

利用者投稿型動画サイトにおける動画利用分析

飯田委哉, 伊東栄典

九州大学

近年, YouTubeやニコニコ動画などの利用者投稿型動画サイトが人気である。これらのCGM (Consumer Generated Media) サイトには毎日多数の動画が投稿されており, また膨大な利用者が動画を閲覧している。利用者のアクセス動向を知ることはサイトの質の向上につながる。利用者投稿型サイトニコニコ動画を対象に、利用者の視聴モデルの確立を目指す。本研究では、アクセス数である再生数を対象に統計的分析を行った。

背景

サービス開始から数年経過した利用者投稿型コンテンツ、いわゆるCGM (Consumer Generated Media) では、コンテンツの多様性喪失 (画一化) が進んでいる。画一化が進み、多様性を失った文化からは、新たな文化的イノベーションが発生しにくいと思われる。本研究では、CGMサイトであるニコニコ動画を対象に、多様性が減少している理由を分析するためにニコニコ動画視聴者による動画選択モデルの推定を目指す、その前段階として、ニコニコ動画内の動画利用分析を行う。

多様な動画?

ニコニコデータセットを用いた

ニコニコ動画には様々な情報が付与されている → Pythonで加工、Rで分析・可視化。



項目	内容
件数	830万件
内容	動画のメタデータ
形式	テキスト形式・JSON形式
期間	2006年12月~2012年10月に投稿された動画
時点	2013年4月時点のデータ

【巡音ルカ】Just Be Friends【オリジナル曲PV】



video_id	view_counter	comment_counter
nm19251588	2	1
sm88823	13521	154
sm7528841	3542025	142635

多様性が失われ、画一化されつつある? ランキング上位が同じようなコンテンツ?

分析① (動画再生数分布の推定)

