近似サンプリングの精度保証

2009/01/22
[SCIS 2009 3D1-1]

沼山晃/**草川恵太**/河内亮周/田中圭介（東京工業大学）

暗号と非一様分布

	分布
[Reg05]	ガウス分布
[GPV08]	離散ガウス分布
[HS08]	ポワソン分布
[NIT08]	二項分布

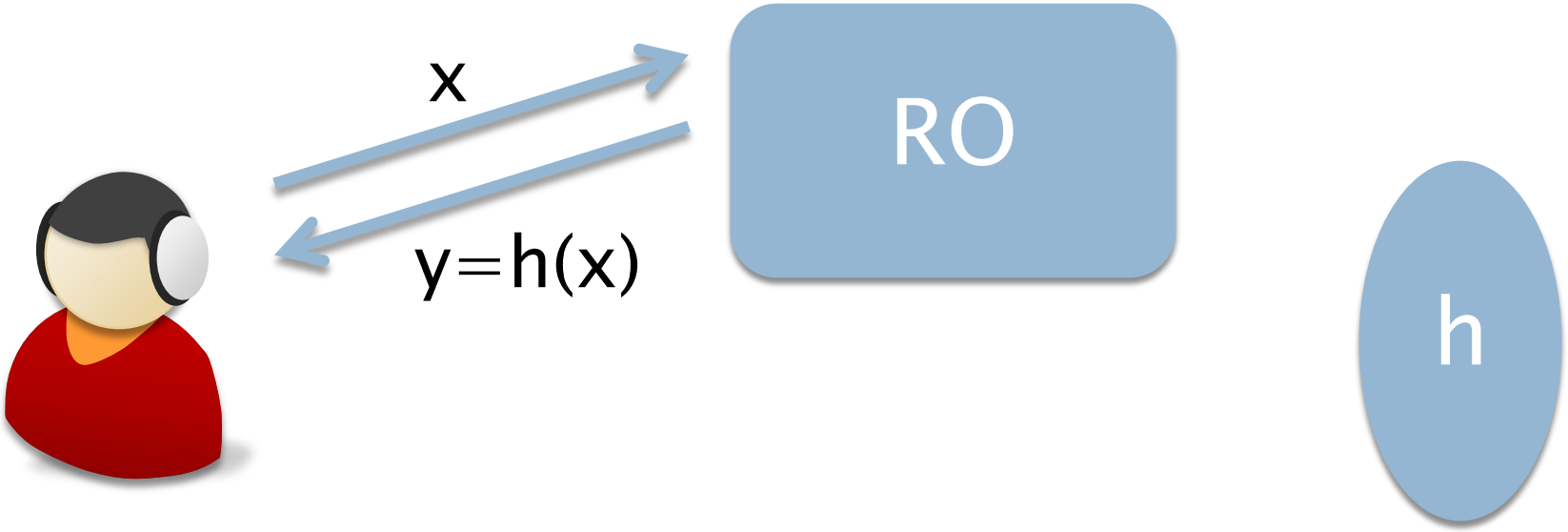
暗号と非一様分布

	分布	アルゴリズム
[Reg05]	ガウス分布	“標準的手法”
[GPV08]	離散ガウス分布	提案あり
[HS08]	ポワソン分布	言及なし
[NIT08]	二項分布	仮定

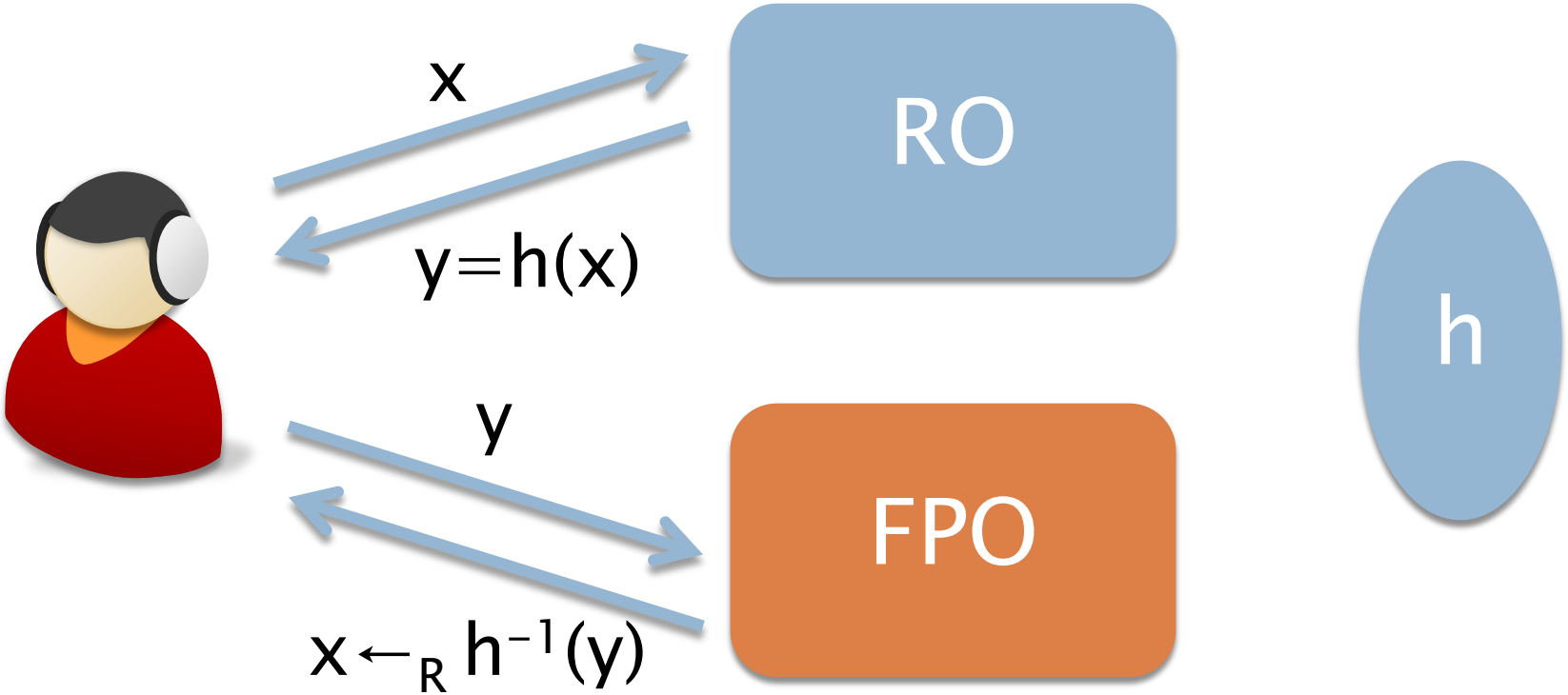
暗号と非一様分布

	分布	アルゴリズム
[Reg05]	ガウス分布	“標準的手法”
[GPV08]	離散ガウス分布	提案あり
[HS08]	ポワソン分布	言及なし
[NIT08]	二項分布	仮定

ROと二項分布 [NIT08]



ROと二項分布 [NIT08]

