Lecture 1: Nonparametric Estimation of Distribution Functions and Quantiles

Applied Statistics 2014
About the course

- 12 weeks: 2 hours a week
- Mainly on nonparametric methods in statistics
- **Exercises**: Some are theoretical, and others are practical and involve some computational methods
- **Prerequisites**: Basic knowledge of probability and statistics, and a certain level of mathematical maturity.
About the course

- 12 weeks: 2 hours a week
- Mainly on nonparametric methods in statistics
- **Exercises**: Some are theoretical, and others are practical and involve some computational methods
- **Prerequisites**: Basic knowledge of probability and statistics, and a certain level of mathematical maturity.

Instructor: **Juan Juan Cai** email: j.j.cai@tudelft.nl
Teaching assistant: **Aurelius Zilko** email: a.a.zilko@tudelft.nl
course website: [http://dutiosb.twi.tudelft.nl/~cai/teaching.html](http://dutiosb.twi.tudelft.nl/~cai/teaching.html)
Evaluation

- Each week except the first week, there will be one or two group presentations. Each group has 15 minutes at most.
  - Presenting a given paper, to explain what you think is most important about the paper, like main contribution, novelty, findings etc.
  - Or demonstrating the implementation of a R code constructed by your group to solve an exercise.
  - Please send your PDF slides and R code to me and Zilko.
- A final exam (written, open book): 10 points (maximum).
- Suppose you get $a$ points (out of 10) from the presentation and $b$ points from the exam. Your final grade equals $\min\{a/10 + b, 10\}$. 
Aim of this course

- Obtain some knowledge of nonparametric methods in statistics
- Study several modern and popular nonparametric methods in statistics, and understand their pros and cons
- Understand some of the theory behind nonparametric models.
Topics for this course...

**Tentative list:**

- Nonparametric estimation of distribution functions and quantiles (notes and Ch. 2 of Wasserman *All of Nonparametric Statistics*).
- Goodness of fit (notes).
- Permutation tests (article + notes).
- Bootstrapping
Topics for this course...

Tentative list:

- Nonparametric estimation of distribution functions and quantiles (notes and Ch. 2 of Wasserman *All of Nonparametric Statistics*).
- Goodness of fit (notes).
- Permutation tests (article + notes).
- Bootstrapping
- Kernel density estimator
- Smoothing: general concepts (Ch. 4 of Wasserman).
- Nonparametric regression (Ch. 5 of Wasserman).
- Extreme value theory (notes).
What is nonparametric statistics?

Wolfowitz (1942):

*We shall refer to this situation (where a distribution is completely determined by the knowledge of its finite parameter set) as the parametric case, and denote the opposite case, where the functional forms of the distributions are unknown, as the nonparametric case.*
What is nonparametric statistics?

Wolfowitz (1942):

*We shall refer to this situation (where a distribution is completely determined by the knowledge of its finite parameter set) as the parametric case, and denote the opposite case, where the functional forms of the distributions are unknown, as the nonparametric case.*


*Nonparametric statistics can and should be broadly defined to include all methodology that does not use a model based on a single parametric family.*
What is nonparametric statistics?

Wolfowitz (1942):

*We shall refer to this situation (where a distribution is completely determined by the knowledge of its finite parameter set) as the parametric case, and denote the opposite case, where the functional forms of the distributions are unknown, as the nonparametric case.*


*Nonparametric statistics can and should be broadly defined to include all methodology that does not use a model based on a single parametric family.*

Wasserman (2005)

*The basic idea of nonparametric inference is to use data to infer an unknown quantity while making as few assumptions as possible.*
Parametric models

Definition

If the data $X$ has a probability distribution $P = P_\theta$ where $\theta \in \Theta \subset \mathbb{R}^d$, then we speak of a \textit{parametric} model.
Parametric models

Definition

If the data $X$ has a probability distribution $P = P_\theta$ where $\theta \in \Theta \subset \mathbb{R}^d$, then we speak of a parametric model.

Example: 799 Waiting times at an IT helpdesk:

![Histogram of waiting times]

Requirement: (at least) 99% of the waiting times should be below one minute.

