「Leftover Hash Lemmaの証明」の和訳

小笠原　琢磨

茨城大学・工学部・情報工学科
黒澤馨研究室 M1

論文要旨

本文は、以下の論文の和訳である。

Universal hash families and the leftover hash lemma, and applications to cryptography and computing

D.R.Stinson
Department of Combinatorics and Optimization
University of Waterloo
Waterloo Ontario, N2L 3G1, Canada
dstinson@uwaterloo.ca

January 15, 2002

1 序論

ユニバーサルハッシュの技術は、1979 年に Carter と Wegman[8] によって発表され、非確率化、擬似乱数生成器、プライバシー増幅を含んだ 3 つの特有の応用に言及したことで、コンピューターサイエンスの多くの分野で不可欠な道具となった。ユニバーサルハッシュ郡は直交配列 [1] や error-correctiong code[9] のような組み合わせの構成と綿密に関係していて、頻繁にこれらの関係を活用している。(概説は [6])

それぞれの強いユニバーサルハッシュ族と関係している乱数生成器は望ましい疑似ランダム特性も持っていることを示している。[14](疑似ランダム性はどのように厳密に確率分布を一様分布に近づけるかの方法を提供している。) 私たちはこの理論の基本的な処理を内蔵したものを提供する。疑似ランダム性の範囲を漸近的ではなく正確に示す。

また、Impagliazzo,Levin,Luby らによって証明 [17] された、いわゆる「leftover hash lemma」の完全な考察も提供する。(「leftover hash lemma」という言葉は Impagliazzo と Zuckerman によって作られた。[16]) 私たちは、「smoothing entropy theorem」として知られる、この結果の単純な組み合わせの証明を提供する。(確率的アルゴリズムの非確率化に用いられる) extractor と、(暗号作成に用いられる) 乱数生成器の構築はもちろん、プライバシー増幅 (他の暗号作成アプリケーション) の技術を含んだ、この主題のそれぞれの結論を調査する。最後に、どのようなコードと直行行列が、興味のあるさまざまな状況を単純な構築として提供するのに向いているか考察する。

この論文は主に特性の解説となっており、ほとんど全ての結果の証明を含んでいる。この後の論文は以下のようにまとめられている。2 章ではそれぞれの定義とユニバーサルハッシュ族の異なる特色の基本的な特徴を紹介する。3 章ではそれぞれの便利なハッシュ族の構築を記す。4 章では、ユニバーサルハッシュ族の 3 つの応用 (疑似乱数生成器、プライバシー増幅、非確率化) を形式張らない方法で示す。5 章では基本的な定理と、距離と確率分布、確率分布の判別性、確率分布の疑似ランダム性、衝突確率、エントロピーと異なるタイプとの間にある概念の関係を定義する。6 章では、強いユニバーサルハッシュ族の疑似ランダム特性に重要な関係がある、基本的な組み合わせの主題の説明と証明をし、疑似乱数生成器が必ずできる種類に対する、この主題の技術の応用を調べる。7 章では、以前の章と似た考え方の、δ-ユニバーサルハッシュ族の疑似ランダム特性に関係がある leftover hash lemma を紹介する。leftover hash lemma の応用は 8 章で考えられている。すなわち、**BPP** クラス内の確率的アルゴリズムの部分的な非確率化に利用される、extractor の考え方である。他の応用のプライバシー増幅は 9 章で考えられている。最後に、10 章で結論を述べる。

2 ユニバーサルハッシュ族

ハッシュ族のいくつかの定義を記す。

- $(D; N, M)$ ハッシュ族は、$|X| = N$、$|Y| = M$ のとき、$f \in \mathcal{F}$ ごとに $f : X \to Y$ となるような、関数 D の集合 $\mathcal{F}$。

- $(D; N, M)$-ハッシュ族$\mathcal{F}$がδ-*universal*のとき、任意の2つの異なる要素$x_1, x_2 \in X$において、$f(x_1) = f(x_2)$となるような関数$f \in \mathcal{F}$が高々δD存在することを言う。パラメーターδはたびたびハッシュ族の*collision probability*と呼ばれる。δ-universalを省略してδ-**U**と表記する。

- $(D; N, M)$-ハッシュ族$\mathcal{F}$が*strongly universal*のとき、任意の2つの異なる要素$x_1, x_2 \in X$と2つの要素$y_1, y_2 \in Y$(異ならなくても良い)において、

$$| \{f \in \mathcal{F} : f(x_i) = y_i, i = 1, 2\} | = \frac{D}{M^2}.$$

が成り立つ。strongly universalを省略して**SU**と表記する。

しばしば、$(D; N, M)$- ハッシュ族$\mathcal{F}$を、行を$\mathcal{F}$の関数のインデックス、列をXの要素のインデックスとして、M記号の配列$D \times N$の形式で表記する。行fと列xの項目の要素は$f(x)$となる（すべての$f \in \mathcal{F}$とすべての$x \in X$)。それぞれの要素の行は、族の関数のひとつと対応している。この配列は$A(\mathcal{F})$を意味し、ハッシュ族$\mathcal{F}$の*array representation*と呼ぶ。もし$\mathcal{F}$がδ -U$(D; N, M)$ハッシュ族ならば、任意の$A(\mathcal{F})$の2つの列において、与えられた2つの列の要素が等しいような$A(\mathcal{F})$の行は高々δD存在する。

Yをq記号のアルファベットとする。(n, K, d, q)*code*は$\mathcal{C}$内の任意の2つの異なるベクトル間のハミング距離が少なくともdとなるような、Y^n内のKベクトル（*codewords*と呼ばれる）の集合$\mathcal{C}$である。もしコードが*linear*（例えば、もしqが素数累乗、$Y = \mathbb{F}_q$、$\mathcal{C}$は$(\mathbb{F}_q)^n$の部分空間）ならば、$k = \log_q K$がコードの*dimension*のとき、コードは$[n, k, d, q]$*code*と言う。

以下と同値なことはBierbrauer,Johansson,Kabatianskii,Smeetsによって最初に述べられた。[7]

定理 2.1. もし(n, K, d, q)*code*が存在したとすると、$1 - \frac{d}{n}$-U$(n; K, q)$ハッシュ族が存在する。逆に、もしδ-U$(D; N, M)$ハッシュ族が存在するならば、$(D, N, D(1-\delta), M)$*code*が存在する。

定理 2.1は$\mathcal{F}$が決まったハッシュ族のとき、決まったコード内のcordwordsと$A(\mathcal{F})$の列は対応していることを証明している。

例 2.1. 以下の$\frac{1}{3}$-U$(3; 9, 3)$ハッシュ族$\{f_i : i \in \mathbb{Z}_3\}$は$(3, 9, 2, 3)$*code*と等しい。

	$(0,0)$	$(0,1)$	$(0,2)$	$(1,0)$	$(1,1)$	$(1,2)$	$(2,0)$	$(2,1)$	$(2,2)$
f_0 :	0	0	0	1	1	1	2	2	2
f_1 :	0	1	2	1	2	0	2	0	1
f_2 :	0	2	1	1	0	2	2	1	0

∎

orthogonal array OA$_\lambda(N, M)$は、任意の2つの配列の列において、それぞれの記号の2つの組が正確にλ行生じるようなM記号のN配列によって、λM^2になる。もし$\mathcal{F}$が**SU**$(D; N, M)$-ハッシュ族ならば、$\lambda = D/M^2$のとき$A(\mathcal{F})$はOA$_\lambda(N, M)$となる。逆もまた成り立つ。よって、以下の定理を得る。([5])

