Confidence sets

$X$: a sample from a population $P \in \mathcal{P}$.

$\theta = \theta(P)$: a functional from $\mathcal{P}$ to $\Theta \subseteq \mathbb{R}^k$ for a fixed integer $k$.

$C(X)$: a confidence set for $\theta$, a set in $\mathcal{B}_\Theta$ (the class of Borel sets on $\Theta$) depending only on $X$.

$\inf_{P \in \mathcal{P}} P(\theta \in C(X))$: the confidence coefficient of $C(X)$.

If the confidence coefficient of $C(X)$ is $\geq 1 - \alpha$ for fixed $\alpha \in (0, 1)$, then we say that $C(X)$ has confidence level $1 - \alpha$ or $C(X)$ is a level $1 - \alpha$ confidence set.

We focus on

- Various methods of constructing confidence sets.
- Properties of confidence sets.
Pivotal methods
The most popular method of constructing confidence sets is the use of pivotal quantities defined as follows.

Definition 7.1
A known Borel function $\mathcal{R}$ of $(X, \theta)$ is called a pivotal quantity if and only if the distribution of $\mathcal{R}(X, \theta)$ does not depend on $P$.

Remarks
- A pivotal quantity depends on $P$ through $\theta = \theta(P)$.
- A pivotal quantity is usually not a statistic, although its distribution is known.
- With a pivotal quantity $\mathcal{R}(X, \theta)$, a level $1 - \alpha$ confidence set for any given $\alpha \in (0, 1)$ can be obtained.
- If $\mathcal{R}(X, \theta)$ has a continuous c.d.f., then we can obtain a confidence set $C(X)$ that has confidence coefficient $1 - \alpha$. 
Construction
First, find two constants $c_1$ and $c_2$ such that
\[ P(c_1 \leq \mathcal{R}(X, \theta) \leq c_2) \geq 1 - \alpha. \]
Next, define
\[ C(X) = \{ \theta \in \Theta : c_1 \leq \mathcal{R}(X, \theta) \leq c_2 \}. \]
Then $C(X)$ is a level $1 - \alpha$ confidence set, since
\[
\inf_{P \in \mathcal{P}} P(\theta \in C(X)) = \inf_{P \in \mathcal{P}} P(c_1 \leq \mathcal{R}(X, \theta) \leq c_2) \\
= P(c_1 \leq \mathcal{R}(X, \theta) \leq c_2) \\
\geq 1 - \alpha.
\]
The confidence coefficient of $C(X)$ may not be $1 - \alpha$.
If $\mathcal{R}(X, \theta)$ has a continuous c.d.f., then we can choose $c_i$'s such that the equality in the last expression holds and the confidence set $C(X)$ has confidence coefficient $1 - \alpha$. In a given problem, there may not exist any pivotal quantity, or there may be many different pivotal quantities and one has to choose one based on some principles or criteria, which are discussed in §7.2.
Computation

When $\mathcal{R}(X, \theta)$ and $c_i$’s are chosen, we need to compute the confidence set $C(X) = \{c_1 \leq \mathcal{R}(X, \theta) \leq c_2\}$. This can be done by inverting $c_1 \leq \mathcal{R}(X, \theta) \leq c_2$.

For example, if $\theta$ is real-valued and $\mathcal{R}(X, \theta)$ is monotone in $\theta$ when $X$ is fixed, then

$$C(X) = \{\theta : \underline{\theta}(X) \leq \theta \leq \bar{\theta}(X)\}$$

for some $\underline{\theta}(X) < \bar{\theta}(X)$, i.e., $C(X)$ is an interval (finite or infinite). If $\mathcal{R}(X, \theta)$ is not monotone, then $C(X)$ may be a union of several intervals.

For real-valued $\theta$, a confidence interval rather than a complex set such as a union of several intervals is generally preferred since it is simple and the result is easy to interpret.

When $\theta$ is multivariate, inverting $c_1 \leq \mathcal{R}(X, \theta) \leq c_2$ may be complicated.

