QUANTILES

* Quantiles are percentiles but not expressed as a percent; e.g. the 90^{th} percentile is the 0.9 quantile.

* A baby whose head circumference is at the .9 quantile has a head circumference as great as or greater than 90% of all babies.

* Random variables have quantiles.

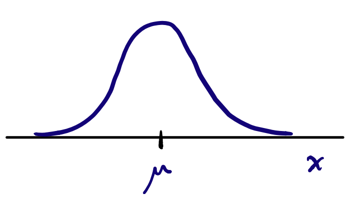
Let X be a continuous r.v. with a strictly increasing cdf F_X. Then the θ^{th} quantile of X is the value q which satisfies

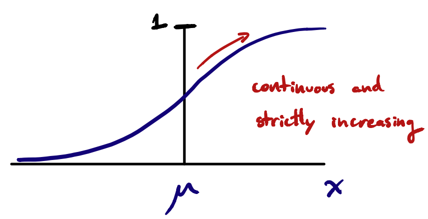
$$F_X(q) = \theta.$$

* So $X \le q$ with probability θ.

E.g. Suppose $X \sim \text{Normal}(\mu, \sigma^2)$.

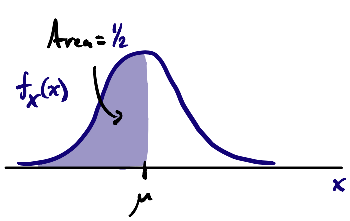
$$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \qquad F_X(x;\mu,\sigma^2) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right]dt$$

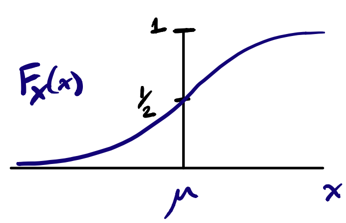




Find the 0.5-quantile (called the median).

We have $q = \mu$, since $P_X(X \le \mu) = \frac{1}{2}$.

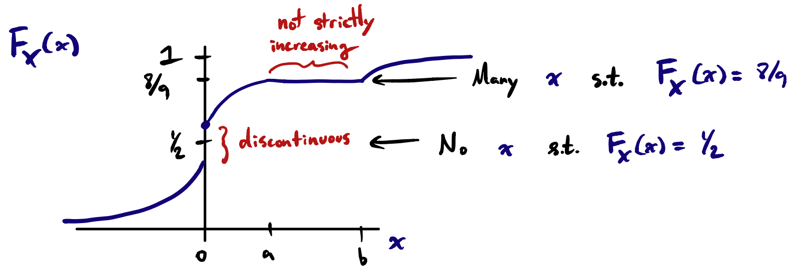




If F_X is not both continuous and strictly increasing, then q which satisfies $F_X(q) = \theta$

- may not exist
- may not be unique

E.g. Consider F_X like:



Better definition comes from the quantile function:

Defn: The quantile function Q_X of a r.v. X with cdf F_X is

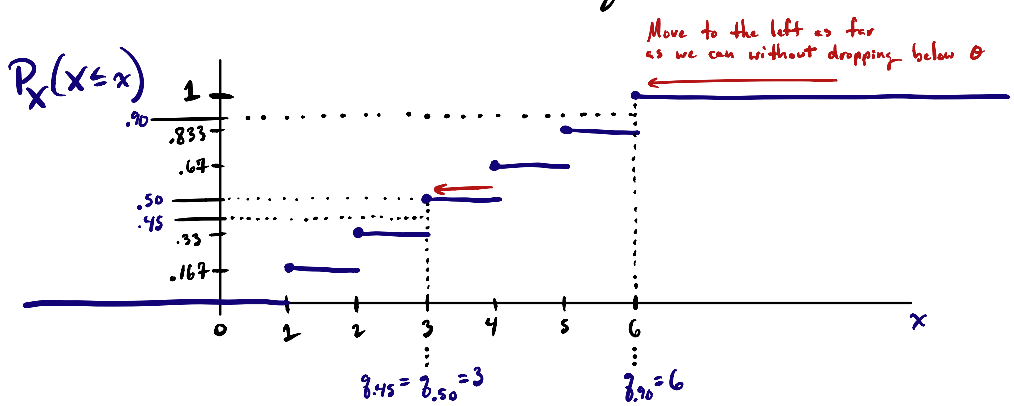
$$Q_X(\theta) = \inf\{x : F_X(x) \geq \theta\}$$

for $\theta \in (0,1)$.