定理 2.2. $\mathbf{SU}(D;N,M)$-ハッシュ族は、$\lambda = D/M^2$ のとき $\mathrm{OA}_\lambda(N,M)$ と等しい。

例 2.2. 以下の $\mathbf{SU}(9;3,3)$-ハッシュ族 $\{f_{i,j} : i,j \in \mathbb{Z}_3\}$ は $\mathrm{OA}_1(3,3)$ と等しい。

	0	1	2
$f_{0,0}:$	0	1	1
$f_{0,1}:$	1	2	2
$f_{0,2}:$	2	0	0
$f_{1,0}:$	1	1	0
$f_{1,1}:$	2	2	1
$f_{1,2}:$	0	0	2
$f_{2,0}:$	1	0	1
$f_{2,1}:$	2	1	2
$f_{2,2}:$	0	2	0

▮

3 いくつかのハッシュ族の構築

この章ではハッシュ族のそれぞれの構築を示す。以下の構築の主なアイディアは orthogonal arrays の言語を使って 1947 年に Rao によって作られた。[4]

定理 3.1. l を正整数、q を素数累乗とする。$X \subseteq (\mathbb{F}_q)^l$ を $\mathbb{F}_q$ 上の任意の一次独立ベクトル対の集合とする。すべての $\vec{r} \in (\mathbb{F}_q)^l$ で、関数 $f_{\vec{r}:X\to\mathbb{F}_q}$ を次のように定義する。

$$f_{\vec{r}}(\vec{x}) = \vec{r}\cdot\vec{x}.$$

最終的に、定義は以下となる。

$$\mathcal{F}(q,l,X) = f_{\vec{r}} : \vec{r} \in (\mathbb{F}_q)^l.$$

その結果、$\mathcal{F}(q,l,X)$ は $\mathbf{SU}(q^l;\mid X \mid,q)$-ハッシュ族となる。

証明. 明らかに $(q^l;\mid X \mid,q)$-ハッシュ族である。よって、これが **SU** であることを証明する。$\vec{x_1},\vec{x_2} \in (\mathbb{F}_q)^l (\vec{x_1} \neq \vec{x_2})$、$y_1,y_2 \in \mathbb{F}_q$ とする。これから、

$$\vec{r}\cdot\vec{x}_1 = y_1$$

かつ

$$\vec{r}\cdot\vec{x}_2 = y_2.$$

となるようなベクトル $\vec{r} \in (\mathbb{F}_q)^l$ の数を数える。今 $\vec{x_1}$ と $\vec{x_2}$ は $\mathbb{F}_q$ 上の一次独立ベクトルである。もし $\vec{r} = (r_1,\cdots,r_l)$ と表すと、未知数が $r_1,\cdots,r_l$ の l 個ある、$\mathbb{F}_q$ の一次独立連立方程式となる。それゆえベクトル $\vec{r}$ の解は正確に q^{l-2} 個あり、ハッシュ族は **SU** となる。 □

定理 3.1 の系を 2 つ示す。系 3.1 により構築できるハッシュ族は、次数 q のクラシカルデザルグアフィン平面と等しくなる。

系 3.1. q は素数累乗とする。$a, b \in \mathbb{F}_q$ において $f_{a,b} : \mathbb{F}_q \to \mathbb{F}_q$ を

$$f_{a,b}(x) = ax + b.$$

と定義する。その結果、$\{f_{a,b} : a, b \in \mathbb{F}_q\}$ は $\mathbf{SU}(q^2; q, q)$-ハッシュ族になる。

証明. $l = 2$ とし、$X = \mathbb{F}_q \times \{1\}$ とする。その後、定理 3.1 を適用する。 □

系 3.2 で示すハッシュ族は [18] で提案されている。

系 3.2. l は正整数で、q は素数累乗とする。$X = \{0, 1\}^l \setminus \{(0, \cdots, 0)\}$ とする。すべての $\vec{r} \in (\mathbb{F}_q)^l$ において、$f_{\vec{r}} : X \to \mathbb{F}_q$ を

$$f_{\vec{r}}(\vec{x}) = \vec{r} \cdot \vec{x}.$$

と定義する。その結果、$\{f_{\vec{r}} : \vec{r} \in (\mathbb{F}_q)^l\}$ は $\mathbf{SU}(q^l; 2^l - 1, q)$-ハッシュ族になる。

今まで提供した構築は、今のところハッシュ関数はすべて一次関数になる。二次関数による **SU** ハッシュ族のための構築は [10] で提供されている。

定理 3.2. q は奇数の素数累乗とする。$a, b \in \mathbb{F}_q$ において $f_{a,b} : \shortparallel \to \mathbb{F}_q$ を

$$f_{a,b}(x) = (x + a)^2 + b.$$

と定義する。その結果、$\{f_{a,b} : a, b \in \mathbb{F}_q\}$ は $\mathbf{SU}(q^2; q, q)$-ハッシュ族になる。

証明. 明らかに、$(q^2; q, q)$-ハッシュ族である。よって、これが **SU** であることを証明する。$\vec{x_1}, \vec{x_2} \in (\mathbb{F}_q)^l (\vec{x_1} \neq \vec{x_2})$、$y_1, y_2 \in \mathbb{F}_q$ とする。これから、

$$(x_1 + a)^2 + b = y_1$$

かつ

$$(x_2 + a)^2 + b = y_2.$$

は定数となるような次数のペア $(a, b) \in (\mathbb{F}_q)^2$ の数を示す。2 つの方程式を引き算して、a の値を解くと

$$a = \frac{y_1 - y_2}{2(x_1 - x_2)} - \frac{x_1 + x_2}{2}$$

よって、a を決めると、b が一意に得られる。 □

ハッシュ族を定理 3.2 の方法で構築すると、次数 q のデザルグアフィン平面と等しくなることも興味深い。例 2.2 で示した $\mathbf{SU}(9; 3, 3)$ ハッシュ族は定理 3.2 の適用である。

定理 2.1 を考慮して、コードから多くの δ-**U** ハッシュ族を構築することができる。例えば、リードソロモン符号を使って、[7] で指摘されているような以下の構築を得る。

定理 3.3. q を a 素数累乗とし、$1 \leq k \leq q-1$ とする。$a \in \mathbb{F}_q$ において、$f_a : (\mathbb{F}_q)^k \to \mathbb{F}_q$ を

$$f_a(x_0, \cdots, x_{k-1}) = x_0 + \sum_{i=1}^{k-1} x_i a^i.$$

と定義する。その時 $\{f_a \in \mathbb{F}_q\}$ は $\frac{k-1}{q}$-$\mathbf{U}(q; q^k, q)$ ハッシュ族になる。

証明. 明らかに $(q; q^k, q)$ ハッシュ族である。$\frac{k-1}{q}$-$\mathbf{U}$ を証明する。まず、2つの異なるベクトルを、

$$(x_0, \cdots, x_{k-1}), (x'_0, \cdots, x'_{k-1}) \in (\mathbb{F}_q)^k$$

とする。

$$\sum_{i=0}^{k-1} x_i a^i = \sum_{i=0}^{k-1} x'_i a'^i$$

のような $a \in \mathbb{F}_q$ 要素の（上限の）数を決定したい。これは、

$$\sum_{i=0}^{k-1} (x_i - x'_i) a^i = 0$$

と等しい。

体 $\mathbb{F}_q$ 上の全次数が高々 $k-1$ の非ゼロ多項式は根が高々 $k-1$ だけあるので、それはこの等式がもつ a の値が高々 $k-1$ あることになる。それゆえ、ハッシュ族は $\frac{k-1}{q}$-$\mathbf{U}$ となる。 □