In most cases where explicit forms of $C(X)$ do not exist, $C(X)$ can still be obtained numerically.
Example 7.2
Let $X_1, \ldots, X_n$ be i.i.d. random variables from the uniform distribution $U(0, \theta)$.
Consider the problem of finding a confidence set for $\theta$.
Note that the family $\mathcal{P}$ in this case is a scale family so that the results in Example 7.1 can be used.
But a better confidence interval can be obtained based on the sufficient and complete statistic $X_{(n)}$ for which $X_{(n)}/\theta$ is a pivotal quantity (Example 7.13).
Note that $X_{(n)}/\theta$ has the Lebesgue p.d.f. $nx^{n-1}I_{(0,1)}(x)$.
Hence $c_i$’s should satisfy $c_2^n - c_1^n = 1 - \alpha$.
The resulting confidence interval for $\theta$ is
\[
[c_2^{-1}X_{(n)}, c_1^{-1}X_{(n)}].
\]
Choices of $c_i$’s are discussed in Example 7.13.
Example 7.3 (Fieller's interval)

Let \((X_{i1}, X_{i2}), i = 1, \ldots, n\), be i.i.d. bivariate normal with unknown \(\mu_j = E(X_{1j}), \sigma_j^2 = \text{Var}(X_{1j}), j = 1, 2\), and \(\sigma_{12} = \text{Cov}(X_{11}, X_{12})\). Let \(\theta = \mu_2/\mu_1\) be the parameter of interest \((\mu_1 \neq 0)\). Define \(Y_i(\theta) = X_{i2} - \theta X_{i1}\). Then \(Y_1(\theta), \ldots, Y_n(\theta)\) are i.i.d. from \(N(0, \sigma_2^2 - 2\theta\sigma_{12} + \theta^2\sigma_1^2)\).

Let
\[
S^2(\theta) = \frac{1}{n-1} \sum_{i=1}^{n} [Y_i(\theta) - \bar{Y}(\theta)]^2 = S_2^2 - 2\theta S_{12} + \theta^2 S_1^2,
\]
where \(\bar{Y}(\theta)\) is the average of \(Y_i(\theta)\)'s and \(S_i^2\) and \(S_{12}\) are sample variances and covariance based on \(X_{ij}\)'s.

It follows from Examples 1.16 and 2.18 that \(\sqrt{n} \bar{Y}(\theta)/S(\theta)\) has the t-distribution \(t_{n-1}\) and, therefore, is a pivotal quantity.

Let \(t_{n-1,\alpha}\) be the \((1 - \alpha)\)th quantile of the t-distribution \(t_{n-1}\). Then
\[
C(X) = \{\theta : n[\bar{Y}(\theta)]^2/S^2(\theta) \leq t_{n-1,\alpha/2}^2\}
\]
is a confidence set for \(\theta\) with confidence coefficient \(1 - \alpha\).
Example 7.3 (continued)

Note that \( n[\bar{Y}(\theta)]^2 = t_{n-1,\alpha/2}^2 \frac{S^2(\theta)}{2} \) defines a parabola in \( \theta \).
Depending on the roots of the parabola, \( C(X) \) can be a finite interval, the complement of a finite interval, or the whole real line (exercise).

Proposition 7.1 (Existence of pivotal quantities in parametric problems)

Let \( T(X) = (T_1(X), \ldots, T_s(X)) \) and \( T_1, \ldots, T_s \) be independent statistics. Suppose that each \( T_i \) has a continuous c.d.f. \( F_{T_i,\theta} \) indexed by \( \theta \). Then \( \mathcal{R}(X, \theta) = \prod_{i=1}^s F_{T_i,\theta}(T_i(X)) \) is a pivotal quantity.

Proof

The result follows from the fact that \( F_{T_i,\theta}(T_i) \)'s are i.i.d. from the uniform distribution \( U(0,1) \).

When \( \theta \) and \( T \) in Proposition 7.1 are real-valued, we can use the following result to construct confidence intervals for \( \theta \) even when the c.d.f. of \( T \) is not continuous.
Theorem 7.1
Suppose that $P$ is in a parametric family indexed by a real-valued $\theta$. Let $T(X)$ be a real-valued statistic with c.d.f. $F_{T,\theta}(t)$ and let $\alpha_1$ and $\alpha_2$ be fixed positive constants such that $\alpha_1 + \alpha_2 = \alpha < \frac{1}{2}$.

