

統計的距離を考慮した サンプリング

2008/07/23
11:50-12:10
[夏のLA2008-15]

沼山晃 / 草川恵太
河内亮周 / 田中圭介（東工大）

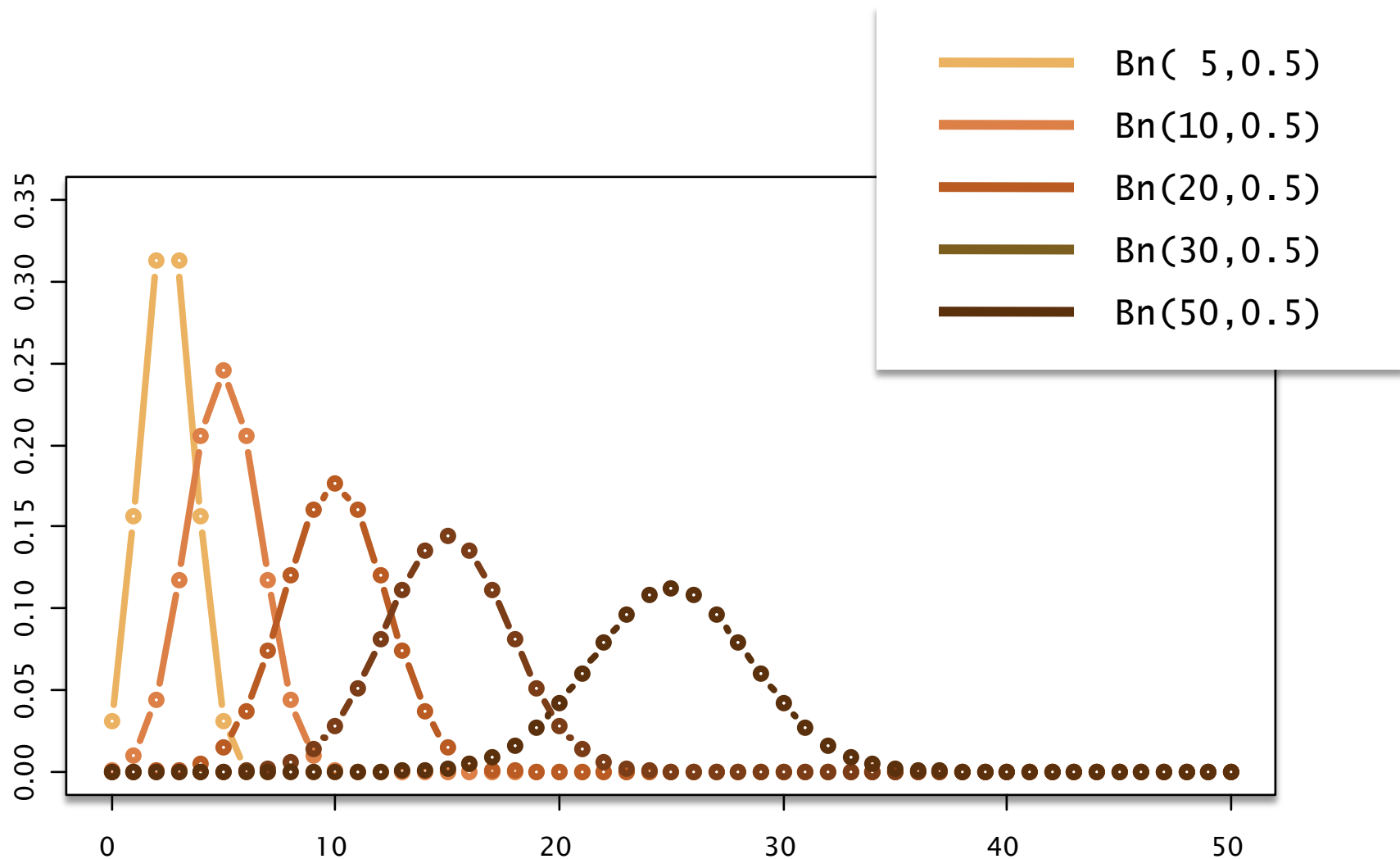
とある日の話

- 暗号の帰着を作っている最中,
Bn(K,p) からサンプリングしたくなった
 - ▣ $1 \leq K \leq 2^n$
- できるのか?