以下の構築は [6] 内で与えられている。（実際には Bose と Bush[15] によって異なった行列で 1952 年に構築されたものに基づいている。）

定理 3.4. q は a 素数累乗とし、$s \geq t$ となるような正整数を s, t とする。$\phi : \mathbb{F}_{q^s} \to (\mathbb{F}_q)^t$ は任意の全射の q-線形写像とする。（すなわち、すべての $x, y \in \mathbb{F}_{q^s}$ で $\phi(x+y) = \phi(x) + \phi(y)$ かつ、すべての $a \in \mathbb{F}_q, x \in \mathbb{F}_{q^s}$ で $\phi(ax) = a\phi(x)$）すべての $a \in \mathbb{F}_{q^s}$ において、$f_a : \mathbb{F}_{q^s} \to (\mathbb{F}_q)^t$ を

$$f_a(x) = \phi(ax)$$

と定義する。そのとき、$\{f_a : a \in \mathbb{F}_{q^s}\}$ は $\frac{1}{q^t}$-$\mathbf{U}(q^s; q^s, q^t)$ ハッシュ族である。

証明. 明らかに $(q^s; q^s, q^t)$ ハッシュ族である。$\frac{1}{q^t}$-$\mathbf{U}$ を証明する。$x_1, x2 \in \mathbb{F}_{q^s}, x_1 \neq x_2$ とする。

$$\phi(ax_1) = \phi(ax_2)$$

のような $a \in \mathbb{F}_q$ 要素の（上限の）数を決定したい。ϕ は線形なので、次の式と等しくなる。

$$\phi(a(x_1 - x_2)) = 0$$

今、ϕ は全射で線形なので、$|\ker(\phi)| = q^{s-t}$ となる。$x_1 - x_2 \neq 0$ なので、$a(x_1 - x2) \in \ker(\phi)$ のような a の値は正確に q^{s-t} ある。それゆえ、要望通り $\frac{1}{q^t}$-$\mathbf{U}$ ハッシュ族になる。 □

一つの族の中の関数の定義域が他の族の中の関数の値域と同じであるときはいつも、２つのハッシュ族を合成して使用することが出来る。今、[20] から δ_2-**U** ハッシュ族と δ_1-**U** ハッシュ族を合成し、$(\delta_1+\delta_2)$-**U** ハッシュ族を得る、合成構築を記す。（この手続きは concatenated coce の構築として考えることができる。）

定理 3.5. $\mathcal{F}_1$ は X から Y_1 への関数の δ_1-$\mathbf{U}(D_1; N, M_1)$ ハッシュ族、$\mathcal{F}_2$ は Y_1 から Y_2 への関数の δ_2-$\mathbf{U}(D_2; M_1, M_2)$ ハッシュ族と仮定する。任意の $f_1 \in \mathcal{F}_1$,$f_2 \in \mathcal{F}_2$ において、$f_1 \circ f_2 : X \to Y_2$ を $f_1 \circ f_2(x) = f_2(f_1(x))$ と定義する。その結果、

$$\{f_1 \circ f_2 : f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2\}$$

は $(\delta_1+\delta_2)$-$\mathbf{U}(D_1D_2; N, M_2)$ ハッシュ族となる。

証明. 任意の２つの要素 $x, x' \in X$ を固定する。$f_2(f_1(x)) = f_2(f_1(x'))$ となるような (f_1, f_2) のペアの数の上限を計算する。$\mathcal{G} = \{f_1 \in \mathcal{F}_1 : f_1(x) = f_1(x')\}$. と仮定する。明らかに $|\mathcal{G}| \leq \delta_1 D_1$、そして任意の $f_1 \in \mathcal{G}$ において、すべての $f_2 \in \mathcal{F}_2$ で $f_2(f_1(x)) = f_2(f_1(x'))$ が成り立つ。

今、もし $f_1 \in \mathcal{F}_1 \setminus \mathcal{G}$ ならば $f_1(x) \neq f_1(x')$ である。すべての $f_1 \in \mathcal{F}_1 \setminus \mathcal{G}$ において、$f_2(f_1(x)) = f_2(f_1(x'))$ のような $f_2 \in \mathcal{F}$ 関数は高々$\delta_2 D_2$ ある。それゆえ、$f_2(f_1(x)) = f_2(f_1(x'))$ のような (f_1, f_2) のペアの合計は高々

$$\begin{aligned} |\mathcal{G}| D_2 + (D_1 - |\mathcal{G}|)\delta_2 D_2 &\leq |\mathcal{G}| D_2 + D_1\delta_2 D_2 \\ &\leq \delta_1 D_1 D_2 + D_1\delta_2 D_2 \\ &= (\delta_1+\delta_2) D_1 D_2. \end{aligned}$$

それゆえ、ハッシュ族は $(\delta_1+\delta_2)$-**U** になる。 □

4　3つのアプリケーション

このセクションでは、ユニバーサルハッシュ族の３つの興味深いアプリケーションを提供する。

4.1　疑似乱数生成器

最初のアプリケーションは疑似乱数生成器のために強いユニバーサルハッシュ族を使う。$\mathcal{F}$ は X から Y への関数の $\mathbf{SU}(D; N, M)$ ハッシュ族と仮定する。特定の関数 $f \in \mathcal{F}$ はランダムに選ばれ、秘密を保持される。次に、一つ以上の要素 $x \in X$ を X 上の特定の確率分布 p にしたがって選ばれる。それぞれの選ばれた x について、値 $y = f(x)$ を計算し出力する。目的は値 y の分布の結果の出力が「一様に近い」ことである。定理を使い、多くの $f \in \mathcal{F}$ の選択に応じられる、パラメーター D, N, M の適切な選び方と確率分布 p がやや一様に近くなるよう設定する強いユニバーサルプロパティを示し、論文の後で明らかにする。

BPV 生成器 ([18]) は上記の方法のよい例を提供している。この生成器は系 3.2 を使い、以下のようになっている。p を素数とする。集合 X はすべての $2^l - 1$ のゼロを除いたバイナリの l 組から成る。今、ベクトル $\vec{r} \in (\mathbb{Z}_p)^l$ はランダムに選ばれる。これは固定された関数 $f_{\vec{r}} \in \mathcal{F}(p, l, X)$ を決定する。その時、$\vec{x}$ をランダムに選択して作り、そして、それぞれの $\vec{x}$ について選択し、値 $f_{\vec{r}}(\vec{x})$ を計算し出力する。

BPV 生成器は乱数生成を素早くするための前計算の便利な方法を認めるため、役に立つ。これは例えばスマートカードなど制約があるデバイス上で署名スキームを実装するという状況で役に立つことがある。α がいくつかの素数 p において、有限体 $\mathcal{F}_p$ の巡回部分群を生成するとき、秘密のランダム値 y のために、(y, α^y) のペアを計算する必要があるエルガマルタイプの署名スキームを実装すると仮定する。BPV 生成器を必須の y の値を生成するために使っても良い。例えば、$\vec{x} \in 0, 1^l$ をランダムに選び、$\vec{r} = (r_1, \cdots, r_l) \in (\mathbb{F}_p)^l$ を固定したとき、$y = f_{\vec{r}}(\vec{x}) = \vec{r} \cdot \vec{x}$ とする。もし値 $\alpha^{r_1}, \cdots, \alpha^{r_l}$ を前計算して記憶しておけるならば、

$$\alpha^y = \prod_{\{i:x_i=1\}} \alpha^{r_i}$$

は前計算した値 l の部分集合と一緒に掛け算することで計算することが出来る。これは法 p の累乗法を、$\mathbb{F}_p$ 上の（高々）$l - 1$ 乗算に置き換える。