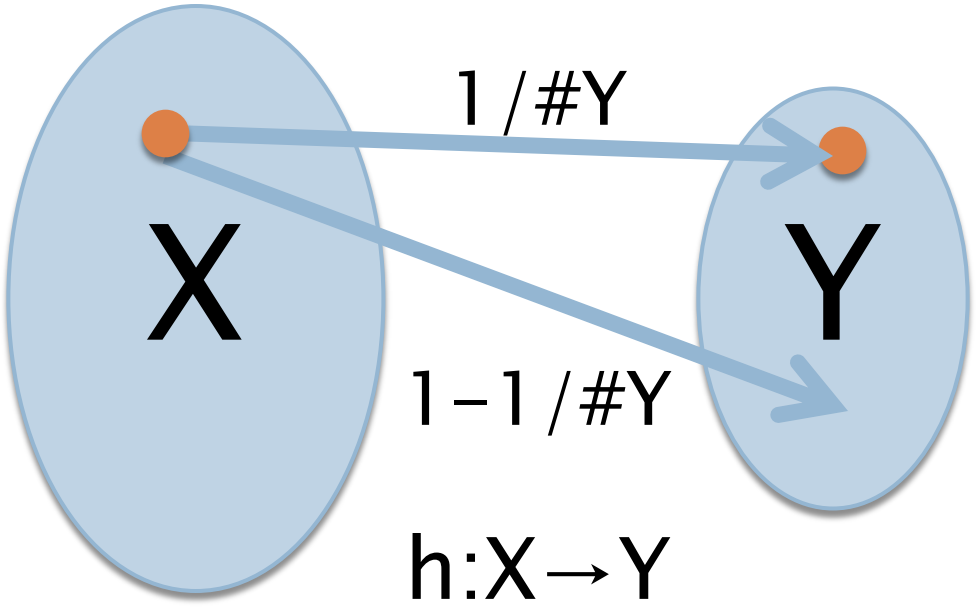


証明技法

- RO+FPOをシミュレートする
  - ROでのリスト管理とは違う方法が必要
  - 原像の数を管理する

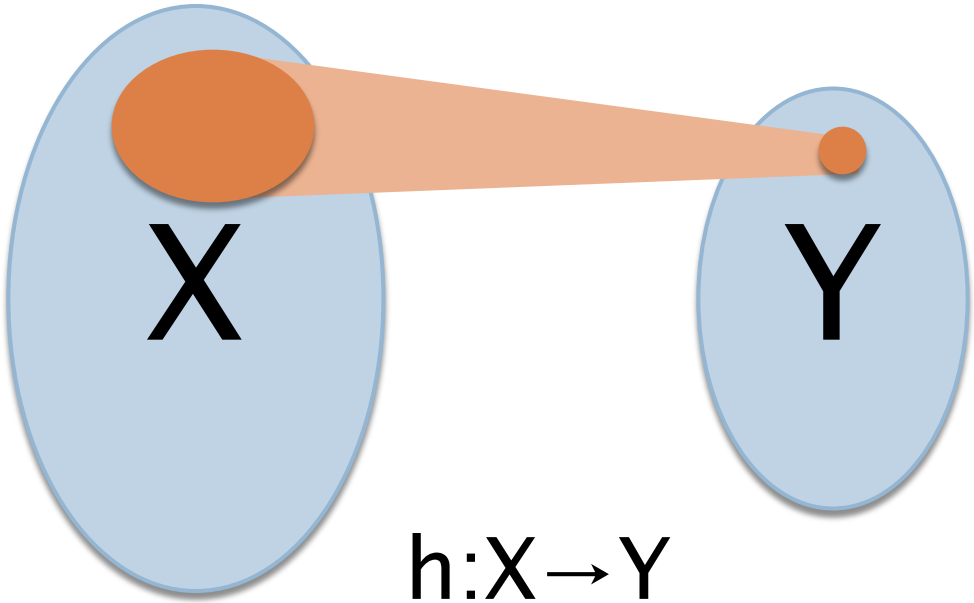
証明技法

- RO+FPOをシミュレートする
  - ROでのリスト管理とは違う方法が必要
  - 原像の数を管理する



証明技法

二項分布Bn(#X,1 /#Y)に従う



宿題

二項分布からサンプリング⇒
ROとFPOをシミュレート可能

出来るの?

仮定にします

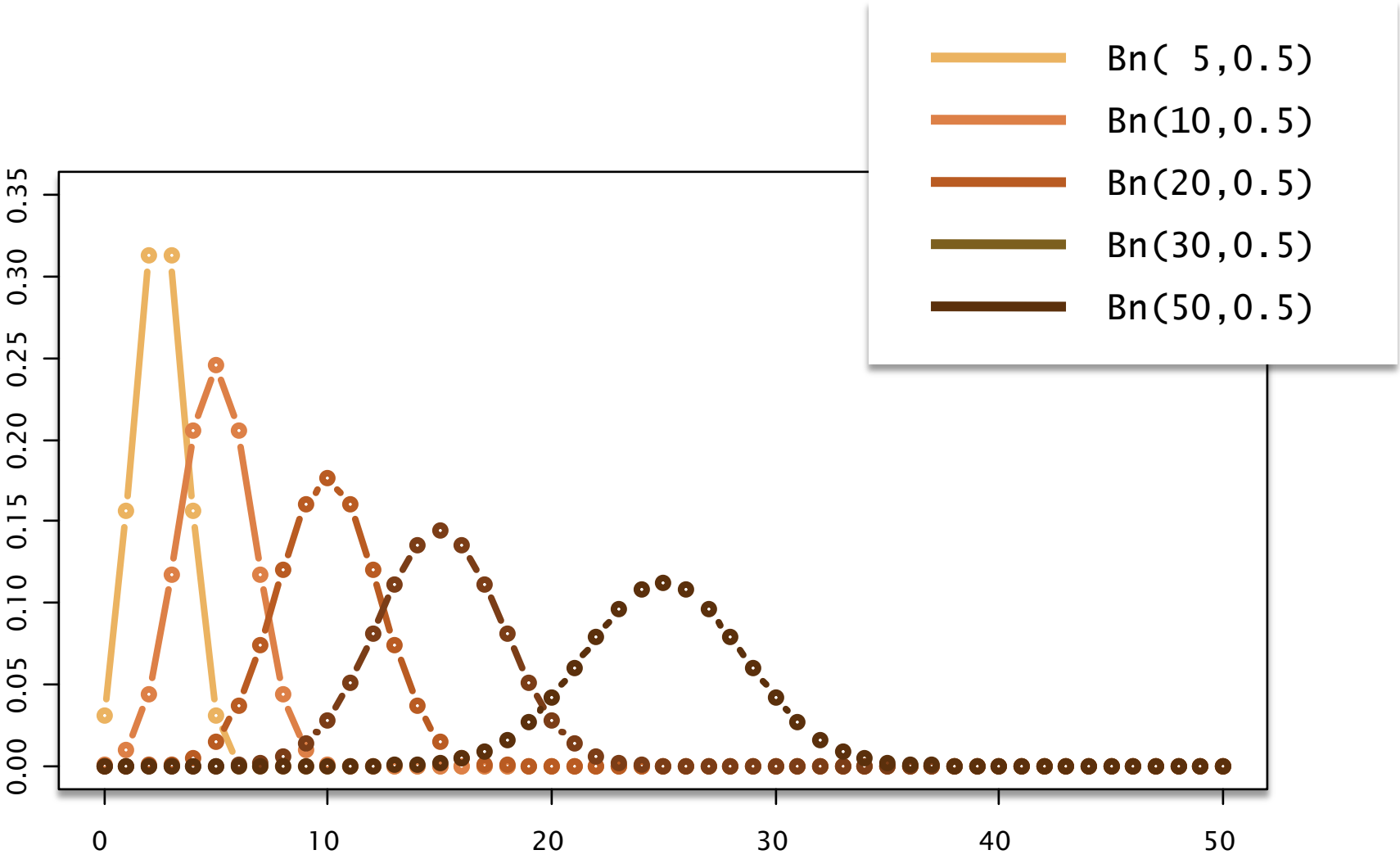
前回の宿題

- 二項分布Bn(K,p)からサンプリング
  - 例: $K=2^{1024}$, $p=2^{-128}$
    - 平均: $Kp=2^{896}$, 分散: $Kp(1-p)=$ほぼ2^{896}
- poly(log K, log p)時間で
  サンプリングできるのか?
- 近似でも可

