

**JANOG29 -Lightning Talk-**  
**10:1と2:1の不思議**  
**(Work in Progress)**

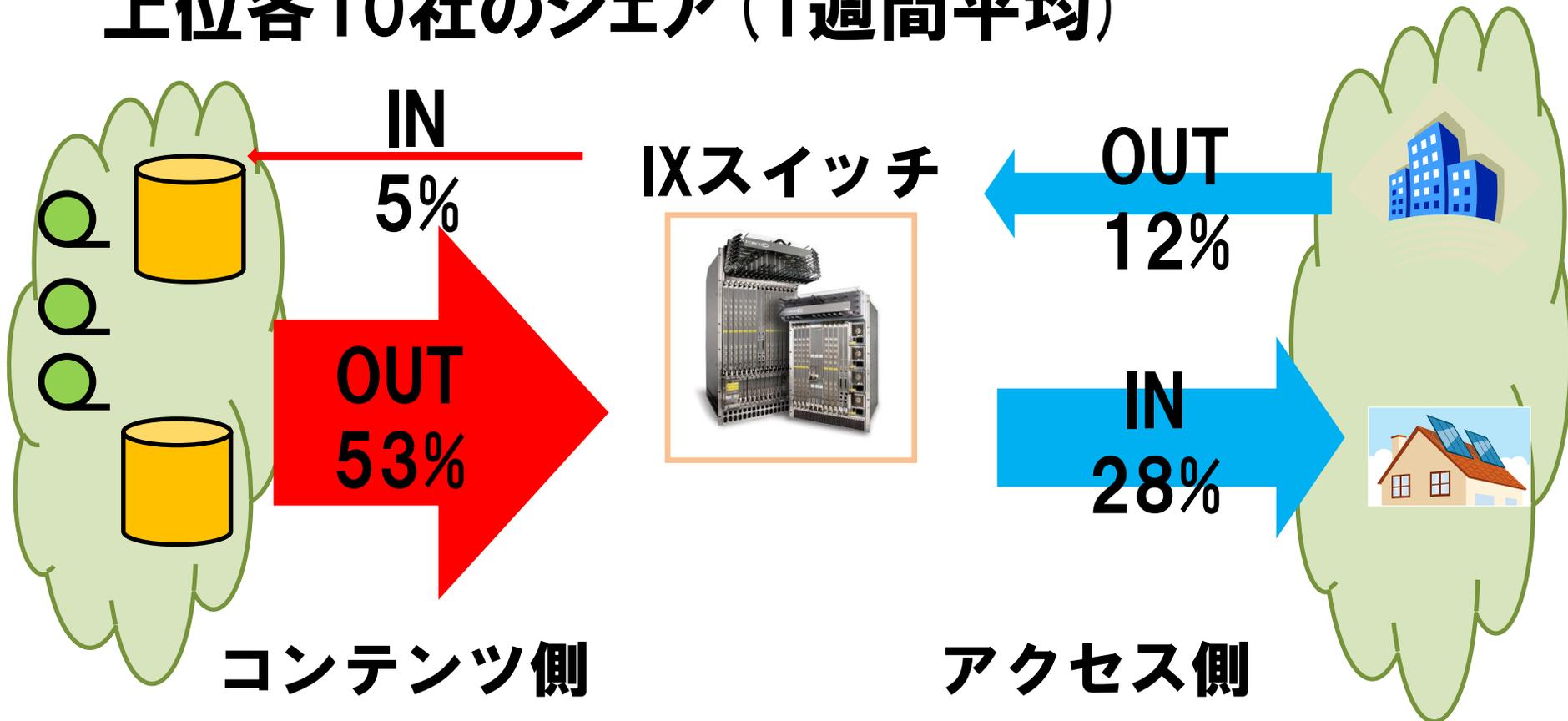
2012/1/20  
日本インターネットエクスチェンジ  
石田慶樹

# インターネット・トラヒック

- **インターネットって何に使ってますか？**
  - 普通のコンシューマは？
  - Janogerは？
- **想像するに**
  - Webとかメールとか
  - 動画を見たり
  - スマホでアプリとか←これってWebだよな
  - 他は？
    - ファイルや動画をアップロードしたり
    - Ustream/ニコ生とかもするかもね

