

Math 4261: Mathematical Statistics I

Lecture 7: Order statistics and quantiles

References: HMC 4.4, 5.1–5.3

DeGroot–Schervish 7.8–7.9, 10.6–10.7

Wasserman, *All of Statistics*, 7.1–7.2

Fall 2026

Outline

Finishing up CLT, Slutsky

Order statistics

Statistical estimation of quantiles

Recap: Modes of convergence

Convergence in probability

$$T_n \xrightarrow{p} T \iff \mathbb{P}[|T_n - T| > \epsilon] \longrightarrow 0 \quad \text{for every } \epsilon > 0.$$

Convergence in distribution

$$T_n \xrightarrow{d} T \iff F_{T_n}(x) \longrightarrow F_T(x)$$

at every x where F_T is continuous.

- ▶ $T_n \xrightarrow{p} T$ implies $T_n \xrightarrow{d} T$. Converse false in general.
- ▶ For a **constant** c , $T_n \xrightarrow{d} c$ is equivalent to $T_n \xrightarrow{p} c$.

Recap: Law of large numbers

Let $X_1, \dots, X_n$ be i.i.d., with mean μ and variance $\sigma^2 < \infty$.

Law of large numbers (LLN)

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu.$$

Statistical implication: The sample mean is a **consistent estimator** of the population mean.

Example: consistency of the sample variance

Assuming additionally $\mathbb{E}[X_1^4] < \infty$, we showed

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 \xrightarrow{p} \sigma^2.$$

Recap: Central limit theorem

Let $X_1, \dots, X_n$ be i.i.d., with mean μ and variance $0 < \sigma^2 < \infty$.

Central limit theorem (CLT)

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Equivalently,

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

For large n , $\bar{X}_n$ is approximately $\mathcal{N}(\mu, \sigma^2/n)$ distributed.

The LLN says the sampling error vanishes. The CLT describes its approximately Gaussian shape and its $1/\sqrt{n}$ scale.

Outline

Finishing up CLT, Slutsky

Order statistics

Statistical estimation of quantiles

Example: Running proportion of heads

Let $X_1, X_2, \dots$ be i.i.d. Bernoulli(1/2). LLN tells us

$$\bar{X}_n \xrightarrow{n \rightarrow \infty} \frac{1}{2} \quad \text{in probability}$$

CLT gives more. Use CLT to approximate the chance that $\bar{X}_n$ will be in the interval $[0.4, 0.6]$.

$$\text{CLT: } \bar{X}_n \stackrel{d}{\approx} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) = \mathcal{N}\left(\frac{1}{2}, \frac{1}{4n}\right)$$

$$\mu = \mathbb{E}(X) = \frac{1}{2}, \quad \sigma^2 = \text{Var } X = \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4}$$

Suppose $n=100$, then

$$\text{SD} = \frac{1}{20}$$

$$P\left[0.4 \leq \bar{X}_{100} \leq 0.6\right] \approx P\left[0.4 \leq \mathcal{N}\left(\frac{1}{2}, \frac{1}{400}\right) \leq 0.6\right]$$

$$= P\left[\frac{0.4 - 0.5}{1/20} \leq Z \leq \frac{0.6 - 0.5}{1/20}\right]$$

$$= P[-2 \leq Z \leq 2] \approx 95\%$$

check that LLN

$$\text{giving } \leq \frac{\sigma^2}{n(0.1)^2} = \frac{1}{400(0.1)^2}$$

= 0.25 is too conservative

Explanation of CLT

CLT: Let $X_1, \dots, X_n$ i.i.d. with common mean $\mu = 0$ and variance $\sigma^2 = 1$. Then

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Fact: Closeness of MGFs implies closeness of distribution.

Goal:

$$M_{\frac{S_n}{\sqrt{n}}}(t) \approx M_{\mathcal{N}(0,1)}(t) = e^{t^2/2}$$

Explanation of CLT

$$S_n = X_1 + \dots + X_n$$

$$\text{CLT: } M_{\frac{S_n}{\sqrt{n}}}(t) \approx e^{t^2/2}$$

① Independence of X_i 's $\Rightarrow$ MGF of sum factorizes.