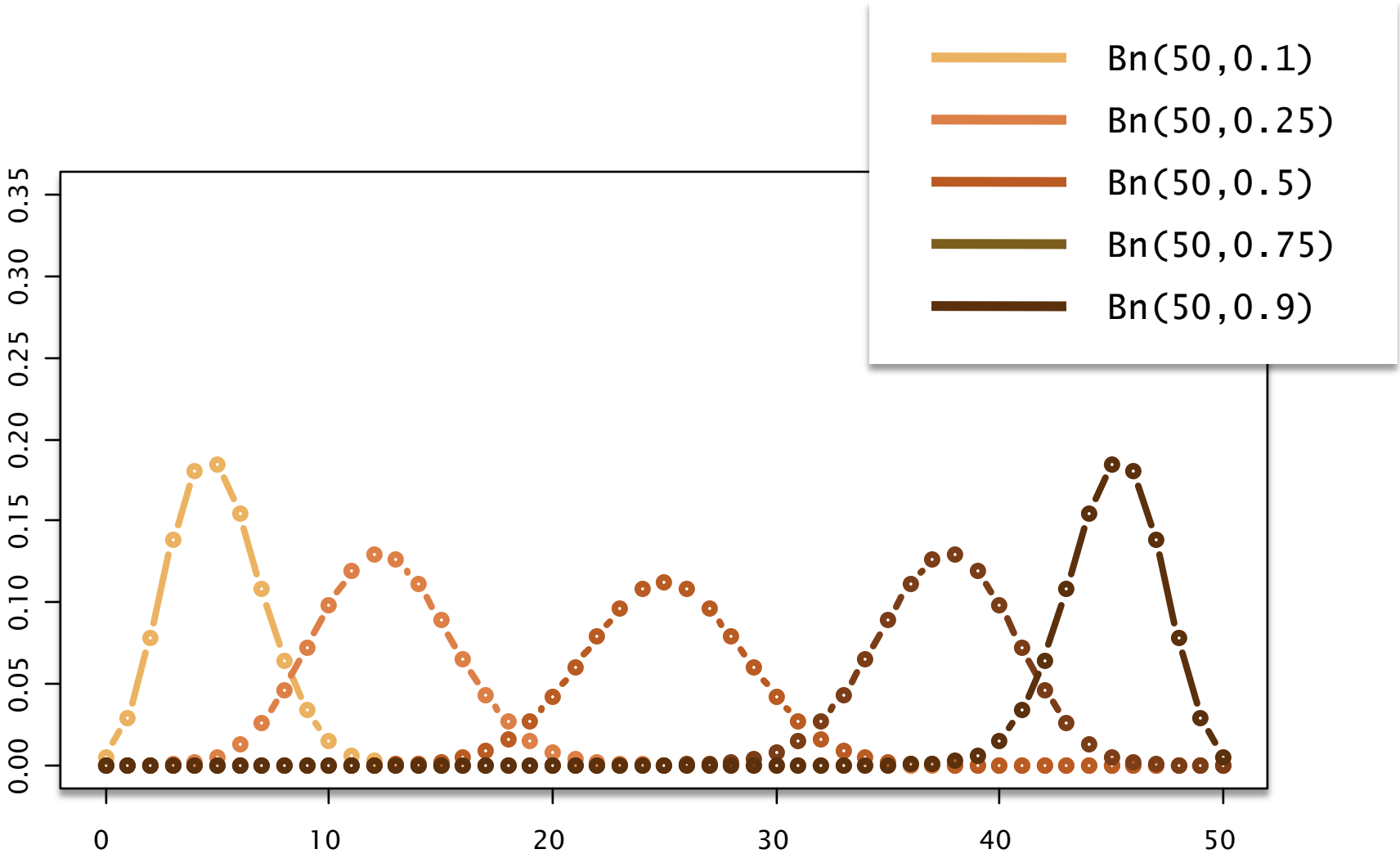
問題の確認 – 二項分布

- Bn(K,p)
  - 二項分布
  - (p,1−p)のコインをK回振ったときの表の数
  - 確率関数: $f(x)={}_{K}C_{x}p^{x}(1-p)^{K-x}$
  - 平均: Kp
  - 分散: Kp(1−p)

二項分布 Kを変化


Bn(5,0.5)
Bn(10,0.5)
Bn(20,0.5)
Bn(30,0.5)
Bn(50,0.5)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
10
20
30
40
50

二項分布 pを変化


Bn(50,0.1)
Bn(50,0.25)
Bn(50,0.5)
Bn(50,0.75)
Bn(50,0.9)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
10
20
30
40
50

問題を正確に

- 以下を満たすアルゴリズムを構成する

- 入力
  - $K = O(2^n)$, $0 \leqq p \leqq 1$ (poly(n) bit)
- 出力
  - $X \in \{0,1,\ldots,K\}$
- 仕様
  - Δ(Xの分布, $Bn(K,p)) \leqq 2^{-n}$
  - poly(n)時間で動作

既存研究

- Knuth
  - “The Art of Computer Programming, Vol. 2”
- “Handbook of Simulation”
- Devroye
  - “Non-Uniform Random Variate Generation”

既存研究

- 統計/シミュレーション分野
  - 実数計算可能
  - [0,1)上の一様乱数を生成可能
- アルゴリズム/暗号分野
  - ビット計算可能
  - {0,1}上の一様乱数を生成可能

問題再掲

- 以下を満たすアルゴリズムを構成する

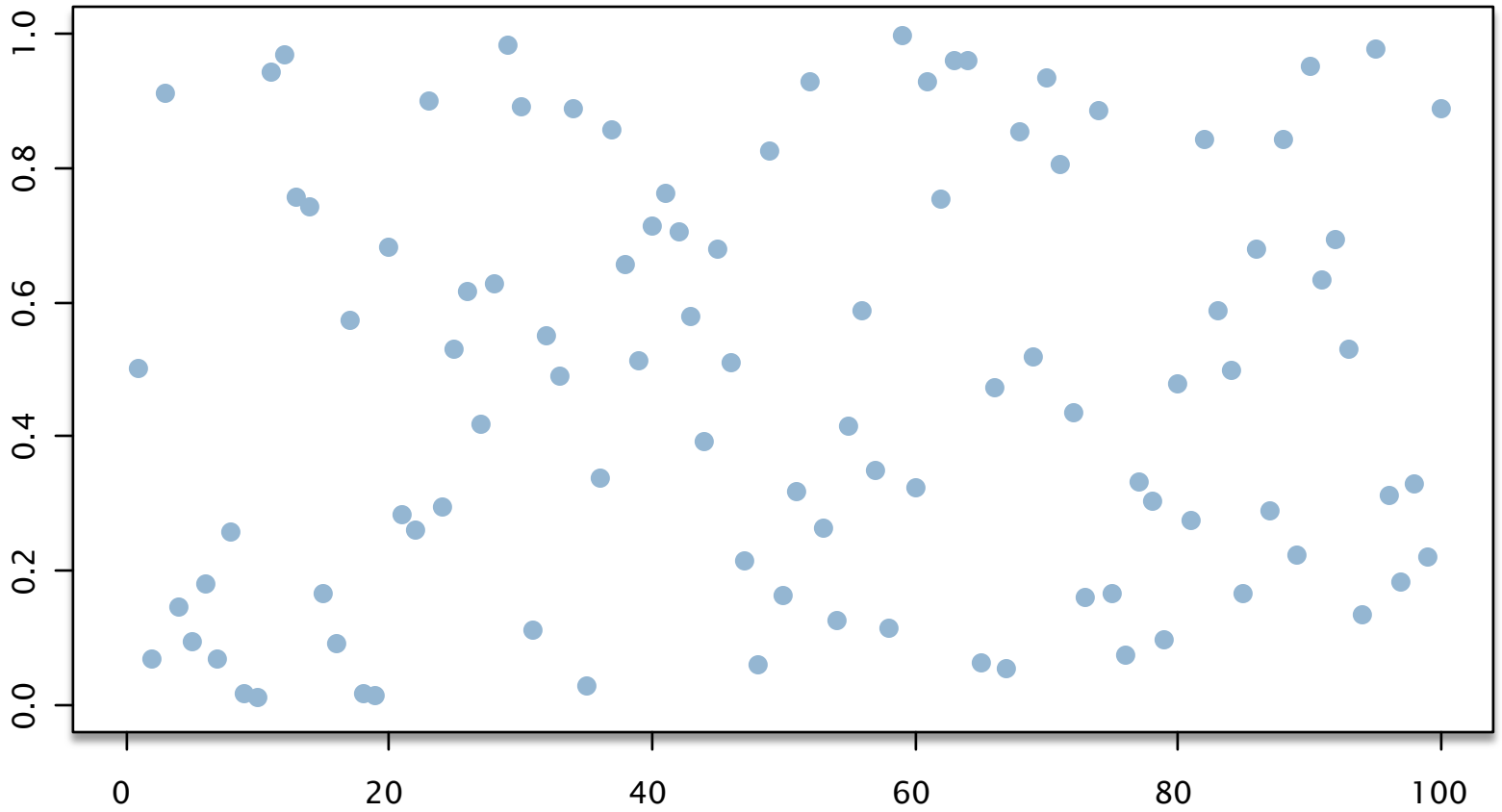
- 入力
  - $K = O(2^n)$, $0 \leqq p \leqq 1$ (poly(n) bit)
- 出力
  - $X \in \{0,1,\ldots,K\}$
- 仕様
  - Δ(Xの分布, $Bn(K,p)) \leqq 2^{-n}$
  - poly(n)時間で動作

