曖昧な周期をもつ作業の発生パターン検出手法

北垣 千拡
岡山大学 工学部 情報工学科
平成26年2月14日

研究背景

＜作業＞

同様の作業が繰返し発生 (例：研究ミーティング，講義)

→ 周期がわかれば予測可能

＜曖昧な周期性＞

作業の発生にはさまざまな要因が影響

例：定例会議 (1) 約2週間に1度

(2) 参加者の都合に左右

(3) 長期休暇の存在

作業の発生は単純な一定間隔の周期ではない

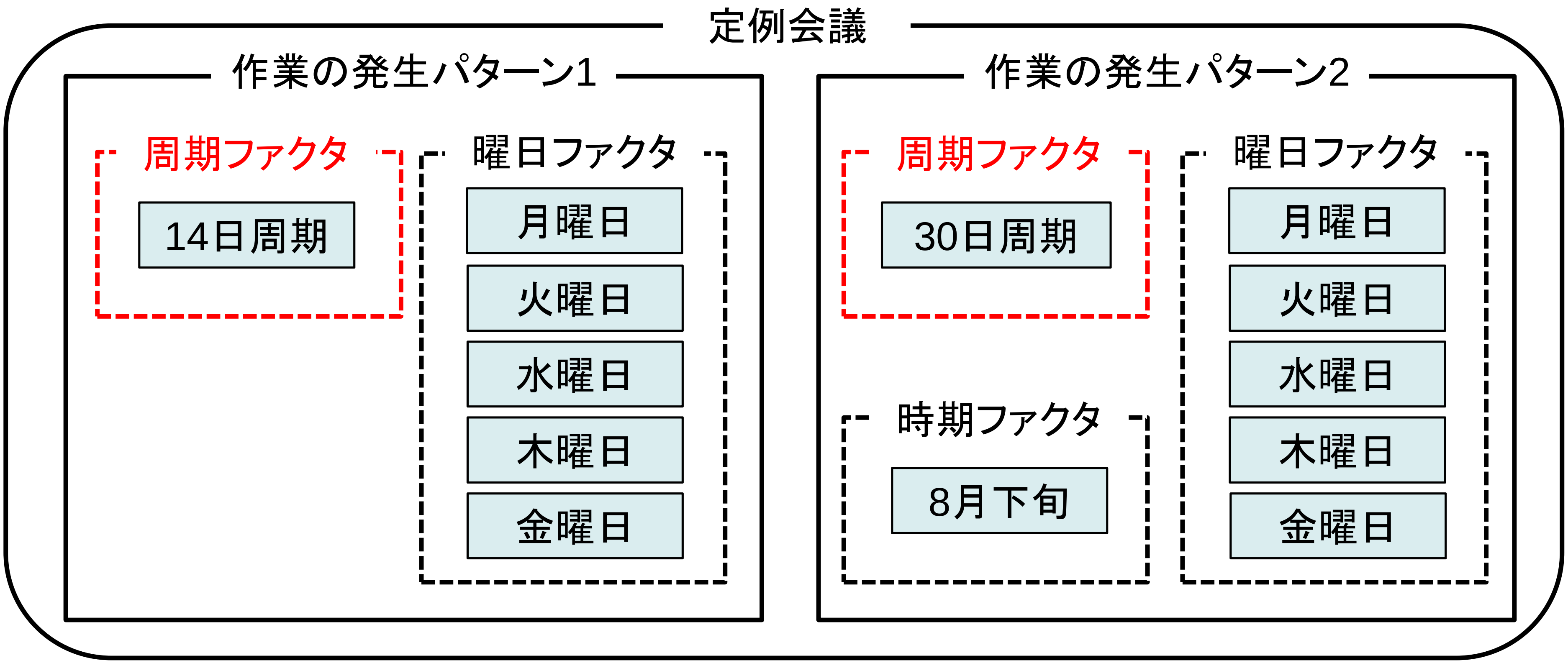
→ 予測のためには曖昧な周期性を考慮する必要

作業の発生パターン

曖昧な周期を表現するために作業の発生パターンを定義

作業の発生パターン1：約2週間周期での発生を表現

作業の発生パターン2：長期休暇を表現



実験

作業の発生を波とみなす
自己相関関数を用いて波を解析することで周期ファクタを検出

＜目的＞

実際の作業の周期ファクタを検出

＜実際の作業＞

対象 : 定例会議
期間 : 2011年度4月から2013年度9月まで
特徴 : (1) 約30日に1度発生
(2) 土日祝日には発生しない
(3) 年末年始には発生しない

