

# 分布は点である Distributions are points

ryamada

2017年2月24日

- 1 分布は点である Distributions are points
  - 1.1 簡単な例 二項分布 Basic example; binomial distribution
  - 1.2 簡単な例 2 三項分布 Basic example 2; Three-nomial distribution
  - 1.3 連続分布の場合 Continuous distributions
- 2 分布をパラメタで表す 分布が多様体の一部になる Parameterize distributions, distribution corresponds to a part of larger manifold
  - 2.1 二項分布の場合 Binomial distribution
  - 2.2 三項分布の場合 Three-nomial distribution
    - 2.2.1 部分多様体 A submanifold, apart of manifold
  - 2.3 正規分布の場合 Normal distribution
- 3 分布の違いを定量する～距離/divergenceを測る Quantitate dissimilarity among distributions with distance/divergence.
  - 3.1 単純な差 Simple “difference”
  - 3.2 二乗してみる Sum of squares
  - 3.3 球面に乗せてみる Put distributions on a sphere surface
    - 3.3.1 弧の長さを測る Measuring arc length
      - 3.3.1.1 角度で測る Angle
      - 3.3.1.2 内積を使う Inner product is also useful
      - 3.3.1.3 積分する Integral
    - 3.3.2 球面上の座標とその伸び縮み Coordinates on the sphere and local elongation/shrinkage
- 4 まとめ：球面上の座標 球面上の距離の二乗 座標変換 微分・積分 Summary: Coordinates on sphere, square of distance on the sphere, coordinate transformation, differentiation/integration.

## 1 分布は点である Distributions are points

情報幾何では、分布を空間の点とみなし、分布同士が遠いか近いかなどを問題にする。

Information geometry considers each distribution as a point in a space and measures similarity among multiple distributions with diversity, similar to distance, in the space

### 1.1 簡単な例 二項分布 Basic example; binomial distribution

確率分布は次のように表される。Its distribution is;

$$\begin{aligned}
 P &= \{p_1, p_2\} \\
 p_1 + p_2 &= 1 \\
 p_1, p_2 &\geq 0
 \end{aligned}$$

$p_1, p_2$  の値はいろいろな値をとるので、このままでは、特定の分布に確定しない。

The values  $p_1, p_2$  vary and the above definition does not corresponds to any particular distribution.

$P = \{0.3, 0.7\}$  と具体的に定めると、特定の分布として確定する。

The specification  $P = \{0.3, 0.7\}$  makes it one particular distribution.

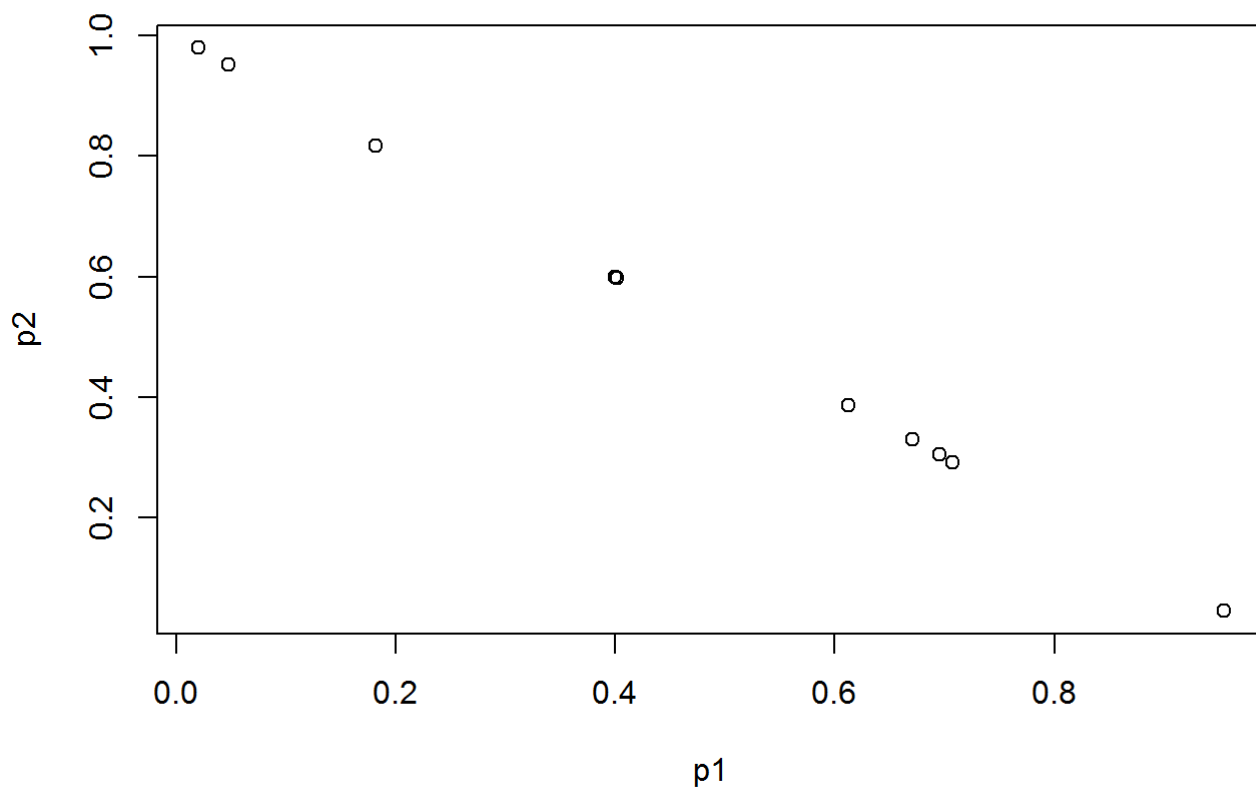
この  $P = \{0.3, 0.7\}$  のような具体的な分布を点とみる。