4.2 計算複雑性理論

（省略）

4.3 プライバシー増幅

プライバシー増幅の考えは Bennett,Brassard,Robert([2]) による。Alice と Bob の 2 人はいくつかの要素 $x \in X$ の値をそれぞれ知った後で、量子暗号を使った鍵認証プロトコルを実行すると仮定する。盗聴者 Eve は X 上の確率分布 p によって指定される x の一部分の情報を持っている。Alice と Bob は確率分布 p を知らない。しかし、例えば分布 p の不均一性について、指定された衝突確率についていくつかの情報は持っている。(この考えは §5.5 で定義される。)

今、$\mathcal{F}$ は X から Y への関数の δ-$\mathbf{U}(D; N, M)$ ハッシュ族と仮定する。特定の関数 $f \in \mathcal{F}$ は Alice と Bob によってランダムに選ばれ、秘密は保持される。Alice と Bob は値 $y = f(x)$ を両方計算することができる。Eve の目的は y の値についてのどんな小さな情報でも得ることである。§9 の中で、パラメーター D, N, M とある程度不均一性な p を与えられて、どのように Eve が知る y から適切に上限を定めることが出来るか説明する。

5 統計の距離と擬似ランダム

$\mathcal{F}$ は X から Y への関数の $\mathbf{SU}(D; N, M)$ ハッシュ族と仮定する。一つの主な結果として,（関数 $f \in \mathcal{F}$ と関係する）高い確率で、$x \in X$ を一様に近い分布を使って選んだ

とき、値 $f(x)$ は一様に近い分布を持っていることを証明した。このセクションでは、「統計距離」と「擬似ランダム」を用いて、「一様分布に近い」概念を計る。また、いくつかの基本的な定義と衝突確率や確率分布のエントロピーのようなトピックスの結果も提供する。

5.1 確率空間とランダム変数

確率空間とランダム変数に関係する少しスタンダードな定義から始める。「有限確率空間」は Y を有限な集合、p を Y 上の確率分布としたとき、(Y, p) のペアとなる。Y 上の一様確率分布は u_Y と書き、すべての要素 $y \in Y$ に確率 $1/ \mid Y \mid$ を割り当てる。有限確率空間 (Y, p) 上の「ランダム変数」は関数 $\mathbf{Y} : Y \to \mathbb{R}$ である。

$\mathbf{Y}$ の期待値は $\mathbf{E}(\mathbf{Y})$ と書き、

$$\mathbf{E}(\mathbf{Y}) = \sum_{y \in Y} p(y)\mathbf{Y}(y)$$

と定義する。

$\mathbf{Y}$ の分散は $\mathbf{var}(\mathbf{Y})$ と書き、

$$\mathbf{var}(\mathbf{Y}) = \mathbf{E}(\mathbf{Y}^2) - (\mathbf{E}(\mathbf{Y}))^2 = \mathbf{E}((\mathbf{Y} - \mathbf{E}(\mathbf{Y}))^2).$$

と定義する。

もし、特別な確率分布 p に依存していることを強調するときは、$\mathbf{E}_p(\mathbf{Y})$ や $\mathbf{var}_p(\mathbf{Y})$ を使う。

証明なしで確率定理に由来する原理を以下に述べる。

系 5.1. (シェビチェフの不等式) 任意のランダム変数 $\mathbf{Y}$ において、

$$\mathbf{Pr}[\mid \mathbf{Y}(y) - \mathbf{E}(\mathbf{Y}) \mid \geq \epsilon] \leq \frac{\mathbf{var}(\mathbf{Y})}{\epsilon^2}.$$

を持つ。

系 5.2. (イェンゼンの不等式) $\mathcal{I} \subseteq \mathbb{R}$ の区間、$\mathbf{Y}$ は $\mathcal{I}$ 上の値をもつランダム変数、$f : \mathcal{I} \to \mathbb{R}$ は区間 $\mathcal{I}$ 上において、完全に凹型と仮定する。そのとき

$$\mathbf{E}(f(\mathbf{Y})) \leq f(\mathbf{E}(\mathbf{Y})).$$

を持つ。

$f(x) = -x^2$ で $\mathcal{I} = \mathbb{R}$ を持つとき、以下の系を得る。

系 5.3. 任意のランダム変数 $\mathbf{Y}$ において、

$$(\mathbf{E}(\mathbf{Y}))^2 \leq \mathbf{E}(\mathbf{Y}^2).$$

を持つ。

$f(x) = \log x$ で $\mathcal{I} = (0, \infty)$ を持つとき、以下の系を得る。

系 5.4. 正の値を持つ任意のランダム変数 $\mathbf{Y}$ において、

$$\log \mathbf{E}(\mathbf{Y}) \geq \mathbf{E}(\log \mathbf{Y}).$$

を持つ。

5.2 統計の距離

p, q を集合 Y 上の 2 つの確率分布とする。p, q の「統計の距離」を $\mathbf{d}(p, q)$ と表し、

$$\mathbf{d}(p, q) = \frac{1}{2} \sum_{y \in Y} \mid p(y) - q(y) \mid .$$

と定義する。

すべての確率分布 p, q において、$0 \leq \mathbf{d}(p, q) \leq 1$ となる。他の要素のプロパティは以下の系で与えられる。

補題 5.1. p, q を集合 Y 上の 2 つの確率分布とする。そのとき、以下の式が成り立つ。

$$\sum_{y \in Y} \max\{p(y), q(y)\} = \mathbf{d}(p, q) + 1$$

証明. $Y_p = \{y \in Y : p(y) \geq q(y)\}$ とする。そのとき、

$$\begin{aligned}
\mathbf{d}(p, q) &= \frac{1}{2} \sum_{y \in Y_p} (p(y) - q(y)) + \frac{1}{2} \sum_{y \in Y \setminus Y_p} (q(y) - p(y)) \\
&= \frac{1}{2} \sum_{y \in Y_p} p(y) - \frac{1}{2} \sum_{y \in Y \setminus Y_p} p(y) - \frac{1}{2} \sum_{y \in Y_p} q(y) + \frac{1}{2} \sum_{y \in Y \setminus Y_p} q(y) \\
&= \frac{1}{2} \sum_{y \in Y_p} p(y) - \frac{1}{2} \left(1 - \sum_{y \in Y_p} p(y) \right) - \frac{1}{2} \sum_{y \in Y_p} q(y) + \frac{1}{2} \left(1 - \sum_{y \in Y_p} q(y) \right) \\
&= \sum_{y \in Y_p} p(y) - \sum_{y \in Y_p} q(y)
\end{aligned}$$

となる。一方、

$$\begin{aligned}
\sum_{y \in Y} \max\{p(y), q(y)\} &= \sum_{y \in Y_p} p(y) + \sum_{y \in Y \setminus Y_p} q(y) \\
&= \sum_{y \in Y_p} p(y) + 1 - \sum_{y \in Y_p} q(y) \qquad (5.1)
\end{aligned}$$