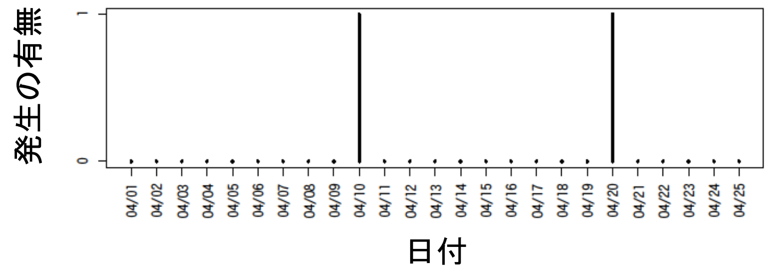
＜手順＞

(1) 作業の発生の波のデータを用意
(2) 自己相関関数を用いて周期ファクタを検出

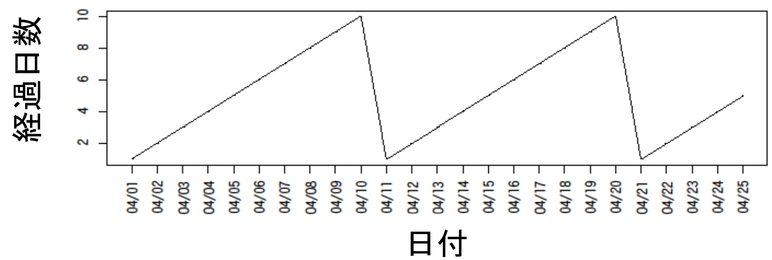
データ形式

データ形式として次の3つの形式を提案

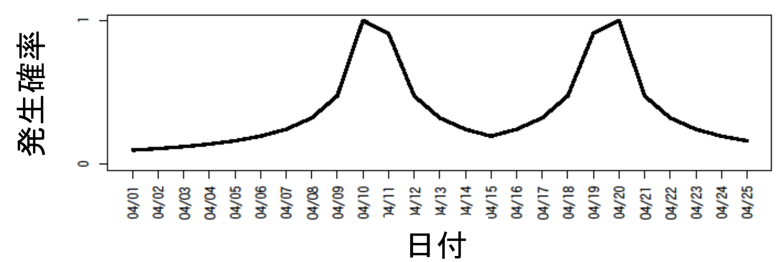
＜形式1＞
　横軸: 日付
　縦軸: 発生の有無



＜形式2＞
　横軸: 日付
　縦軸: 経過日数



＜形式3＞
　横軸: 日付
　縦軸: 発生確率



各データ形式は以下の3つの観点を考慮
　(1) 曖昧さに対する寛容さ
　(2) ノイズの影響の受けにくさ
　(3) 長期休暇を重ね合わせる精度

観点

(1) 曖昧さに対する寛容さ
自己相関関数による重ね合わせのずれを容認するかどうか
(例) 約14日周期なら13,15日間隔も約14日周期として容認

(2) ノイズの影響の受けにくさ
例外的な発生間隔の影響を受けにくいかどうか
(例) ある年にだけ存在する長期休暇の影響を無視

(3) 長期休暇を重ね合わせる精度
長期休暇を重視するかどうか
(例) 14日間隔と30日間隔があれば, 30日間隔を優先

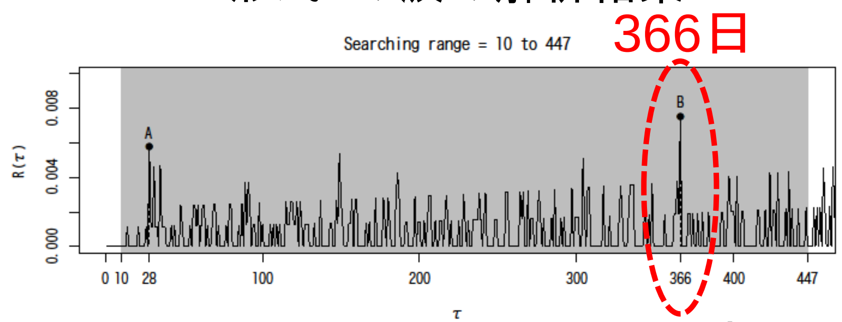
実験結果

すべての形式で約365日周期が算出

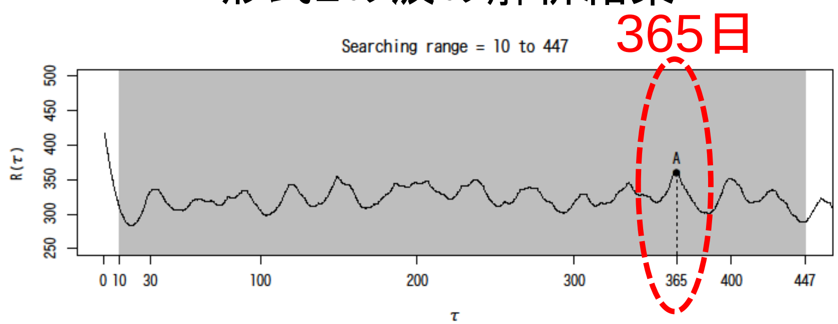
(1) 約30日周期で発生
(2) 1年に1度の長期休暇の存在

定例会議は1年周期で発生

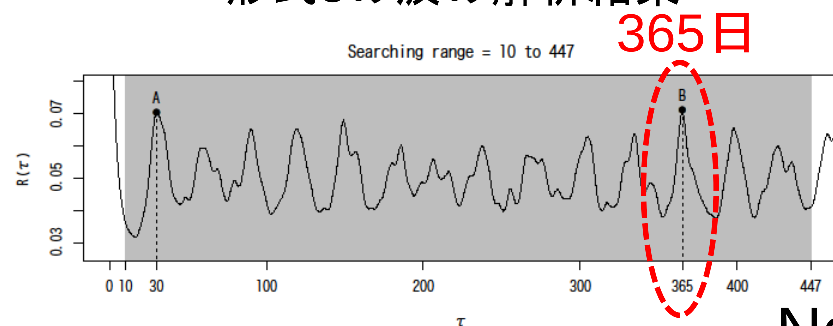
形式1の波の解析結果



形式2の波の解析結果



形式3の波の解析結果



曖昧さに対する寛容さの考察

	形式1		形式2		形式3	
順位	τ	$R(\tau)$	τ	$R(\tau)$	τ	$R(\tau)$
1	366	0.0076	365	360	365	0.071
2	28	0.0058	364	360	30	0.070
3	149	0.0054	366	359	31	0.070
4	304	0.0051	363	356	366	0.069
5	35	0.0047	149	355	364	0.069
6	31	0.0046	367	353	29	0.069
7	435	0.0043	150	352	32	0.069
8	428	0.0043	399	352	149	0.068
9	423	0.0042	400	352	33	0.067
10	186	0.0042	398	350	150	0.067