$$M_{\frac{S_n}{\sqrt{n}}}(t) = \mathbb{E} \exp\left(t \frac{S_n}{\sqrt{n}}\right) = M_{S_n}\left(\frac{t}{\sqrt{n}}\right)$$

$$= M_{X_1 + \dots + X_n}\left(\frac{t}{\sqrt{n}}\right) = M_{X_1}\left(\frac{t}{\sqrt{n}}\right) \cdot M_{X_2}\left(\frac{t}{\sqrt{n}}\right) \cdots M_{X_n}\left(\frac{t}{\sqrt{n}}\right)$$

$$= \left(M_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Explanation of CLT

$$\mathbb{E}X = 0,$$

s.s.

$$1 = \text{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 \xrightarrow{0}$$

$$\Rightarrow \mathbb{E}X^2 = 1$$

$$\text{CLT: } M_{\frac{S_n}{\sqrt{n}}}(t) \approx e^{t^2/2}$$

$$\textcircled{1} M_{\frac{S_n}{\sqrt{n}}}(t) = \left(M_{X_1}\left(\frac{t}{\sqrt{n}}\right) \right)^n$$

② Taylor expand M_{X_1} around 0:

$$M_{X_1}\left(\frac{t}{\sqrt{n}}\right) = \underbrace{M_{X_1}(0)}_1 + \underbrace{\frac{t}{\sqrt{n}} M'_{X_1}(0)}_{\mathbb{E}X=0} + \frac{t^2}{2n} \underbrace{M''_{X_1}(0)}_{\mathbb{E}X^2=1} + \underbrace{O\left(\frac{t^3}{n^{3/2}}\right)}_{o\left(\frac{1}{n}\right)}$$

$$\textcircled{3} M_{\frac{S_n}{\sqrt{n}}}(t) = \left(1 + \frac{t^2/2}{n} + o\left(\frac{1}{n}\right) \right)^n$$
$$\rightarrow e^{t^2/2} = M_{N(0,1)}(t)$$

Calculus:

$$\left(1 + \frac{a}{n}\right) \rightarrow e^a$$

Slutsky's lemma

Slutsky's lemma

Suppose $T_n \xrightarrow{d} T$ and $A_n \xrightarrow{d} a$, where a is a **constant**. Then

$$T_n + A_n \xrightarrow{d} T + a,$$

$$A_n T_n \xrightarrow{d} aT,$$

$$\frac{T_n}{A_n} \xrightarrow{d} \frac{T}{a} \quad \text{if } a \neq 0.$$

- ▶ Independence is not required between T_n and A_n
- ▶ Convergence in distribution is **not** generally closed under sum or product.

Example

Let $X_1, \dots, X_n$ be i.i.d. with mean μ and variance $0 < \sigma^2 < \infty$. The CLT gives

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Here we have scaled by the **standard error** of $\overline{X}_n$: its standard deviation $\sigma/\sqrt{n}$.

Suppose the standard deviation is **unknown**. We can estimate it from the data by

$$\widehat{\text{SE}}(\overline{X}_n) = \frac{\hat{\sigma}_n}{\sqrt{n}}, \quad \text{where} \quad \hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X}_n)^2.$$

Does replacing σ by $\hat{\sigma}_n$ preserve the limiting normal distribution?

Example

For i.i.d. data with $0 < \sigma^2 < \infty$, we have proved

$$\hat{\sigma}_n \xrightarrow{p} \sigma.$$

CMT gives $\sigma/\hat{\sigma}_n \xrightarrow{p} 1$. Hence CLT and Slutsky give

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{\sigma}_n} = \frac{\sigma}{\hat{\sigma}_n} \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1).$$

In other words, it does preserve the normal limit!

$$\boxed{\frac{\bar{X}_n - \mu}{\widehat{\text{SE}}(\bar{X}_n)} \xrightarrow{d} \mathcal{N}(0, 1).}$$

Outline

Finishing up CLT, Slutsky

Order statistics

Statistical estimation of quantiles

Order statistics

Let $X_1, \dots, X_n$ be an i.i.d. sample with CDF F . Arrange the observations in increasing order:

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}.$$

$X_{(k)}$ is the *kth order statistic*.

- ▶ $X_{(1)} = \min_i X_i$
- ▶ $X_{(n)} = \max_i X_i$.
- ▶ If $n = 2m + 1$, the sample median is $X_{(m+1)}$.

The ordered observations are generally *dependent*, even though the original observations are independent.

Distributions of $X_{(1)}$ and $X_{(n)}$

Let $X_1, \dots, X_n$ be i.i.d. with CDF F .

