Lecture 2, part 1:

Estimating the CDF and statistical functionals

Estimating the CDF and statistical functionals

- We will consider the problem of nonparametric estimation of the CDF
- Then we will estimate statistical functionals (functions of the CDF) – such as mean, variance, and correlation
- This non-parametric method for estimating functionals is called the **plug-in method**

The Empirical Distribution Function

- Let $X_1, \ldots, X_n \sim F$ be an IID sample from F – some unknown distribution function on $\mathbb{R}$

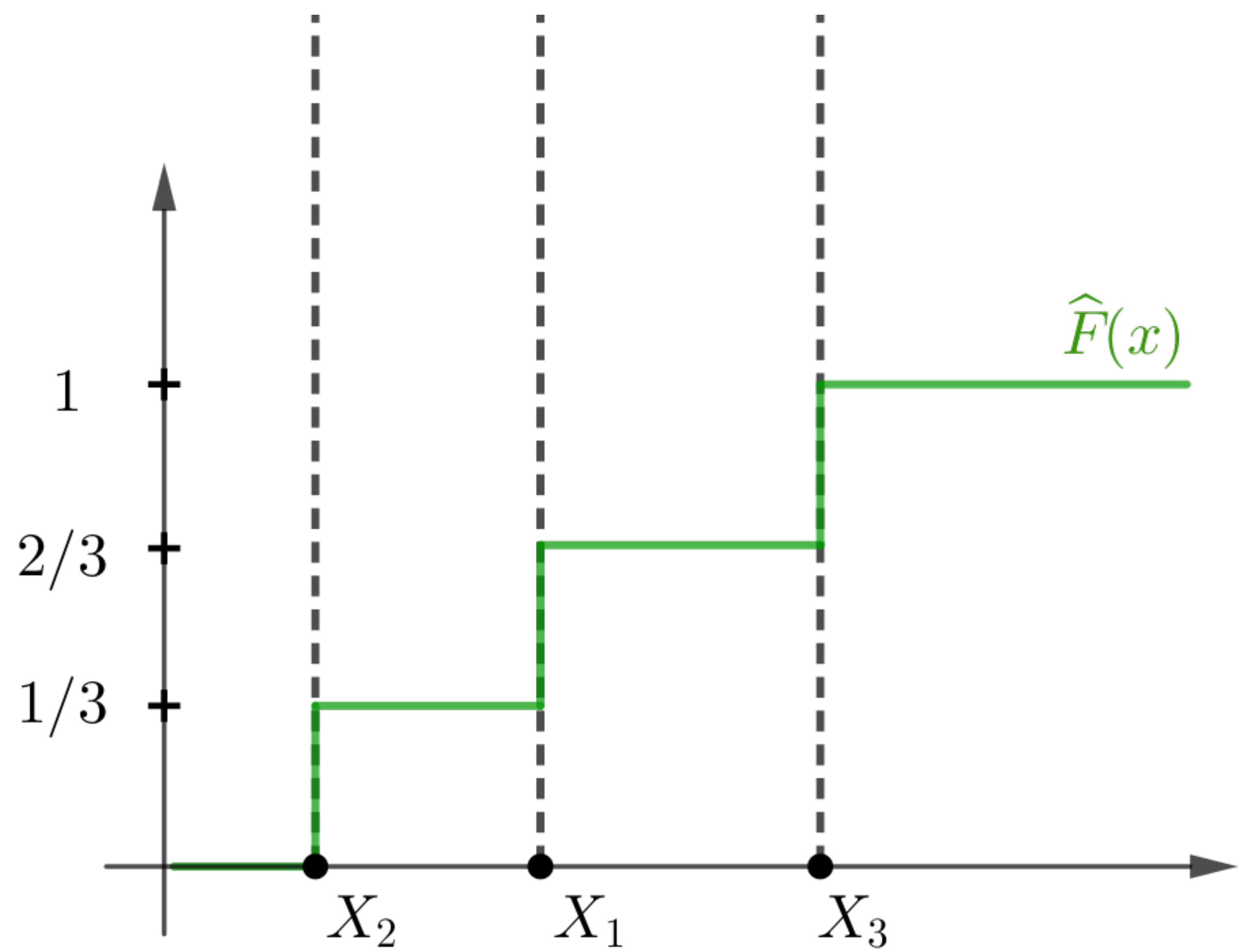
- The **empirical distribution function** $\widehat{F_n}$ is the CDF that puts mass $1/n$ at each data point X_i

