

# 今までの分析結果からわかっている傾向

1. 周期性はなく各周波数成分が万遍なく含まれている
2. 平均や分散に曜日や時間帯、場所に応じた顕著な傾向はない
3. **分布が二つ、三つの分布帯に分かれる**  
➤ **3 時台が三つの分布帯に分かれやすい様に見えたが未検証**
4. 場所やイベント事の有無による顕著な傾向は見られない
5. **実測値に対する時系列モデル回帰では、細かな変動を捉えられないが、中期的な傾向を捉えられる**
6. 変動値に対する時系列回帰では、平均的に実測値よりも良い回帰精度を示す
7. 実測値に対するモデルパラメータを基としたクラスタリングでは曜日や時間帯に応じたモバイル通信トラヒックの変化に応じた傾向がある

# 今後の方針

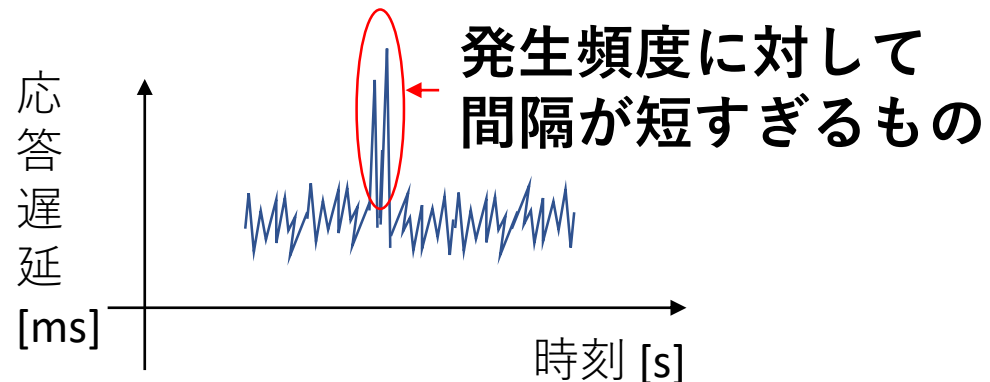
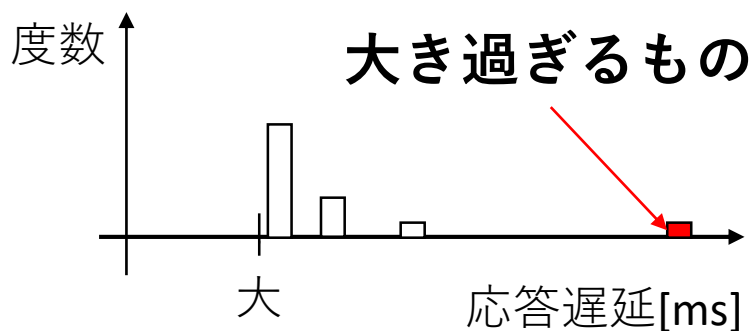
1. 長期的もしくは通常とは異なるネットワーク利用状況での計測データに対して分析を行う．これに向けて計測実験を追加で行う．計測対象のサーバは要検討
2. **他の手法と組み合わせることによる異常検知手法の精度向上**
  - I. 時系列モデルの回帰精度の向上
    - 「データの前処理として Box-Cox 変換や移動平均を取る」「スパイク状の応答遅延を取り除いて回帰を行う」など、さまざまな方法を実施してきたが、あまり良い成果は得られていない
  - II. **時系列モデルで捉えられない傾向を捉える手法の設計**
3. 異常検知手法の設計と検証
  - 異常を明確に定めることと、その異常が発生している状況で計測したデータが必要