365日周期付近の周期が，形式1は1つ，形式2は5つ，形式3は3つ算出

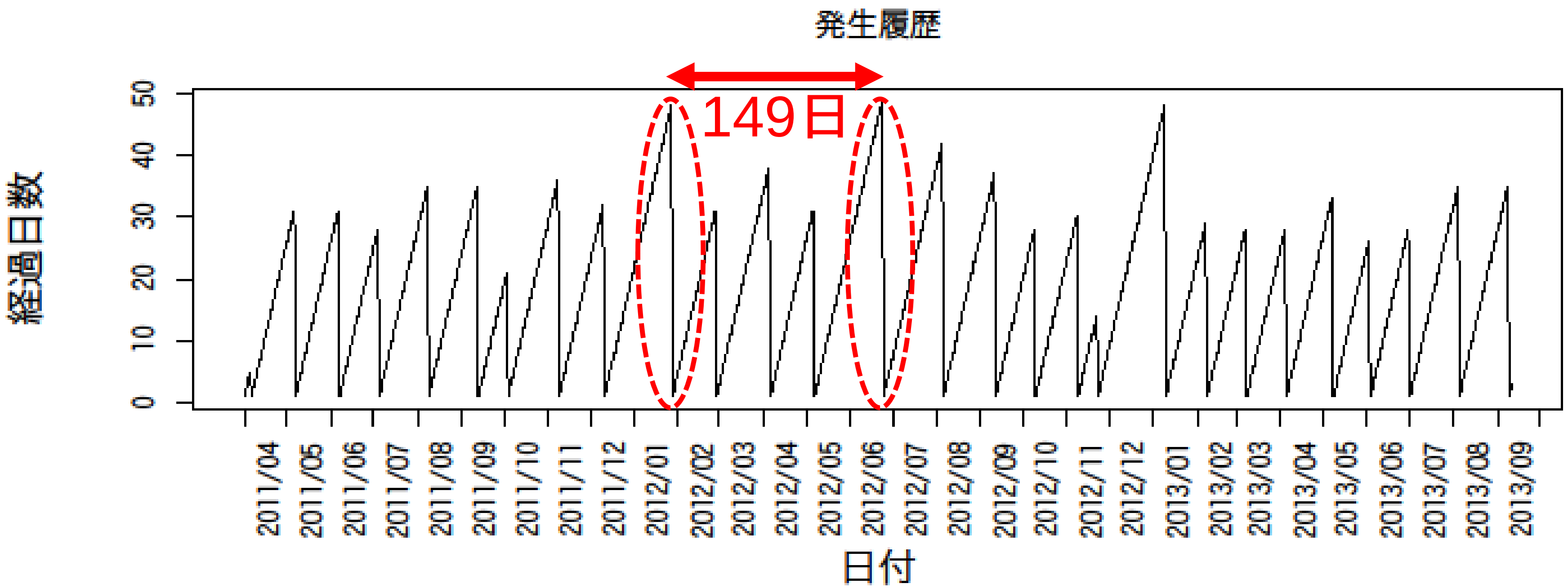
⇒ 形式1と比べて形式2，3は曖昧さに対する寛容さが高い

ノイズの影響の受けにくさの考察

各形式の自己相関の上位10位以内に149日という周期が算出

➡ 各形式でノイズの影響を受けている

➡ 各形式で「ノイズの影響の受けにくさ」に差はない



長期休暇を重ね合わせる精度の考察

長期休暇を重ね合わせる精度

⇕ トレードオフ

ノイズの影響の受けにくさ

各形式で「ノイズの影響の受けにくさ」に差はない

➡ 各形式で「長期休暇を重ね合わせる精度」に差はない

まとめ

＜実績＞

(1) 曖昧な周期を表現するモデルの定義