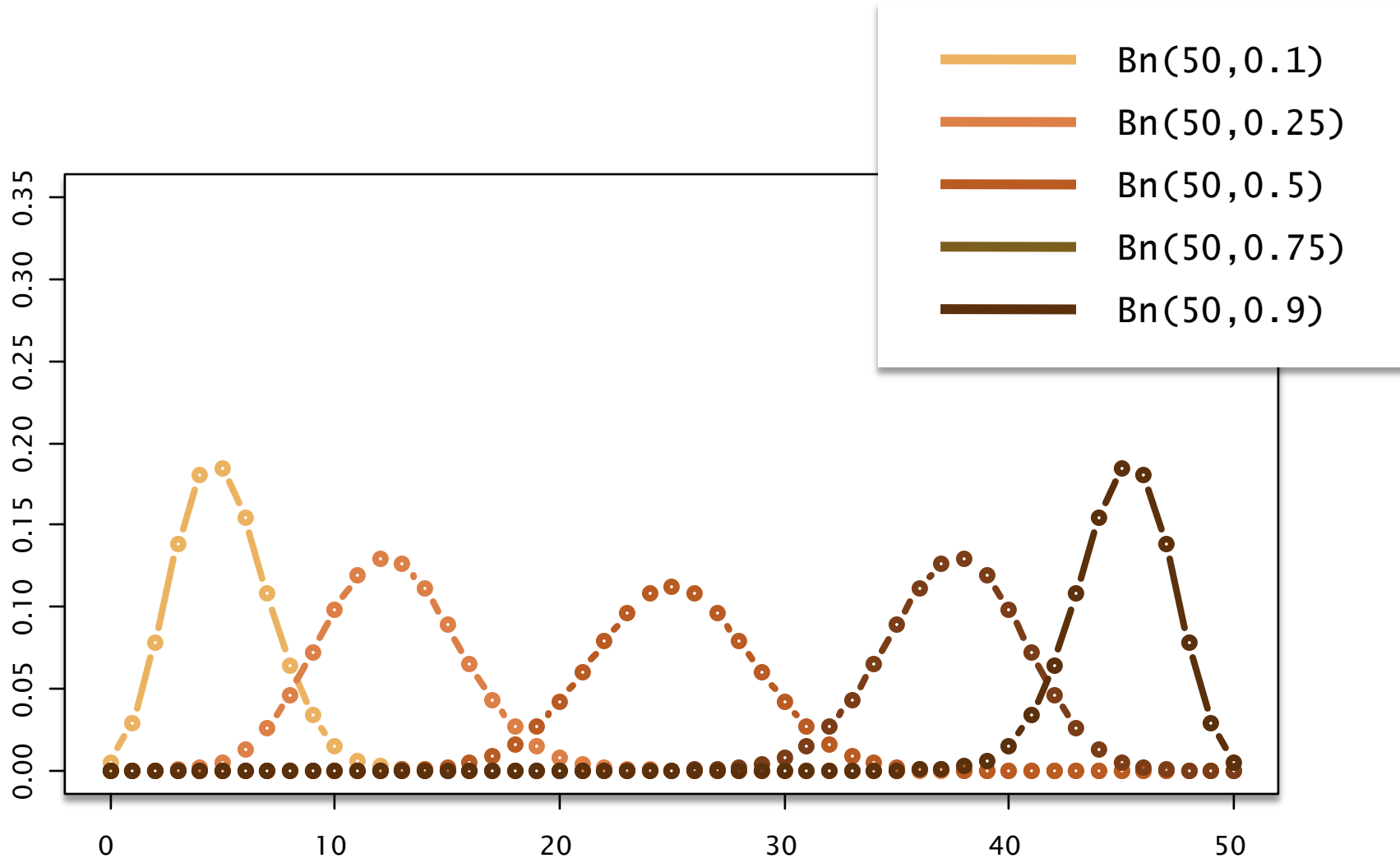
問題の確認 – 二項分布

- $Bn(K, p)$
 - 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



二項分布 p を変化



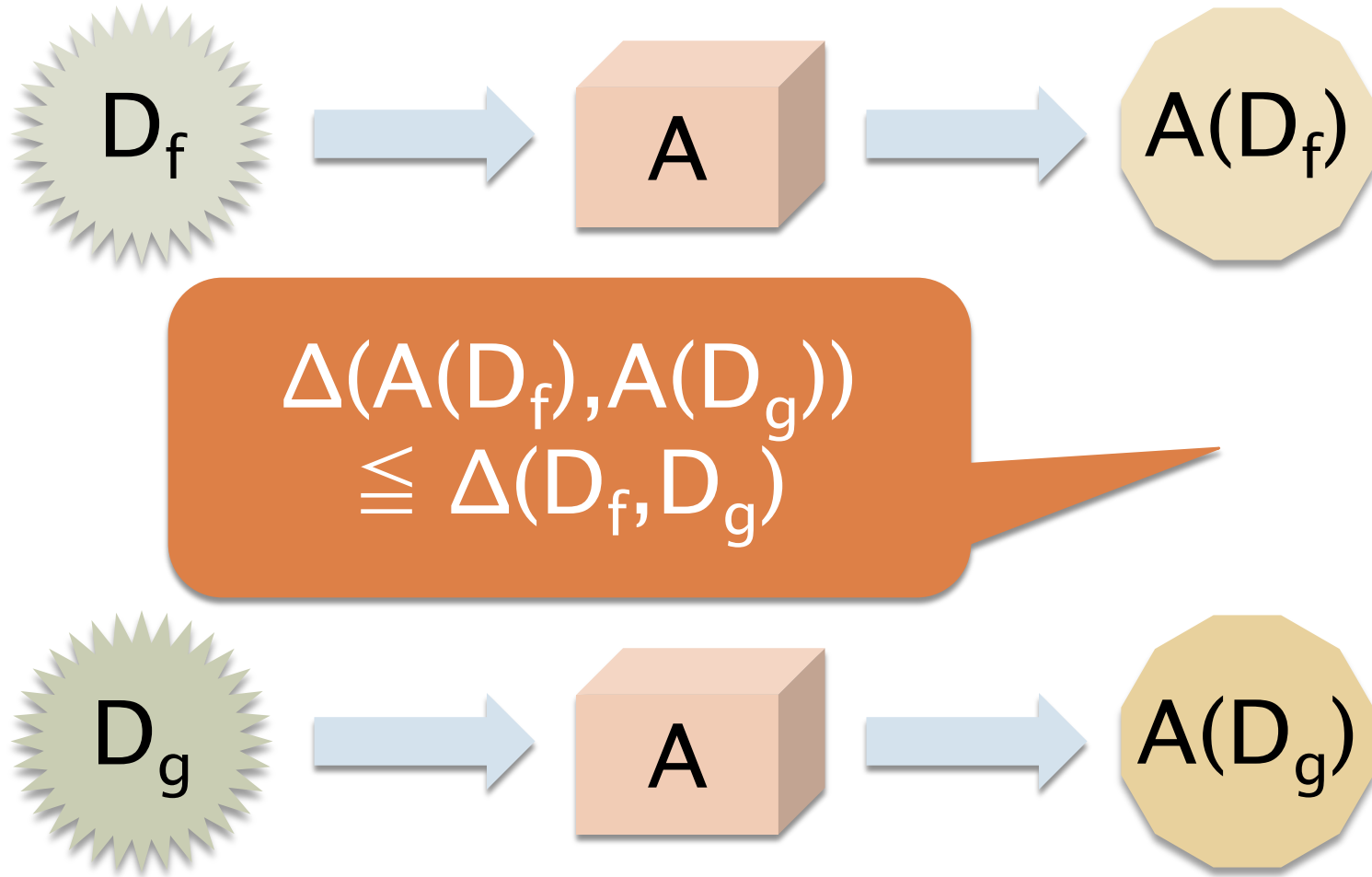
目標の確認

- $K = O(2^n)$ の $B_n(K, p)$ から
poly(n)時間でサンプリング
- 直接的にサンプリングするのは難しそう
→ 近似した分布からのサンプリング
- 近似精度の測り方
→ 統計的距離
→ 帰着において敵の動作を保証するため

統計的距離

D_f, D_g : 確率関数が f, g の離散分布
 $\Delta(D_f, D_g) := \frac{1}{2} \sum_x |f(x) - g(x)|$

統計的距離の性質