- So, formally,

$$\widehat{F_n}(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x) \quad \text{where} \quad I(X_i \leq x) = \begin{cases} 0 & \text{if} \quad x < X_i \\ 1 & \text{if} \quad x \geq X_i \end{cases}$$

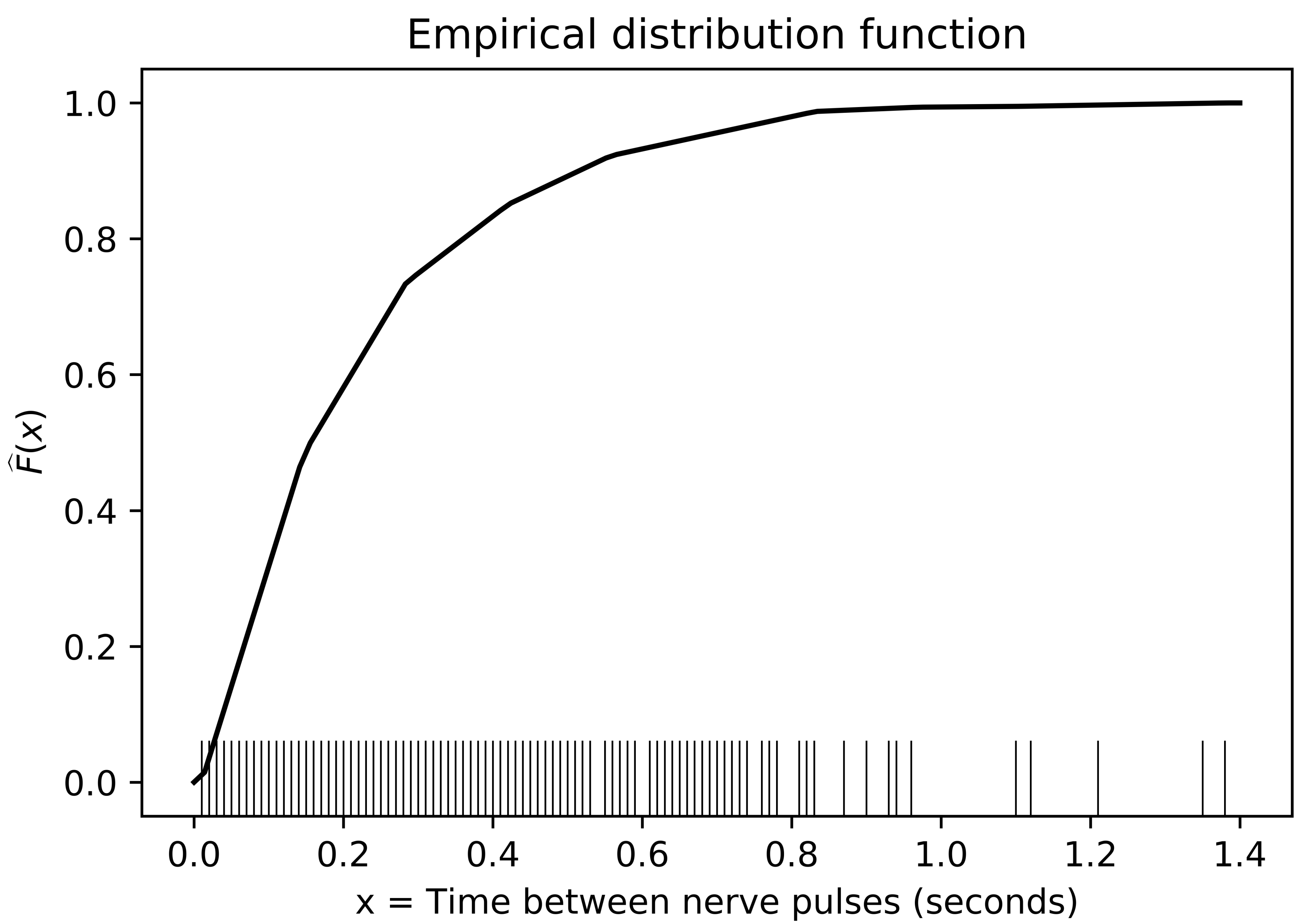
The Empirical Distribution Function

- The **empirical distribution function** $\widehat{F}_n$ is the CDF that puts mass $1/n$ at each data point X_i



The Empirical Distribution Function

- **Example 1:** Here's some data for ~800 waiting times between successive pulses along a nerve fiber (Cox and Lewis, 1966).



The Empirical Distribution Function

For the empirical CDF, one has:

- **Theorem 1**: For any fixed value of x,

$$\mathbb{E}\left(\widehat{F_n}(x)\right) = F(x)$$

$$\mathbb{V}\left(\widehat{F_n}(x)\right) = \frac{1}{n}F(x)\left(1 - F(x)\right)$$

$$\text{MSE} = \frac{1}{n}F(x)\left(1 - F(x)\right) \to 0 \quad \text{as} \quad n \to \infty$$

so $\widehat{F_n}(x) \xrightarrow{P} F(x)$

The Empirical Distribution Function

Moreover,

- **Theorem 2:** (Glivenko-Cantelli).

$$\sup_x \left| \widehat{F}_n(x) - F(x) \right| \overset{a.s.}{\to} 0 \quad \text{(which implies convergence in P)}$$

And, more practically important,

- **Theorem 3:** (Dvoretzky-Kiefer-Wolfowitz, DKW) – for $\varepsilon > 0$

$$\mathbb{P}\left(\sup_x \left| \widehat{F}_n(x) - F(x) \right| > \varepsilon \right) \leq 2\, e^{-2n\varepsilon^2}$$

The Empirical Distribution Function

The DKW inequality (Theorem 3) allows to construct **confidence intervals** for $\widehat{F}_n(x)$