The θ-quantile of X is defined as $Q_X(\theta)$.

E.g. Let X = up-face after rolling a 6-sided die.

Find the 0.45, 0.50, and 0.90 quantiles.



Interpretation: $q_{.90} = 6$: At least 90% of rolls are ≤ 6

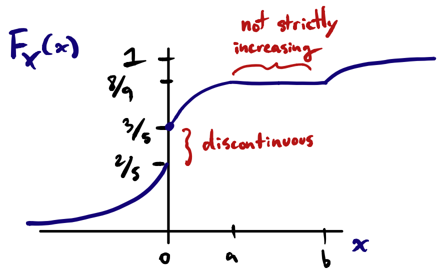
$q_{.50} = 3$: At least 50% of rolls are ≤ 3

$q_{.45} = 3$: At least 45% of rolls are ≤ 3

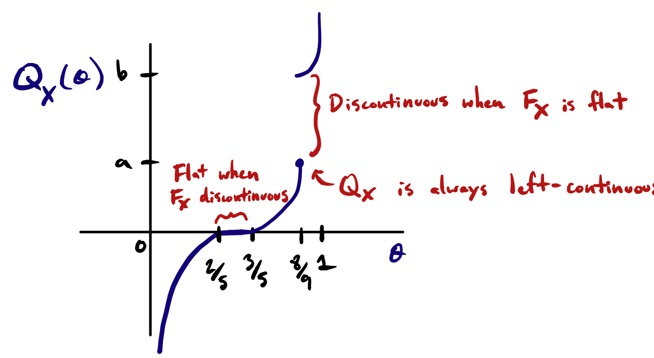
Remark: If F_X is continuous and strictly increasing, then $Q_X(\theta) = F_X^{-1}(\theta)$.

E.g. A cdf with a flat part and a discontinuity.

Cumulative distribution function



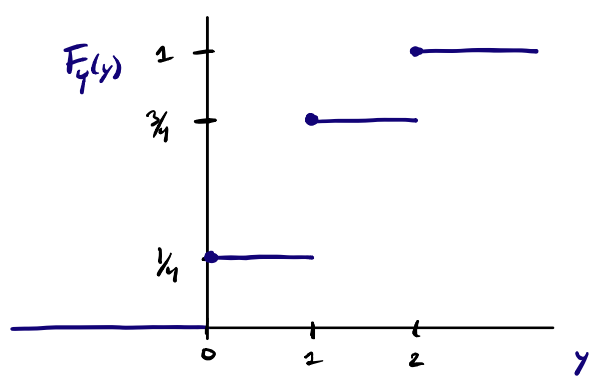
Quantile function



E.g. Find quantile function of $Y \sim \text{Binomial}(2, 1/2)$.

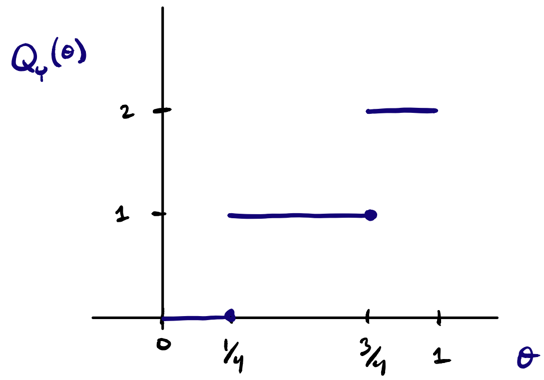
The cdf of Y is given by

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ 1/4, & 0 \leq y < 1 \\ 3/4, & 1 \leq y < 2 \\ 1, & 2 \leq y \end{cases}$$



The quantile function is given by

$$Q_Y(\theta) = \begin{cases} 2, & \frac{3}{4} < \theta < 1 \\ 1, & \frac{1}{4} < \theta \le \frac{3}{4} \\ 0, & 0 < \theta \le \frac{1}{4} \end{cases}$$



Note: $Q_X : (0,1) \to \mathcal{X}$, and Q_X is left-continuous.

E.g. $\underline{Q_X \text{ for Exponential}(\lambda)}$:

Let X = time in years between eruptions of a volcano

Assume $X \sim \frac{1}{\lambda} \exp\left[-\frac{x}{\lambda}\right] \mathbb{1}(0 < x < \infty)$.