Is the requirement satisfied, based on the data?
Build a probability model. Assume $X_1, \ldots, X_n$ are a sample from either

A. a Gamma distribution

$$p_{\theta,m}(x) = \frac{\theta^e^{-\theta x}(\theta x)^{m-1}}{(m-1)!}, \quad x > 0;$$

B. or a Lognormal distribution

$$p_{\mu,\sigma}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp \left( \frac{(\log x - \mu)^2}{2\sigma^2} \right)^2, \quad x > 0.$$
Parametric models

Parametric fits for waiting times

waiting time

0.0 0.5 1.0 1.5

gamma
lognorma
Parametric models

$P_A(W \leq 1) \approx 0.993$

$P_B(W \leq 1) \approx 0.968.$

Conclusions can be very sensitive to parametric assumptions.
Advantages and disadvantages of parametric models

Advantages:

- **Convenience**: parametric models are generally easier to work with.
- **Efficiency**: If a parametric model is correct, then P methods are more efficient than NP methods. (However, the loss in efficiency of NP methods is often small.)
- **Interpretation**: Sometimes parametric models are easier to interpret.
Advantages and disadvantages of parametric models

Advantages:

- **Convenience**: parametric models are generally easier to work with.
- **Efficiency**: If a parametric model is correct, then P methods are more efficient than NP methods. (However, the loss in efficiency of NP methods is often small.)
- **Interpretation**: Sometimes parametric models are easier to interpret.

Disadvantages:

- Sometimes it is **hard to find a suitable parametric model**.
- High risk of **Misspecification**: assuming a (the?) wrong model.
Estimating a distribution function

Let $X_1, \ldots, X_n$ be a random sample (i.i.d.) from some unknown distribution function $F$ (discrete or continuous).

**Empirical distribution function**

The empirical (cumulative) distribution function (EDF) is defined by

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

Here:

$$1\{X_i \leq x\} = \begin{cases} 
1 & \text{if } X_i \leq x \\
0 & \text{otherwise}
\end{cases}$$
Estimating the cumulative distribution function (CDF)

Properties:

- Since $n \hat{F}_n(x) \sim \text{Bin}(n, F(x))$, easy to see that
  
  $$
  E \left[ \hat{F}_n(x) \right] = F(x), \quad \text{Var} \left( \hat{F}_n(x) \right) = \frac{F(x)(1 - F(x))}{n}
  $$

- Glivenko-Cantelli Theorem (fundamental theorem of statistics)
  For any distribution function $F$,
  
  $$
  \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \to 0 \text{ a.s. (} n \to \infty \text{)}.
  $$

- Dvoretzky-Kiefer-Wolfowith (DKW) inequality: for any $\varepsilon > 0$, for any $n$,
  
  $$
  P \left( \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| > \varepsilon \right) \leq 2e^{-2n\varepsilon^2}.
  $$
A confidence-band for $F$

Using the DKW inequality we can construct a confidence band for $F$. Let the significance level $\alpha \in (0, 1)$:

$$L_n(x) = \max \left\{ \hat{F}_n(x) - \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}}, 0 \right\},$$

$$U_n(x) = \max \left\{ \hat{F}_n(x) + \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}}, 1 \right\},$$

**DKW confidence band**

For any CDF $F$ and all $n$

$$P(L_n(x) \leq F(x) \leq U_n(x) \text{ for all } x) > 1 - \alpha.$$
2. Estimating the cdf and Statistical Functionals

FIGURE 2.1. Nerve data. Each vertical line represents one data point. The solid line is the empirical distribution function. The lines above and below the middle line are a 95 percent confidence band.