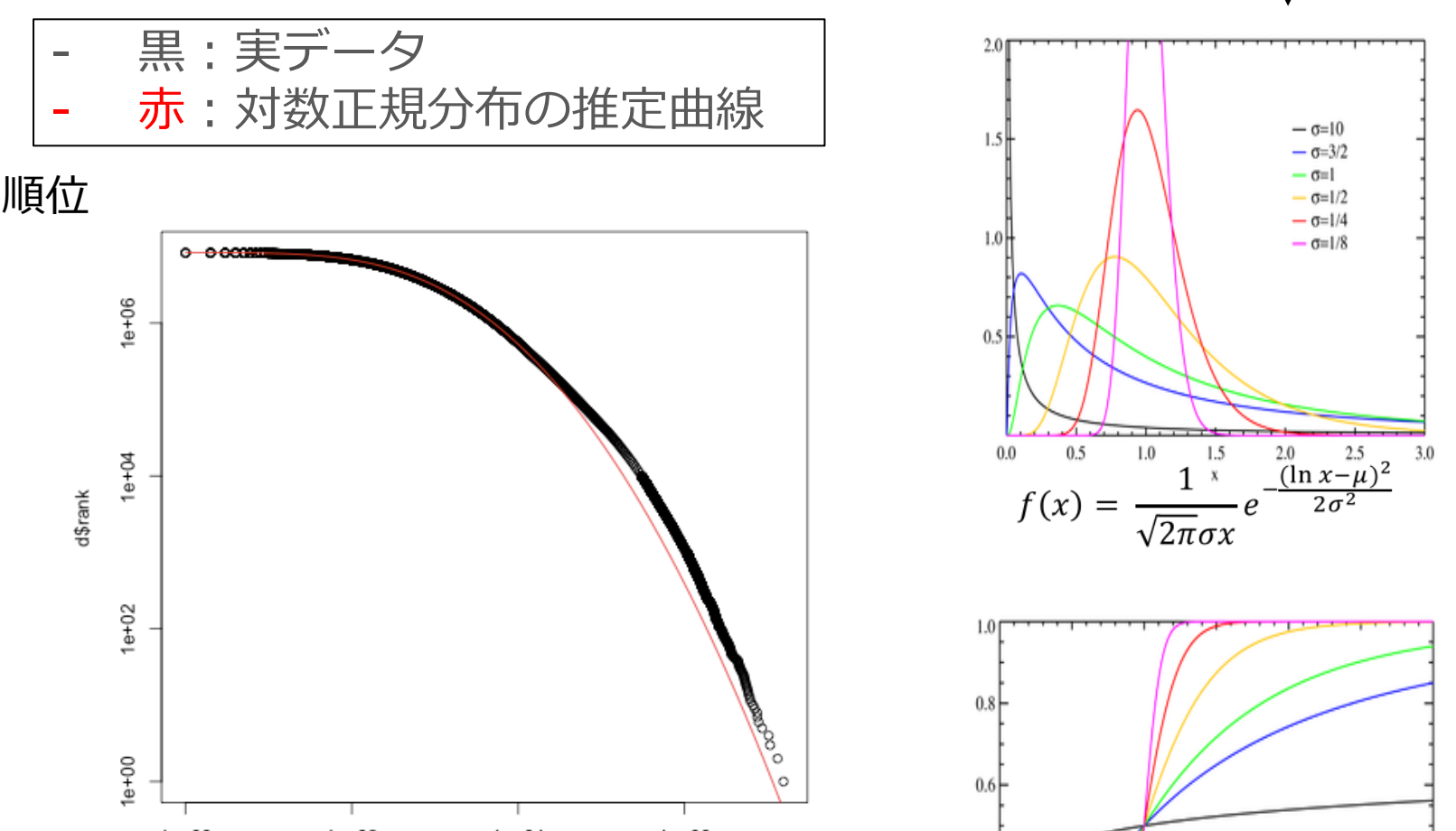


高
↓
人気
↓
低

- ランキングを見て動画を選ぶことが多い
- ランキングは再生数ベース
- 再生数の高い動画は再生されやすい? →再生数の分布をみている

再生数と再生数順位から各分布(累積分布関数)のパラメータを推定