Information geometry treats these particular distribution instances as points.

実際この分布を  $(x_1 = 0.3, x_2 = 0.7)$  という二次元空間の点に対応させることができる。

We can give two dimensional coordinate  $(x_1 = 0.3, x_2 = 0.7)$  to the distribution,

```
n <- 10
p1 <- runif(n)
p2 <- 1-p1
plot(p1, p2)
```



点の座標は二次元空間に置かれるが、 $p_1 + p_2 = 1$  という制約があるため、線分上に並ぶ。

Multiple distribution instances are plot in a straight line in 2d space.  $p_1 + p_2 = 1$  is the restriction making them in the line.

$p_1$  の値が決まると  $p_2$  の値は決まってしまうので、自由度が1。

Due to the restriction,  $p_2$  value is fixed when  $p_1$  value is given. This is called “df = 1”.

2次元空間に置かれた、1次元多様体。

They make one-dimensional manifold in 2d space.

- 二項分布は2次元空間の点の集まりであって、2-1次元多様体になっている。Binomial distribution is a set of points in 2d space and the set is a 2-1 dimensional manifold.

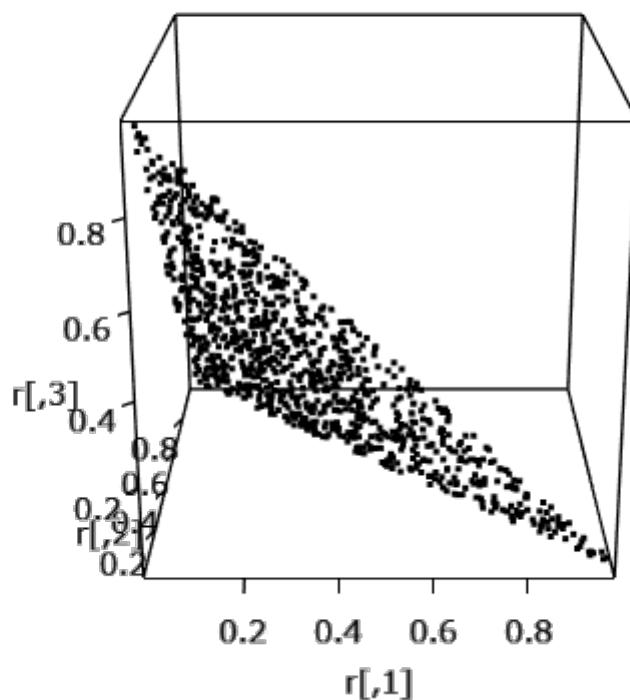
## 1.2 簡単な例 2 三項分布 Basic example 2; Three-nomial distribution