(i) Suppose that $F_{T,\theta}(t)$ and $F_{T,\theta}(t-)$ are nonincreasing in $\theta$ for each fixed $t$.

Define

$$\bar{\theta} = \sup\{\theta : F_{T,\theta}(T) \geq \alpha_1\} \quad \text{and} \quad \underline{\theta} = \inf\{\theta : F_{T,\theta}(T-) \leq 1 - \alpha_2\}.$$

Then $[\underline{\theta}(T), \bar{\theta}(T)]$ is a level $1 - \alpha$ confidence interval for $\theta$.

(ii) If $F_{T,\theta}(t)$ and $F_{T,\theta}(t-)$ are nondecreasing in $\theta$ for each $t$, then the same result holds with

$$\theta = \inf\{\theta : F_{T,\theta}(T) \geq \alpha_1\} \quad \text{and} \quad \bar{\theta} = \sup\{\theta : F_{T,\theta}(T-) \leq 1 - \alpha_2\}.$$

(iii) If $F_{T,\theta}$ is a continuous c.d.f. for any $\theta$, then $F_{T,\theta}(T)$ is a pivotal quantity and the confidence interval in (i) or (ii) has confidence coefficient $1 - \alpha$. 
Proof
We only need to prove (i).
Under the given condition, \( \theta > \bar{\theta} \) implies \( F_{T,\theta}(T) < \alpha_1 \) and \( \theta < \underline{\theta} \) implies \( F_{T,\theta}(T-) > 1 - \alpha_2 \). Hence,

\[
P(\underline{\theta} \leq \theta \leq \bar{\theta}) \geq 1 - P(F_{T,\theta}(T) < \alpha_1) - P(F_{T,\theta}(T-) > 1 - \alpha_2).
\]

The result follows from

\[
P(F_{T,\theta}(T) < \alpha_1) \leq \alpha_1 \quad \text{and} \quad P(F_{T,\theta}(T-) > 1 - \alpha_2) \leq \alpha_2.
\]

The proof of this inequality is left as an exercise.

Discussion
When the parametric family in Theorem 7.1 has monotone likelihood ratio in \( T(X) \), it follows from Lemma 6.3 that the condition in Theorem 7.1(i) holds; in fact, it follows from Exercise 2 in \S 6.6 that \( F_{T,\theta}(t) \) is strictly decreasing for any \( t \) at which \( 0 < F_{T,\theta}(t) < 1 \).
If \( F_{T,\theta}(t) \) is also continuous in \( \theta \), \( \lim_{\theta \to \theta^-} F_{T,\theta}(t) > \alpha_1 \), and \\
\( \lim_{\theta \to \theta^+} F_{T,\theta}(t) < \alpha_1 \), where \( \theta^- \) and \( \theta^+ \) are the two ends of the \\
parameter space, then \( \bar{\theta} \) is the unique solution of \( F_{T,\theta}(t) = \alpha_1 \).
A similar conclusion can be drawn for \( \underline{\theta} \).

Theorem 7.1 can be applied to obtain the confidence interval for \( \theta \) 
in Example 7.2 (exercise).
The following example concerns a discrete \( F_{T,\theta} \).

**Example 7.5**

Let \( X_1, \ldots, X_n \) be i.i.d. random variables from the Poisson 
distribution \( P(\theta) \) with an unknown \( \theta > 0 \) and 
\( T(X) = \sum_{i=1}^{n} X_i \).
Note that \( T \) is sufficient and complete for \( \theta \) and has the Poisson 
distribution \( P(n\theta) \).
Thus
\[
F_{T,\theta}(t) = \sum_{j=0}^{t} \frac{e^{-n\theta}(n\theta)^j}{j!}, \quad t = 0, 1, 2, \ldots.
\]
Since the Poisson family has monotone likelihood ratio in $T$ and $0 < F_{T,\theta}(t) < 1$ for any $t$, $F_{T,\theta}(t)$ is strictly decreasing in $\theta$. Also, $F_{T,\theta}(t)$ is continuous in $\theta$ and $F_{T,\theta}(t)$ tends to 1 and 0 as $\theta$ tends to 0 and $\infty$, respectively. Thus, Theorem 7.1 applies and $\overline{\theta}$ is the unique solution of $F_{T,\theta}(T) = \alpha_1$.