実データの散布図と推定曲線を視覚的に比較
AICも算出

- 正規分布
- べき分布
- 指数分布
- 対数正規分布



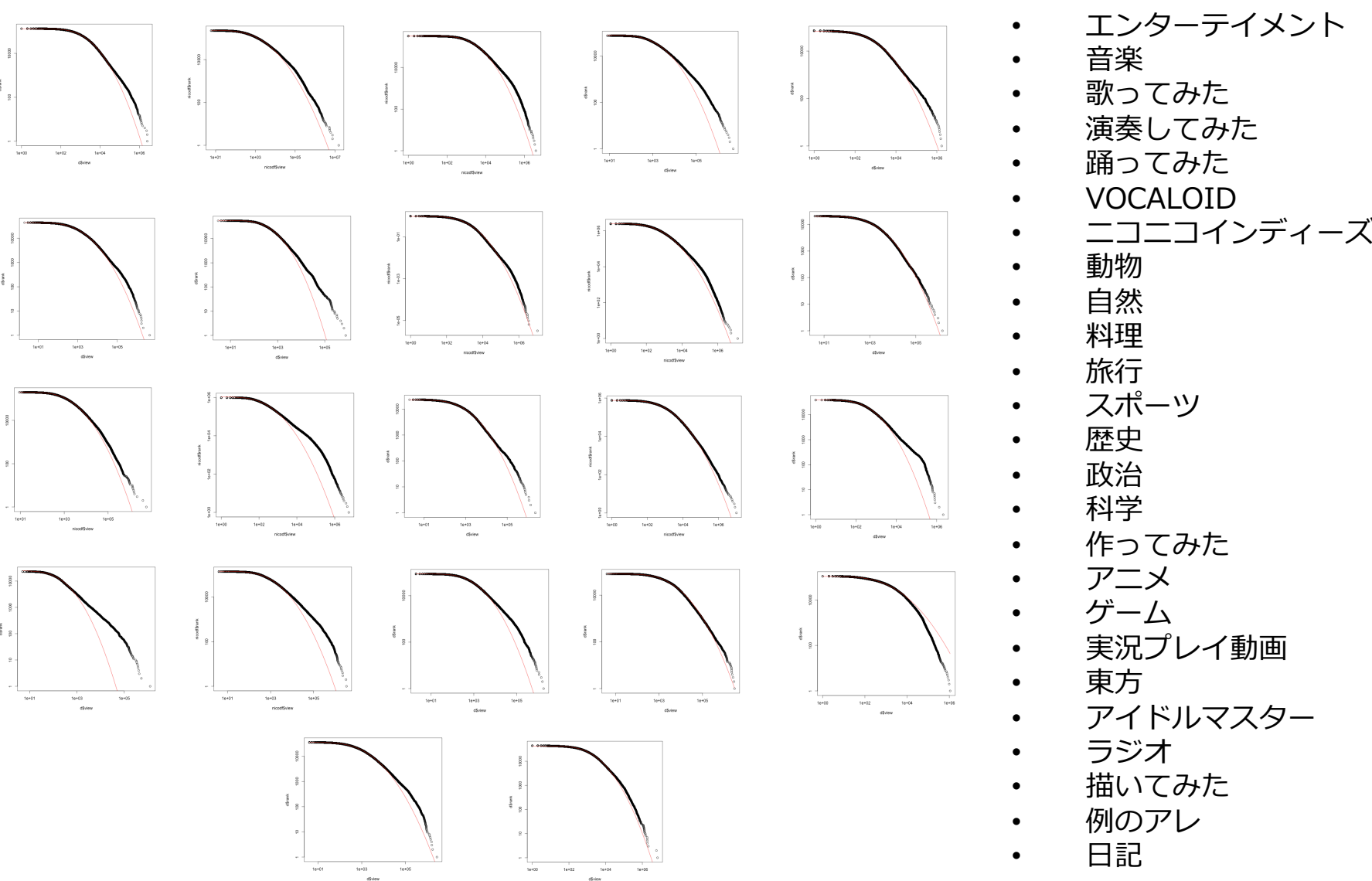
対数正規分布と合致! 回数

分布	AIC
正規分布	242972635