$$P = \{p_1, p_2, p_3\}$$

$$p_1 + p_2 + p_3 = 1$$

$$p_1, p_2, p_3 \geq 0$$

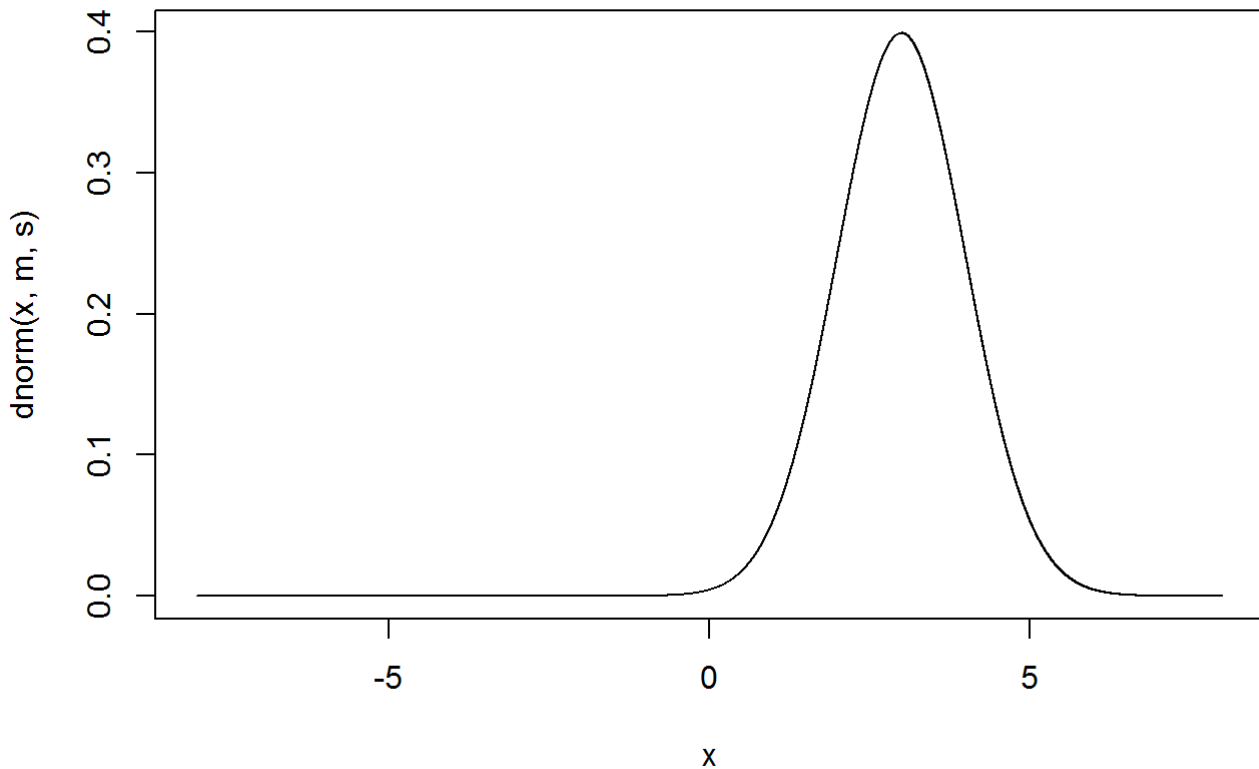
```
n <- 1000
r <- rdirichlet(n, rep(1, 3))
plot3d(r)
```



- 三項分布は3次元空間の点の集まりであって、3-1次元多様体になっている。The instances of three-nomial distribution is a set of points in 3d space and the set is a 2d manifold.
- k項分布はk次元空間の点の集まりであって、k-1次元多様体になっている。k-nomial distribution makes a k-1 dimensional manifold in k dimensional space.

## 1.3 連続分布の場合 Continuous distributions

```
m <- 3
s <- 1
x <- seq(from=-8, to=8, length=10000)
plot(x, dnorm(x, m, s), type="l")
```



- $k$ の値をどんどん大きくしていった無限大にすると、実数全体を台とする正規分布も  
 $P = \{p_{-\infty}, p_{-\infty+\delta}, p_{-\infty+2\delta}, \dots, p_{-\delta}, p_0, p_\delta, \dots, p_\infty\}$  と考えて、 $k = \infty$ 次元空間の点の集まりであって、 $k - 1 = \infty - 1$ 次元多様体(の一部)になっているとみなせる。The value  $k$  can be increased and it gets  $\infty$  eventually; for example all values in a real number line for normal distribution, where  $P = \{p_{-\infty}, p_{-\infty+\delta}, p_{-\infty+2\delta}, \dots, p_{-\delta}, p_0, p_\delta, \dots, p_\infty\}$ . These make a fraction of  $k - 1 = \infty - 1$  dimensional manifold in  $k = \infty$  dimensional space.

## 2 分布をパラメタで表す 分布が多様体の一部になる Parameterize distributions, distribution corresponds to a part of larger manifold

### 2.1 二項分布の場合 Binomial distribution

- $P = \{p_1, p_2\}$ を  $\theta = p_1$  というひとつのパラメタで表すこともできるし  $P = \{p_1, p_2\}$  can be specified with one parameter  $\theta = p_1$ .
- $\theta = p_1 - p_2$  という別の表し方もできる。Another one parameter  $\theta = p_1 - p_2$  can be used instead.
- どちらの表し方でも1つのパラメタを用いて、二項分布全体を表せている。Either case, one parameter expresses the whole binomial distribution.

### 2.2 三項分布の場合 Three-nomial distribution

- $P = \{p_1, p_2, p_3\}$  を  $\theta_1 = p_1, \theta_2 = p_2$  という2つのパラメタで表すこともできるし、  
 $P = \{p_1, p_2, p_3\}$  is expressed with two parameters,  $\theta_1 = p_1, \theta_2 = p_2$
- $\theta_1 = p_1 - p_3, \theta_2 = p_2 - p_3$  という別の表し方もできる。  $\theta_1 = p_1 - p_3, \theta_2 = p_2 - p_3$  can be used.
- どちらの表し方でも2つのパラメタを用いて、三項分布全体を表せている。 Either case, one parameter expresses the whole three-nomial distribution.