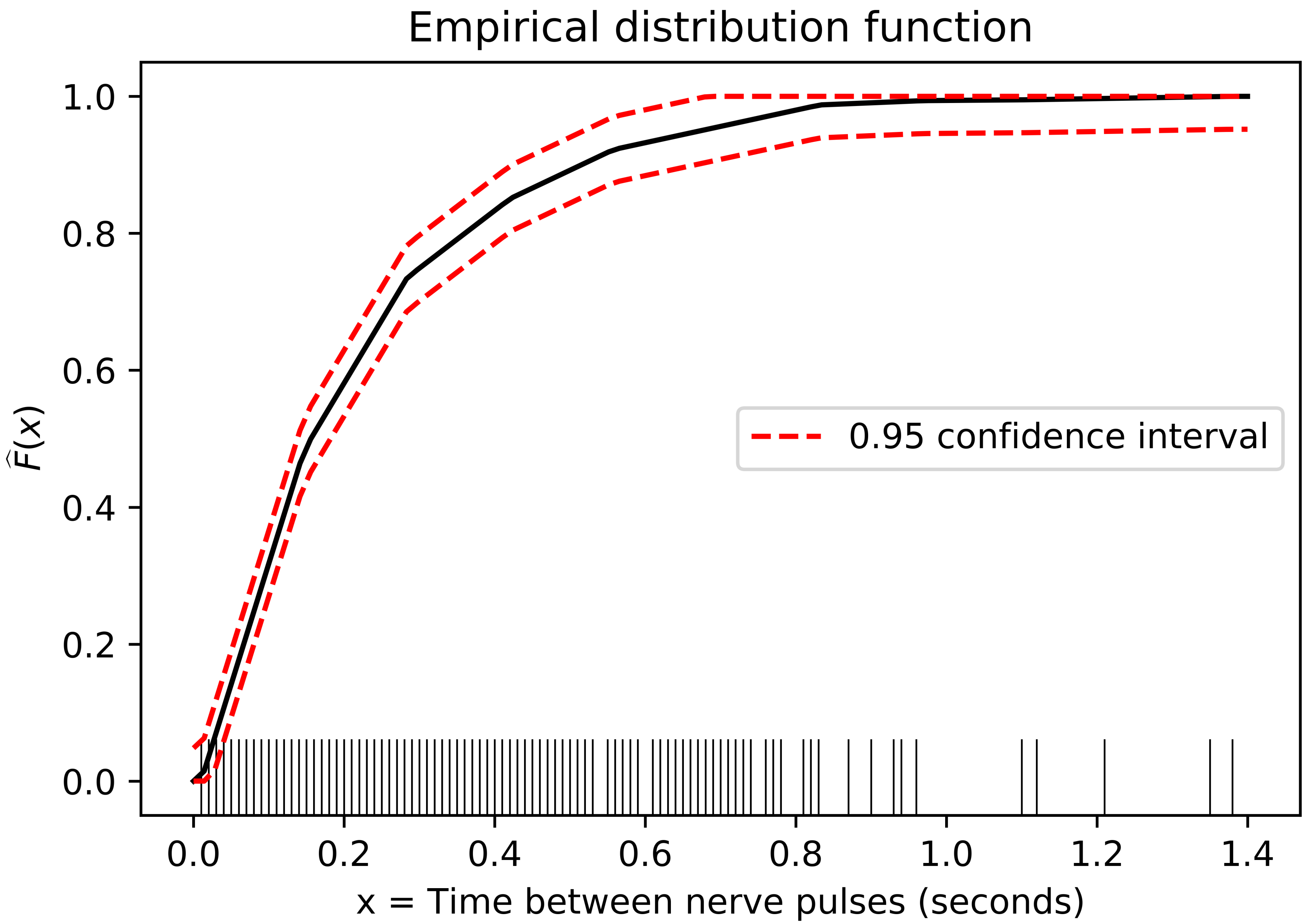
For a given $1 - \alpha$ confidence level, pick $\varepsilon_n = \sqrt{\dfrac{\log(2/\alpha)}{2n}}$

Then for the lower and upper bounds: $L(x) = \max\{\widehat{F}_n(x) - \varepsilon_n, 0\}$ and $U(x) = \min\{\widehat{F}_n(x) + \varepsilon_n, 1\}$ we have:

$$\mathbb{P}\Big(L(x) \leq F(x) \leq U(x) \text{ for all } x\Big) \geq 1 - \alpha$$

The Empirical Distribution Function

Here are the 95% confidence intervals for the nerve waiting times data from our Example 1:



Statistical Functionals

- A **statistical functional** $T(F)$ is any function F, the distribution function. Examples are the mean $\mu = \int x\, dF(x)$, the variance $\sigma^2 = \int (x-\mu)^2\, dF(x)$, the median $m = F^{-1}(1/2)$

- **Definition:** The **plug-in estimator** of $\theta = T(F)$ is defined by

$$\widehat{\theta_n} = T(\widehat{F_n})$$ – so, just plug in $\widehat{F_n}$ for the unknown F

Statistical Functionals

An important class of statistical functionals are linear functionals:

- **Definition:** If $T(F) = \int r(x)\, dF(x)$ for some function $r(x)$, then

  T is called a **linear functional**

  This is because integration is linear w.r. to its arguments, so
  $T(aF + bG) = aT(F) + bT(G)$

Statistical Functionals

Recall that the empirical CDF $\widehat{F_n}$ is just putting mass $1/n$ at each X_i

- **Theorem 4**: For a linear functional $T(F) = \int r(x)\, dF(x)$, its plug-in estimator is

$$T(\widehat{F_n}) = \int r(x)\, d\widehat{F_n}(x) = \frac{1}{n}\sum_{i=1}^{n} r(X_i)$$