Find the CDFs of the **maximum** $X_{(n)}$ and the **minimum** $X_{(1)}$.

for x in the support of F ,

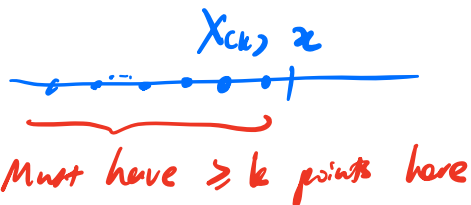
$$\begin{aligned} P[X_{(n)} \leq x] &= P[X_1 \leq x, \dots, X_n \leq x] \\ &\stackrel{\text{indep}}{=} P[X_1 \leq x] \cdot \dots \cdot P[X_n \leq x] \\ &= (F(x))^n \end{aligned}$$

$$\begin{aligned} P[X_{(1)} > x] &= P[X_1 > x, \dots, X_n > x] \\ &= [1 - F(x)]^n \end{aligned}$$

$$\text{Hence } F_{X_{(1)}}(x) = 1 - [1 - F(x)]^n$$

Distribution of $X_{(k)}$: CDF

Let $X_1, \dots, X_n$ be i.i.d. with CDF F . Fix a threshold x and let

$$N_x = \sum_{i=1}^n \mathbf{1}\{X_i \leq x\}.$$


Find the distribution of N_x , then use it to find $\mathbb{P}[X_{(k)} \leq x]$.

$$N_x \sim \text{Bin}(n, F(x))$$

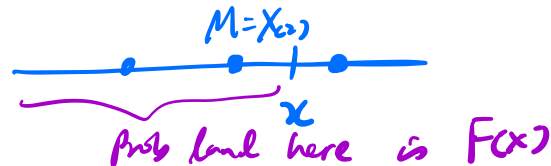
$$\text{The link: } \{X_{(k)} \leq x\} = \{N_x \geq k\}$$

$$\mathbb{P}[X_{(k)} \leq x] = \mathbb{P}[N_x \geq k] = \sum_{j=k}^n \binom{n}{j} F(x)^j [1 - F(x)]^{n-j}$$

Example: Median of three

Let $M = X_{(2)}$ be the **sample median** of three independent observations from a population with continuous CDF F and density f .

Find the CDF and density of M .



$M \leq x \iff$ at least two obs are $\leq x$

$$\begin{aligned} P[M \leq x] &= F(x)^3 + \binom{3}{2} (F(x))^2 (1 - F(x)) \\ &= 3F(x)^2 - 2F(x)^3 \end{aligned}$$

Differentiating

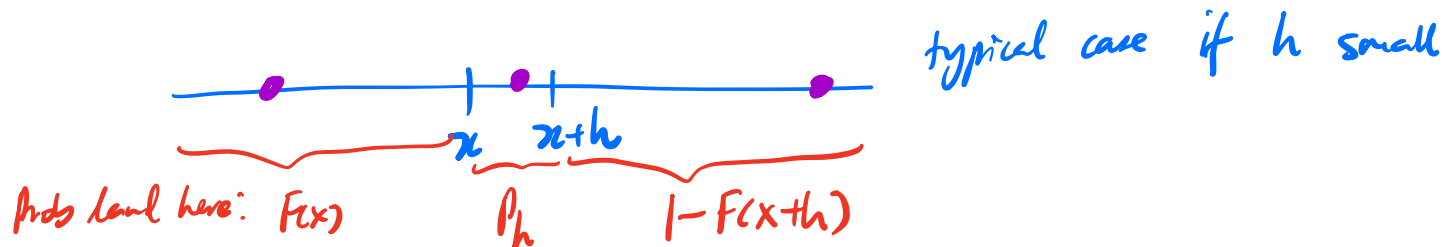
$$f_M(x) = 6F(x)[1 - F(x)]f(x)$$

Example: Median of three, Proof #2

OPTIONAL

Again, $M = X_{(2)}$ is the median of three i.i.d. observations with CDF F and continuous density f . Write F_M for the CDF of M . Then

$$f_M(x) = \lim_{h \downarrow 0} \frac{F_M(x+h) - F_M(x)}{h} \stackrel{= P_h}{=} \lim_{h \downarrow 0} \frac{\mathbb{P}[x < M \leq x+h]}{h}.$$