Since $F_{T,\theta}(t-) = F_{T,\theta}(t-1)$ for $t > 0$, $\underline{\theta}$ is the unique solution of $F_{T,\theta}(t-1) = 1 - \alpha_2$ when $T = t > 0$ and $\underline{\theta} = 0$ when $T = 0$.

In fact, in this case explicit forms of $\underline{\theta}$ and $\overline{\theta}$ can be obtained from

\[
\frac{1}{\Gamma(t)} \int_{\lambda}^{\infty} x^{t-1} e^{-x} dx = \sum_{j=0}^{t-1} \frac{e^{-\lambda} \lambda^j}{j!}, \quad t = 1, 2, \ldots
\]

Using this equality, it can be shown (exercise) that

\[
\overline{\theta} = (2n)^{-1} \chi_{2(T+1),\alpha_1}^2 \quad \text{and} \quad \underline{\theta} = (2n)^{-1} \chi_{2T,1-\alpha_2}^2
\]

where $\chi_{r,\alpha}^2$ is the $(1 - \alpha)$th quantile of the chi-square distribution $\chi_r^2$ and $\chi_{0,a}^2$ is defined to be 0.
Methods of inverting acceptance regions of tests

Another popular method of constructing confidence sets is to use a close relationship between confidence sets and hypothesis tests. For any test $T$, the set $\{x : T(x) \neq 1\}$ is called the acceptance region. This terminology is not precise when $T$ is a randomized test.

**Theorem 7.2**

For each $\theta_0 \in \Theta$, let $T_{\theta_0}$ be a test for $H_0 : \theta = \theta_0$ (versus some $H_1$) with significance level $\alpha$ and acceptance region $A(\theta_0)$. For each $x$ in the range of $X$, define

$$C(x) = \{\theta : x \in A(\theta)\}.$$  

Then $C(X)$ is a level $1 - \alpha$ confidence set for $\theta$. If $T_{\theta_0}$ is nonrandomized and has size $\alpha$ for every $\theta_0$, then $C(X)$ has confidence coefficient $1 - \alpha$. 
Proof
We prove the first assertion only. The proof for the second assertion is similar.
Under the given condition,

$$\sup_{\theta = \theta_0} P(X \not\in A(\theta_0)) = \sup_{\theta = \theta_0} P(T_{\theta_0} = 1) \leq \alpha,$$

which is the same as

$$1 - \alpha \leq \inf_{\theta = \theta_0} P(X \in A(\theta_0)) = \inf_{\theta = \theta_0} P(\theta_0 \in C(X)).$$

Since this holds for all $\theta_0$, the result follows from

$$\inf_{P \in \mathcal{P}} P(\theta \in C(X)) = \inf_{\theta_0 \in \Theta} \inf_{\theta = \theta_0} P(\theta_0 \in C(X)) \geq 1 - \alpha.$$

The converse of Theorem 7.2 is partially true.
Proposition 7.2
Let \( C(X) \) be a confidence set for \( \theta \) with confidence level (or confidence coefficient) \( 1 - \alpha \).
For any \( \theta_0 \in \Theta \), define a region \( A(\theta_0) = \{ x : \theta_0 \in C(x) \} \).
Then the test \( T(X) = 1 - I_{A(\theta_0)}(X) \) has significance level \( \alpha \) for testing \( H_0 : \theta = \theta_0 \) versus some \( H_1 \).

