using `qreg` command in Stata.