(2) 周期ファクタの検出手法の提案

(3) 周期ファクタの検出に用いる3つのデータ形式の提案

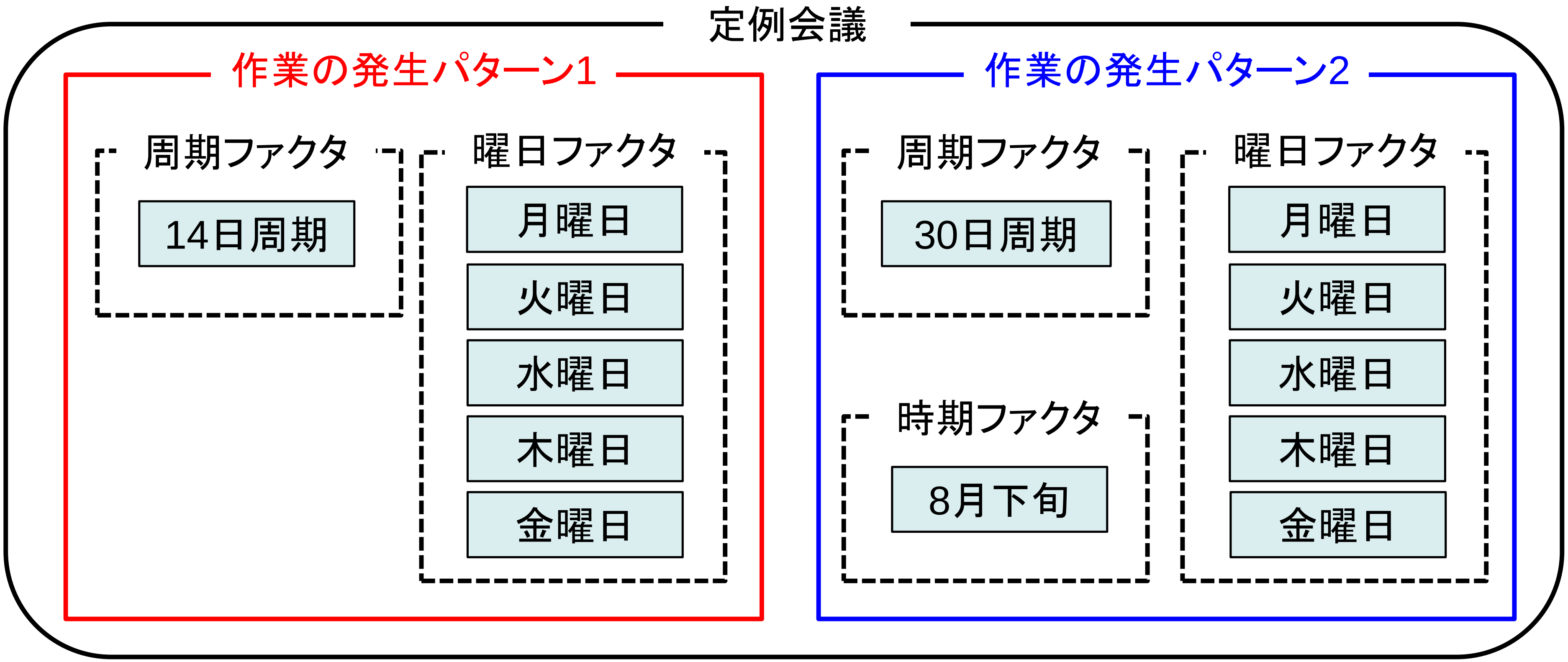
＜残された課題＞

(1) 周期ファクタの検出に用いるデータ形式の選択

(2) 周期ファクタ以外のファクタの検出

予備スライド

作業の発生パターン



作業の発生パターン1 : 約2週間周期での発生を表現

作業の発生パターン2 : 長期休暇を表現

周期ファクタの検出

作業の発生を波とみなして解析

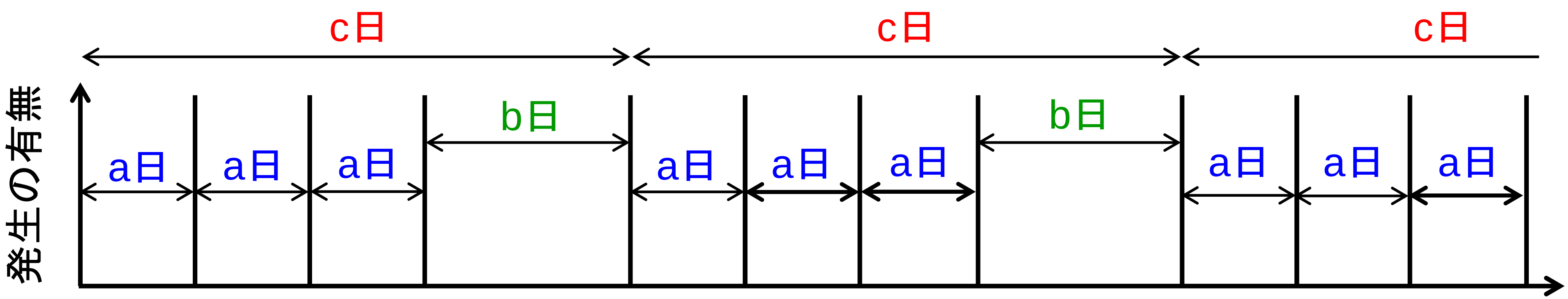
＜作業発生の波＞

(1) 小さな波
- (A) ある発生から次の発生までの間の波を1周期とした波