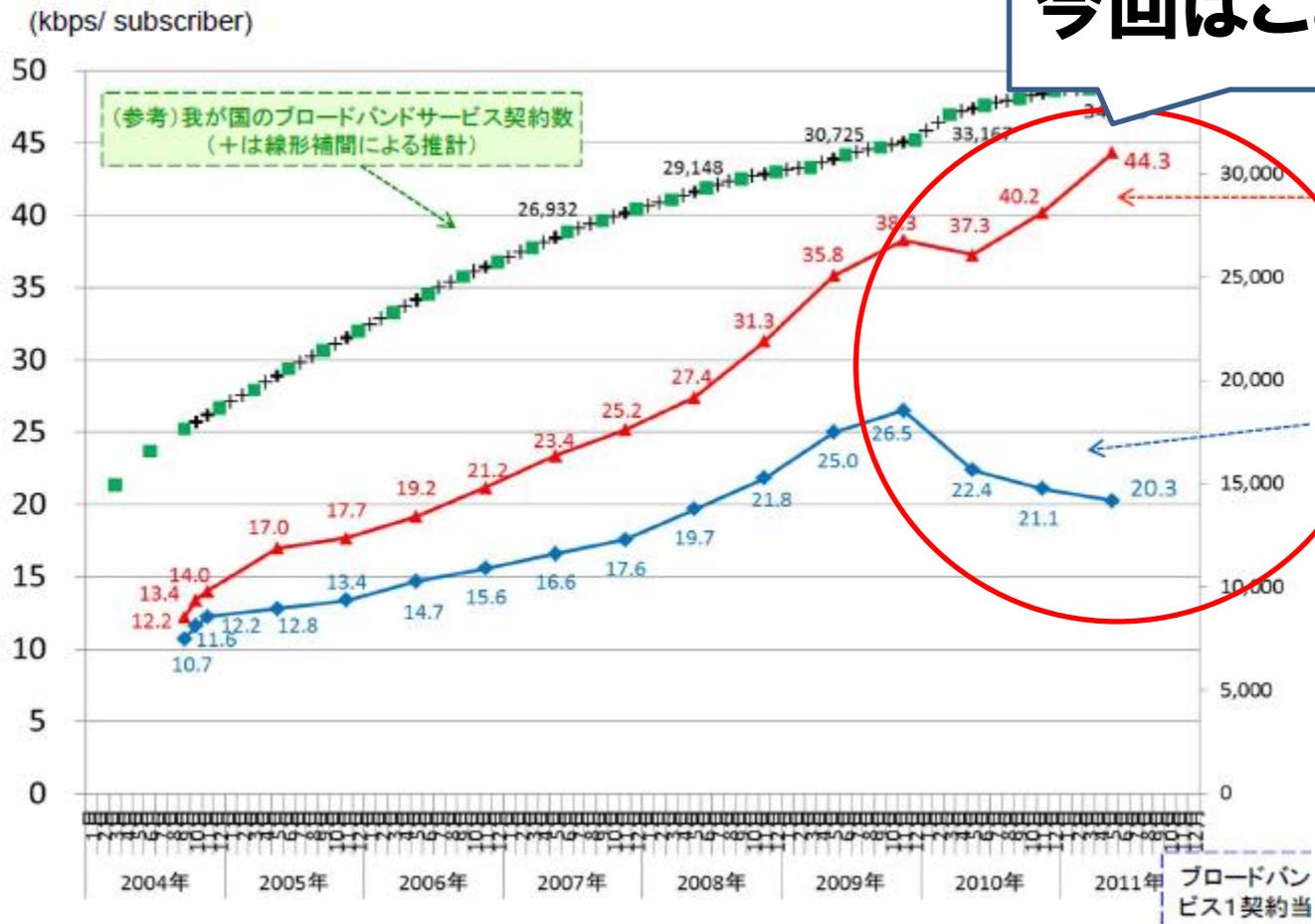
# トラフィック傾向の2分化

- JPIXにおけるアクセス系／コンテンツ系事業者の上位各10社のシェア（1週間平均）



# 総務省のトラフィックデータ