(a) Find Q_X.

We have

$$F_X(x;\lambda) = \begin{cases} 1 - \exp\left[-\frac{x}{\lambda}\right] & 0 < x < \infty \\ 0 & -\infty < x \le 0 \end{cases} \leftarrow$$

Just focus on support. Here the cdf is continuous and strictly increasing.

Solve $F_X(z\,;\lambda) = \theta$ for z :

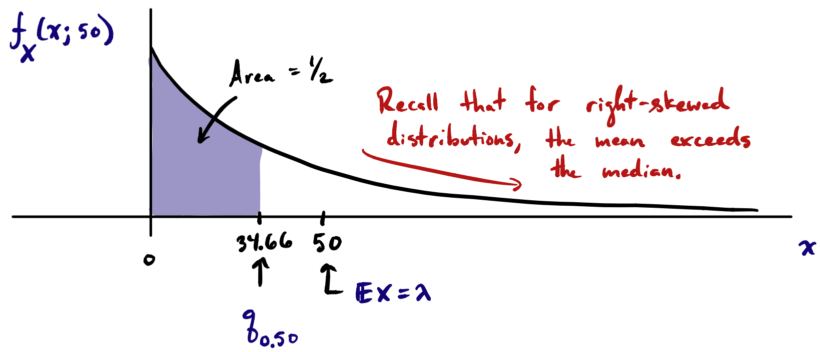
$$1 - \exp\left[-\frac{z}{\lambda}\right] = \theta \iff 1 - \theta = \exp\left[-\frac{z}{\lambda}\right]$$

$$\Leftrightarrow \quad -\lambda \log(1-\theta) = q$$

So $\quad Q_X(\theta; \lambda) = -\lambda \log(1-\theta) \quad$ for $\quad \theta \in (0,1)$.

(b) Suppose $\lambda = 50$ yrs and find the median time between eruptions.

$$Q_X(.50; 50) = -50 \log(1-.50) = 34.66$$



E.g. $\underline{Q_X \text{ for Empirical Distribution } (x_1, \ldots, x_n)}$:

Recall the pmf

$$p_X(x; x_1, \ldots, x_n) = \begin{cases} \frac{1}{n} & \text{for } x \in \{x_1, \ldots, x_n\} \\ 0 & \text{otherwise,} \end{cases}$$

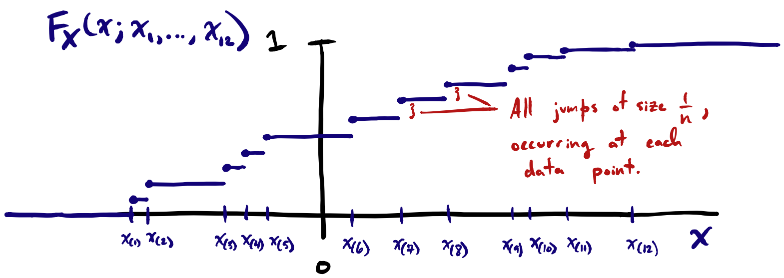
which gives the cdf

$$F_X(x; x_1, \ldots, x_n) = \frac{\#\{x_1, \ldots, x_n \le x\}}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(x_i \le x).$$

Denote by $x_{(1)}, \ldots, x_{(n)}$ the data points when reordered such that

$$x_{(1)} < x_{(2)} < \cdots < x_{(n)}.$$

The cdf of the empirical distribution based on 12 "data points" looks like:



Then $$Q_X(\theta) = \inf\left\{ x : \frac{\#\{x_1, \dots, x_n \le x\}}{n} \ge \theta \right\}$$

$$= \min\left\{ x_i : \frac{\#\{x_1, \dots, x_n \le x_i\}}{n} \ge \theta \right\}$$

$$= x_{(\lceil n\theta \rceil)} \qquad \text{for} \qquad \theta \in (0,1),$$

where $\lceil \cdot \rceil$ is the ceiling function: $\lceil x \rceil = j$ if $j-1 \; x \le j$, j an integer.