- (B) 繰返しだけでは作業発生の波を表現できない波

(2) 大きな波
- (A) 繰返しだけで作業発生の波を表現できる波
- (B) (A)を満たす波のうち最小の周期をもつ波

作業発生の波は大きな波の繰返しのみで表現可能

大きな波の周期の算出 = 周期ファクタの検出



日付

小さな波 : a日周期の波, b日周期の波

大きな波 : c日周期の波

各データ形式の特徴

データ形式	曖昧さに 対する寛容さ	ノイズの影響の 受けにくさ	長期休暇を 重ね合わせる精度
形式1	×	○	×
形式2	○	×	○
形式3	○	○	×

(1) 曖昧さに対する寛容さ
自己相関関数による重ね合わせのずれを容認するかどうか
(例) 約14日周期なら13,15日間隔も約14周期として容認

(2) ノイズの影響の受けにくさ
例外的な発生間隔の影響を受けにくいかどうか
(例) ある年にだけ存在する長期休暇の影響を無視

(3) 長期休暇を重ね合わせる精度
長期休暇を重視するかどうか
(例) 14日間隔と30日間隔があれば，30日を優先

自己相関関数

波に含まれる繰返しパターンの検出に有用

地点 t における作業発生の波を $f(t)$ としたとき
自己相関関数 $R(\tau)$ は次のように表される

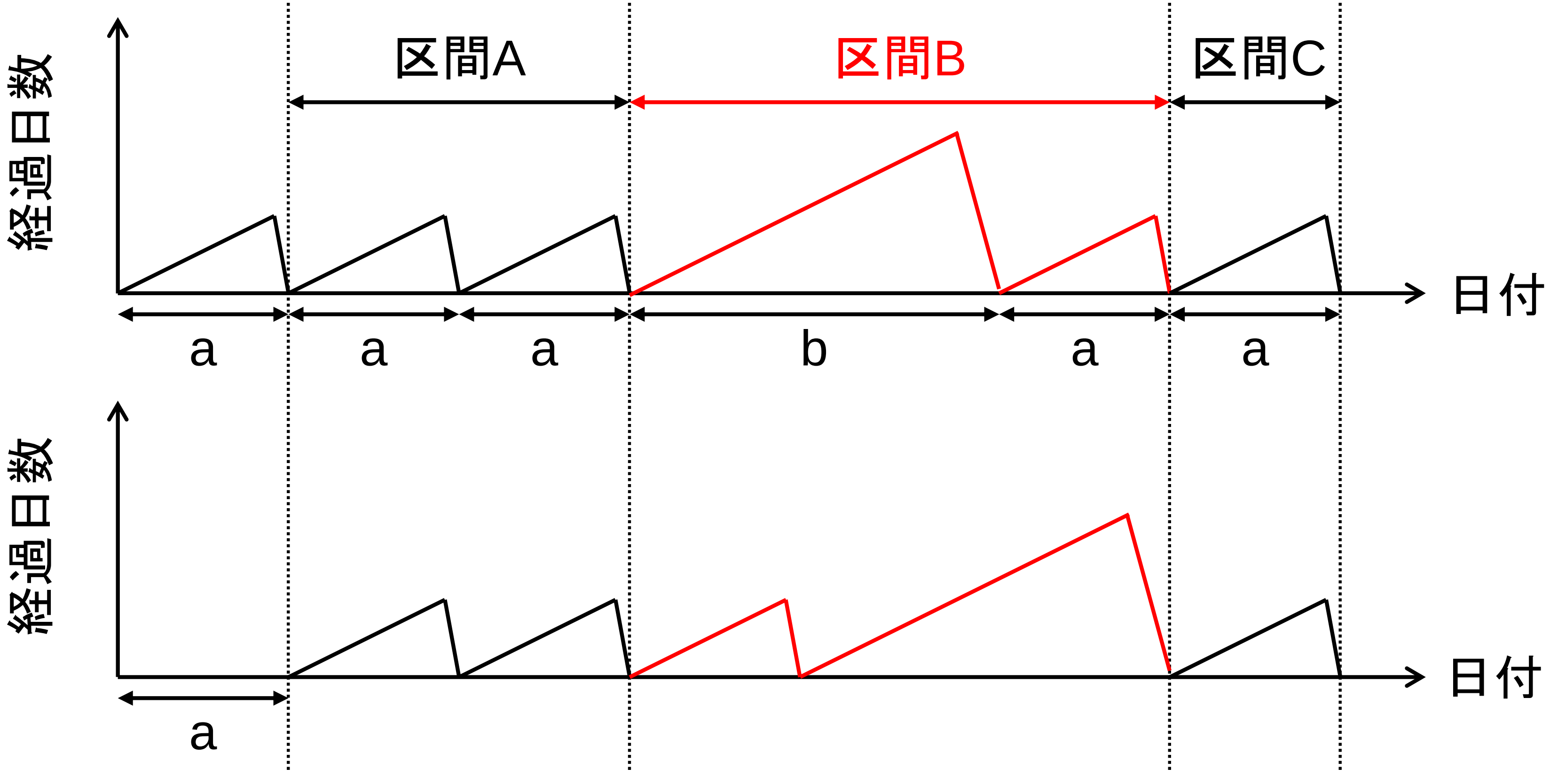
$$R(\tau) = \sum f(t) f(t-\tau)$$

このとき，最大の$R(\tau)$を与えるτが大きな波の周期となる

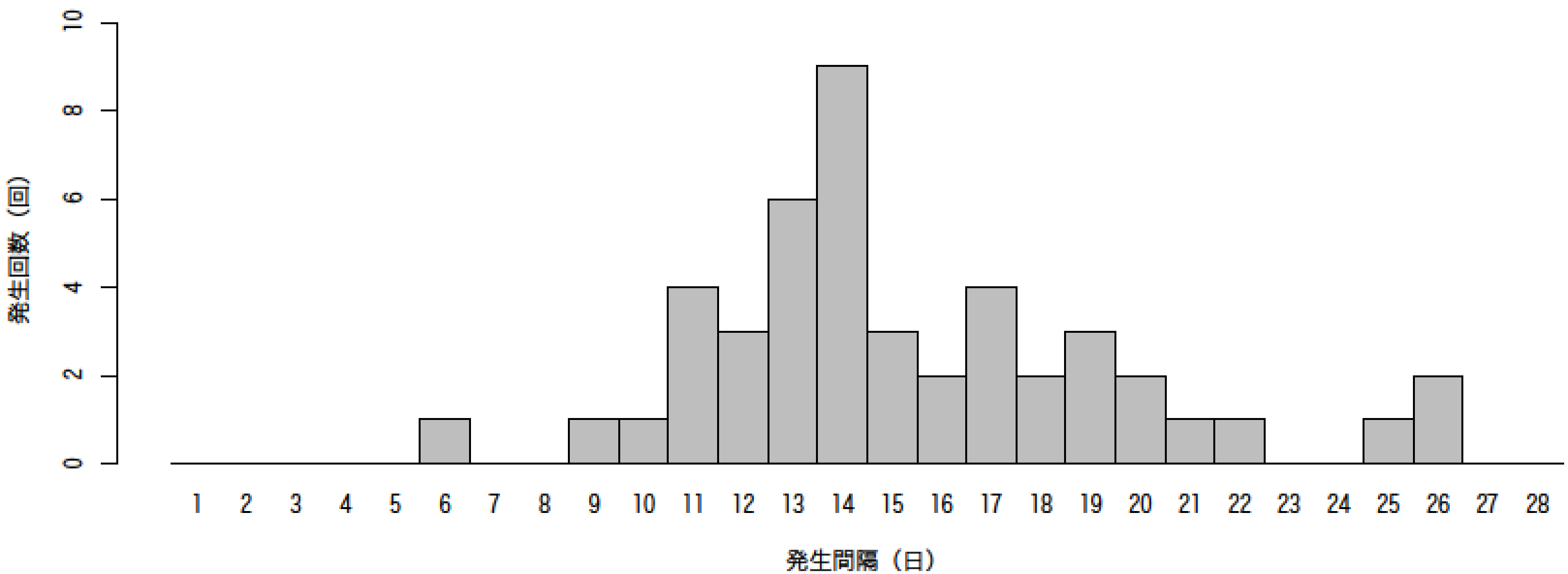