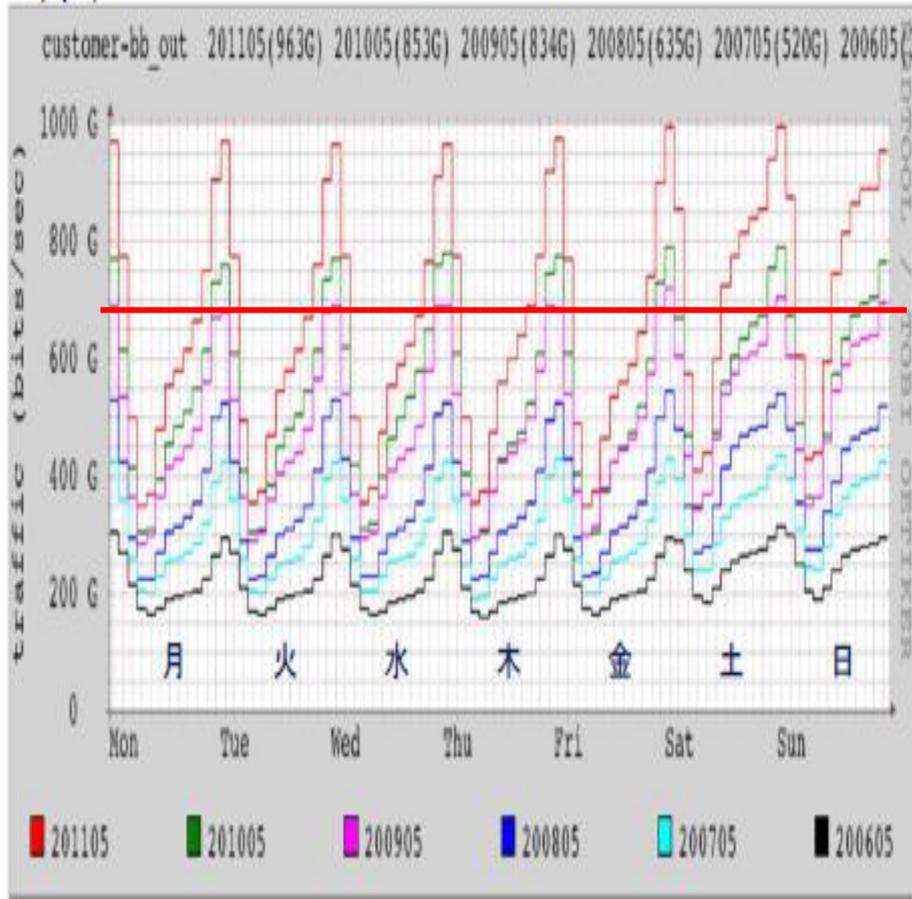
今回はこの話



# 総務省のトラフィックデータ

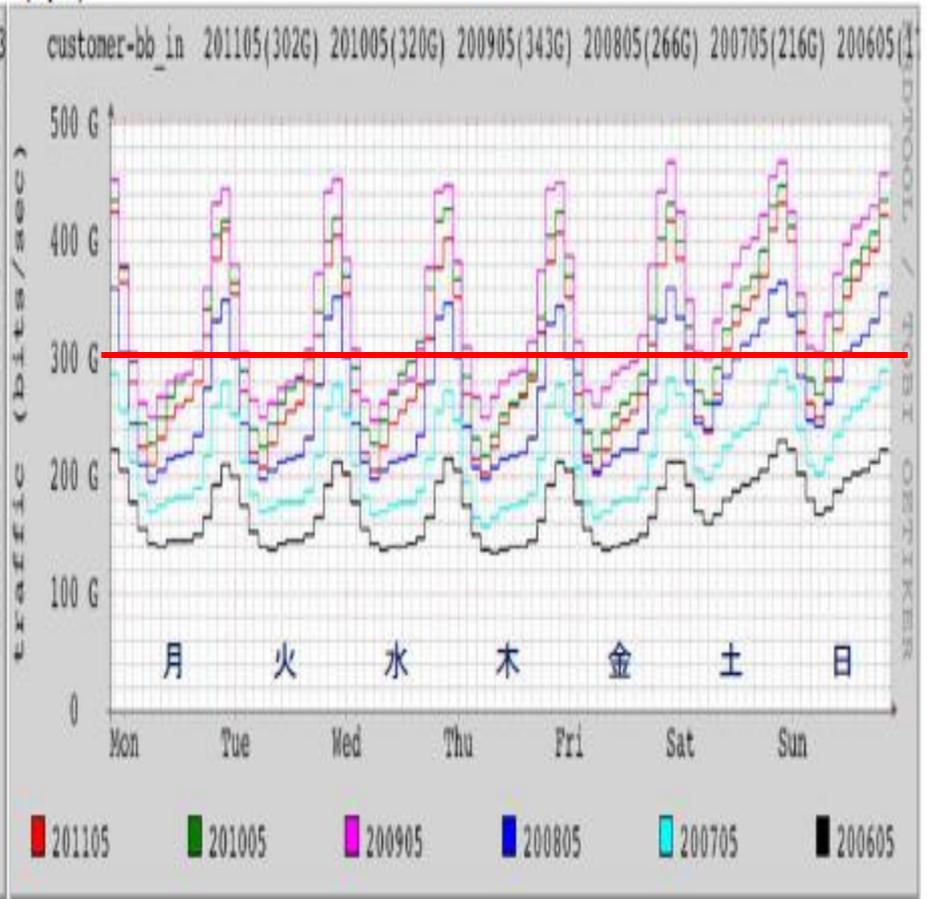
## ダウンロード

(bps)