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
 - ▣ $K = O(2^n)$, $0 \leq p \leq 1$ (poly(n) bit)
- 出力
 - ▣ $X \in \{0, 1, \dots, K\}$
- 仕様
 - ▣ $\Delta(X \text{の分布}, B_n(K, p)) \leq 2^{-n}$
 - ▣ 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”
- Knuth-Yao

既存研究

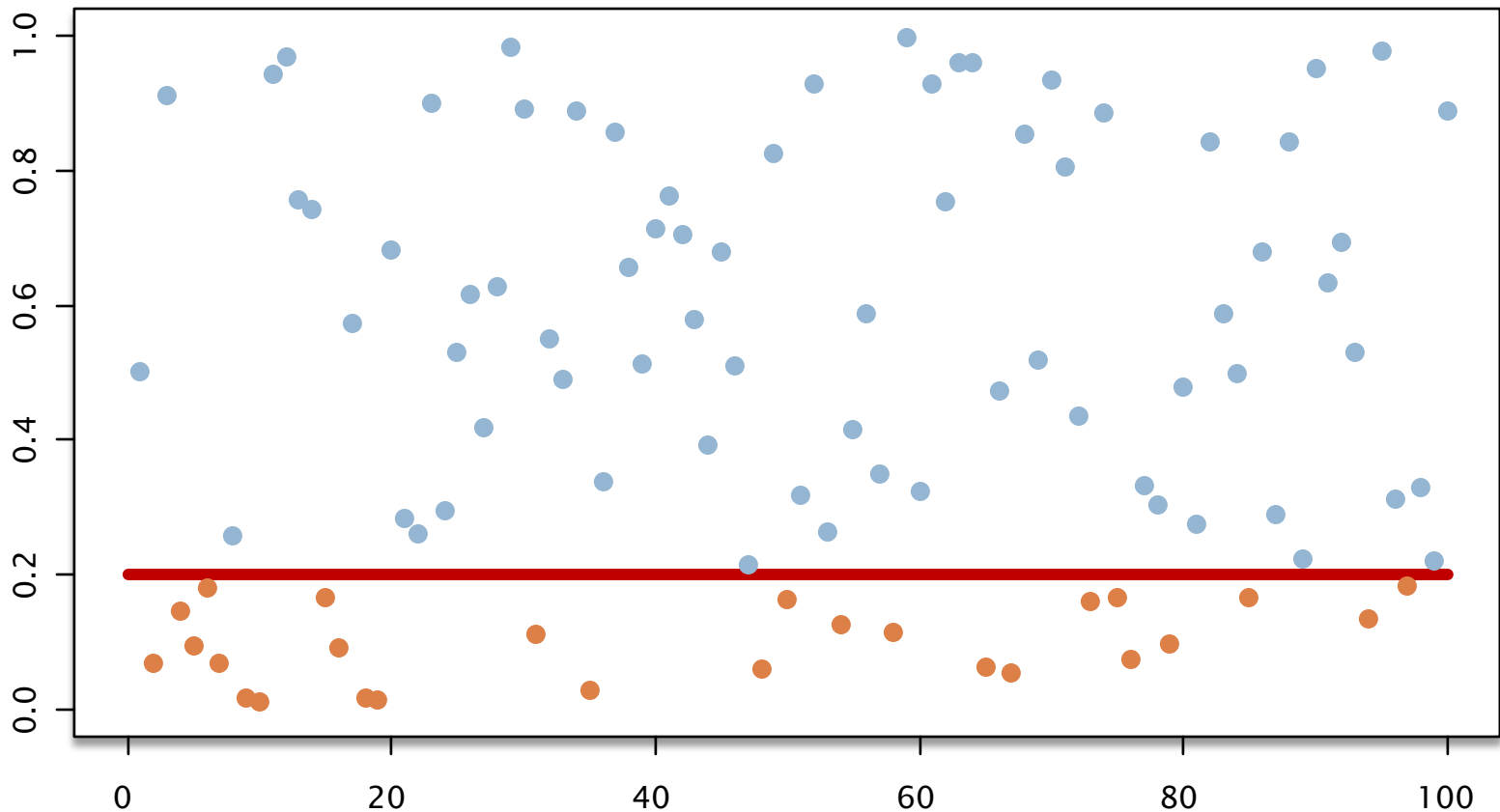
- 統計/シミュレーション分野
 - ▣ 実数計算可能
 - ▣ $[0,1)$ 上の一様乱数を生成可能
- アルゴリズム/暗号分野
 - ▣ ビット計算可能
 - ▣ $\{0,1\}$ 上の一様乱数を生成可能