それゆえ、

$$\sum_{y \in Y} \max\{p(y), q(y)\} = \mathbf{d}(p, q) + 1$$

□

集合 Y 上の任意の確率分布を p とする。任意の $Y_0 \subseteq Y$ において、

$$p(Y_0) = \sum_{y \in Y_0} p(y).$$

と定義する。

補題 5.2. p, q を集合 Y 上の 2 つの確率分布とする。そのとき、以下の式が成り立つ。

$$\mathbf{d}(p,q) = max\{\mid p(Y_0) - q(Y_0) \mid : Y_0 \subseteq Y\}.$$

証明. 補題 5.1 の証明で用いた Y_p を定義する。すると、

$$\mid p(Y_p) - q(Y_p) \mid = \sum_{y \in Y_p} (p(y) - q(y)).$$

となる。

補題 5.1 の証明より、

$$\sum_{y \in Y_p} (p(y) - q(y)) = \mathbf{d}(p,q)$$

それゆえ、$\mid p(Y_p) - q(Y_p) \mid = \mathbf{d}(p,q)$。これから、

$$\mid q(Y \setminus Y_p) - p(Y \setminus Y_p) \mid = \mathbf{d}(p,q)$$

となる。

これからの証明により、すべての $Y_0 \subseteq Y$ において、$\mid p(Y_p) - q(Y_p) \mid \geq \mid p(Y_0) - q(Y_0) \mid$ を示す。$Y_0 \subseteq Y$ とし、$Y_1 = Y_0 \cap Y_p$、$Y_2 = Y_0 \cap (Y \setminus Y_p)$ と表す。$p(Y_1) - q(Y_1) > 0$、$p(Y_2) - q(Y_2) < 0$ に注意する。そのとき、

$$\begin{aligned} p(Y_0) - q(Y_0) &= p(Y_1) - q(Y_1) + (p(Y_2) - q(Y_2)) \\ &\leq p(Y_1) - q(Y_1) \\ &\leq p(Y_p) - q(Y_p) \qquad since Y_1 \subseteq Y_p \\ &= \mid p(Y_p) - q(Y_p) \mid . \end{aligned}$$

同様に、

$$\begin{aligned} q(Y_0) - p(Y_0) &= q(Y_2) - p(Y_2) + (q(Y_1) - p(Y_1)) \\ &\leq q(Y_2) - p(Y_2) \\ &\leq q(Y \setminus Y_p) - p(Y \setminus Y_p) \qquad since Y_2 \subseteq Y \setminus Y_p \\ &= \mid p(Y_p) - q(Y_p) \mid . \end{aligned}$$

それゆえ、

$$\mid p(Y_0) - q(Y_0) \mid \leq \mid p(Y_p) - q(Y_p) \mid .$$

が成り立つ。 □

例 5.1. 集合 $\{y_1, y_2, y_3, y_4\}$ 上の 2 つの確率分布を p, q を以下と考慮する。

	$p(y_i)$	$q(y_i)$
y_1	1/3	1/4
y_2	1/3	1/4
y_3	1/6	1/4
y_4	1/6	1/4

前述の 3 つの方法のうち任意の一つを使うと、$\mathbf{d}$(p,q) を計算することができる。距離の定義を使うと、

$$\mathbf{d}(p,q) = \frac{1}{2} \times 4\frac{1}{12} = \frac{1}{6}$$

と計算できる。

もし、補題 5.1 を使うと、

$$\mathbf{d}(p,q) = 2 \times \frac{1}{3} + 2 \times \frac{1}{4} - 1 = \frac{1}{6}.$$

最後に、補題 5.2 を使うと、

$$\mathbf{d}(p,q) = p(\{y_1, y_2) - q(\{y_1, y_2\}) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}.$$

それぞれの方法で同じ答えを得ることができた。

5.3 識別性

確率分布の「統計の距離」は識別性の考えと関係がある。p_0, p_1 を Y 上の確率分布とする。確率分布 q を集合 Y において、$i \in 0, 1$ を一様にランダムに選び、後に $y \in Y$ を選んだ時の確率 $p_i(y)$ と定義する。これは、

$$q(y) = \frac{p_0 + p_1(y)}{2}$$

と書ける。

「判別器」は関数 $f : Y \to 0, 1$ とする。直感的に、値 $y \in Y$ を上記の方法（すなわち、確率分布 q）によって選び与え、判別器は $i = 0$ か $i = 1$ かどうかを推測する。$corr(f)$ を判別器 f に y を与えたとき、i を正しく推測する確率と表す。判別器 f が $y \in Y$ を与えたときに正しい確率は

$$\frac{p_{f(y)}(y)}{p_0(y) + p_1(y)}$$

となる。これより、

$$corr(f) = \sum_{y \in Y} \frac{p_0(y) + p_1(y)}{2} \times \frac{p_{f(y)}(y)}{p_0(y) + p_1(y)} = \sum_{y \in Y} \frac{p_{f(y)}(y)}{2}.$$

となる。

それぞれの $y \in Y$ において、判別器は値 i を正しく推測する確率が最も高くなるようにする。つまり、

$$p_i(y) \geq y_{1-i}(y)$$

となるようにする。

それゆえ、「最善な判別器」を f^* と表すと、以下のように定義できる。

$$f^*(y) = \begin{cases} 0 & \text{if} p_1 < p_0(y) \\ 1 & \text{if} p_1 \geq p_0(y) \end{cases}$$

判別器 f^* が正しく推測する確率は、

$$corr(f^*) = \sum_{y \in Y} \frac{\max\{p_0(y), p_1(y)\}}{2} = \frac{\mathbf{d}(p_0, p_1) + 1}{2}$$

となる。補題 5.1 から最後の等式は成り立つ。

両方の確率空間 (Y, p_0) と (Y, p_1) 上の確率変数と、f と f^* をみなす事ができる。以下に、2 つの確率空間上の f, f^* の期待値から、p_0, p_1 の「統計の距離」の関係比較を簡単に証明することができる。

系 5.5. p_0, p_1 を集合 Y 上で定義される確率分布と仮定する。そのとき、任意の $f : Y \to 0, 1$ で以下の式が成り立つ。

$$\mid \mathbf{E}(f_{p_1}) - \mathbf{E}(f_{p_0}) \mid \leq \mathbf{E}(f^*_{p_1}) - \mathbf{E}(f^*_{p_1}) = \mathbf{d}(p_0, p_1)$$

証明. 最初に、$\mathbf{E}(f^*_{p_1}) - \mathbf{E}(f^*_{p_0}) = \mathbf{d}(p_0, p_1)$ を示す。

$$\begin{aligned} \mathbf{E}(f^*_{p_1}) &= \sum_{y \in Y} f^*(y) p_1(y) \\ &= \sum_{\{y \in Y : p_1(y) \geq p_1(y)\}} . \end{aligned}$$

同様に、

$$\mathbf{E}(f^*_{p_0}) = \sum_{\{y \in Y : p_1(y) \geq p_0(y)\}} .$$

その時、補題 5.1 の証明により、$\mathbf{E}(f^*_{p_1}) - \mathbf{E}(f^*_{p_0}) = \mathbf{d}(p_0, p_1)$ を示せる。