## アップロード

(bps)



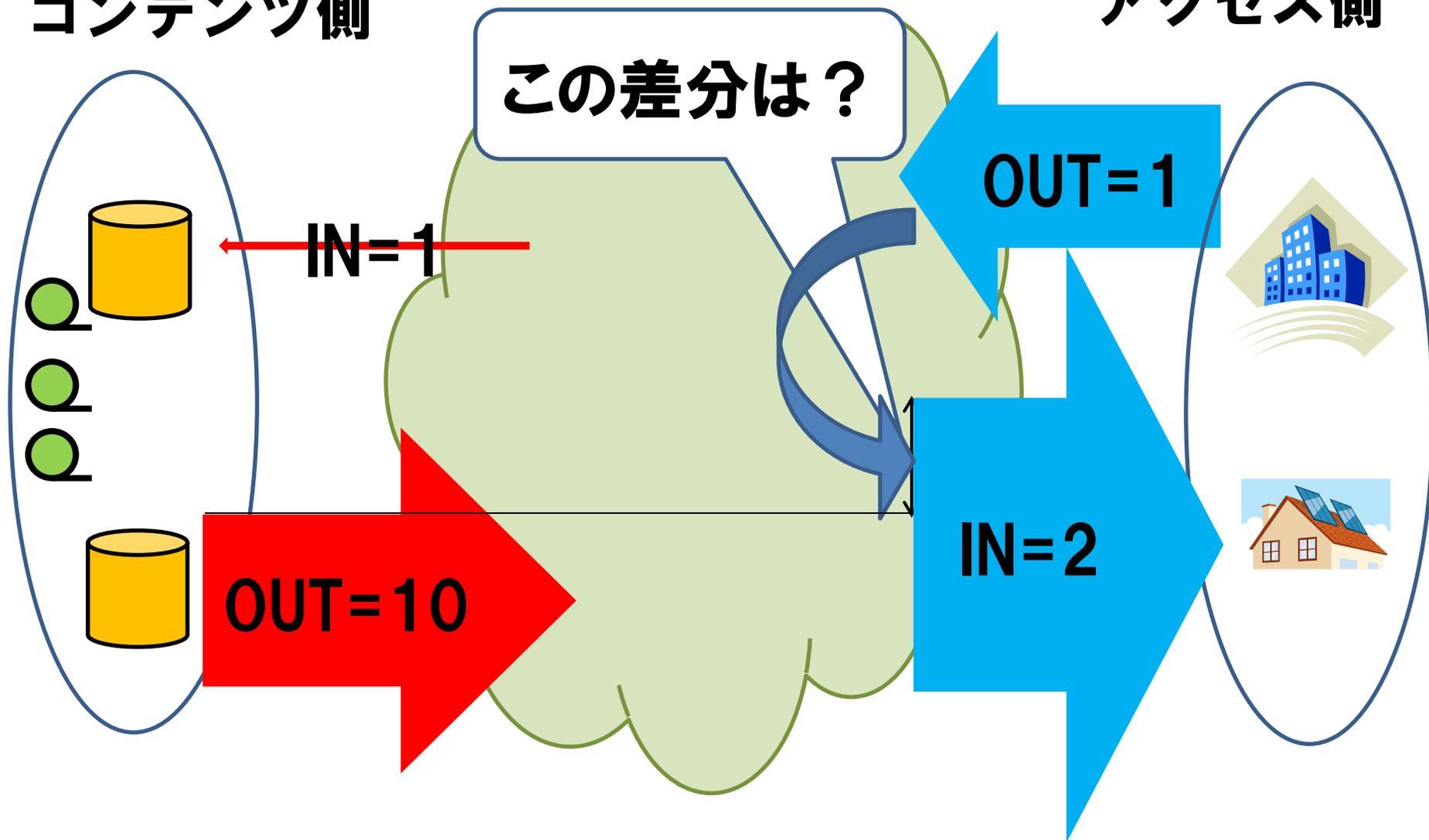
# 総務省の調査から読みとれること

- 一回線（帯域100Mbps）当たり
- ダウンロード側は平均で40kbps（44.3kbps）
  - ピークでは60kbps～80kbps程度
  - 一日あたりにデータ転送量は～0.5G byte（478M byte）
- アップロード側の平均は20Kbps（20.3Kbps）
  - ピークでは26Kbps程度
  - 一日あたりのデータ転送量は～0.2G byte（219M byte）