問題再掲

- 以下を満たすアルゴリズムを構成する
- 入力
 - ▣ $K = O(2^n)$, $0 \leq p \leq 1$ ($\text{poly}(n)$ bit)
- 出力
 - ▣ $X \in \{0, 1, \dots, K\}$
- 仕様
 - ▣ $\Delta(X \text{の分布}, B_n(K, p)) \leq 2^{-n}$
 - ▣ $\text{poly}(n)$ 時間で動作

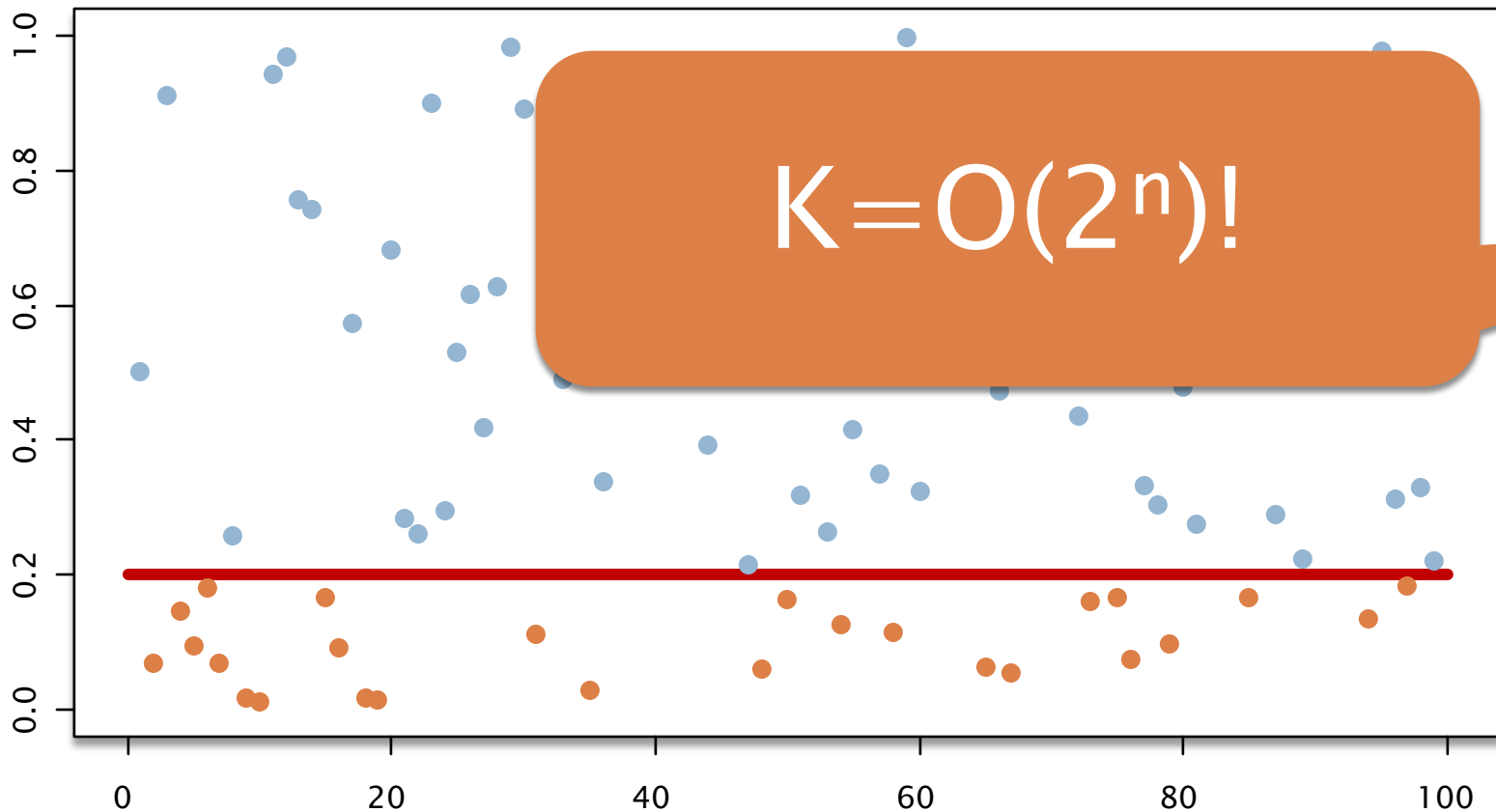
方法案1 - ベルヌーイ試行

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2 – 中心極限定理

- 中心極限定理

- ▣ $K \rightarrow \infty$ で, $B_n(K, p)$ と $N(Kp, Kp(1-p))$ が一致

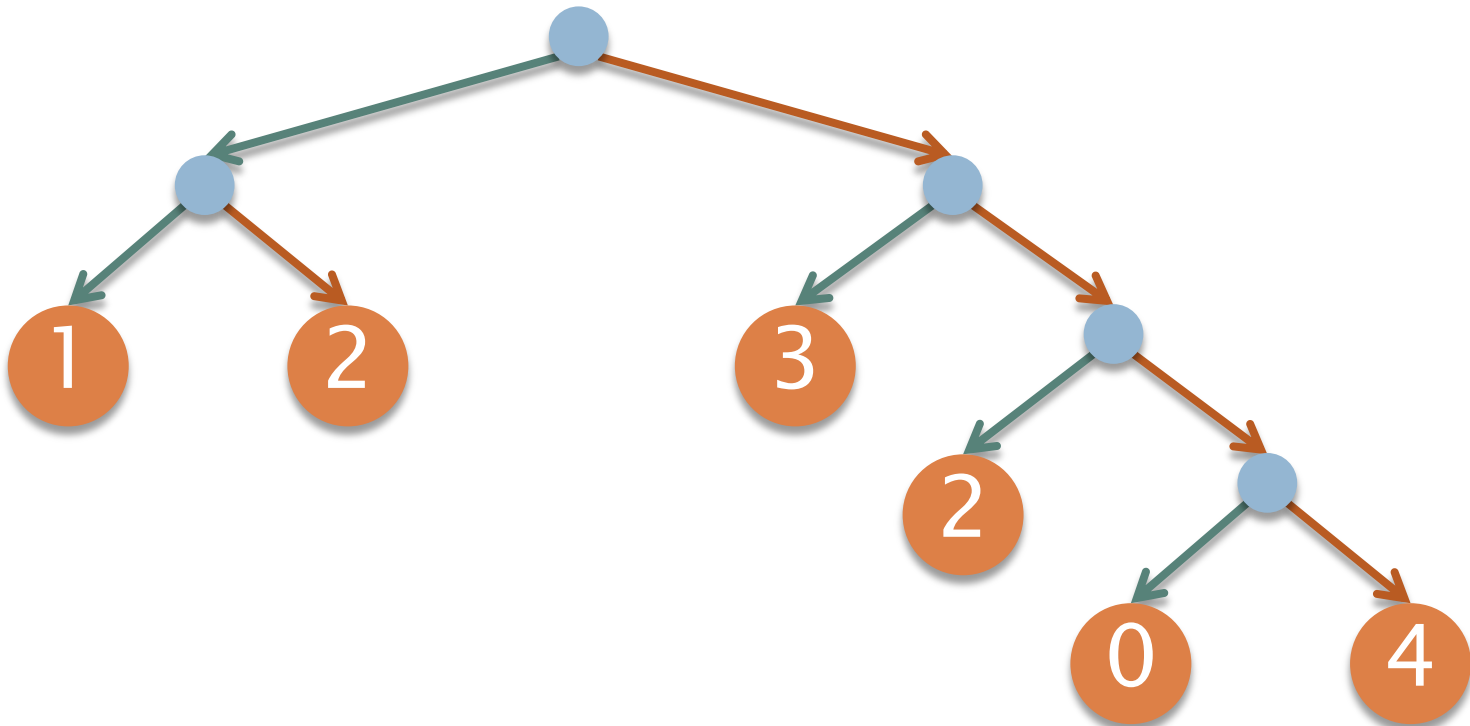
Kが有限だと?