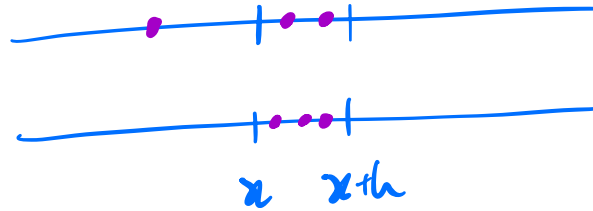


$$\mathbb{P}[x < M \leq x+h] = 3! \cdot F(x) P_h [1 - F(x+h)] + \text{Remainder}(h)$$

Example: Median of three, Proof #2

OPTIONAL

Here



Remainder(h)

$$= p_h^3 + \binom{3}{2} p_h^2 (1-p_h) = p_h^3 + 3p_h^2 - 3p_h^3 \leq 4p_h^2$$

$$\text{Note } \frac{\text{Remainder}(h)}{h} \leq \frac{4p_h^2}{h} = 4 \underbrace{\left(\frac{p_h}{h}\right)^2}_{\rightarrow (f(x))^2} \frac{1}{h} \rightarrow 0$$

Hence

$$f_M(x) = \lim_{h \downarrow 0} \frac{P[x \leq M \leq x+h]}{h} = 6 F(x) f(x) [1-F(x)]$$

Distribution of $X_{(k)}$: PDF

Suppose $X_1, \dots, X_n$ are i.i.d. with continuous density f and CDF F .

PDF of $X_{(k)}$

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} F(x)^{k-1} [1 - F(x)]^{n-k} f(x).$$

This follows by **differentiating the CDF**, or by the same **counting argument** used for the median.

Outline

Finishing up CLT, Slutsky

Order statistics

Statistical estimation of quantiles

Overview

Under a nice i.i.d. setting, LLN and tools like continuous mapping theorem gave

$$\overline{X}_n \xrightarrow{p} \mu, \quad \text{and} \quad \widehat{\sigma}_n^2 \xrightarrow{p} \sigma^2.$$

The CLT describes more: the error fluctuations of the sample mean.

Other statistics of interest are:

- ▶ The **maximum** and **minimum** of the sample
- ▶ The **median** of the sample

All these are trying to estimate a **population quantile**: e.g. q_1 for the max, and $q_{0.5}$ for the median.

We ask the same two questions:

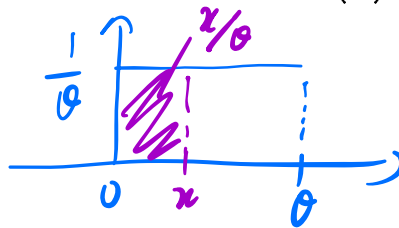
1. Do these sample quantiles converge to its target?
2. How large is its sampling error?

Their sampling distributions will coincide with the order- statistic laws we just found!

Example: Uniform

Suppose $X_1, \dots, X_n$ are i.i.d. $\text{Unif}(0, \theta)$, where $\theta > 0$ is unknown. The upper endpoint is $q_1 = \theta$. A natural estimator is $\hat{\theta}_n = X_{(n)}$.

Find the CDF of $\hat{\theta}_n$.



From $X_{(n)}$ CDF ,

$$P[\hat{\theta}_n \leq x] = [F(x)]^n = \begin{cases} 0 & , x < 0 \\ (x/\theta)^n & , 0 \leq x \leq \theta \\ 1 & , x > \theta \end{cases}$$

Example: Uniform

Continue with $\hat{\theta}_n = X_{(n)}$ for a $\text{Unif}(0, \theta)$ sample. Since $\hat{\theta}_n \leq \theta$ with probability one, its error is $\theta - X_{(n)}$. $\hat{\theta}_n \leq \theta$ always

Show that $\hat{\theta}_n$ is **consistent** for θ .

$$\begin{aligned} P[|\hat{\theta}_n - \theta| > \varepsilon] &= P[\hat{\theta}_n < \theta - \varepsilon] \\ &= \left(\frac{\theta - \varepsilon}{\theta}\right)^n = \left(1 - \frac{\varepsilon}{\theta}\right)^n \longrightarrow 0 \quad \text{for } 0 < \varepsilon < \theta \end{aligned}$$

For $\varepsilon \geq \theta$, the prob. is 0. Thus $\hat{\theta}_n \xrightarrow{P} \theta$.