# ?トラヒックの不思議?

コンテンツ側

アクセス側



# ?トラヒックの不思議?

- コンテンツ事業者側のIN/OUT比率は1:10程度なのに、ブロードバンド回線利用者におけるIN/OUT比率は2:1になっている！
- 利用者の活動がコンテンツ (Web) の閲覧が主だとするとIN/OUT比率が10:1になっても不思議ではないのにOUT側が非常に多い
- 利用者側から出てくるトラヒックって何？

# トラヒックのモデル化の必要性

- **統計データによるトラヒック・モデルの構築**
  - 計測することは目的ではなくて手段
- **トラヒックをサンプリング計測することにより、**
  - 短期・中期・長期のトラヒックの傾向およびその変化を観測
  - 今後のネットワーク・アーキテクチャを考察
  - 定量的データに基づく課金モデルの検討
- そもそもサンプリング調査を行うにしても、どの程度のサンプル数があればユーザ動向が把握できるのかかもよくわかっていない

# トラヒックに関する仮説

- **トラヒックは「対数正規分布」しているのではないか**
- **「対数正規分布」とは**
  - **「確率変数の対数値が正規分布をするような統計分布」**
  - **「この対数正規分布の特色は、規模の大きい個体は数が少なく、規模の小さい例は数が多いということであり、また、それにもかかわらず大きい方からある程度の数の個体の規模を合計すると全体合計規模の大部分のシェアを占めるということである。」**  
<http://www.weblio.jp/content/%E5%AF%BE%E6%95%B0%E6%AD%A3%E8%A6%8F%E5%88%86%E5%B8%83> より)
- **(例) 恒星の密度、鋳床の規模、市町村の人口、年間所得の分布、JANOGの県別参加者数**

# 対数正規分布

- 対数正規分布

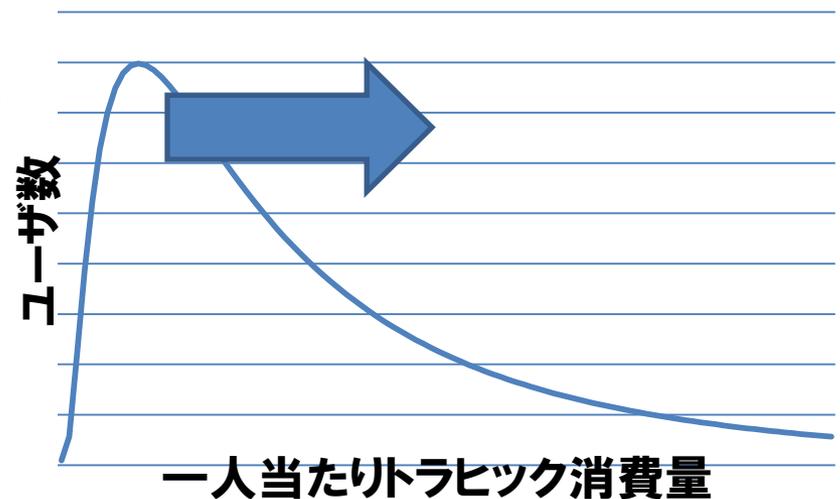
母数は  $\mu$  と  $\sigma$

確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi\sigma x}} e^{-\frac{(\ln t - \mu)^2}{2\sigma^2}}$$

平均  $E(X) = e^{\mu + \frac{\sigma^2}{2}}$

分散  $V(X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$



# トラヒックの分布

- **ブロードバンドサービス契約者のトラヒック**
  - **Download側は対数正規分布に従うだろう**
    - 平均値は44.1Kbps
    - 分散 (標準偏差) は不明
    - 長期的なパラメータの変化を把握したい
  - **Upload側は不思議**
    - 平均値は20.8Kbps
    - Downloadの1/10 (4Kbpsあたりに山があって  
然るべき

# トラヒックの分布 (cont)

- Upload側のトラヒック分布について
  - 仮説1:対数正規分布で分散が非常に大きい
  - 仮説2:大半は対数正規分布に従うがごく一部が大量にトラヒックを吐き出している

# まとめ

- **緩募:トラヒックの状況について一緒に動いてもらえる方**
  - サンプルングによる計測が可能な人
  - 数学、特に統計学に強い人
- **目的**
  - トラヒックの短期・中期・長期の動向の把握
  - データの流れに基づくインターネットアーキテクチャについて検討
  - 課金モデルの定量的検証
- **まずは自ネットワークのサンプルングによる観測からはじめてみませんか？**