- 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

方法案3 Knuth-Yao

- ラベル付き木上の乱歩を使う

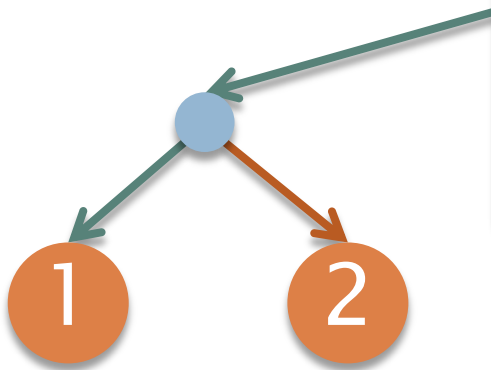
$K=4, p=0.5$



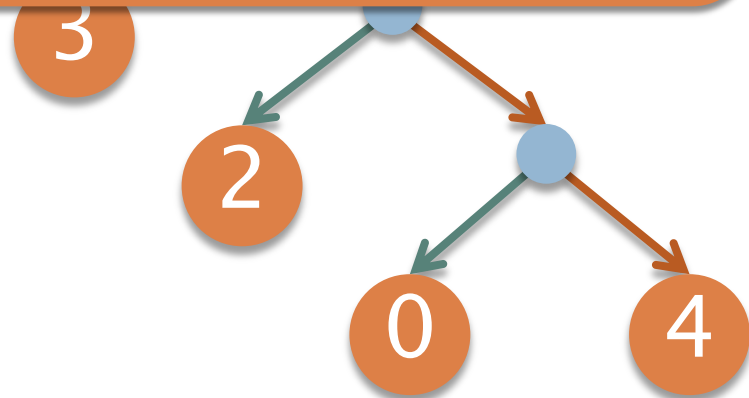
方法案3 Knuth-Yao

- ラベル付き木上の乱歩を使う

$K=4, p=0.5$



木を作るコスト
 $=\Omega(K)!$



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?
- Knuth-Yao
 - ▣ 木を作るコストが $\Omega(K)$ なのでダメ

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: $\Delta(X\text{の分布}, Bn(K,p)) \leq 2^{-n}$