Statistical Functionals

How does one estimate the **confidence intervals** for functionals? Assume we know **s.e.** (we'll find it with **bootstrap** in the next part)

- It is often the case that $T(\widehat{F_n}) \approx \mathcal{N}(\mu = T(F), \sigma = \hat{\text{se}})$ – so we can use our knowledge of the normal distribution and set a $1 - \alpha$ confidence interval to be

$$T(\widehat{F_n}) \pm z_{\alpha/2}\,\hat{\text{se}}$$ – **normal-based confidence interval**

here, $z_{\alpha/2}$ is such (a positive number) that for $z \sim \mathcal{N}(0,1)$ we have $\mathbb{P}(-z_{\alpha/2} \leq z \leq z_{\alpha/2}) = 1 - \alpha$

from **68-95-99 rule** we know that for $1 - \alpha = 0.95$, $z_{\alpha/2} \approx 2$

Statistical Functionals: Examples

- **Example 1:** The mean, $\mu = T(F) = \int x\, dF(x)$,

  the plug-in estimator is $\widehat{\mu} = \int x\, d\widehat{F_n}(x) = \overline{X}_n$

  the standard error is $\mathrm{se} = \sqrt{\mathbb{V}(\overline{X}_n)} = \sigma/\sqrt{n}$, if an estimate for σ is $\widehat{\sigma}$, then $\widehat{\mathrm{se}} = \widehat{\sigma}/\sqrt{n}$

  normal-based confidence interval for μ is $\overline{X}_n \pm z_{\alpha/2}\widehat{\mathrm{se}}$

Statistical Functionals: Examples

- **Example 2:** The variance,

$$\sigma^2 = T(F) = \mathbb{V}(X) = \int x^2\, dF(x) - \left(\int x\, dF(x) \right)^2$$

the plug-in estimator is just $\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}_n \right)^2$

or $S_n^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n \right)^2$, if we consider there to be $n - 1$ degrees of freedom

Statistical Functionals: Examples

- **Example 3:** The skewness, $\kappa = \dfrac{\mathbb{E}(X-\mu)^3}{\sigma^3}$

  (measuring the lack of symmetry of the distribution)

  plug-in estimate is $\widehat{\kappa} = \dfrac{\frac{1}{n}\sum_i (X_i - \widehat{\mu})^3}{\widehat{\sigma}^3}$

Statistical Functionals: Examples

- **Example 4:** Quantiles. If true F is strictly increasing, for $0 < p < 1$, the p-th quantile is given by $T(F) = F^{-1}(p)$ (median is an example with $p = 1/2$).