Discussions
In general, \( C(X) \) in Theorem 7.2 can be determined numerically, if it does not have an explicit form.
Suppose \( A(\theta) = \{ Y : a(\theta) \leq Y \leq b(\theta) \} \) for a real-valued \( \theta \) and statistic \( Y(X) \) and some nondecreasing functions \( a(\theta) \) and \( b(\theta) \).
When we observe \( Y = y \), \( C(X) \) is an interval with limits \( \underline{\theta} \) and \( \overline{\theta} \), which are the \( \theta \)-values at which the horizontal line \( Y = y \) intersects the curves \( Y = b(\theta) \) and \( Y = a(\theta) \) (Figure 7.1), respectively.
If \( y = b(\theta) \) (or \( y = a(\theta) \)) has no solution or more than one solution, \( \underline{\theta} = \inf \{ \theta : y \leq b(\theta) \} \) (or \( \overline{\theta} = \sup \{ \theta : a(\theta) \leq y \} \)).
\( C(X) \) does not include \( \underline{\theta} \) (or \( \overline{\theta} \)) if and only if at \( \underline{\theta} \) (or \( \overline{\theta} \)), \( b(\theta) \) (or \( a(\theta) \)) is only left-continuous (or right-continuous).
Example 7.7

Suppose that $X$ has the following p.d.f. in a one-parameter exponential family:

$$f_\theta(x) = \exp\{\eta(\theta) Y(x) - \xi(\theta)\} h(x),$$

where $\theta$ is real-valued and $\eta(\theta)$ is nondecreasing in $\theta$.

First, we apply Theorem 7.2 with $H_0 : \theta = \theta_0$ and $H_1 : \theta > \theta_0$.

By Theorem 6.2, the acceptance region of the UMP test of size $\alpha$ is

$$A(\theta_0) = \{x : Y(x) \leq c(\theta_0)\},$$

where $c(\theta_0) = c$ in Theorem 6.2.

It can be shown that $c(\theta)$ is nondecreasing in $\theta$.

Inverting $A(\theta)$ according to Figure 7.1 with $b(\theta) = c(\theta)$ and $a(\theta)$ ignored, we obtain

$$C(X) = [\underline{\theta}(X), \infty) \quad \text{or} \quad (\underline{\theta}(X), \infty),$$

a one-sided confidence interval for $\theta$ with confidence level $1 - \alpha$.

$\underline{\theta}(X)$ is called a lower confidence bound for $\theta$ in §2.4.3.

When the c.d.f. of $Y(X)$ is continuous, $C(X)$ has confidence coefficient $1 - \alpha$. 
If $H_0 : \theta = \theta_0$ and $H_1 : \theta < \theta_0$ are considered, then $\mathcal{C}(X) = \{\theta : Y(X) \geq c(\theta)\}$ and is of the form

$$(-\infty, \overline{\theta}(X)] \text{ or } (-\infty, \overline{\theta}(X)).$$

$\overline{\theta}(X)$ is called an upper confidence bound for $\theta$.

Consider next $H_0 : \theta = \theta_0$ and $H_1 : \theta \neq \theta_0$.

By Theorem 6.4, the acceptance region of the UMPU test of size $\alpha$ is given by $A(\theta_0) = \{x : c_1(\theta_0) \leq Y(x) \leq c_2(\theta_0)\}$, where $c_i(\theta)$ are nondecreasing (exercise).

A confidence interval can be obtained by inverting $A(\theta)$ according to Figure 7.1 with $a(\theta) = c_1(\theta)$ and $b(\theta) = c_2(\theta)$.

Let us consider a specific example in which $X_1, \ldots, X_n$ are i.i.d. binary random variables with $p = P(X_i = 1)$.

Note that $Y(X) = \sum_{i=1}^n X_i$.

Suppose that we need a lower confidence bound for $p$ so that we consider $H_0 : p = p_0$ and $H_1 : p > p_0$. 
From Example 6.2, the acceptance region of a UMP test of size \( \alpha \in (0, 1) \) is 

\[ A(p_0) = \{ y : y \leq m(p_0) \} \]

where \( m(p_0) \) is an integer between 0 and \( n \) such that

\[
\sum_{j=m(p_0)+1}^{n} \binom{n}{j} p_0^j (1 - p_0)^{n-j} \leq \alpha < \sum_{j=m(p_0)}^{n} \binom{n}{j} p_0^j (1 - p_0)^{n-j}.
\]

Thus, \( m(p) \) is an integer-valued, nondecreasing step-function of \( p \).

Define

\[
p = \inf\{ p : m(p) \geq y \} = \inf \left\{ p : \sum_{j=y}^{n} \binom{n}{j} p^j (1 - p)^{n-j} \geq \alpha \right\}.
\]

Then a level \( 1 - \alpha \) confidence interval for \( p \) is \( \left( p, 1 \right] \) (exercise). One can compare this confidence interval with the one obtained by applying Theorem 7.1 (exercise).