次に $\mid \mathbf{E}(f_{p_1}) - \mathbf{E}(f_{p_0}) \mid \leq \mathbf{d}(p_0, p_1)$ によって証明を示す。これは補題 5.2 を使うと、以下のようになる。

$$\mid \mathbf{E}(f_{p_1}) - \mathbf{E}(f_{p_0}) \mid \leq \mid p_1(f^{-1}(1)) - p_0(f^{-1}(1)) \mid = \mathbf{d}(p_0, p_1)$$

□

5.4 擬似ランダム性

(Y, p) を有限確率空間、$Y_0 \subseteq Y$、$\epsilon > 0$ の実数とすると、p は Y_0 に関係した ϵ 内の擬似乱数生成器を

$$\left| p(Y_0) - \frac{\mid Y_0 \mid}{\mid Y \mid} \right| \leq \epsilon.$$

の式で作れる。

さらに、p は ϵ 内の擬似乱数生成器をすべての $Y_0 \subseteq Y$ で

$$\left| p(Y_0) - \frac{\mid Y_0 \mid}{\mid Y \mid} \right| \leq \epsilon.$$

の式で作れる。

$u_Y(Y_0) = \mid Y_0 / \mid Y \mid$ という事実を使うと、以下に補題 5.2 の系を示せる。

補題 5.3. u_Y を Y 上の一様確率分布とする。そのとき任意の Y 上の確率分布 p は ϵ 内で擬似ランダムであることと、$\mathbf{d}(p, u_Y) \leq \epsilon$ が必要十分条件となる。

5.5 衝突確率

(Y,p) は確率空間とする。確率分布 p の衝突確率は、

$$\Delta_p = \sum_{y\in Y}(p(y))^2.$$

と定義する。

もし、$p = u_Y$ ならば、$\Delta_p 1/|Y|$ となることに注意する。衝突確率と確率分布の擬似ランダム性のあいだの関連性は Impagliazzo と Zuckerman[16] に起因して証明する。

補題 5.4. (Y,p) は確率空間とする。p は $\sqrt{\Delta_p|Y|-1}/2$ 内で擬似乱数生成器となる。

証明. $|Y| = M$ とする。$\Delta_p = \sum(p(y))^2$ を使うと、以下が導ける。

$$\sum_{y\in Y}\left(p(y)-\frac{1}{M}\right)^2 = \Delta_p - \frac{1}{M}$$

確率空間 (Y,u_Y) 上のランダム変数 $\mathbf{Y}$ を式 $\mathbf{Y}(y) = |p(y) - (1/M)|$ と定義する。その時、

$$\mathbf{E}(\mathbf{Y}) = \frac{1}{M}\left(\Delta_p - \frac{1}{M}\right) = \frac{\Delta_p M - 1}{M^2}.$$

となる。

系 5.3 を適用すると、

$$\mathbf{E}(\mathbf{Y}) \le \sqrt{\mathbf{E}(\mathbf{Y})} = \frac{\sqrt{\Delta_p M - 1}}{M}.$$

となる。

よって、

$$\mathbf{d}(p,u_Y) = \frac{1}{2}\sum_{y\in Y}\left|p(y)-\frac{1}{M}\right| = \frac{M}{2}\times\mathbf{E}(\mathbf{Y}) \le \frac{\sqrt{\Delta_p M-1}}{2}.$$

が導ける。 □

5.6 Shannon,Renyi,Min エントロピー

(Y,p) を確立空間とする。(Y,p) の Renyi entropy を $h_{\mathrm{Ren}}(p)$ と表し、

$$h_{\mathrm{Ren}}(p) = -\log_2 \Delta_p.$$

と定義する。

(Y,p) の min entropy を $h_{\min}(p)$ と表し、

$$h_{\min}(p) = \min\{-\log_2 p(y) : y\in Y\} = -\log_2(\max\{p(y) : y\in Y\}).$$

と定義する。

(Y,p) の Shannon entropy を $h(p)$ と表し、

$$h(p) = -\sum_{y\in Y} p(y)\log_2 p(y).$$

と定義する。一様分布 u_Y は $h(u_Y) = h_{\mathrm{Ren}}(u_Y) = h_{\min}(u_Y) = \log_2|Y|$ となることに注意する。

以下の補題は簡単に証明できる。

補題 5.5. (Y,p) を確立空間とする。そのとき、$h_{\mathrm{Ren}}(p)/2 \le h_{\min}(p) \le h_{\mathrm{Ren}}(p) \le h(p)$ となる。

証明. まず、

$$(\max\{p(y) : y\in Y\})^2 \le \sum (p(y))^2.$$

から、$h_{\mathrm{Ren}}(p)/2 \le h_{\min}(p)$ となる。

次に、

$$\sum (p(y))^2 \le \sum (p(y)\times \max\{p(y) : y\in Y\}) = \max\{p(y) : y\in Y\}.$$

から、$h_{\min}(p) \le h_{\mathrm{Ren}}(p)$ となる。

最後に、確率空間 (Y,p) 上の確率変数 $\mathbf{Y}$ を $\mathbf{Y}(y) = p(y)$ と定義する。
$E(\mathbf{Y}) = \sum (p(y))^2$ と、$E(\log \mathbf{Y}) = \sum p(y)\log p(y)$ になることに注意する。
今、系 5.4 を適用すると、以下の式が得られる。

$$\log\Big(\sum (p(y))^2\Big) \ge \sum p(y)\log p(y).$$

よって、$h_{\mathrm{Ren}}(p) \le h(p)$ となる。 □

§4 で示されたアプリケーションに関する文献によるいくつかの結果は、これらの min entropy か Renyi entropy の特定の条件のために確率分布に作用する。上記の補題の意味は、これらの 2 つの数は高々2 つの因数によって異なるので、基本的に（定数項次第で）交互に用いることができるということである。この論文では一般的に、衝突確率を単位として成果を述べる。

6 SU ハッシュ族の擬似ランダム性

この章では、**SU** ハッシュ族の擬似ランダム性について述べ、いくつかの成果を証明する。[5] からいくつかの定理の少し単純な処理を提供する。私たちのアプローチは [13] で使われているものに似ている。

(X,p) を有限確率空間、$\mathcal{F}$ を任意の X から Y への関数の $\mathbf{SU}(D;N,M)$ ハッシュ族とする。任意の $f\in\mathcal{F}$ において、Y 上に展開される q_f の上の帰納確率分布を、すべての $y\in Y$ において、

$$q_f(y) = \sum_{x\in f^{-1}(y)} p(x)$$

と定義する。

q_f は関数 f の出力が値 y を取る確率で、$x \in X$ は確率分布 p を使って選び与えられる。

任意の $y \in Y$ において、確率空間 $(\mathcal{F}, u_{\mathcal{F}})$ 上の確率変数 χ_y は、すべての $f \in \mathcal{F}$ において、

$$\chi_y(f) = q_f(y)$$

と定義する。

これより、

$$\sum_{f \in \mathcal{F}} \chi_y(f) = \frac{D}{M}.$$

となる。

故に、

$$\mathbf{E}(\chi_y) = \frac{1}{M}. \tag{6.1}$$

今、以下に重要な組合わせの補題を証明する。

補題 6.1. (X, p) は有限確率空間、$\mathcal{F}$ を任意の X から Y への関数の $\mathbf{SU}(D; N, M)$ ハッシュ族とする。$y \in Y$ とし、$\mathcal{F}$ 上の確率変数 χ_y を上記のように定義する。すると、

$$\sum_{f \in \mathcal{F}} (\chi_y(f))^2 = \frac{D(1 + (M-1)\Delta_p)}{M^2}.$$

となる。

証明.