形式2で小さな波が含まれない理由

形式1, 3において区間Bの差分は小さい

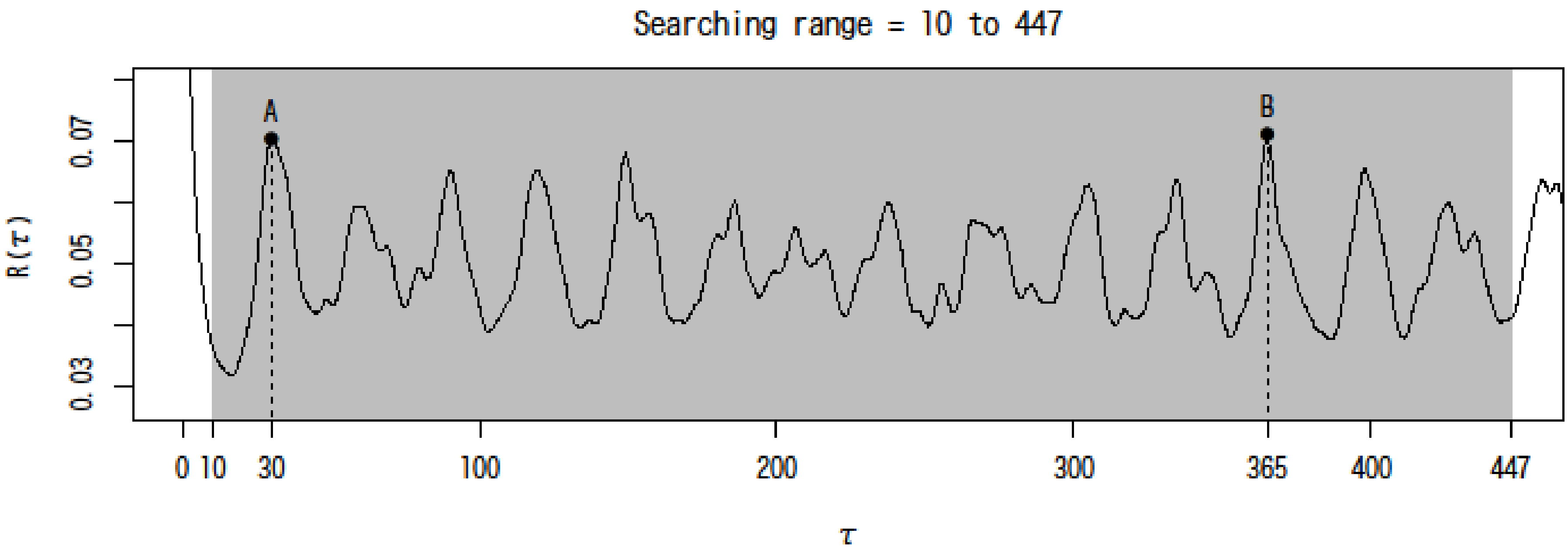
形式2において区間Bの差分は大きい



発生間隔ごとの発生確率の分布



形式3による自己相関R(τ)



大きな波の繰返し

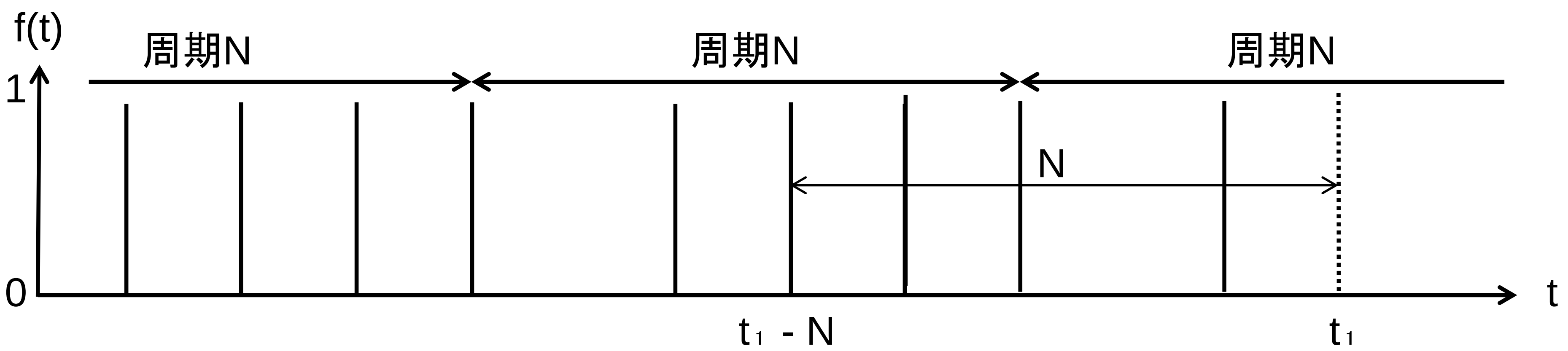
地点 t における作業発生の波を f(t), 大きな波の周期をNとする

作業発生の波は大きな波の繰返しのみで表現可能

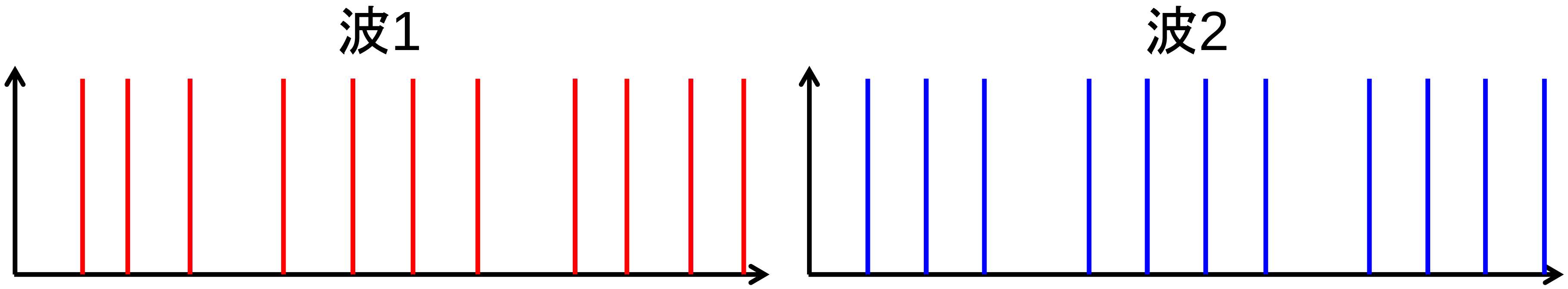
➡ ある地点 t_1 の作業発生の波は $f(t_1) = f(t_1 - N)$ となる