べき分布	251853728
指数分布	253249632
対数正規分布	182385398

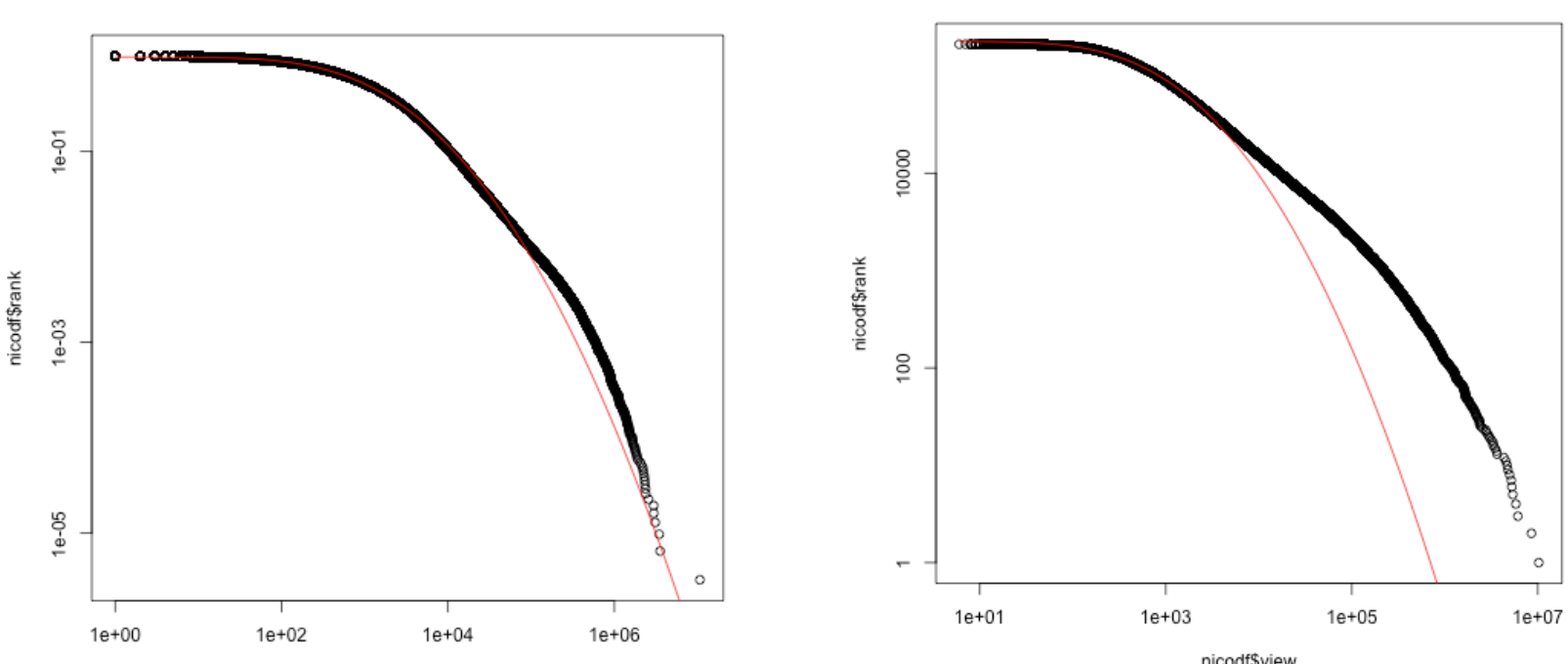
- ロングテール型の分布
- 数字が大きいほど増加しやすい傾向がある
- 再生数の高い動画は再生されやすい