How large must n be to have at least 95% probability of estimating θ to within 5% of its true value?

$$\begin{aligned} P[|\hat{\theta}_n - \theta| \leq 0.05\theta] &= P[0.95\theta \leq \hat{\theta}_n \leq 1.05\theta] \\ &= P[\hat{\theta}_n \geq 0.95\theta] = 1 - 0.95^n \stackrel{\text{Goal}}{\geq} 0.95 \end{aligned}$$

Solve: $\boxed{n \geq 59}$

Example: Uniform

For a $\text{Unif}(0, \theta)$ population, $\mathbb{E}[X_i] = \theta/2$. The LLN gives

$$2\overline{X}_n \xrightarrow{p} \theta.$$

We just saw that $X_{(n)} \xrightarrow{p} \theta$ as well.

Thus there are two consistent estimators for θ . Consistency alone does not tell us which estimator has smaller error.

Understanding the **sampling distributions** of $2\overline{X}_n$ and $X_{(n)}$ would let us further compare their efficiency.

Uniform order statistics and the Beta distribution

Suppose $U_1, \dots, U_n$ are i.i.d. $\text{Unif}(0, 1)$. Then

$$f_{U_{(k)}}(u) = \frac{n!}{(k-1)!(n-k)!} u^{k-1} (1-u)^{n-k}, \quad 0 < u < 1,$$

and the density is zero elsewhere.

Beta distribution

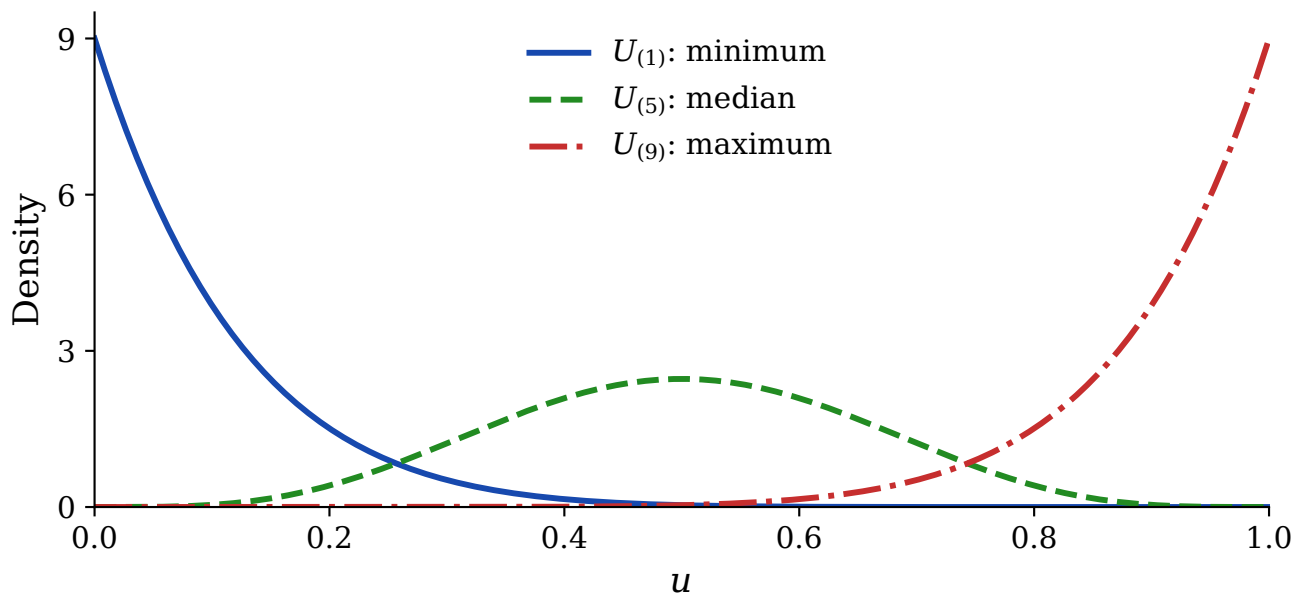
$$U_{(k)} \sim \text{Beta}(k, n+1-k).$$

The $\text{Beta}(a, b)$ density is proportional to $u^{a-1}(1-u)^{b-1}$ on $(0, 1)$, for $a, b > 0$.