方法案1 – ベルヌーイ試行

1. K回一様乱数を得てp以下のものを数える

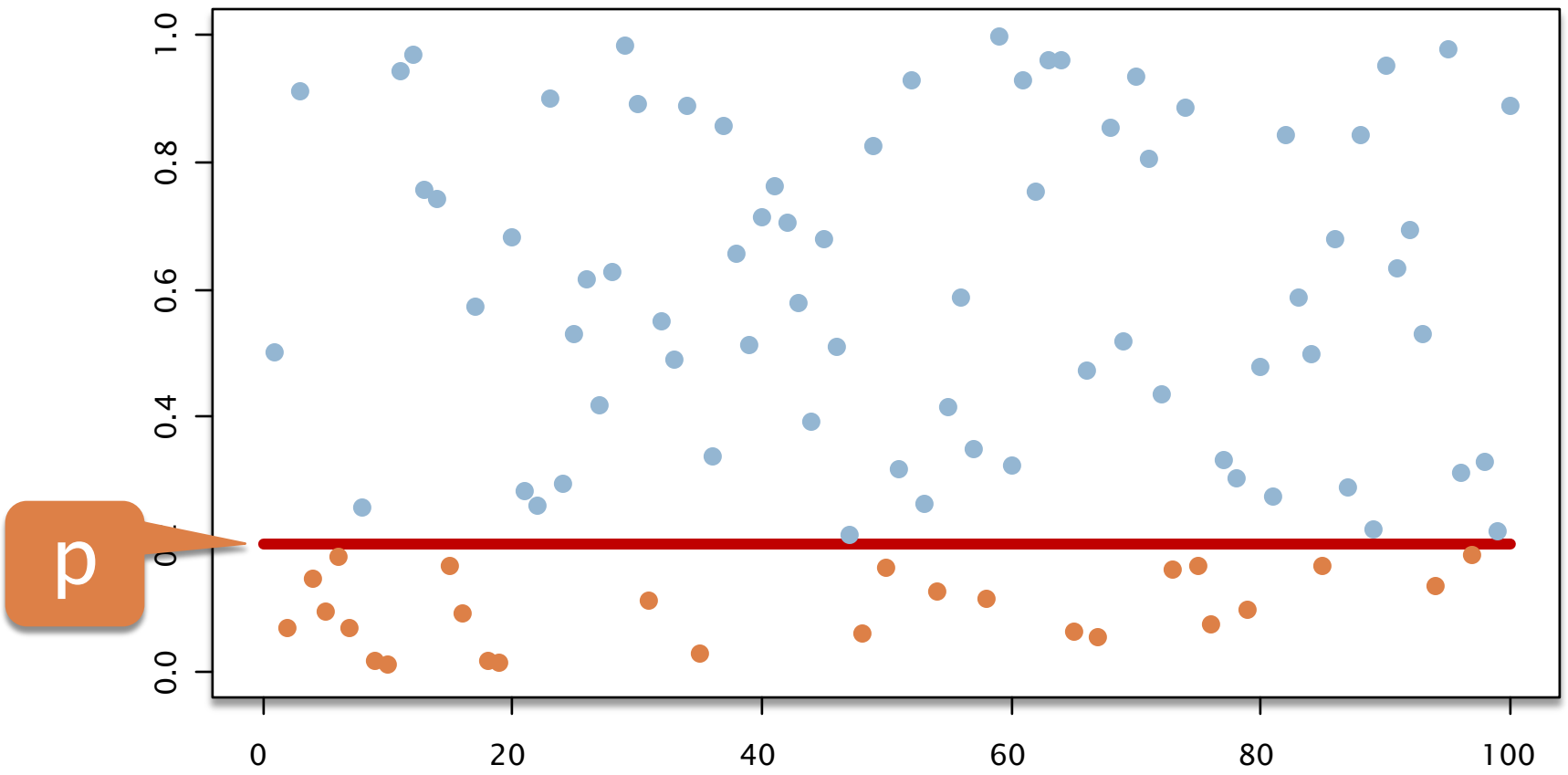
方法案1 – ベルヌーイ試行

1. K回一様乱数を得てp以下のものを数える



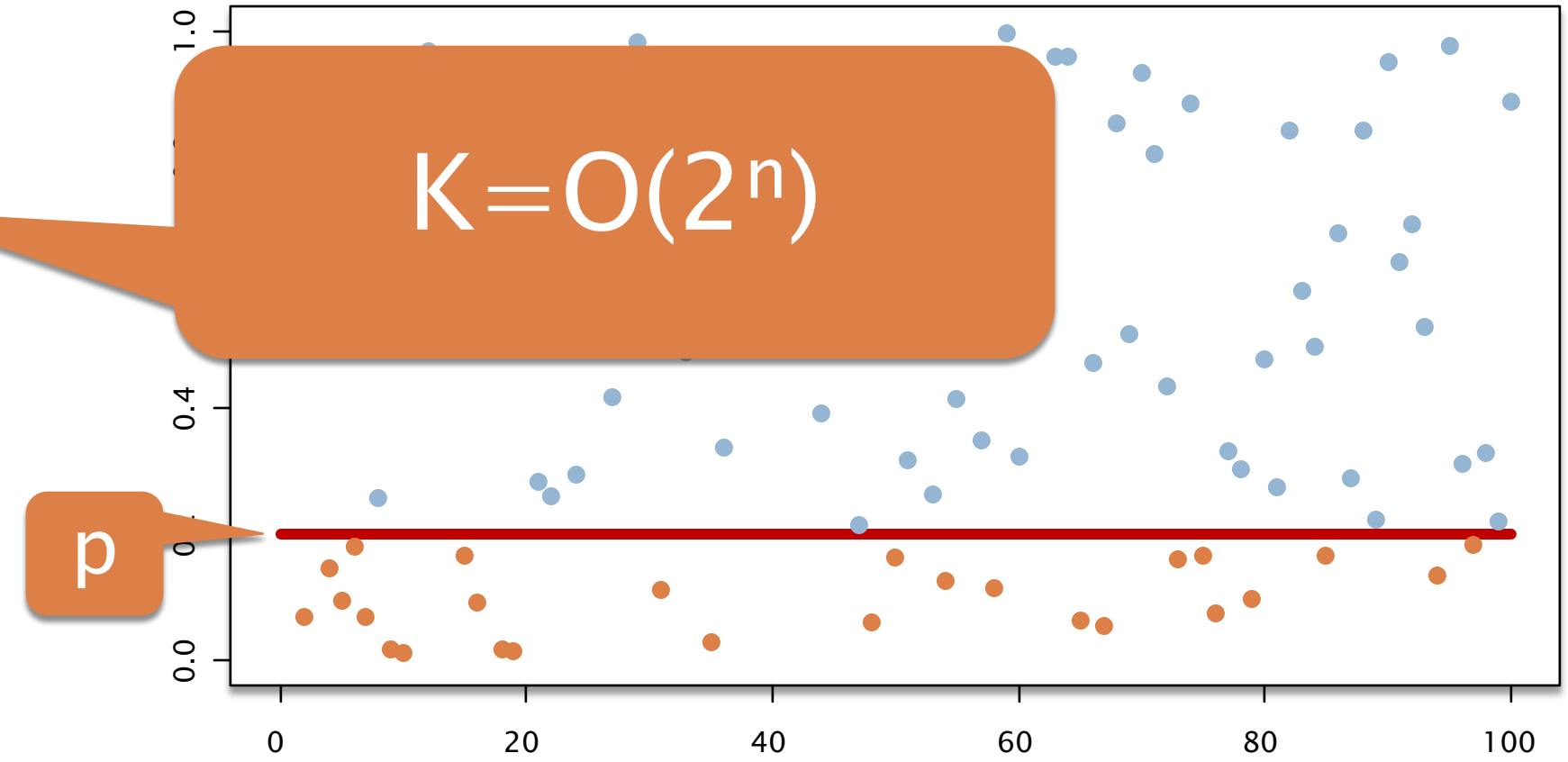
方法案1 – ベルヌーイ試行

1. K回一様乱数を得てp以下のものを数える



方法案1 – ベルヌーイ試行

1. K回一様乱数を得てp以下のものを数える



方法案2 – 中心極限定理

- 中心極限定理
  - K→∞で, Bn(K,p)とN(Kp,Kp(1−p))が一致

Kが有限だと?