$$\begin{aligned}
\sum_{f \in \mathcal{F}} (\chi_y(f))^2 &= \sum_{f \in \mathcal{F}} \left(\sum_{x \in f^{-1}(y)} p(x) \right)^2 \\
&= \sum_{f \in \mathcal{F}} \sum_{x_1 \in f^{-1}(y)} \sum_{x_2 \in f^{-1}(y), x_2 \neq x_1} p(x_1)p(x_2) + \sum_{f \in \mathcal{F}} \sum_{x_1 \in f^{-1}(y)} (p(x))^2 \\
&= \frac{D}{M^2} \sum_{x_1 \in X} \sum_{x_2 \in X, x_2 \neq x_1} p(x_1)p(x_2) + \frac{D}{M} \sum_{x \in X} (p(x))^2 \\
&= \left(\frac{D}{M^2} \right)(1 - \Delta_p) + \left(\frac{D}{M} \right)\Delta_p \\
&= \frac{D(1 + (M-1)\Delta_p)}{M^2}.
\end{aligned}$$

□

例 6.1. 特定の確率分布 p で与えられる X をもつ、$\mathbf{SU}(9; 3, 3)$ ハッシュ族を示す。以下のテーブルの最後の列は、χ_0 の値である。

	$p(0)=1/2$	$p(1)=1/4$	$p(2)=1/4$	χ_0
$f_{0,0}$	0	1	1	$\frac{1}{2}$
$f_{0,1}$	1	2	2	0
$f_{0,2}$	2	0	0	$\frac{1}{2}$
$f_{1,0}$	0	1	0	$\frac{1}{4}$
$f_{1,1}$	1	2	1	0
$f_{1,2}$	2	0	2	$\frac{3}{4}$
$f_{2,0}$	0	1	1	$\frac{1}{4}$
$f_{2,1}$	1	2	2	0
$f_{2,2}$	2	0	0	$\frac{3}{4}$

そのとき、補題 6.1 により、

$$\sum(\chi_0(f_{a,b}))^2 = \frac{7}{4} = \frac{9\Big(1+(3-1)\Big((\frac{1}{2})^2+\frac{1}{4})^2+\frac{1}{4})^2\Big)\Big)}{3^2}$$

を導くことが出来る。 ▮

補題 6.1 から、

$$\mathbf{E}({\chi_y}^2) = \frac{1+(M-1)\Delta_p}{M^2}. \tag{6.2}$$

今、(6.1) と (6.2) を用いて、以下の式を得ることが出来る。

$$\mathbf{var}(\chi_y) = \mathbf{E}({\chi_y}^2) - (\mathbf{E}(\chi_y))^2 = \frac{(M-1)\Delta_p}{M^2}.$$

そのとき、Chebyshev' の不等式 (系 5.1) を適用して、

$$\mathbf{Pr}[|\chi_y(f) - \mathbf{E}(\chi_y)| \geq \epsilon] \leq \frac{(M-1)\Delta_p}{\epsilon^2 M^2}.$$

となり、最後に、

$$|\chi_y(f) - \mathbf{E}(\chi_y)| = \left| q_f(y) - \frac{1}{M} \right|.$$

それゆえ、q_f が y に関して疑似ランダムであることと、

$$|\chi_y(f) - \mathbf{E}(\chi_y)| \leq \epsilon$$

は必要十分条件となる。[13] のわずかな一般化と [14] に証明されている。

定理 6.1. (X,p) を有限確率空間、$\mathcal{F}$ を任意の X から Y への関数の $\mathbf{SU}(D;N,M)$ ハッシュ族とする。$y \in Y$ は固定で、$f \in \mathcal{F}$ をランダムに選ぶ。そのとき、q_f が ϵ 内で y に関して疑似ランダムにならない確率は高々

$$\frac{(M-1)\Delta_p}{\epsilon^2 M^2}.$$

となる。

前の結果は、任意の $Y_0 \subseteq Y$ に関しての疑似ランダム性へ一般化できる。以下は似たような方法で簡単に証明できる。

定理 6.2. (X,p) を有限確率空間、$\mathcal{F}$ を任意の X から Y への関数の $\mathbf{SU}(D;N,M)$ ハッシュ族とする。$Y_0 \in Y$ は固定で、$f \in \mathcal{F}$ をランダムに選ぶ。そのとき、q_f が ϵ 内で Y_0 に関して疑似ランダムにならない確率は高々

$$\frac{|Y_0|(M-|Y_0|)\Delta_p}{\epsilon^2 M^2}.$$

となる。

定理 6.3. (X,p) を有限確率空間、$\mathcal{F}$ を任意の X から Y への関数の $\mathbf{SU}(D;N,M)$ ハッシュ族とする。そして、$f \in \mathcal{F}$ をランダムに選ぶ。そのとき、q_f が ϵ 内で疑似ランダムにならない確率は高々

$$\frac{\Delta_p M(M-1)}{4\epsilon^2}$$

となる。

証明. 初めに、もしすべての $y \in Y$ において

$$\left| q_f(y) - \frac{1}{M} \right| < \frac{2\epsilon}{M}$$

ならば、$\mathbf{d}(q_f, u_Y) \leq \epsilon$ かつ、補題 5.3 より q_f は ϵ 内で疑似ランダムとなる。$f \in \mathcal{F}$ をランダムに選ぶとする。任意の $y \in Y$ において、

$$\left| q_f(y) - \frac{1}{M} \right| < \frac{2\epsilon}{M}$$

となる確率は高々

$$\frac{\Delta_p(M-1)}{(2\epsilon/M)^2 M^2} = \frac{\Delta_p(M-1)}{4\epsilon^2}$$

となる。

よって、$y \in Y$ に対して M の選択肢があり、ある $y \in Y$ において $|q_f(y) - \frac{1}{M}| > 2\epsilon/M$ となる確率は高々 $\Delta_p M(M-1)/4\epsilon$ となる。 □

7 Leftover Hash Lemma

この章では、いわゆる「leftover hash lemma」の一般形を記し、証明する。この初期の型は [17] で証明されている。[3][13][16][11] もまた詳しく関係した結果となっている。ここで、集合 $\mathcal{F} \times \mathcal{Y}$ を $f \in \mathcal{F}$ をランダムに選び、それから確率分布 p を使って $x \in X$ を選んだ時に $f(x)$ の値を求めることで定義され、その確率分布を r とする。leftover hash lemma は、確率分布 r の疑似ランダム性と関係がある。それゆえ、r は以下で定義される。

$$r(f,y) = \frac{q_f(y)}{D} = \frac{\chi_y(f)}{D}.$$

以下の補題は、$\mathcal{F}$ が δ-$\mathbf{U}$ ハッシュ族という弱い仮定の下、補題 6.1 と似た結果を証明する。

補題 7.1. (X,p) は有限確率空間、$\mathcal{F}$ を任意の X から Y への関数の δ-$\mathbf{U}(D;N,M)$ ハッシュ族とする。すべての $y\in Y$ と $f\in\mathcal{F}$ において、$\chi_y(f)=q_f(y)$ とすると、以下の式が成り立つ。

$$\sum_{y\in Y}\sum_{f\in\mathcal{F}}(\chi_y(y))^2\le D(\delta+(1-\delta)\Delta_p).$$

例 7.1. 特定の確率分布 p で与えられる X をもつ、$\frac{1}{2}$-$(4;4,2)$ ハッシュ族を示す。以下のテーブルの最後の列は、χ_0 と χ_1 の値である。