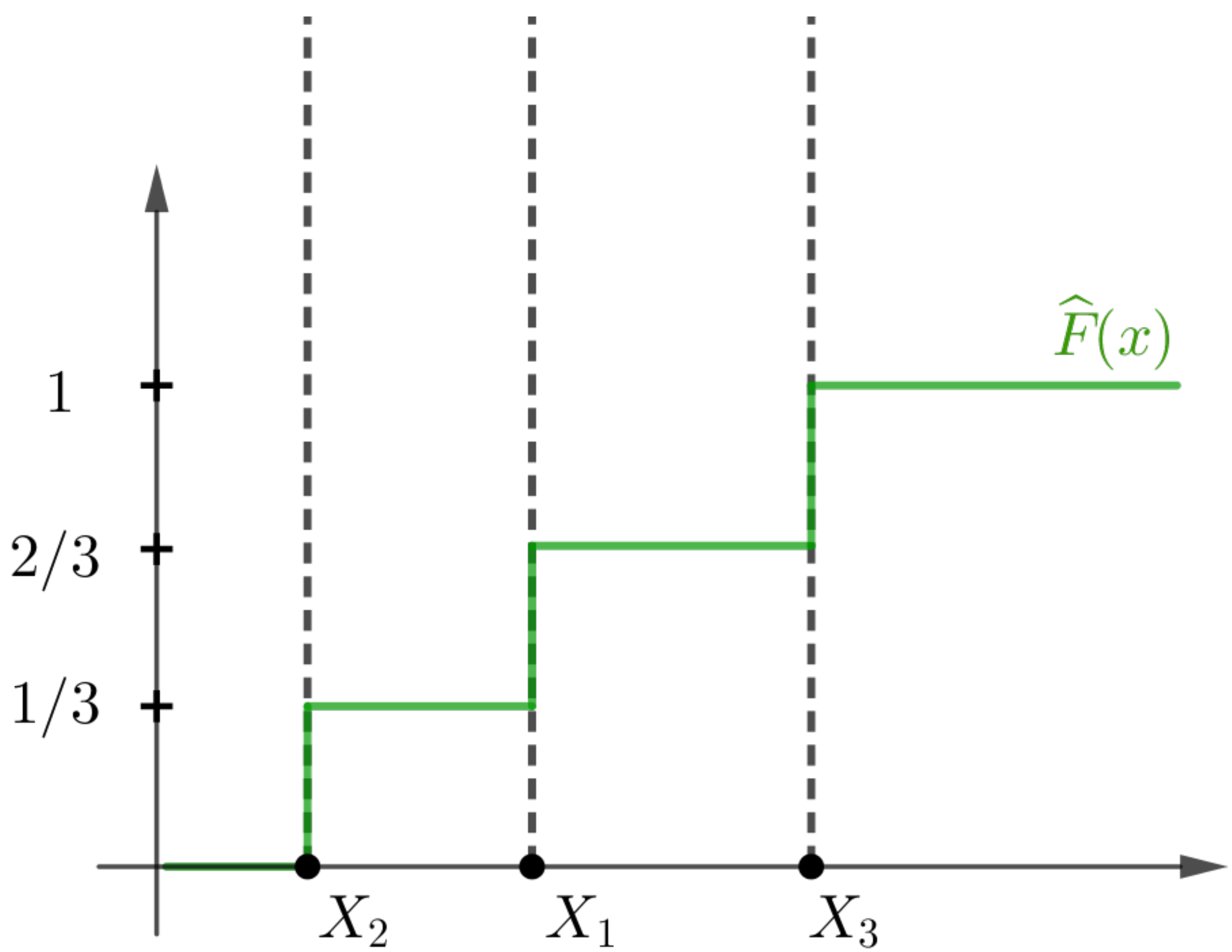
  With a plug-in estimate, we have to be careful, since the empirical $\widehat{F}_n$ is not invertible, so a proper definition of **sample quantile** is:

$$T(\widehat{F}_n) = \widehat{F}_n^{-1}(p) = \inf\{x : \widehat{F}_n(x) \geq p\}$$

Statistical Functionals: Examples

- **Example 4: Sample quantile** is:

$$T(\widehat{F}_n) = \widehat{F}_n^{-1}(p) = \inf\{x : \widehat{F}_n(x) \geq p\}$$



<- here we have

$$\widehat{F}_n^{-1}(2/3) = X_1$$

and

$$\widehat{F}_n^{-1}(1/3) = X_2$$

Statistical Functionals: Examples

- **Example 5:** Correlation. For X and Y – real random variables,

$$\rho = \frac{\mathbb{E}(X - \mu_X)(Y - \mu_Y)}{\sigma_x \sigma_y}$$ – Pearson's correlation coefficient

can be shown to have a plug-in estimate

$$\widehat{\rho} = \frac{\sum_i (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sqrt{\sum_i (X_i - \overline{X}_n)^2}\sqrt{\sum_i (Y_i - \overline{Y}_n)^2}}$$

Statistical Functionals: Examples

- Only in the first example did we compute the s.e. and the confidence intervals – but how do we do this for other examples?

  If we were doing parametric inference, one could derive formulae for these, but in a nonparametric setting we need something else.

  In the next part we'll see how bootstrap solves this problem!