- 中心極限定理の証明をよく読むとKが有限の時に統計的距離を抑えられない

ここまでのまとめ

- ベルヌーイ試行
  - K回コインを投げるのでダメ
- 正規分布で近似 (中心極限定理)
  - 離散と離散の場合でも統計的距離は怪しい
  - 離散版の正規分布を取るには?

我々の結果

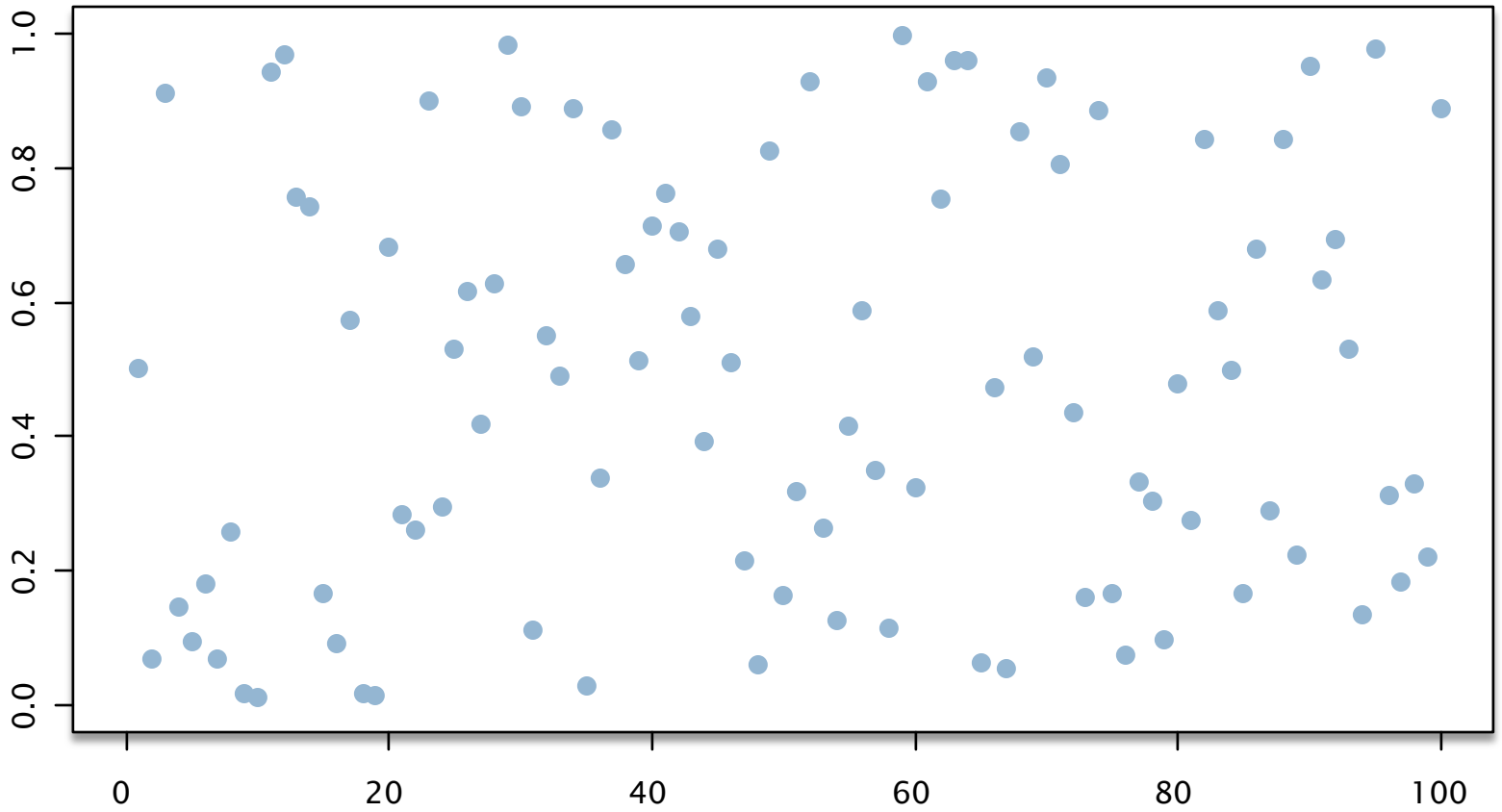
- アルゴリズム BB
  - 入力: $K=O(2^n)$, p
  - 出力: X
  - 仕様
    - 統計的距離: $\Delta(X\text{の分布}, Bn(K,p)) \leqq 2^{-n}$
    - 動作時間: poly(n)
  - 統計/シミュレーション分野のアルゴリズムを離散化して解析