E.g. Consider the empirical distribution based on the data points

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1.98	1.01	0.86	−0.15	1.55	0.20	−0.90	1.10	−0.37	−0.77
$x_{(10)}$	$x_{(7)}$	$x_{(6)}$	$x_{(4)}$	$x_{(9)}$	$x_{(5)}$	$x_{(1)}$	$x_{(8)}$	$x_{(3)}$	$x_{(2)}$

Let $X \sim p_X(x; x_1, \dots, x_n)$.

Then $$q_{0.25} = x_{(\lceil 10 \cdot 0.25 \rceil)} = x_{(3)} = -0.37$$

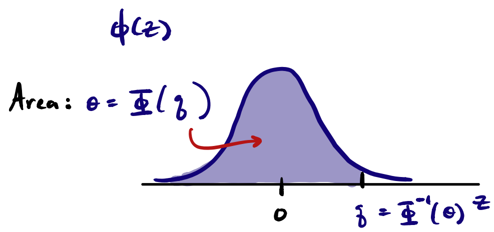
$$q_{0.50} = x_{(\lceil 10 \cdot 0.5 \rceil)} = x_{(5)} = 0.20$$

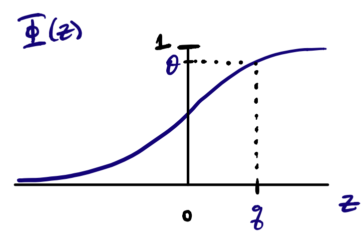
$$q_{0.75} = x_{(\lceil 10 \cdot 0.75 \rceil)} = x_{(8)} = 1.10$$

E.g. Q_X for Normal(μ, σ^2):

Begin with $Z \sim \text{Normal}(0,1)$.

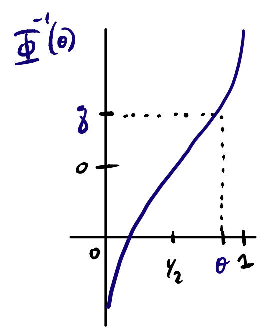
The cdf of Z is $\Phi(z) = \int_{-\infty}^{z} \phi(t)\,dt$, $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$.





* $\Phi(z)$ is continuous and strictly increasing, so

$$Q_Z(\theta) = \Phi^{-1}(\theta)$$



* $\Phi^{-1}(\theta)$ does not have a simple expression.
  We use tables / computer.

  In R: qnorm(theta)

* If $X \sim \text{Normal}(\mu, \sigma^2)$ then $Q_X(\theta) = \mu + \sigma Q_Z(\theta)$

Derivation: Let q be the θ^{th} quantile of X. Then

$$\theta = P_X(X \le q) = P_X\left(\frac{X-\mu}{\sigma} \le \frac{q-\mu}{\sigma}\right)$$

$$= P_Z\left(Z \le \frac{q-\mu}{\sigma}\right)$$

$$= \Phi\left(\frac{q-\mu}{\sigma}\right)$$

$$\iff \quad \Phi^{-1}(\theta) = \frac{q-\mu}{\sigma}$$

$$\iff \quad q = \mu + \sigma\,\Phi^{-1}(\theta)$$

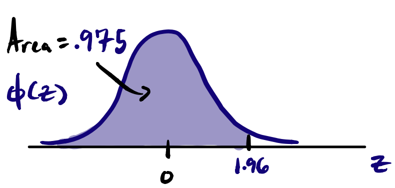
So $\quad Q_X(\theta) = \mu + \sigma\, Q_Z(\theta).$

* Therefore, to get quantiles of X, we transform quantiles of Z.

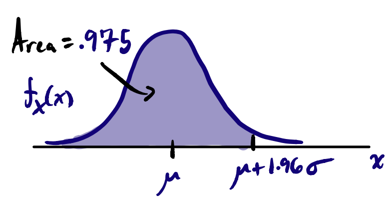
$$Q_X(0.975) = \mu + Q_Z(0.975)\,\sigma = \mu + \Phi^{-1}(0.975)\,\sigma = \mu + 1.96\sigma$$

Obtain from a table or qnorm(0.975) in R

(i) Get Normal(0,1) quantile



(ii) scale and shift to get Normal(μ, σ^2) quantile



COMPUTING QUANTILES

R has for every distribution the four functions

- r___ : generate realizations from ___ distribution
- d___ : compute value of pmf or pdf of ___ distribution
- p___ : compute value of cdf of ___ distribution
- q___ : compute quantile of ___ distribution

For quantiles