2.4 Theorem. Let \( X_1, \ldots, X_n \sim F \) and let \( \hat{F}_n \) be the empirical cdf. Then:

1. At any fixed value of \( x \),
   \[
   \mathbb{E}(\hat{F}_n(x)) = F(x) \quad \text{and} \quad \mathbb{V}(\hat{F}_n(x)) \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.
   \]

2. (Glivenko–Cantelli Theorem)
   \[
   \sup_x |\hat{F}_n(x) - F(x)| \rightarrow 0.
   \]

3. (Dvoretzky–Kiefer–Wolfowitz (DKW) inequality) For any \( \epsilon > 0 \),
   \[
   \mathbb{P}(\sup_x |F(x) - \hat{F}_n(x)| > \epsilon) \leq 2e^{-2n\epsilon^2}.
   \]
Confidence intervals for $F(x)$ at a fixed point $x$

Find $l_n, u_n$ such that

$$P(l_n \leq F(x) \leq u_n) \geq 1 - \alpha.$$ 

Since

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

the problem can be restated:
Confidence intervals for $F(x)$ at a fixed point $x$

Find $l_n, u_n$ such that

$$
P(l_n \leq F(x) \leq u_n) \geq 1 - \alpha.
$$

Since

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

the problem can be restated:

Observe $Y \sim \text{Bin}(n, p)$, find a CI for $p$.

1. Exact (Clopper-Pearson)
2. Asymptotic (Wald)
3. Wilson method
4. Asymptotic using variance stabilizing transformation
1. Exact (Clopper-Pearson)

- \( P_p(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, \ldots, n. \)
- For an observed value \( y \), the Clopper-Pearson interval (equal-tail) is defined by

\[ [p_L, p_U], \]

where \( p_L \) is the solution of

\[ \frac{\alpha}{2} = P_{p_L}(Y \geq y) = \sum_{i=y}^{n} \binom{n}{i} p_L^i (1 - p_L)^{n-i}; \]

and \( p_U \) is the solution of

\[ \frac{\alpha}{2} = P_{p_U}(Y \leq y) = \sum_{i=0}^{y} \binom{n}{i} p_U^i (1 - p_U)^{n-i}. \]
1. Exact (Clopper-Pearson)

- \( P_p(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, \ldots, n. \)

- For an observed value \( y \), the Clopper-Pearson interval (equal-tail) is defined by

\[
[p_L, p_U],
\]

where \( p_L \) is the solution of

\[
\frac{\alpha}{2} = P_{p_L}(Y \geq y) = \sum_{i=y}^{n} \binom{n}{i} p^i_L (1 - p_L)^{n-i};
\]

and \( p_U \) is the solution of

\[
\frac{\alpha}{2} = P_{p_U}(Y \leq y) = \sum_{i=0}^{y} \binom{n}{i} p^i_U (1 - p_U)^{n-i}.
\]

This interval is in general conservative, with coverage probability always \( \geq 1 - \alpha \) (due to the discreteness of \( Y \), exact coverage is not possible).
2. Asymptotic (Wald)

Let \( \hat{p}_n = Y/n \). By the central limit theorem

\[
\sqrt{n} \frac{\hat{p}_n - p}{\sqrt{p(1 - p)}} \xrightarrow{D} \mathcal{N}(0, 1).
\]
2. Asymptotic (Wald)

- Let $\hat{p}_n = Y/n$. By the central limit theorem

$$\sqrt{n} \frac{\hat{p}_n - p}{\sqrt{p(1 - p)}} \xrightarrow{D} \mathcal{N}(0, 1).$$

- Strong law of large numbers $\hat{p}_n \xrightarrow{a.s.} p$

- Slutsky’s lemma:

$$\sqrt{n} \frac{\hat{p}_n - p}{\sqrt{\hat{p}_n(1 - \hat{p}_n)}} \xrightarrow{D} \mathcal{N}(0, 1).$$
2. Asymptotic (Wald)

- Let $\hat{p}_n = Y/n$. By the central limit theorem

$$\sqrt{n} \frac{\hat{p}_n - p}{\sqrt{p(1-p)}} \xrightarrow{D} \mathcal{N}(0, 1).$$

- Strong law of large numbers $\hat{p}_n \xrightarrow{a.s.} p$

- Slutsky’s lemma:

$$\sqrt{n} \frac{\hat{p}_n - p}{\sqrt{\hat{p}_n(1 - \hat{p}_n)}} \xrightarrow{D} \mathcal{N}(0, 1).$$

- Isolating $p$, we obtain the following CI for $p$

$$\left[ \hat{p}_n - z_{\alpha/2} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}}, \hat{p}_n + z_{\alpha/2} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}} \right],$$

where $z_{\alpha/2}$ denotes the upper $\alpha/2$-quantile of the standard normal distribution.
3. Wilson Method

- In the derivation of the Wald interval, forget about the plug-in step for \( \hat{p}_n \).
- Asymptotically, for large \( n \)

\[
P\left(-z_{\alpha/2} \leq \sqrt{n} \frac{\hat{p}_n - p}{\sqrt{p(1-p)}} \leq z_{\alpha/2}\right) \approx 1 - \alpha.
\]
3. Wilson Method

- In the derivation of the Wald interval, forget about the plug-in step for \( \hat{p}_n \).