分析② (ジャンル別再生数分布の推定)

- 全動画では対数正規分布
- カテゴリランキングでも動画を見ることがある
- カテゴリごとでも対数正規分布? →カテゴリごとの再生数の分布をみている

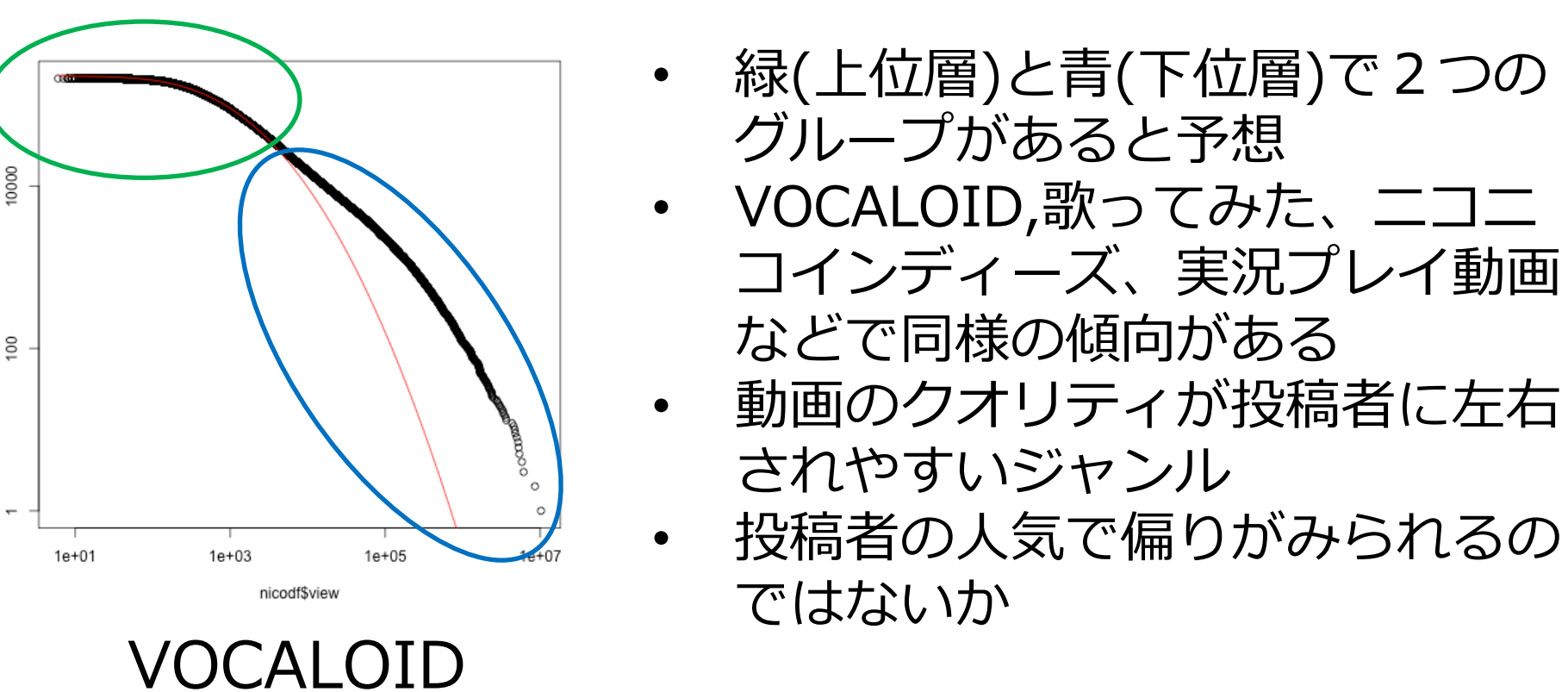


大きく2パターンに



エンターテイメント
動画集合全体同様、推定曲線と実データが一致

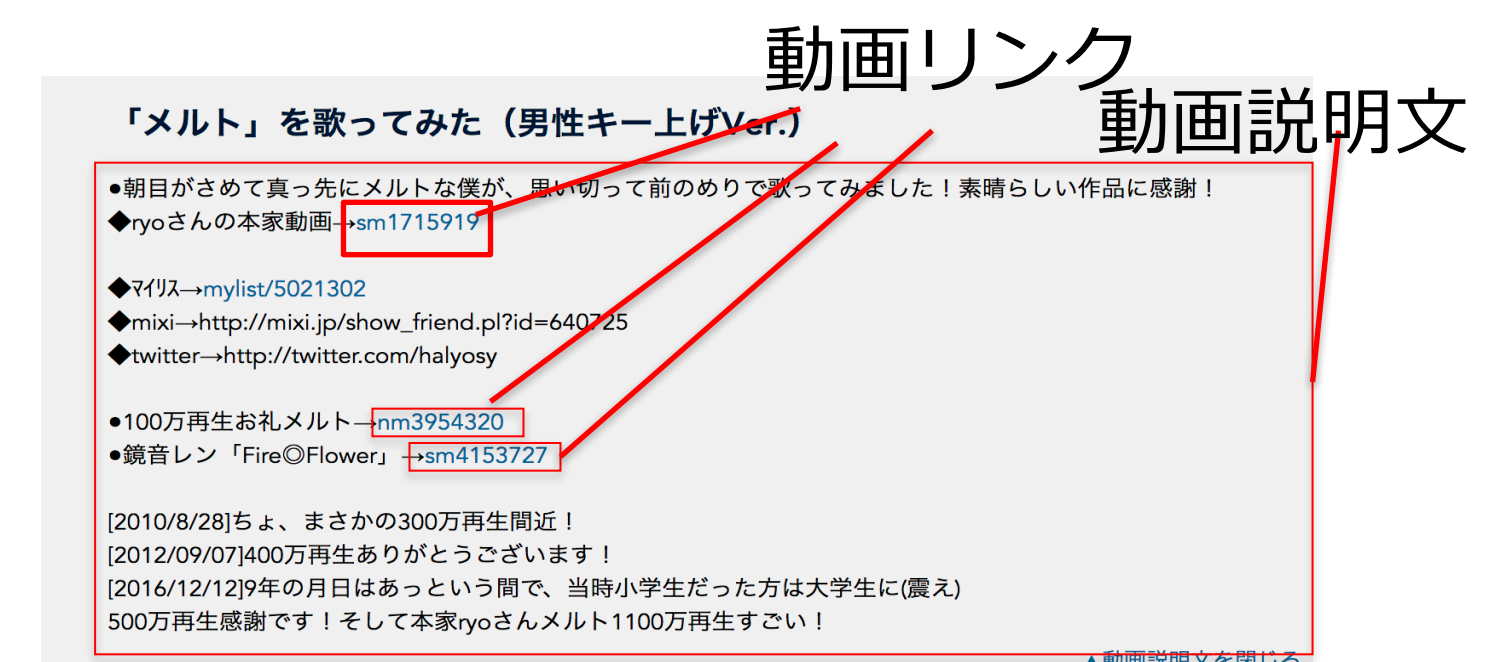
VOCALOID
途中から推定ランクが実データよりも過大に見積もっている



- 緑(上位層)と青(下位層)で2つのグループがあると予想
- VOCALOID, 歌ってみた、ニコニコインディーズ、実況プレイ動画などで同様の傾向がある
- 動画のクオリティが投稿者に左右されやすいジャンル
- 投稿者の人気で偏りがみられるのではないかと

分析③ (動画説明文リンク数)