See also Example 7.16.
Example 7.8
Suppose that $X$ has the following p.d.f. in a multiparameter exponential family:

$$f_{\theta, \varphi}(x) = \exp \{\theta Y(x) + \varphi^T U(x) - \zeta(\theta, \varphi)\}$$

By Theorem 6.4, the acceptance region of a UMPU test of size $\alpha$ for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$ or $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ is

$$A(\theta_0) = \{(y, u) : y \leq c_2(u, \theta_0)\}$$

or

$$A(\theta_0) = \{(y, u) : c_1(u, \theta_0) \leq y \leq c_2(u, \theta_0)\},$$

where $c_i(u, \theta)$, $i = 1, 2$, are nondecreasing functions of $\theta$. Confidence intervals for $\theta$ can then be obtained by inverting $A(\theta)$ according to Figure 7.1 with $b(\theta) = c_2(u, \theta)$ and $a(\theta) = c_1(u, \theta)$ or $a(\theta) \equiv -\infty$, for any observed $u$. 
Consider more specifically the case where $X_1$ and $X_2$ are independently distributed as the Poisson distributions $P(\lambda_1)$ and $P(\lambda_2)$, respectively, and we need a lower confidence bound for the ratio $\rho = \lambda_2/\lambda_1$.

From Example 6.11, a UMPU test of size $\alpha$ for testing $H_0 : \rho = \rho_0$ versus $H_1 : \rho > \rho_0$ has the acceptance region

$$A(\rho_0) = \{(y, u) : y \leq c(u, \rho_0)\},$$

where $c(u, \rho_0)$ is determined by the conditional distribution of $Y = X_2$ given $U = X_1 + X_2 = u$.

Since the conditional distribution of $Y$ given $U = u$ is the binomial distribution $Bi(\rho/(1 + \rho), u)$, we can use the result in Example 7.7, i.e., $c(u, \rho)$ is the same as $m(\rho)$ in Example 7.7 with $n = u$ and $p = \rho/(1 + \rho)$. 
Then a level $1 - \alpha$ lower confidence bound for $p$ is $\underline{p}$ given by

$$\underline{p} = \inf \{ p : m(p) \geq y \} = \inf \left\{ p : \sum_{j=y}^{u} \binom{u}{j} p^j (1 - p)^{u-j} \geq \alpha \right\}$$

Since $\rho = p/(1 - p)$ is a strictly increasing function of $p$, a level $1 - \alpha$ lower confidence bound for $\rho$ is $\underline{p}/(1 - \underline{p})$.

**Example 7.9**

Consider the normal linear model $X = N_n(Z\beta, \sigma^2 I_n)$ and the problem of constructing a confidence set for $\theta = L\beta$, where $L$ is an $s \times p$ matrix of rank $s$ and all rows of $L$ are in $\mathcal{R}(Z)$. The LR test of size $\alpha$ for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ has the acceptance region

$$A(\theta_0) = \{ X : W(X, \theta_0) \leq c_\alpha \},$$
where \( c_\alpha \) is the \((1 - \alpha)\)th quantile of the F-distribution \( F_{s, n-r} \),

\[
W(X, \theta) = \frac{[\|X - Z\hat{\beta}(\theta)\|^2 - \|X - Z\hat{\beta}\|^2]}{\|X - Z\hat{\beta}\|^2/(n-r)},
\]

\( r \) is the rank of \( Z \), \( r \geq s \), \( \hat{\beta} \) is the LSE of \( \beta \) and, for each fixed \( \theta \), \( \hat{\beta}(\theta) \) is a solution of

\[
\|X - Z\hat{\beta}(\theta)\|^2 = \min_{\beta: L\beta = \theta} \|X - Z\beta\|^2.
\]

Inverting \( A(\theta) \), we obtain the following confidence set for \( \theta \) with confidence coefficient \( 1 - \alpha \): \( C(X) = \{ \theta : W(X, \theta) \leq c_\alpha \} \), which forms a closed ellipsoid in \( \mathcal{R}^s \).