Relles ’72

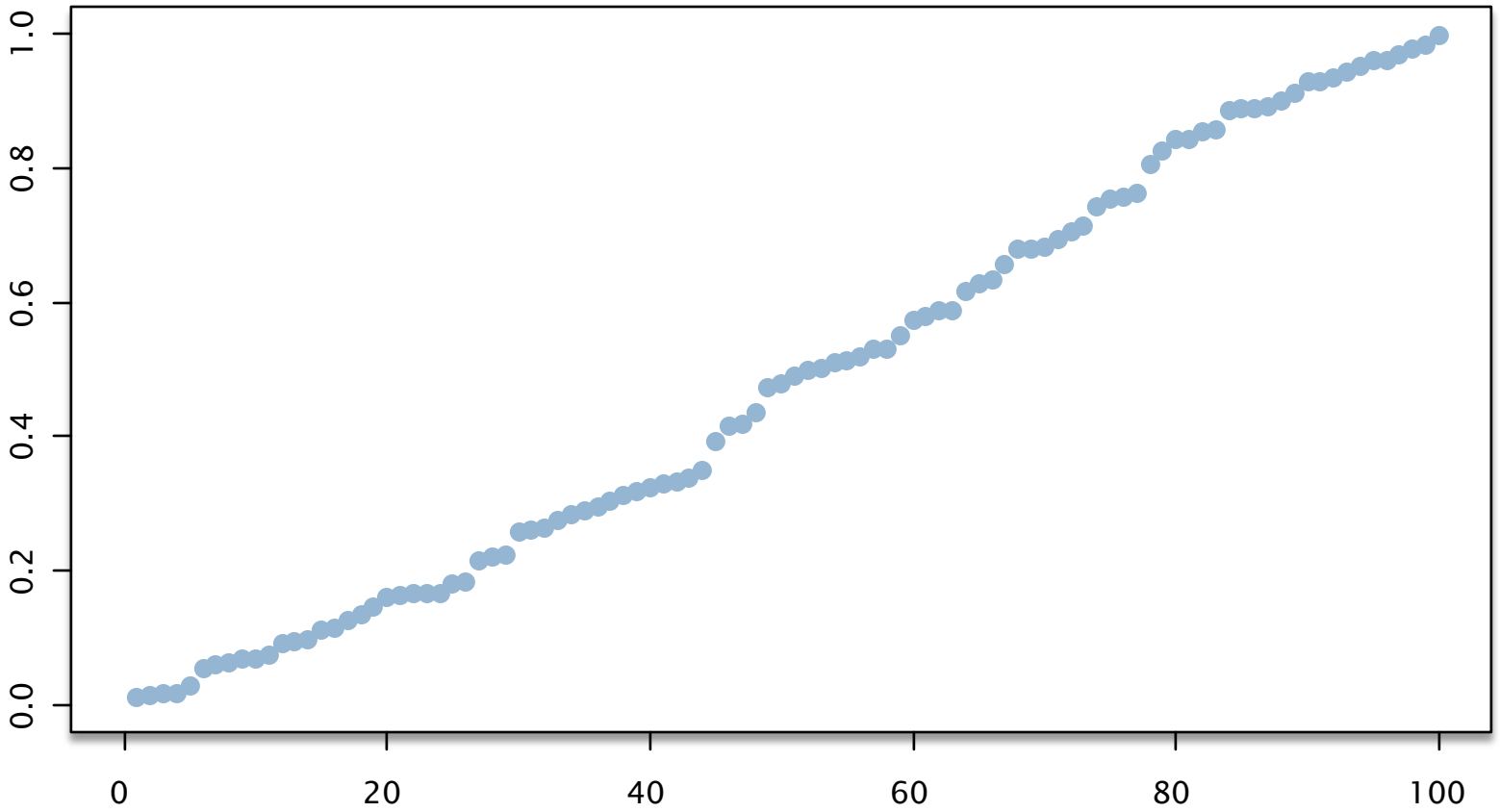
- 二分探索
- K=2a−1とする
- 一度にa枚のコインの表裏を確定したい

Relles – アイデア

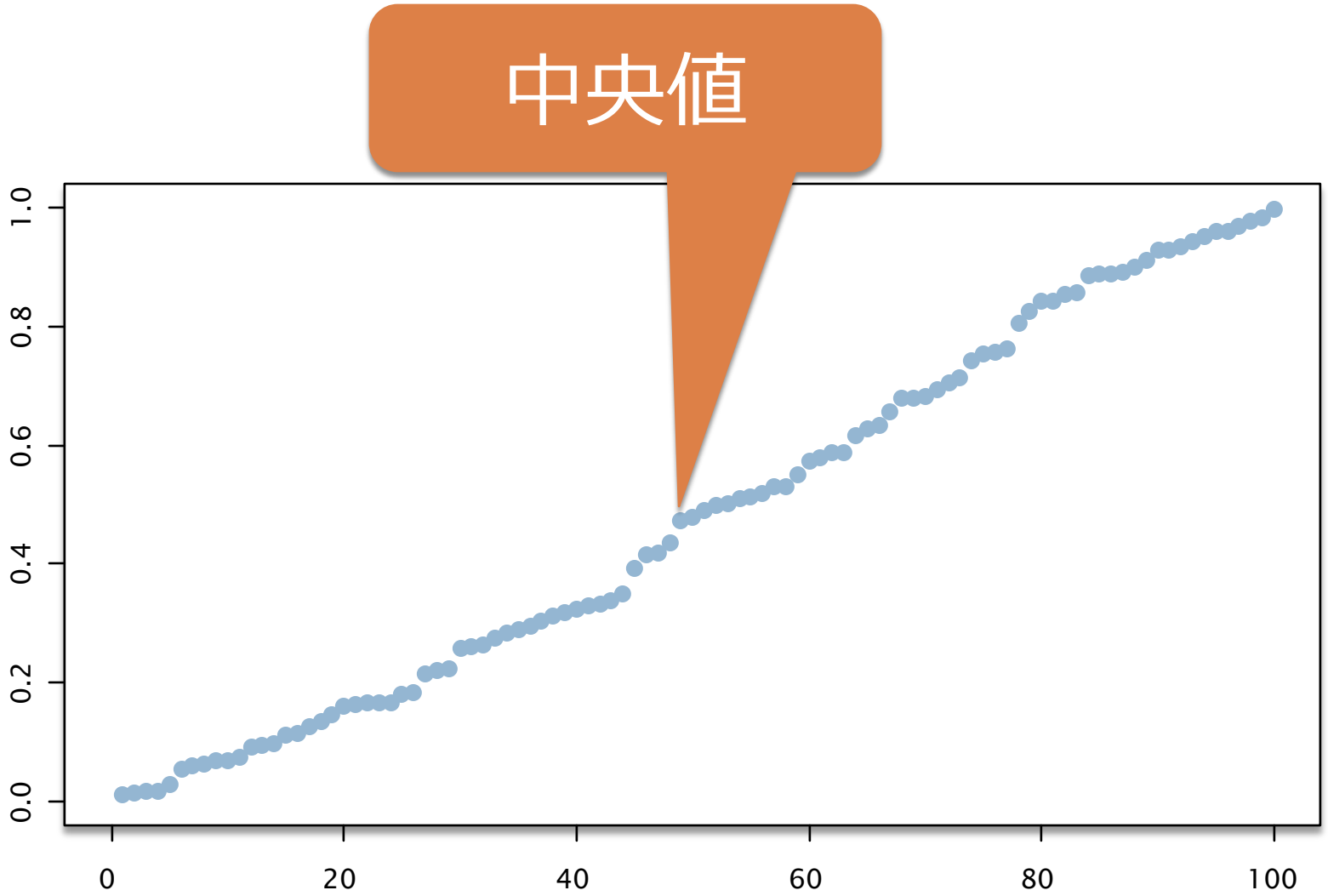
- K個の一様乱数を考える (実際には振らない)