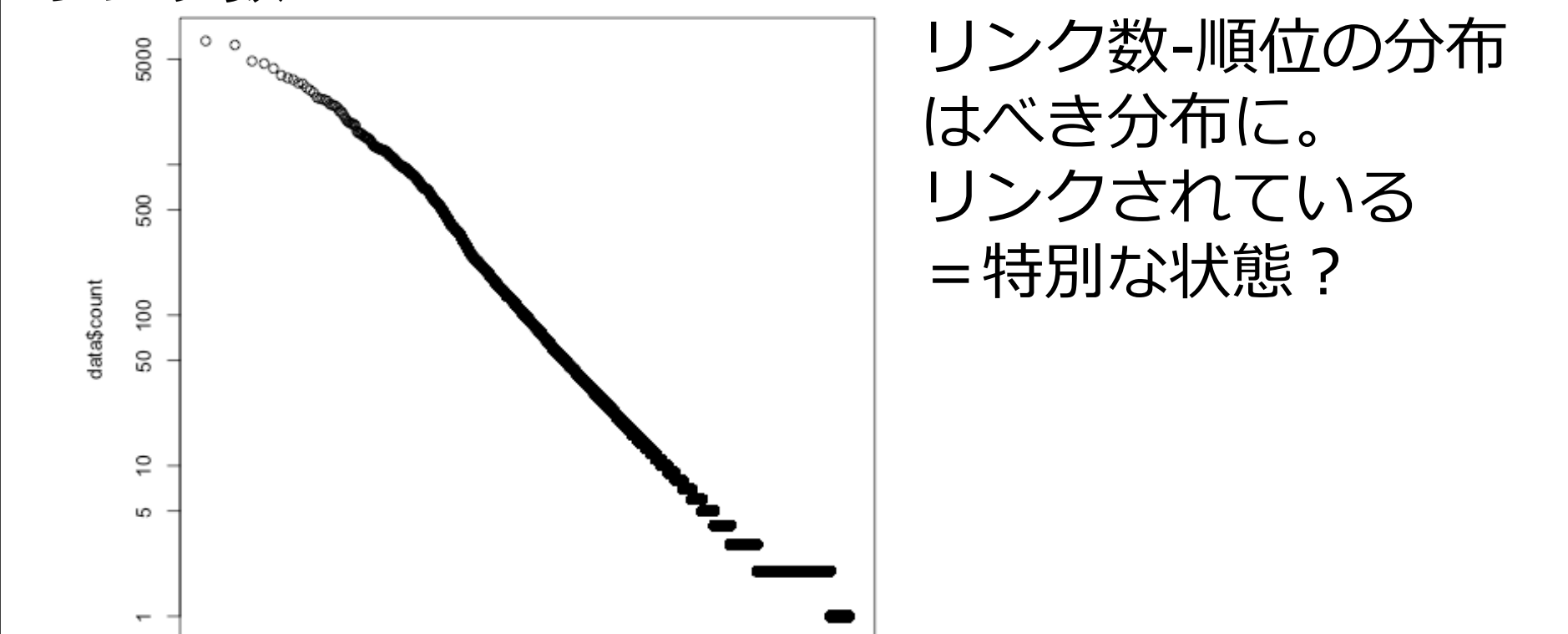
- 動画説明文以外にもリンクを辿って動画を探す
- 説明文中のリンクには動画に関連の強い動画がある(いわゆるn次創作) →リンクについて調べることで動画選択の傾向が見えるのでは



