基礎数理統計学　中間レポート

1114230018

原田雄大

**１,２．実験（シミュレーション）**

本シミュレーションはPython（NumPy, SciPy, pandas, matplotlib）を用いて実施した。

サンプルサイズはとし，試行回数はとし，乱数シードはより乱数生成の再現性を確保した．一様分布、指数分布のそれぞれの場合に対して標本平均と標準偏差推定量について不偏推定量を用いて求めた。その後、両側信頼区間95%となるように自由度 n−1 の t 分布上側97.5%点を用いて上限Uと下限Lを求めた．そして，それぞれの母平均 に対して が成り立つかどうかをカウントして繰り返したのちに試行回数Kで割ることで被覆率（Coverage Rate）を求めた．以下ではグラフと表のレイアウトの問題で１と２を並べている。

　まずサンプルサイズ（Size）ごとの被覆率（Coverage Rate）の関係を表にした。

表１.左：一様分布（Uniform）右：指数分布：（Exponential）



　表1によると，一様分布ではn=５から93％を超えているが，n=10～100では単調に増加するわけでもなかった。指数分布では初め88%となっており比較的低い被覆率を示しているが徐々にその値を増やし最終的にn=200で94.5％となっている。n=200ではどちらも95%に近い値をとっている。またこれら被覆率の95％からの誤差の絶対値をグラフにした。

グラフ, 折れ線グラフ

AI 生成コンテンツは誤りを含む可能性があります。グラフ, 折れ線グラフ

AI 生成コンテンツは誤りを含む可能性があります。

図１.左：一様分布（Uniform）右：指数分布：（Exponential）

　図1によると，一様分布ではn=30の時からとなっている．一方で指数分布ではn=200の時を除いての中に収まっていない．

３．考察

　まず今回のt分布による信頼区間の求め方は、母集団が正規分布であり、分布が対称であることに留意する。一様分布に関して、分布が対称かつ有界であるためｎが小さい段階からでも被覆率が95%に収束しやすかったのである。指数分布に関して，分布が非対称であるためｎが小さいときは被覆率が95%を大きく下回った。これはnが小さい場合は特にｔ分布の前提である正規性が崩れているためである。一方で，被覆率の95％からの誤差のグラフ（図2）を見ると，サンプルサイズの増加に伴い誤差は約の範囲に収束する傾向がある。これは中心極限定理より両分布とも標本平均が正規に近づくためである。この「」という水準は，本シミュレーションがモンテカルロ法の一種であることに由来する。すなわち，母分布からランダムに標本を作製して信頼区間の被覆可否を 5000 回繰り返し計算するため，まず被覆率そのものに約±0.31％の誤差が生じる。この推定値に対して95％信頼区間を構成するときにはの範囲でばらつくため、実際に誤差が約±0.5％前後に収束するのは妥当な結果と言える。

グラフ, 折れ線グラフ

AI 生成コンテンツは誤りを含む可能性があります。

図２.一様分布（青）と指数分布（黄）の誤差

本シミュレーションで用いたコードについては

GitHub（<https://github.com/yu531deve/FundamentalMathematicalStatistics>）

で公開している。