

<p>データサイエンス</p>	<p>【代表的な研究テーマ】</p> <p>□ 高次元データに対する手法の開発</p>												
<p>key word</p>	<p>課題解決に役立つシーズの説明</p>												
<p>■ 多変量解析 ■ 高次元データ ■ 漸近理論</p>	<p>【1】高次元データを扱う意義</p> <p>多くの事象、出来事に対し、様々な原因が複雑に絡み合っている現在、物事を単純な統計モデルによって記述してしまうと、その本質をとらえることができなくなってしまう場合もある。一方で、多くの変数を使う場合、古典的な統計手法の多くは利用できなくなってしまう。この場合の一つの解決策としては、有用な変数のみを取り出して分析を行うことであるが、変数を減らしてしまうと重要な情報を失ってしまい、調べたいことを十分に評価できなくなる可能性がある。そこで、古典的な手法を高次元データ(変数の数が多いデータ)でも利用可能な手法に拡張する。これにより、今まで分析が困難であった、標本を多く集められないケースや関連のある変数の絞り込みができないようなケースにおいても統計分析を行い、何らかの特徴を見出すことを可能にする。</p> <p>【2】高次元データに対する手法と従来の手法の違い</p> <p>一般的によく知られている多くの統計手法は、サンプルサイズ n が大きく、次元(変数の数) p が小さいことが想定されている。このようなケースに対する統計手法は、中心極限定理により標本平均の分布が正規分布に近づくことを利用するため、データがどのような分布から発生しているかを考える必要が無く、汎用性が高い。一方、十分なサンプルサイズが得られない場合(ただし、$n > p$ の場合)は、上記の正規近似が使えないため、データを発生させる分布を仮定する必要がある。</p> <p>高次元データの場合($n < p$ の場合も含む)は、上記で挙げた手法がそもそも計算できなかったり、バイアスが生じて誤った結果を導いたりすることもある。よって、高次元データに対して新たな手法が必要となる。高次元データに対する統計手法には、データを発生させる分布を仮定する手法と、分布に対する条件(制約)を与える手法がある。前者は分布があらかじめ分かっている場合は問題ないが、そうでないときは、高次元データの分布を評価することは困難であるため利用できない。一方、後者であれば、制約が弱い手法であれば、ほぼ一般的なデータに対し利用可能となる。</p> <p>このように、高次元データに対する手法はまだ発展途上であり、状況によって使い分ける必要がある。現在は、分布に対する制約を比較的弱くした高次元データに対する手法の開発を行っている。</p>												
 <p>姫野 哲人 Tetsuto Himeno</p> <p>データサイエンス学部 准教授</p>	<p style="text-align: center;">サンプルサイズと次元(サンプルサイズ)の大きさに対する統計手法</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 20%;">n: サンプルサイズ</th> <th style="width: 20%;">p: 次元(変数の数)</th> <th style="width: 60%;">各種統計手法の特徴</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">大</td> <td style="text-align: center;">小</td> <td>特に分布の仮定をほとんど必要とせず、多くの統計手法が利用可能(中心極限定理による正規近似が利用可能なため)</td> </tr> <tr> <td style="text-align: center;">小</td> <td style="text-align: center;">小</td> <td>分布についての厳密な仮定(主に正規性の仮定)が必要</td> </tr> <tr> <td style="text-align: center;">大(または小)</td> <td style="text-align: center;">大</td> <td>古典的な統計手法はほとんど利用不可。変数を削減のための方法や高次元データに対する手法が必要(研究対象)</td> </tr> </tbody> </table>	n: サンプルサイズ	p: 次元(変数の数)	各種統計手法の特徴	大	小	特に分布の仮定をほとんど必要とせず、多くの統計手法が利用可能(中心極限定理による正規近似が利用可能なため)	小	小	分布についての厳密な仮定(主に正規性の仮定)が必要	大(または小)	大	古典的な統計手法はほとんど利用不可。変数を削減のための方法や高次元データに対する手法が必要(研究対象)
n: サンプルサイズ	p: 次元(変数の数)	各種統計手法の特徴											
大	小	特に分布の仮定をほとんど必要とせず、多くの統計手法が利用可能(中心極限定理による正規近似が利用可能なため)											
小	小	分布についての厳密な仮定(主に正規性の仮定)が必要											
大(または小)	大	古典的な統計手法はほとんど利用不可。変数を削減のための方法や高次元データに対する手法が必要(研究対象)											
<p>【プロフィール】</p> <p>●略歴</p> <ul style="list-style-type: none"> ・広島大学理学部卒業 ・同大学院数理学研究科修士 ・九州大学大学院数理学研究 院 学術研究員 ・情報・システム研究機構新領域融合研究センター 特任研究員 ・成蹊大学理工学部 助教 ・滋賀大学経済学部准教授 ・滋賀大学データサイエンス教育研究センター准教授 ・滋賀大学データサイエンス学部准教授 <p>●専門分野</p> <p>数理統計学</p> <p>【主な社会的活動】</p> <ul style="list-style-type: none"> ・日本統計学会公式認定 統計検定 CBT 委員会分科会 3 級副委員長 ・応用統計学会 評議員 <p>【主な著書】</p> <ul style="list-style-type: none"> ・日本統計学会公式認定 統計検定準1級対応 統計学実践ワークブック. 日本統計学会(編). 学術図書出版. 2020. ・データサイエンス入門. 竹村彰通, 姫野哲人, 高田聖治(編). 学術図書出版. 2019. ・データサイエンスのための数学. 清水昌平(編), 椎名洋, 姫野哲人, 保科架風(著). 講談社. 2019 <p>【連絡先】</p> <ul style="list-style-type: none"> ・tetsuto-himeno@biwako.shiga-u.ac.jp 	<p>【3】社会連携</p> <ul style="list-style-type: none"> ・各種公的機関におけるデータサイエンス人材育成に関わる講演を行っている。 ・日本統計学会公式認定 統計検定 CBT 分科会委員として、各種業務に携わっている。また、統計教育の普及促進に関わっている。 ・高等学校における模擬講義を行っている。 ・株式会社日立建機ティアラのコンサルティングを行っている。 												
<p>企業・自治体へのメッセージ</p>	<p>データサイエンス学部として、研究室(学部 3、4 年生)との共同研究を募集しています。また、社員・職員のデータサイエンスリテラシー向上のための講義、各種共同研究等も担当可能です。</p>												