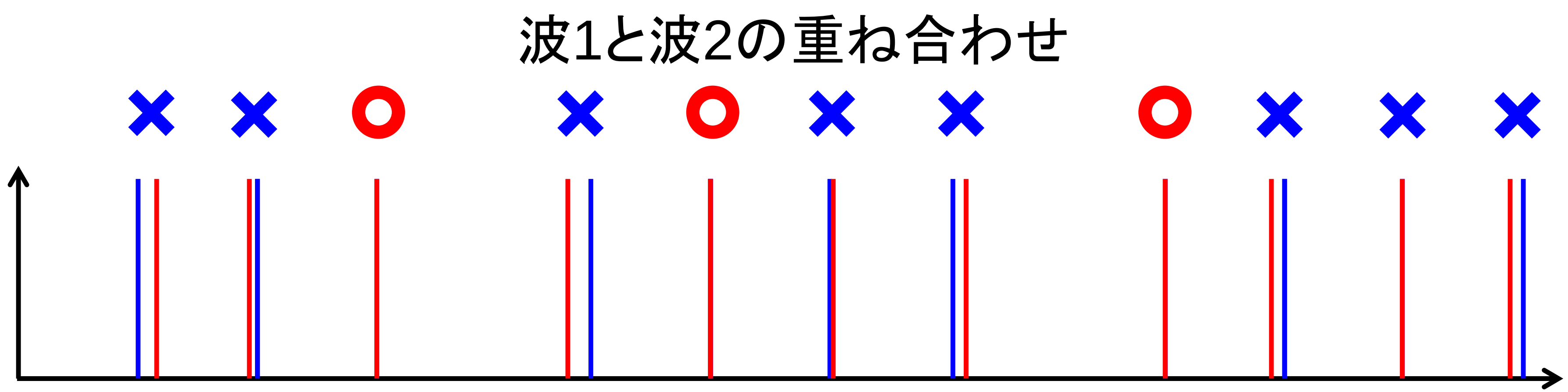
仮定 : 大きな波の周期が既知

結論 : 小さな波の周期を知ることなく発生の予測が可能



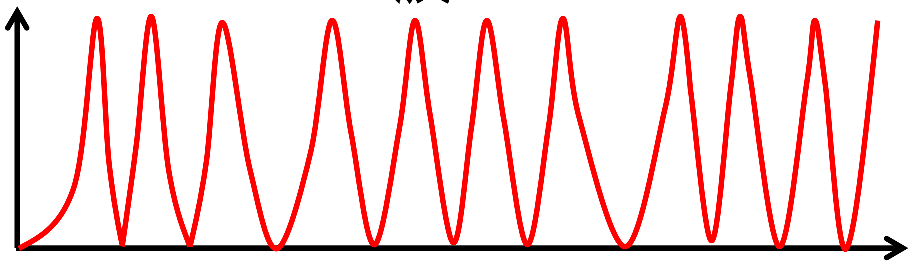
曖昧さに対する寛容さの例(1/2)

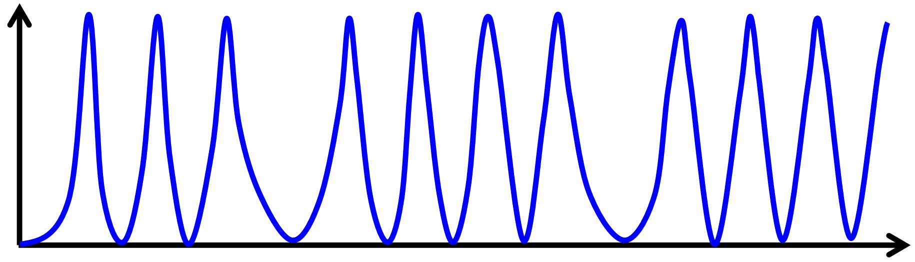




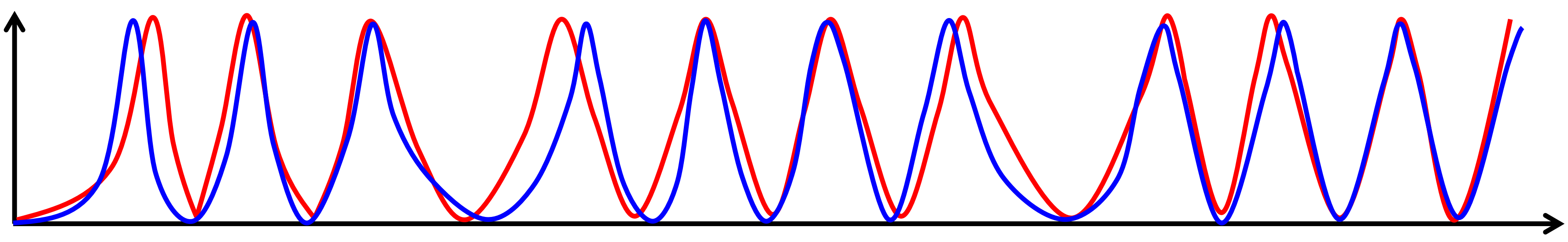
少しのずれで全く一致していないとみなされてしまう

曖昧さに対する寛容さの例(2/2)





波1と波2の重ね合わせ



あらゆる地点での波がほぼ一致

ノイズの影響の受けにくさの例