- Asymptotically, for large \( n \)

\[
P \left( \frac{-z_{\alpha/2} \leq \sqrt{n} \frac{\hat{p}_n - p}{\sqrt{p(1-p)}} \leq z_{\alpha/2}}{\sqrt{p(1-p)}} \right) \approx 1 - \alpha.
\]

- Solving for \( p \) gives the desired confidence interval, which has endpoints

\[
\hat{p}_n + \frac{z_{\alpha/2}^2}{2n} \pm \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\frac{\hat{p}_n(1-\hat{p}_n) + z_{\alpha/2}^2/(4n)}{1 + z_{\alpha/2}^2/n}}.
\]
4. Asymptotic using variance stabilizing transformation

- Apply a variance stabilizing transformation
- Suppose $\varphi : \mathbb{R} \to \mathbb{R}$ is differentiable at $p$ with $\varphi'(p) \neq 0$. By the $\delta$-method

\[ \sqrt{n} (\varphi(\hat{p}_n) - \varphi(p)) \xrightarrow{D} N(0, \varphi'(p)^2). \]

---

\textsuperscript{1}Let $Y_n$ be a sequence of random variables and $\varphi : \mathbb{R} \to \mathbb{R}$ a map that is differentiable at $\mu$ and $\varphi'(\mu) \neq 0$. Suppose $\sqrt{n}(Y_n - \mu) \xrightarrow{D} N(0, 1)$. Then $\sqrt{n}(\varphi(Y_n) - \varphi(\mu)) \xrightarrow{D} N(0, (\varphi'(\mu))^2)$. 

4. Asymptotic using variance stabilizing transformation

- Apply a variance stabilizing transformation
- Suppose \( \varphi : \mathbb{R} \to \mathbb{R} \) is differentiable at \( p \) with \( \varphi'(p) \neq 0 \). By the \( \delta \)-method

\[
\sqrt{n}(\varphi(\hat{p}_n) - \varphi(p)) \xrightarrow{D} \mathcal{N}(0, p(1-p)(\varphi'(p))^2).
\]

- Take \( \varphi \) so that \( \varphi'(p) = 1/(\sqrt{p(1-p)} \). That is take \( \varphi(x) = 2 \arcsin(\sqrt{x}) \).

---

\( ^1 \)Let \( Y_n \) be a sequence of random variables and \( \varphi : \mathbb{R} \to \mathbb{R} \) a map that is differentiable at \( \mu \) and \( \varphi'(\mu) \neq 0 \). Suppose \( \sqrt{n}(Y_n - \mu) \xrightarrow{D} \mathcal{N}(0, 1) \). Then

\[
\sqrt{n}(\varphi(Y_n) - \varphi(\mu)) \xrightarrow{D} \mathcal{N}(0, (\varphi'(\mu))^2).
\]
4. Asymptotic using variance stabilizing transformation

- Apply a variance stabilizing transformation
- Suppose \( \varphi : \mathbb{R} \to \mathbb{R} \) is differentiable at \( p \) with \( \varphi'(p) \neq 0 \). By the \( \delta \)-method \(^1\)

\[
\sqrt{n}(\varphi(\hat{p}_n) - \varphi(p)) \xrightarrow{D} \mathcal{N}(0, p(1 - p)(\varphi'(p))^2).
\]

- Take \( \varphi \) so that \( \varphi'(p) = 1/(\sqrt{p(1 - p)}). \) That is take \( \varphi(x) = 2 \arcsin \sqrt{x}. \)

\[
Z_n(p) := 2\sqrt{n}(\arcsin(\sqrt{\hat{p}_n}) - \arcsin(\sqrt{p})) \xrightarrow{D} \mathcal{N}(0, 1).
\]