# 時系列モデルでは捉えられない傾向とその傾向を捉えるための手法

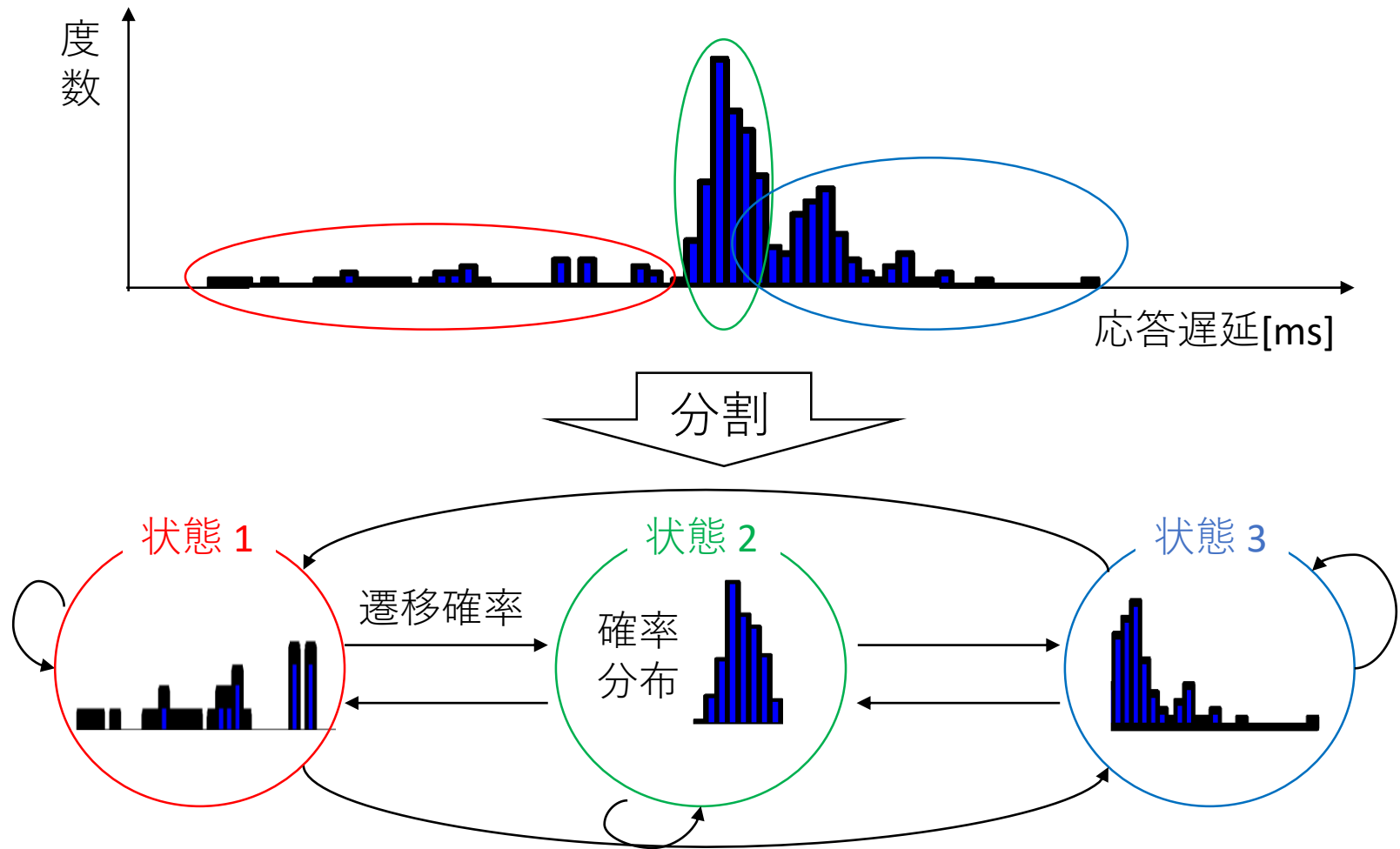
1. 単発的に発生する大きな応答遅延
  - I. 発生頻度やその大きさの分布
2. 短期的な変動
  - I. 各計測時点間での分布帯間の遷移を表した確率モデル
3. 計測時間帯による分布の違い
  - I. 金尻先輩が行っていた分布に基づくクラスタリング

# 1. 単発的に発生する大きな応答遅延の傾向を捉えるための発生頻度や分布の分析

- 大きな応答遅延 := 一定値以上の応答遅延  
100 ms や,  $\alpha * (\text{平均値})$  など  
or := 時系列モデルでの回帰結果において, ノイズ項を加味した信頼区間を上を外れた応答遅延
- 発生頻度 := 一時間の区間データに含まれる大きな応答遅延の比率
- 異常とみなすもの



## 2. 短期的な変動を捉えるための各計測時点間の 応答遅延の確率遷移モデル



変動の異常度  $:= 1 - (\text{遷移確率}) * (\text{遷移後の状態における発生確率})$  など

### 3. 計測時間帯による分布の違いを捉えるための分布に基づくクラスタリング

金尻先輩が行っていた、確率分布間距離に KLダイバージェンスを用いたクラスタリングを行う

クラスタ数やクラスタリング手法はこれまでの経験を活かしながら定量的によいものを用いる