The mean and variances are

$$\mathbb{E}[U_{(k)}] = \frac{k}{n+1}, \quad \text{Var}[U_{(k)}] = \frac{k(n+1-k)}{(n+1)^2(n+2)}.$$

Uniform order statistics and the Beta distribution



$U_{(1)}$, $U_{(5)}$, and $U_{(9)}$ are the minimum, median, and maximum. Their means are 0.1, 0.5, and 0.9.

Estimating quantiles

More generally, consider a population quantile $q_\alpha = F^{-1}(\alpha)$ of interest, such as the median or 90th percentile.

Estimate the unknown $F(x) = \mathbb{P}[X_i \leq x]$ by:

Empirical cumulative distribution function

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_i \leq x\} = \frac{N_x}{n}.$$

The associated distribution puts mass $1/n$ at each observation. A value repeated r times will create a jump in the CDF of size r/n .

Idea: find where $\widehat{F}_n$ first reaches α to estimate the population q_α .

Population and sample quantiles

Recall our definition for quantiles: for $0 < \alpha < 1$,

$$q_\alpha = F^{-1}(\alpha) = \inf\{x : F(x) \geq \alpha\}.$$

We use the *plug-in* estimator $\widehat{F}_n$ to replace the unknown F :

Sample quantile:

$$\widehat{q}_{n,\alpha} = \widehat{F}_n^{-1}(\alpha) = \inf\{x : \widehat{F}_n(x) \geq \alpha\} = X_{(\lceil n\alpha \rceil)},$$

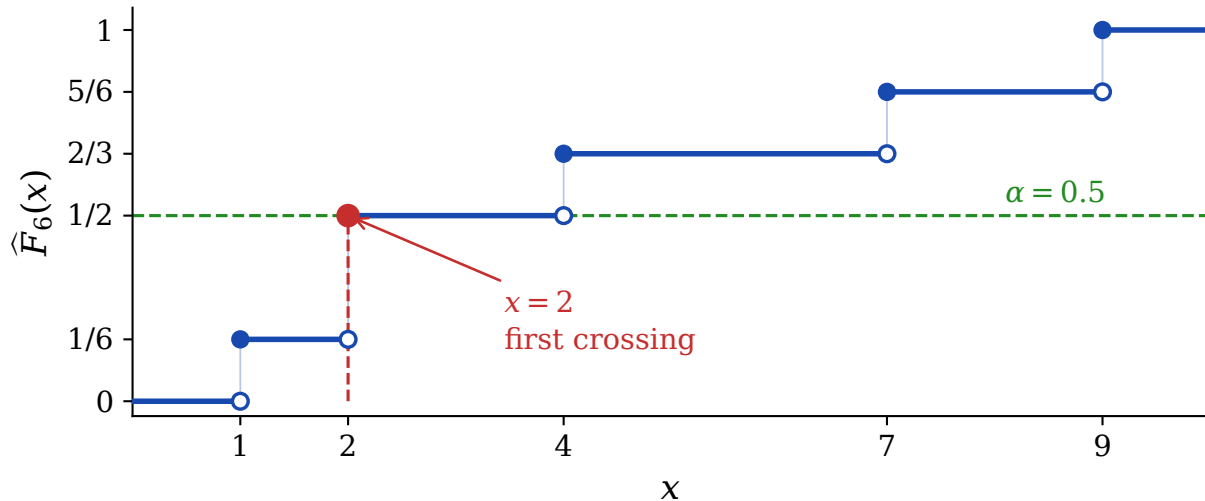
where $\lceil \cdot \rceil$ is the ceiling function.

Thus $\widehat{q}_{n,\alpha}$ is the first time the empirical CDF $\widehat{F}_n$ crosses α .

E.g. For **median**, take $\alpha = 0.5$: $q_{0.5}$ is the population median, and $\widehat{q}_{n,0.5}$ is our estimator. The sampling distribution is exactly the law of $X_{(\lceil n/2 \rceil)}$.

Example

Estimate the population median $q_{0.5}$ from the observations 1, 2, 2, 4, 7, 9.



Scanning from the left, $\hat{F}_6$ first reaches $1/2$ at $x = 2$.¹

$$\hat{q}_{6,0.5} = X_{(\lceil 6/2 \rceil)} = X_{(3)} = 2.$$

¹The usual even-sample convention gives $(X_{(3)} + X_{(4)})/2 = (2 + 4)/2 = 3$. For odd sample sizes, the two conventions agree.