---

\(^1\)Let \( Y_n \) be a sequence of random variables and \( \varphi : \mathbb{R} \to \mathbb{R} \) a map that is differentiable at \( \mu \) and \( \varphi'(\mu) \neq 0 \). Suppose \( \sqrt{n}(Y_n - \mu) \xrightarrow{D} \mathcal{N}(0, 1) \). Then

\[
\sqrt{n}(\varphi(Y_n) - \varphi(\mu)) \xrightarrow{D} \mathcal{N}(0, (\varphi'(\mu))^2).
\]
4. Asymptotic using variance stabilizing transformation

Center $p$ in $\mathbb{P}(-z_{\alpha/2} \leq Z_n(p) \leq z_{\alpha/2}) \approx 1 - \alpha$ with

$$Z_n(p) := 2\sqrt{n}(\arcsin(\sqrt{\hat{p}_n}) - \arcsin(\sqrt{p}))$$
4. Asymptotic using variance stabilizing transformation

Center $p$ in $P(-z_{\alpha/2} \leq Z_n(p) \leq z_{\alpha/2}) \approx 1 - \alpha$ with

$$Z_n(p) := 2\sqrt{n}(\arcsin(\sqrt{\hat{p}_n}) - \arcsin(\sqrt{p}))$$

The resulting confidence interval has endpoints

$$\left[ \sin^2 \left( \arcsin(\sqrt{\hat{p}_n}) - \frac{z_{\alpha/2}}{2\sqrt{n}} \right), \sin^2 \left( \arcsin(\sqrt{\hat{p}_n}) + \frac{z_{\alpha/2}}{2\sqrt{n}} \right) \right].$$
Coverage of binomial confidence intervals

- Study coverage probability of a CI. If the CI takes the form $[l(Y), u(Y)]$, evaluate

  $$\Pr(p \in [l(Y), u(Y)]) \quad \text{for} \quad Y \sim \text{Bin}(n, p).$$

- Not easy, so we conduct a simulation study, where we instead look at the coverage fraction.
Coverage of binomial confidence intervals

- Study **coverage probability** of a CI. If the CI takes the form $[l(Y), u(Y)]$, evaluate

  $$\Pr(p \in [l(Y), u(Y)]) \quad \text{for} \quad Y \sim \text{Bin}(n,p).$$

- Not easy, so we conduct a **simulation study**, where we instead look at the coverage fraction.

**Monte Carlo simulation scheme:**

1. Choose $n$ and $p \in (0, 1)$.
2. Generate 1000 $\text{Bin}(n, p)$ random variables.
3. For each realization, compute the 95% confidence interval (Clopper-Pearson, Wald (asymptotic), Wilson and variance stabilizing).
4. Compute the fraction of times that the CIs contain the true parameter $p$. 
Coverage of binomial confidence intervals
Coverage of binomial confidence intervals

Coverage fractions of binomial C.I., n = 25

coverage fraction

true success prob.

Exact
Asymptotic
Wilson
Varstab
Coverage of binomial confidence intervals

Coverage fractions of binomial C.I., n = 100

coverage fraction

true success prob.
Estimating quantiles using order statistics
Estimating quantiles using order statistics

Quantile function

The quantile function of $F$ is its generalized inverse function given by

$$q_p = F^{-1}(p) = \inf\{u \mid F(u) \geq p\} \quad p \in (0, 1).$$
Estimating quantiles using order statistics

Quantile function

The quantile function of $F$ is its generalized inverse function given by

$$q_p = F^{-1}(p) = \inf\{u \mid F(u) \geq p\} \quad p \in (0, 1).$$

Order statistics

Suppose $X_1, \ldots, X_n$ is a random sample from a continuous distribution, then the order statistics are denoted by

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}.$$
Estimating quantiles using order statistics

- If we estimate $F^{-1}$ by the empirical quantile function $\hat{F}_n^{-1}$, then the estimator of $q_p$ is then the order statistics $X(i)$, where $p \in \left(\frac{i-1}{n}, \frac{i}{n}\right]$.