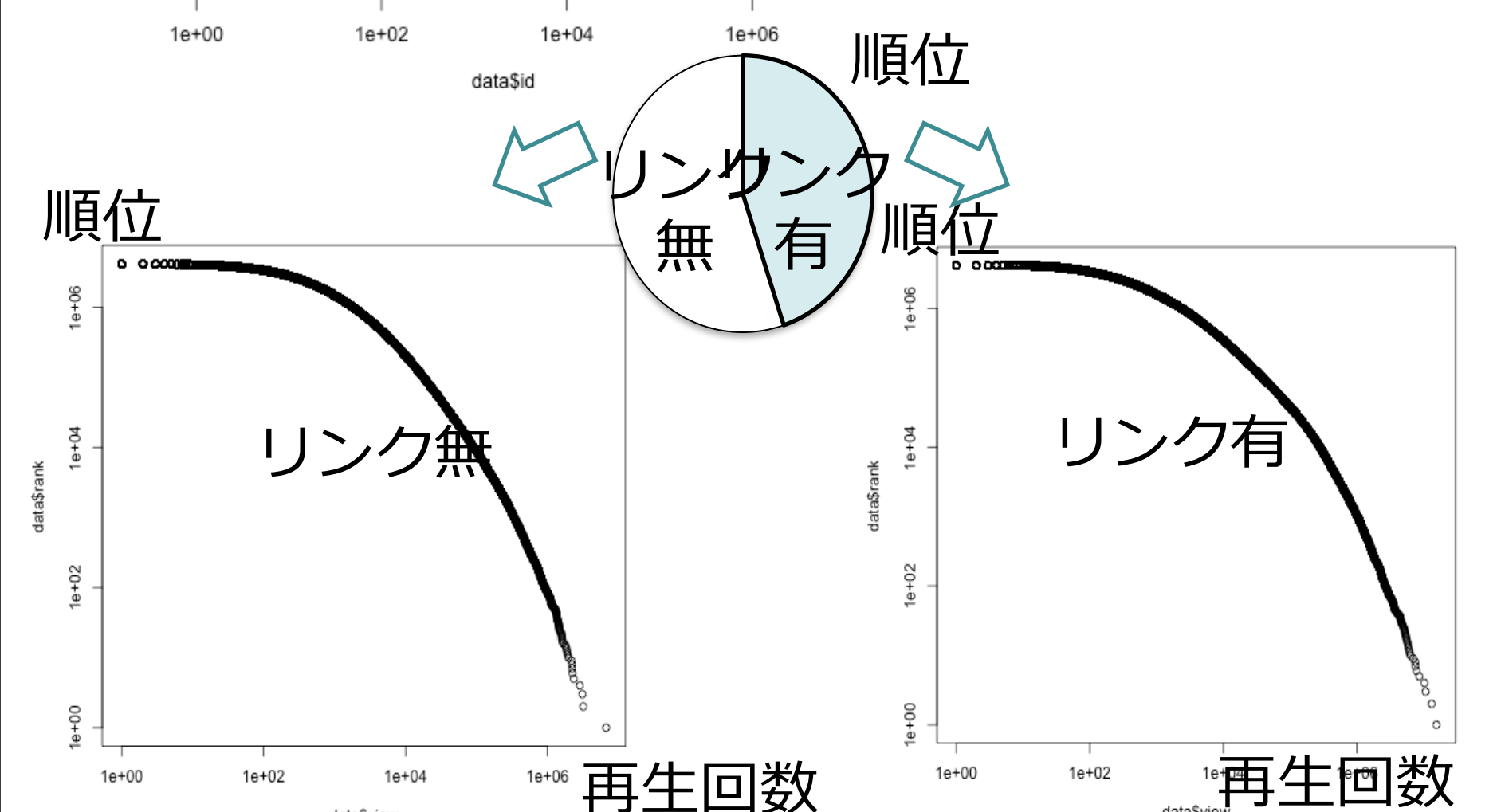
データセットにはリンクの情報がないため、動画説明文から抜き出して出現頻度を数えた。

頻度が多い動画 =多くの動画からリンクされている =再生されやすい?

リンク数



リンク数-順位の分布はべき分布に。リンクされている =特別な状態?



分布の形状は変化なし。単純にリンクをたどっているわけではなさそう。

考察・今後の展望

動画再生数の分布は対数正規分布に従うことから、再生数の高い動画ほど再生されやすいことがわかった。このことから、サイト利用者はランキングをみて動画を見ていることが多いと考えられる。一方、カテゴリで分けると対数正規分布に従うものとそうでないものにわかれる結果が得られた。従わないものは音楽系やゲーム実況動画など、投稿者の人気に左右されやすいものが大きく、投稿者で動画を選んでいると考えられる。最後にリンク数についての分析も行った。リンク数はこれまでと異なりべき分布になったものの、リンクの有無では再生数の分布に変化はなかった。リンク数は再生数に影響を与えないのではないかと考える。

今後は、投稿者による影響の数値的分析や、サムネイル画像など他のデータが再生に与える影響についても調べたい。