 - 動作時間: $\text{poly}(n)$

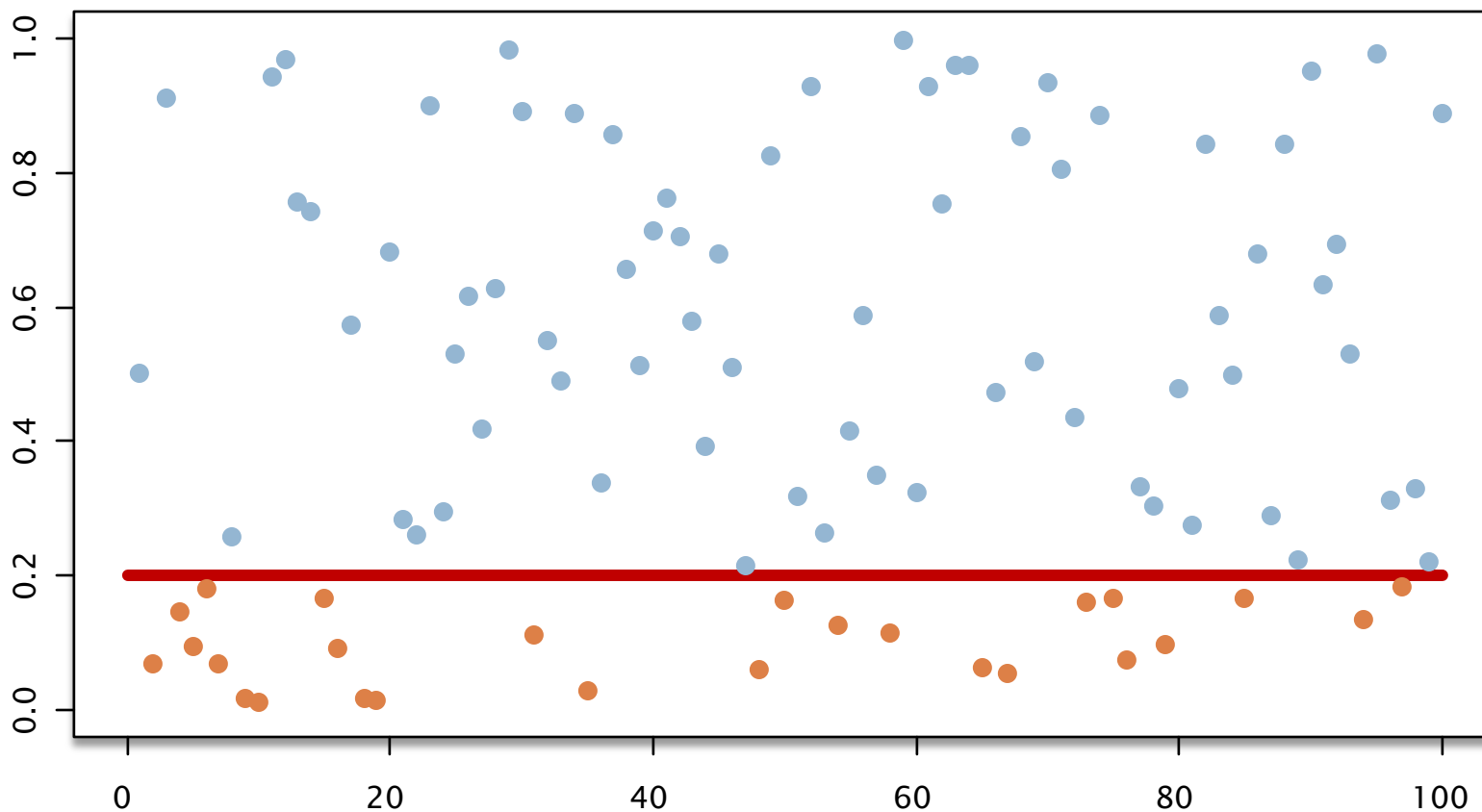
- 統計/シミュレーション分野のアルゴリズムを
離散化して解析

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

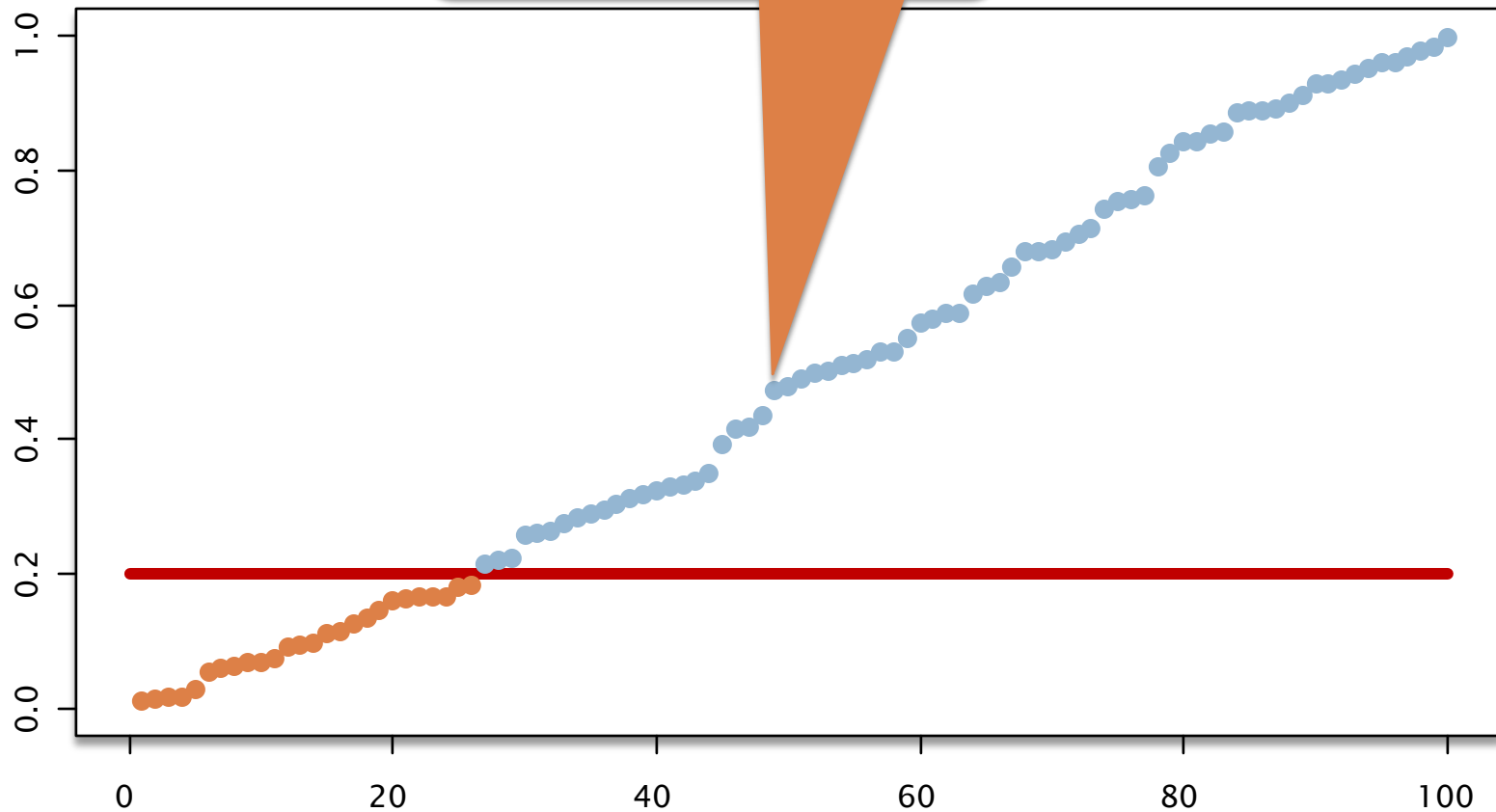
Relles – アイデア

- K個の一樣乱数を考える (実際には振らない)