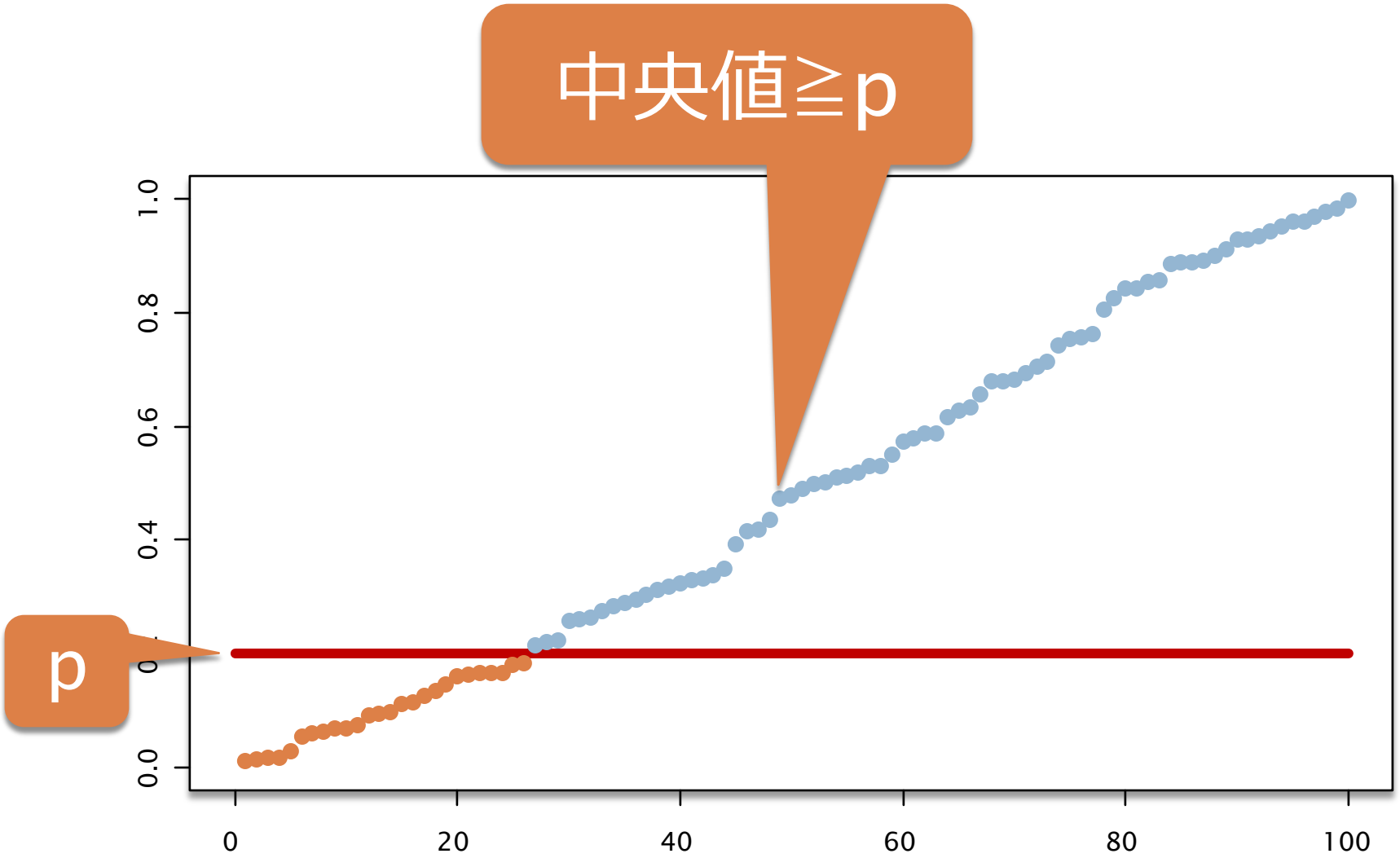
Relles – アイデア



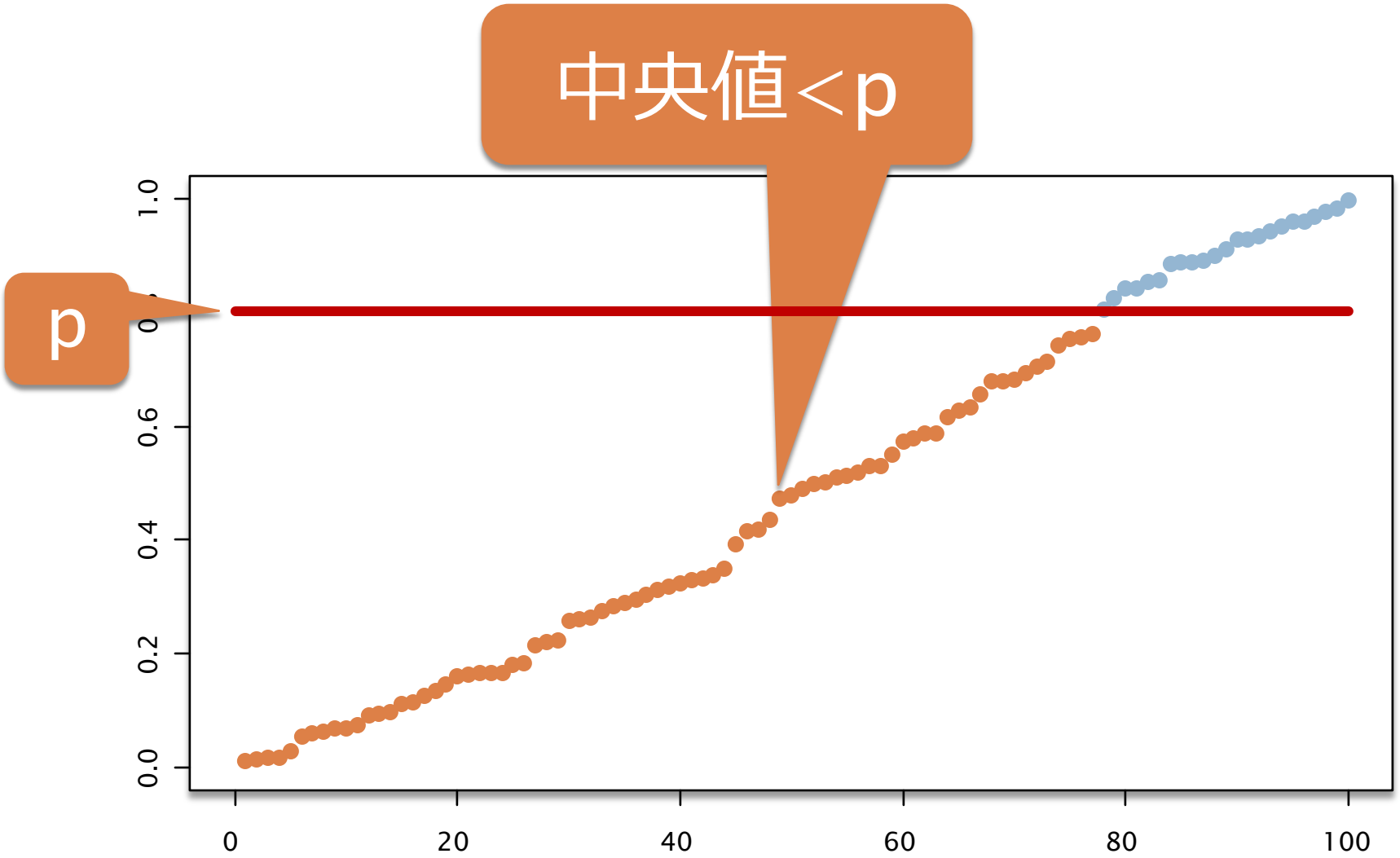
Relles – アイデア



Relles – アイデア

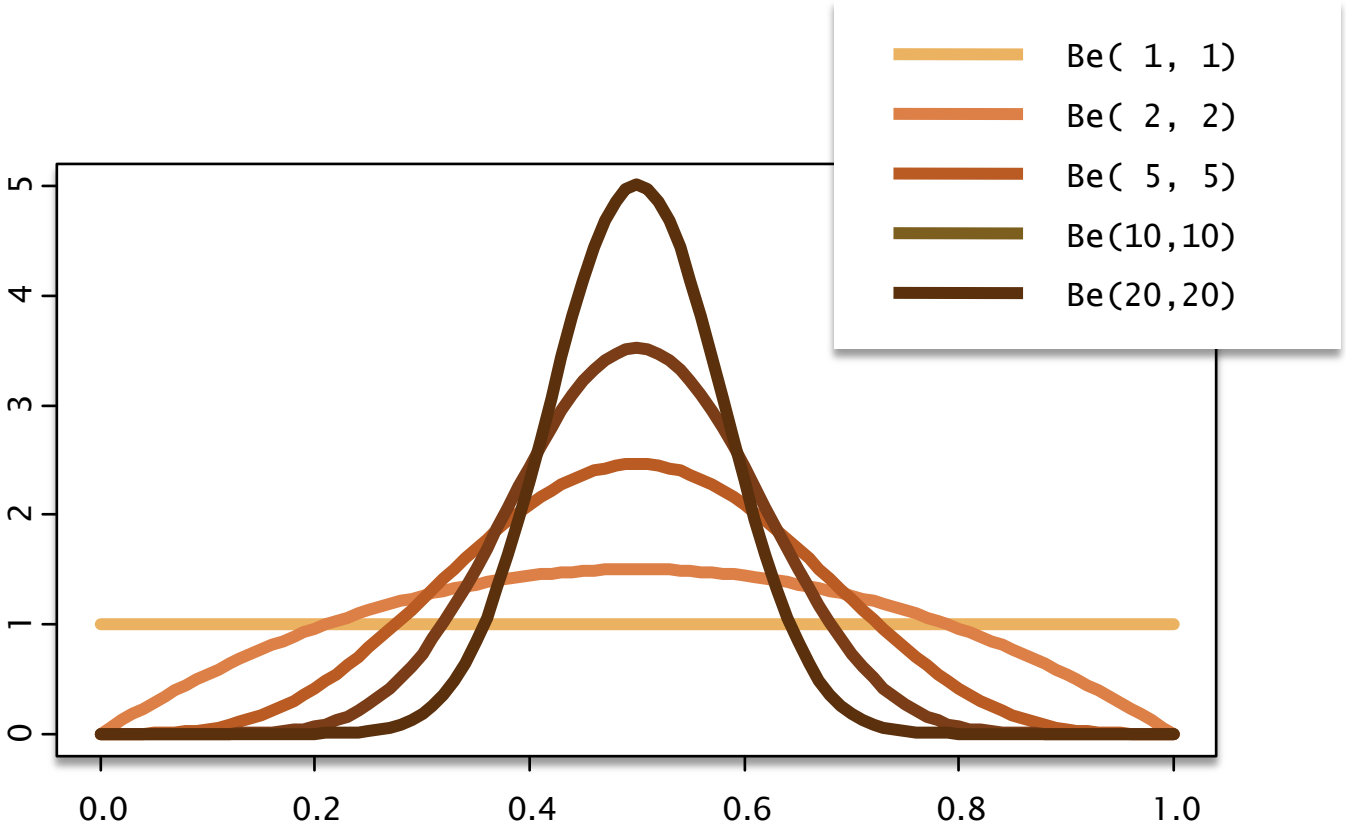


Relles – アイデア



Relles – アイデア

- K個のU(0,1)からのサンプルの中央値
- 中央値はベータ分布 Be(a,a) に従う



Relles ’72

- アルゴリズム
  1. K＝2a−1個
  2. 中央値sをベータ分布Be(a,a)から取る
  3. s≦pならa個は表が出ているとカウント
     a) 残りa−1個を決める
  4. 逆ならa個は裏が出ているとカウント
     a) 残りa−1個を決める

ベータ分布のサンプリング

- ベータ分布←ガンマ分布 [textbook]
- ガンマ分布←コーシー分布 [Ahrens-Dieter '74]
- コーシー分布←一様分布 [textbook]

- 各アルゴリズムを全て解析
- もともと連続なものでしか考えてないので全て離散化

解析の例

- ベータ分布からのサンプリング
- よく知られている定理
  - $x \sim Ga(a),\ y \sim Ga(b) \Rightarrow x/(x+y) \sim Be(a,b)$

解析の例

Be*(a,a)

1. x,y←Ga(a)
2. output R'(x/(x+y))

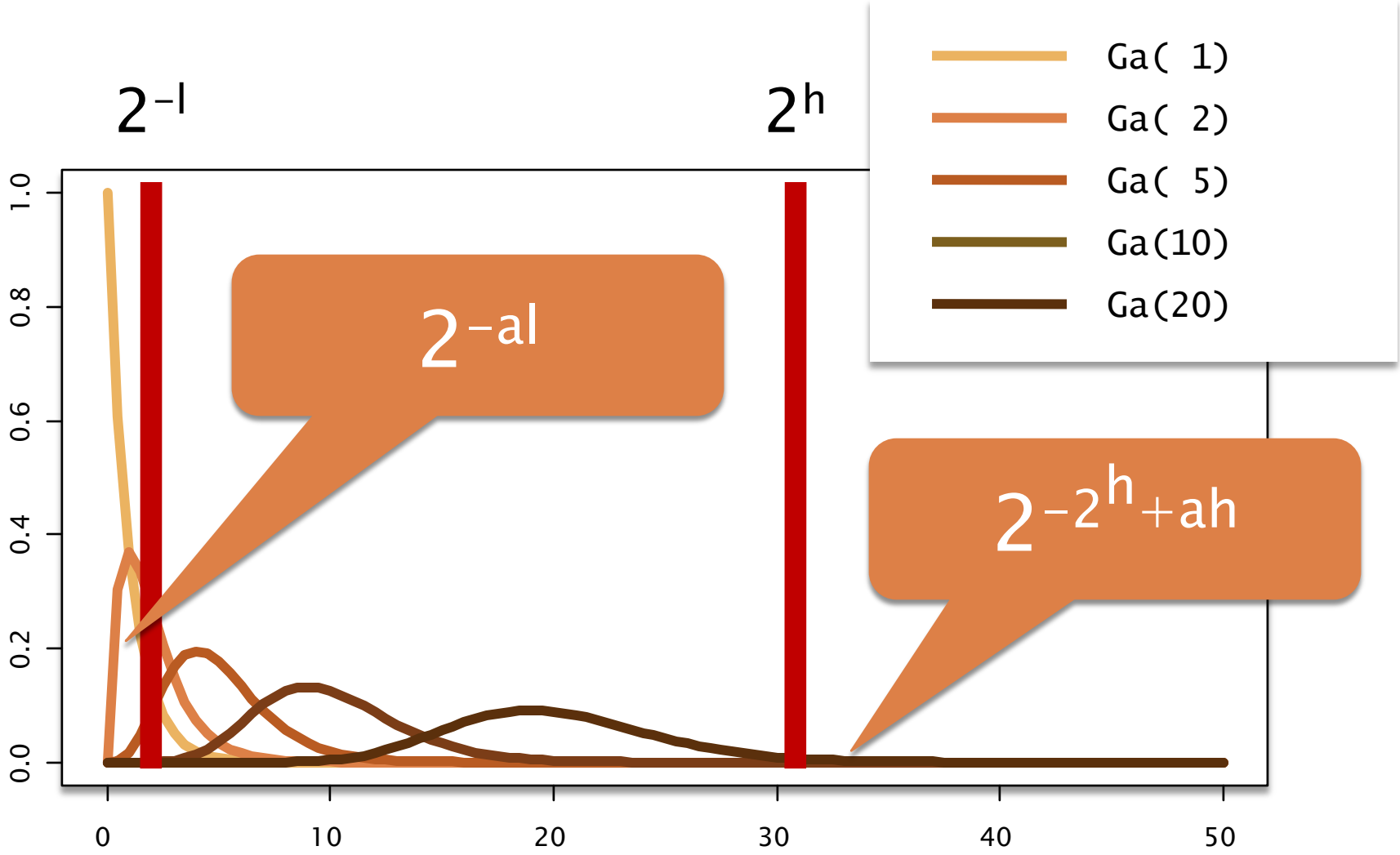
BG'(a,a)

1. x,y←Ga(a)
   a) Set $x^* \leftarrow R(x)$, $y^* \leftarrow R(y)$
   b) If $2^{-l} < x^*, y^* < 2^h$ then go 2
2. output $R'(x^*/(x^*+y^*))$

$$2^{-al+1} + 2^{-2^h+ah+1}$$

注: Rは精度2^{-g}の切り捨て関数
R'は精度2^{-b}の切捨て関数
$-b = -g + 2h + 4l + 2$

ガンマ分布の裾


2^{-l}
2^{h}
Ga(1)
Ga(2)
Ga(5)
Ga(10)
Ga(20)
2^{-al}