	$p(0)=1/2$	$p(1)=1/6$	$p(2)=1/6$	$p(3)=1/6$	χ_0	χ_1
f_1	0	0	1	1	$\frac{2}{3}$	$\frac{1}{3}$
f_2	0	0	0	0	1	0
f_3	0	1	1	0	$\frac{2}{3}$	$\frac{1}{3}$
f_4	0	1	0	1	$\frac{2}{3}$	$\frac{1}{3}$

そのとき、

$$\sum_{i=0}^{1}\sum_{j=0}^{4}(\chi_i(f_j))^2=\frac{8}{3}=4\left(\frac{1}{2}+\left(1-\frac{1}{2}\right)\left(\left(\frac{1}{2}\right)^2+\left(\frac{1}{6}\right)^2+\left(\frac{1}{6}\right)^2+\left(\frac{1}{2}\right)^2\right)\right)$$

これより、補題 7.1 の上限は等式を満たす。 ▌

今、補題 7.1 から、この章の主な成果を述べる。

定理 7.1. $\mathcal{F}$ を任意の X から Y への関数の δ-$\mathbf{U}(D;N,M)$ ハッシュ族とする。X 上の確率分布を p と仮定し、確率分布 r を上記で定義したように $\mathcal{F}\times Y$ 上の帰納とする。その時、以下の式が成り立つ。

$$\Delta_r\le\frac{\delta+(1-\delta)\Delta_p}{D}.$$

系 7.1. $\mathcal{F}$ を任意の X から Y への関数の δ-$\mathbf{U}(D;N,M)$ ハッシュ族とする。X 上の確率分布を p と仮定し、確率分布 r を上記で定義したように $\mathcal{F}\times Y$ 上の帰納とする。その時、以下の式が成り立つ。

$$\mathbf{d}(u_{\mathcal{F}\times Y},r)\le\frac{\sqrt{M(\delta+(1-\delta)\Delta)-1}}{2}.$$

証明. 補題 5.4 と、定理 7.1 を適用する。 □

8 Extractors

定義から始める。$\mathcal{F}$ を任意の X から Y への関数の δ-$\mathbf{U}(D;N,M)$ ハッシュ族とする。X 上の確率分布を p と仮定し、確率分布 r を §7 で定義したように $\mathcal{F}\times Y$ 上の帰納とする。もし $\mathbf{d}(u_{\mathcal{F}\times Y},r)<\epsilon$ で $h_{\mathbf{Ren}}(p)\ge k$ のときはいつも、$\mathcal{F}$ は (k,ϵ)-*extractor* という。以下の結果は与えられたハッシュ族が extractor となるための十分条件を提供する。

定理 8.1. 以下の式を満たす時、δ-$\mathbf{U}(D;N,M)$ ハッシュ族は (k,ϵ)-*extractor* となる。

$$\sqrt{M(\delta+2^{-k})-1} \leq 2\epsilon$$

証明. $\Delta_p = 2^{-k}$ とし、系 7.1 を適用する。そのとき、

$$\mathbf{d}(u_{\mathcal{F}\times Y,r}) \leq \frac{\sqrt{M(\delta+(1-\delta)2^{-k})-1}}{2} < \frac{\sqrt{M(\delta+2^{-k})-1}}{2} \leq \epsilon \qquad \square$$

今、§4.2 で示されたテクニックを使い $(k,1/4)$-*extractor* は **BBP** アルゴリズムのエラー確率を $2^k/N$ に減らせる証明をする。$\mathcal{F}$ を任意の X から Y への関数の δ-$\mathbf{U}(D;N,M)$ ハッシュ族とする。さらに、集合 Y からランダムに選んだ値に依存してそれぞれアルゴリズムが走る、クラス **BBP** 内にあるランダマイズアルゴリズム $\mathbf{A}$ をもつと仮定する。I は任意の problem instance とする。ランダムに要素 $x \in X$ を選ぶ。そして、アルゴリズム $\mathbf{A}(I,f(x))$ を全ての $f \in \mathcal{F}$ で走らせる。$\mathbf{B}(I,x)$ は yes か no で多かったほうを出力すると定義する。以下の結果は、アルゴリズム $\mathbf{B}$ のエラー確率に関係し、[12] からなる。

（以下略）

参考文献

[1] A.S.Hedayat, N.J.A. Sloane, and J.Stufken. Orthogonal arrays:theory and applications. *Springer-Verlag*, 1999.

[2] C.H.Bennett and G.Brassard adn J-M.Robert. Privacy amplification by public discussion. *SIMA Journal on Computing*, No. 17, pp. 210–229, 1988.

[3] C.H.Bennett, G.Brassard, C.Crepeau, and U.Maurer. Generalized privacy amplification. *IEEE Transactions on Information Theory*, No. 41, pp. 1915–1923, 1995.

[4] C.R.Rao. Factorial experiments derivable from combinatorial arrangements of arrays. *Journal of the Royal Statistical Society*, No. 9, pp. 128–139, 1947.

[5] D.R.Stinson. Combinatorial techniques for universal hashing. *Journal of Computer and System Sciences*, No. 48, pp. 337–346, 1994.

[6] D.R.Stinson. On the connections between universal hashing, combinatorial designs and error-correcting codes. *Congressus Numerantium*, No. 114, pp. 7–27, 1996.

[7] J.Bierbrauer, T.Johansson, G.Kabatianskill, and B.Smeets. On families of hash functions via geometric codes and concatenation. *Lecture Notes in Computer Science*, No. 773, pp. 311–342, 1994.

[8] J.L.Carter and M.N.Wegman. Universal classes of hash functions. *Jounal of Computer and System Sciences*, No. 18, pp. 143–154, 1979.

[9] F.J. MacWilliams and N.J.A. Sloane. The theory of error-correctiong codes. *North-Holland*, 1977.

[10] M.Etzel, S.Patel, and Z.Ramzan. Square hash:fast message authenticaiton via optimized universal hash functions. *Lecture Notes in Computer Science*, Vol. CRYPTO '99, No. 1666, pp. 234–251, 1999.

[11] M.Luby. Pseudorandomness and cryptographic applications. *Princeton University Press*, 1996.

[12] N.Nisan and A.Ta-Shma. Extracting randomness:a survey and new constructions. *J. Comput. System Sci*, No. 58, pp. 148–173, 1999.

[13] O.Goldreich. Modern cryptography, probabilistic proofs and pseudorandomness.

[14] P.Nguyen and J.Stern. The hardness of the hidden subset sum problem and its cryptographic application. *Lecture Notes in Computer Science*, Vol. CRYPTO '99, No. 1666, pp. 31–46, 1999.

[15] R.C.Bose and K.A.Bush. Orthogonal arrays of strength two and three. *Annals Math.Statistics*, No. 23, pp. 508–524, 1952.

[16] R.Impagliazzo and D.Zuckerman. How to recycle random bits. *In 30th IEEE Symposium on Foundations of Computer Science*, pp. 12–24, 1989.

[17] R.Impagliazzo, L.Levin, and M.Luby. Pseudo-random generation from one-way functions. *In 21st ACM Symposium on Theory of Computing*, pp. 12–24, 1989.

[18] V.Boyko, M.Peinado, and R.Venkatesan. Speeding up discrete log and factoring based schemes via precomputation. *Lecture Notes in Computer Science*, Vol. EUROCRYPT '98, No. 1403, pp. 221–235, 1998.