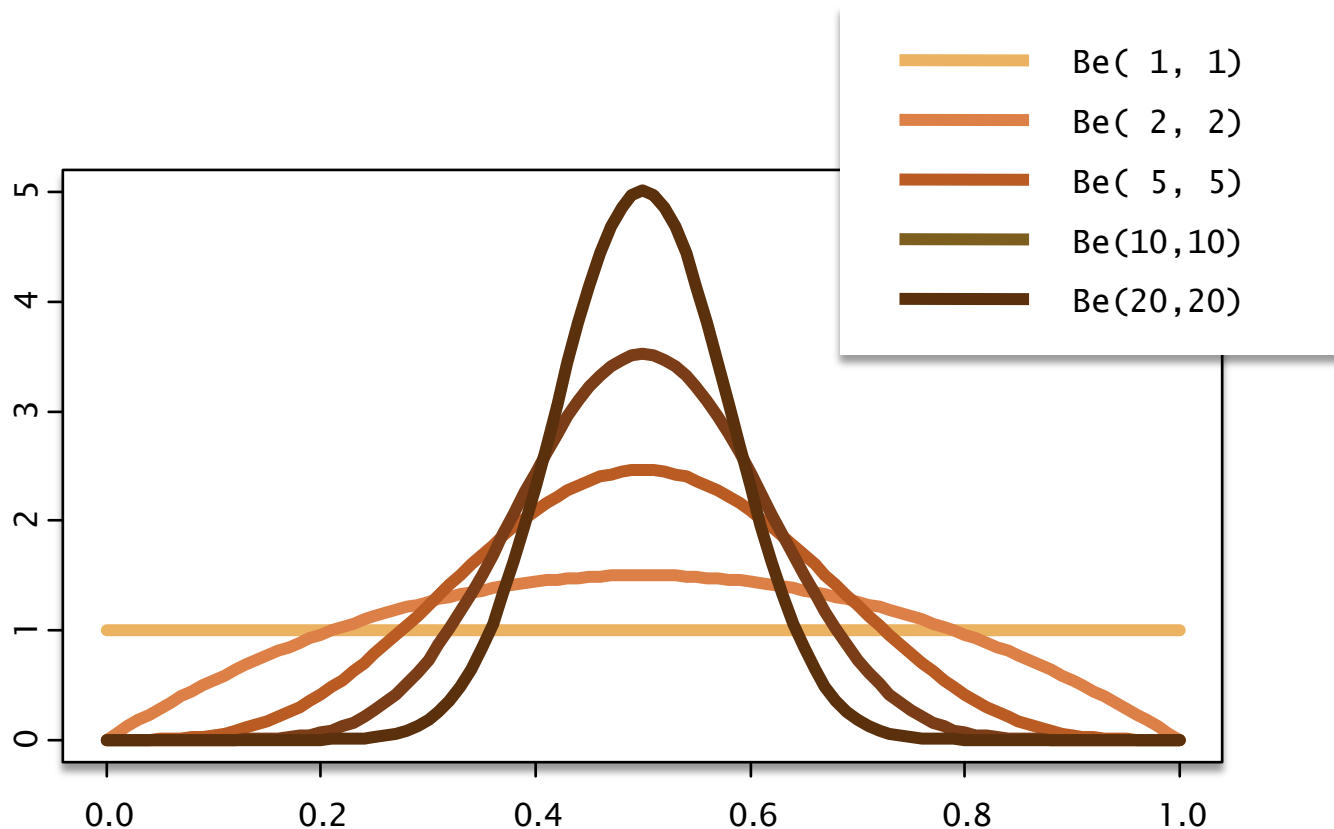
Relles – アイデア

中央値



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. $s \leq p$ 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
- $\text{Be}(a,b)$ (a,b :整数)
 - ▣ pdf: $f(x) = x^{a-1}(1-x)^{b-1} / \beta(a,b)$ (for $0 < x < 1$)
 - $\beta(a,b) = \Gamma(a)\Gamma(b) / \Gamma(a+b) = (a-1)! (b-1)! / (a+b-1)!$
- よく知られている定理
 - ▣ $x \sim \text{Ga}(a), y \sim \text{Ga}(b) \Rightarrow x/(x+y) \sim \text{Be}(a,b)$

解析の例

$Be^*(a,a)$

1. $x, y \leftarrow Ga(a)$
2. output $R'(x/(x+y))$

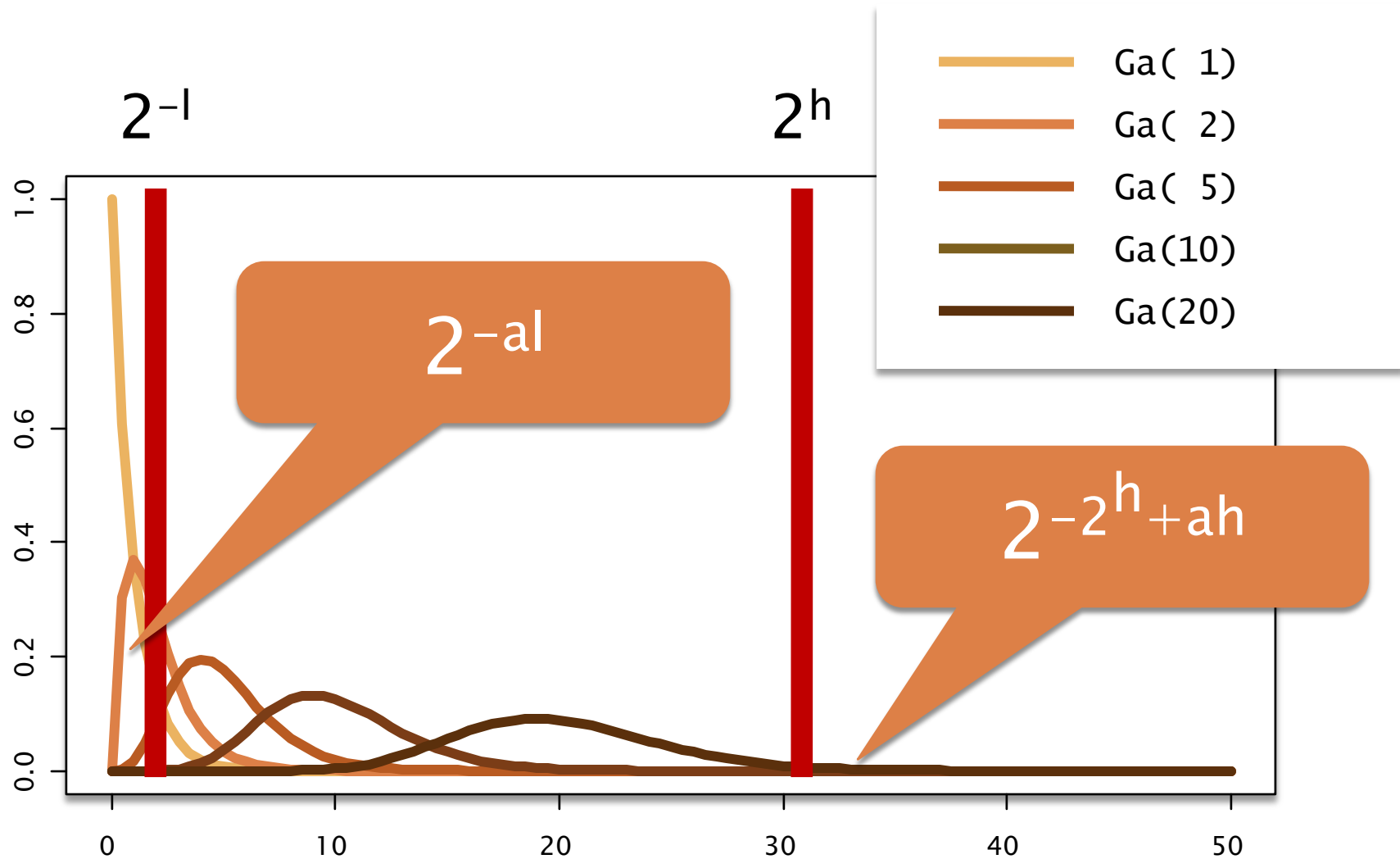
$BG'(a,a)$

1. $x, y \leftarrow Ga(a)$
 - 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} の切り捨て関数
 $-b = -g + 2h + 4l + 2$

ガンマ分布の裾



解析の例

$\text{Be}^*(a,a)$

1. $x, y \leftarrow \text{Ga}(a)$
2. output $R'(x/(x+y))$

$\text{BG}'(a,a)$

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

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

$\text{BG}(a,a)$

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

$$2\Delta_{\text{GC}}$$

まとめ

- $B_n(K, p)$ からのサンプリング ($K = O(2^n)$)
 - ▣ $\text{poly}(n)$ 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
- 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)