P5-18

動的計画法によるリターン分布推定

森村 哲郎†*, 杉山 将††, 鹿島 久嗣‡, 八谷 大岳††, 田中 利幸†‡

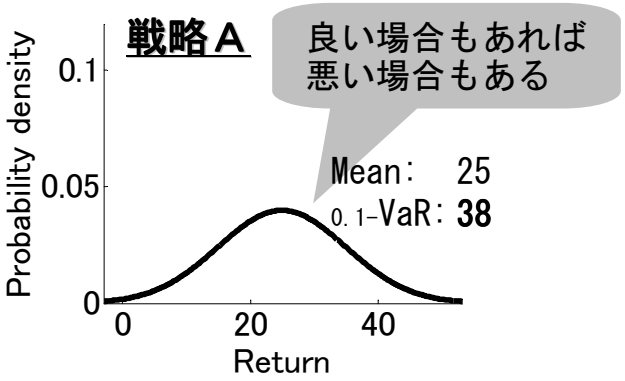
†:IBM東京基礎研究所, ††:東京工業大学, ‡:東京大学, †‡:京都大学

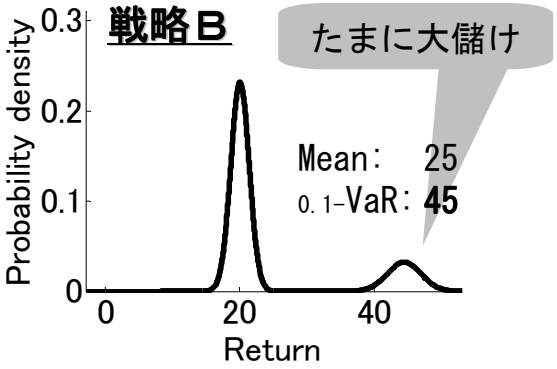
*tetsuro@jp.IBM.com

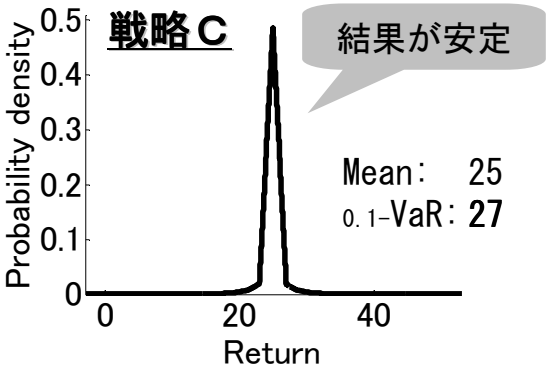
➢リターン(累積報酬和)の分布

$$\Pr(\text{return} \mid \text{state}, \text{policy})$$

応用例: リスク指標(VaR)の算出、
リスク考慮型意思決定など







問題: リターンの観測まで時間遅れがあるため、
単純なモンテカルロ法では推定が困難

P5-18

動的計画法によるリターン分布推定

2/2

森村 哲郎†*, 杉山 将††, 鹿島 久嗣‡, 八谷 大岳††, 田中 利幸†‡

†:IBM東京基礎研究所, ††:東京工業大学, ‡:東京大学, †‡:京都大学

*tetsuro@jp.IBM.com

➢リターン分布推定

- 分布Bellman方程式(リターン分布の再帰式) [IBIS06,中田&田中] を利用
  - **Parametric approach** [UAI10, Morimura et al.]
  - **Nonparametric approach** [ICML10, Morimura et al.]
- これらの手法の**理論的性質は未解明**のまま

➢目標:分布Bellman方程式にもとづく手法の収束性解析 & 効率化

- 動的計画法で分布Bellmanを解いた場合を解析

  ⇒ 初期推定分布よらず、

  常に真のリターン分布に収束

  ⇒ 低次モーメントの収束率の改善が大切
- 解析結果を用いて、

  効率の良い分布推定法を提案