- To find the $(1 - \alpha)$ confidence interval for $q_p$, the main idea is to choose $r < s$ such that

  $$P\left(X(r) < q_p \leq X(s)\right) \geq 1 - \alpha.$$
Estimating quantiles using order statistics

- If we estimate $F^{-1}$ by the empirical quantile function $\hat{F}_{n}^{-1}$, then the estimator of $q_p$ is then the order statistics $X(i)$, where $p \in \left(\frac{i-1}{n}, \frac{i}{n}\right]$.
- To find the $(1 - \alpha)$ confidence interval for $q_p$, the main idea is to choose $r < s$ such that

$$P\left( X(r) < q_p \leq X(s) \right) \geq 1 - \alpha.$$

- First note that

$$P\left( X(r) < q_p \leq X(s) \right) = 1 - \left( P\left( X(r) \geq q_p \right) + P\left( X(s) < q_p \right) \right).$$

- Find $r$ and $s$ such that

$$P\left( X(r) \geq q_p \right) \leq \alpha/2 \quad \text{and} \quad P\left( X(s) < q_p \right) \leq \alpha/2.$$
Computing $P(X_{(s)} < q_p)$

- $\{X_{(s)} < q_p\}$ is equivalent to say that at least $s$ observations are less than $q_p$. 

$$N = \sum_{i=1}^{n} 1\{X_i < q_p\}.$$ 

Suppose that $F$ is continuous. Then $P(X_1 < q_p) = P(X_1 \leq q_p) = p.$ Thus $N \sim \text{bin}(n, p)$ and 

$$P(X_{(r)} \geq q_p) = P(N < r).$$
Computing $P(X_{(s)} < q_p)$

- $\{X_{(s)} < q_p\}$ is equivalent to say that at least $s$ observations are less than $q_p$.

- Write $N = \sum_{i=1}^{n} 1\{X_i < q_p\}$. Suppose that $F$ is continuous. Then $P(X_1 < q_p) = P(X_1 \leq q_p) = p$. Thus $N \sim \text{bin}(n, p)$ and

$$P(X_{(s)} \leq q_p) = P(N \geq s) = \sum_{i=s}^{n} \binom{n}{i} p^i (1-p)^{n-i}.$$ 

- Similarly, we have $P(X_{(r)} \geq q_p) = P(N < r)$. 


A confidence interval for $q_p$

Let $X(0) = -\infty$ and $X(n+1) = \infty$.

The $(1 - \alpha)$ confidence interval for $q_p$ is given by

$$(X(r), X(s)],$$

where $r = \max\{k \in \{0, 1, \ldots, n\} : P(N < k) \leq \alpha/2\}$ and $s = \min\{k \in \{1, \ldots, n + 1\} : P(N \geq k) \leq \alpha/2\}$. 
Back to the waiting time data

\[ P_A(W \leq 1) \approx 0.993 \quad \text{and} \quad P_B(W \leq 1) \approx 0.968. \]
Back to the waiting time data

\[ P_A(W \leq 1) \approx 0.993 \]

\[ P_B(W \leq 1) \approx 0.968. \]

\[ \hat{q}_{0.99} \approx 0.90 \text{ and a 95\% confidence interval for } q_{0.99} \text{ is given by } (0.81, 1.21) \]
## Useful R commands

<table>
<thead>
<tr>
<th>Method</th>
<th>R-function</th>
<th>within package</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical distr. function</td>
<td>ecdf</td>
<td>stats</td>
</tr>
<tr>
<td>Binomial confidence interval</td>
<td>binconf</td>
<td>Hmisc</td>
</tr>
</tbody>
</table>
Group Presentation (February 9)

- Group 1
  Present the paper “Interval Estimation for a Binomial Proportion” (Brown-Cai-DasGupta 2001). Choose two methods and compare the performance by simulation.
Group Presentation (February 9)

Group 2
Write an R-function that implements a confidence interval for the $p$-th quantile based on the order-statistics of the data. Simulate 1000 times a sample of size 10 from a standard normal distribution and compute the coverage of the confidence interval with $p = 0.5$. Repeat for sample sizes 50, 100 and 1000.