$2^{-2^{h}+ah}$
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50

解析の例

Be*(a,a)

1. x,y←Ga(a)
2. output R'(x/(x+y))

2^{-al+1}
$+2^{-2^h+ah+1}$

BG'(a,a)

1. x,y←Ga(a)
   a) Set x*←R(x), y*←R(y)
   b) If $2^{-l}<x^*,y^*<2^h$ then go 2
2. output R'(x*/(x*+y*))

$2\Delta_{GC}$

BG(a,a)

1. x*,y*←**GC**(a)
   1. If $2^{-l}<x^*,y^*<2^h$ then go 2
2. output R'(x*/(x*+y*)))

まとめ

- Bn(K,p)からのサンプリング ($K=O(2^n)$)
  - poly(n)時間で近似的にサンプリング可能
  - 近似精度を統計的距離から保証
  - RO+FPOのシミュレーションに適用可能
- その他の離散分布
  - ポワソン分布もサンプリング可能
    - 内部でガンマ分布と二項分布を使う
  - 幾何分布, 超幾何分布, 負の二項分布は?

参考文献

- [Relles ‘72]
  - “A simple algorithm for generating Binomial random variables when N is large.” (J. of ASA 67(339), 1972)
- [Ahrens-Dieter ’74]
  - “Computer methods for sampling from Gamma, Beta, Poisson and Binomial distributions.” (Computing 12(3), 1974)
- [Ahrens-Dieter ‘80]
  - “Sampling from Binomial and Poisson distributions: A method with bounded computation times.” (Computing, 25(3), 1980)
- [Devroye ‘86]
  - “Non-uniform random variate generation.” (Springer-Verlag, 1986)