## 2.2.1 部分多様体 A submanifold, apart of manifold

- 今、  $\theta_1 = p_1 - p_3, \theta_2 = p_2 - p_3 = 0$  とすると、  $\theta_2 = 0$  は固定値になるから、  $\theta_1$  という1つのパラメタのみが有効となる。 Assume we have distribution instances that meet  $\theta_1 = p_1 - p_3, \theta_2 = p_2 - p_3 = 0$ . Because  $\theta_2 = 0$  is not essentially a parameter, they are expressed with one parameter  $\theta_1 = 0$ .
- この場合は以下に示すように、3次元空間の直線(1次元多様体)になる。 This represents a straight line (1d manifold) in 3d space.
- この直線は、3次元空間にある三角形(2次元多様体)に含まれている。 The line is a subset of a triangle (2d-manifold).
- 1個のパラメタが表す分布は、k次元空間のk-1次元多様体である分布の部分としての1次元多様体となる。 The instances expressed with one parameter corresponds to a submanifold (1d-manifold) as a subset of k-1 manifold in k-dimensional space.
- v個のパラメタが表す分布は、k次元空間のk-1次元多様体である分布の部分としてのv次元多様体となる。 ( $v \leq k - 1$ ) The instances of distributions expressed with v parameters, it is a v dimensional manifold that is a subset of k-1 dimensional manifold in k dimensional space.

```
n <- 100
theta1 <- runif(n) * 2 -1
p1 <- (1+2*theta1)/3
p3 <- p1-theta1
p2 <- p3

plot3d(p1, p2, p3)
```

## 2.3 正規分布の場合 Normal distribution

- 正規分布は、  $P = \{p_{-\infty}, p_{-\infty+\delta}, p_{-\infty+2\delta}, \dots, p_{-\delta}, p_0, p_\delta, \dots, p_\infty\}$  と考えて、  $\infty - 1$  次元多様体の点の集まりの一部である。 Normal distribution is a part of  $\infty - 1$  dimensional manifold in  $\infty$  dimensional space because it is expressed as  $P = \{p_{-\infty}, p_{-\infty+\delta}, p_{-\infty+2\delta}, \dots, p_{-\delta}, p_0, p_\delta, \dots, p_\infty\}$ .
- また正規分布は2個のパラメタ  $m, s$  で表される分布でもある。 Normal distribution is expressed with two parameters  $m, s$ .

$$N(m, s) = \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-m)^2}{2s^2}}$$

- したがって、正規分布は  $\infty$  次元空間の  $\infty - 1$  多様体である分布の一部となっていて、それは2次元多様体となっている、こととなる。 Therefore normal distribution is 2d manifold that is a submanifold of  $\infty - 1$  dimensional manifold in  $\infty$  dimensional space.

# 3 分布の違いを定量する～距離/divergenceを測る Quantitate dissimilarity among distributions with distance/divergence.

## 3.1 単純な差 Simple “difference”

- 分布A  $P_a = \{p_1^a, p_2^a\}$  と分布B  $P_b = \{p_1^b, p_2^b\}$  との違いを数値で表したい。Assume we want to quantitate difference distribution A  $P_a = \{p_1^a, p_2^a\}$  and distribution B  $P_b = \{p_1^b, p_2^b\}$ .

$$\sum_i p_i^a - p_i^b$$

- これは、各要素の差の足し合わせなので、『差』を表しているようだが、 $\sum_i p_i = 1$ なので、さしひき0になる。This is a sum of difference of all elements but it is not useful because  $\sum_i p_i = 1$  makes the sum of difference being 0, all the time.

```
Pa <- c(0.3, 0.7)
```

```
Pb <- c(0.4, 0.6)
```

```
sum(Pa-Pb)
```

```
## [1] -5.551115e-17
```

- それは連続分布でも同様。It is true for continuous distributions.

```
ma <- 0
```

```
sa <- 1
```

```
mb <- 3
```

```
sb <- 2
```

```
x <- seq(from=-20, to=20, length=10000)
```

```
Pa <- dnorm(x, ma, sa)
```

```
Pb <- dnorm(x, mb, sb)
```

```
sum(Pa-Pb)
```

```
## [1] 7.818129e-15
```

## 3.2 二乗してみる Sum of squares

- 単純な差では、0になってしまうから、二乗を取ってみる。The simple sum of difference of elements was 0. How about the sum of squares?

$$\sum_i (p_i^a - p_i^b)^2$$

```
Pa <- c(0.3, 0.7)
```

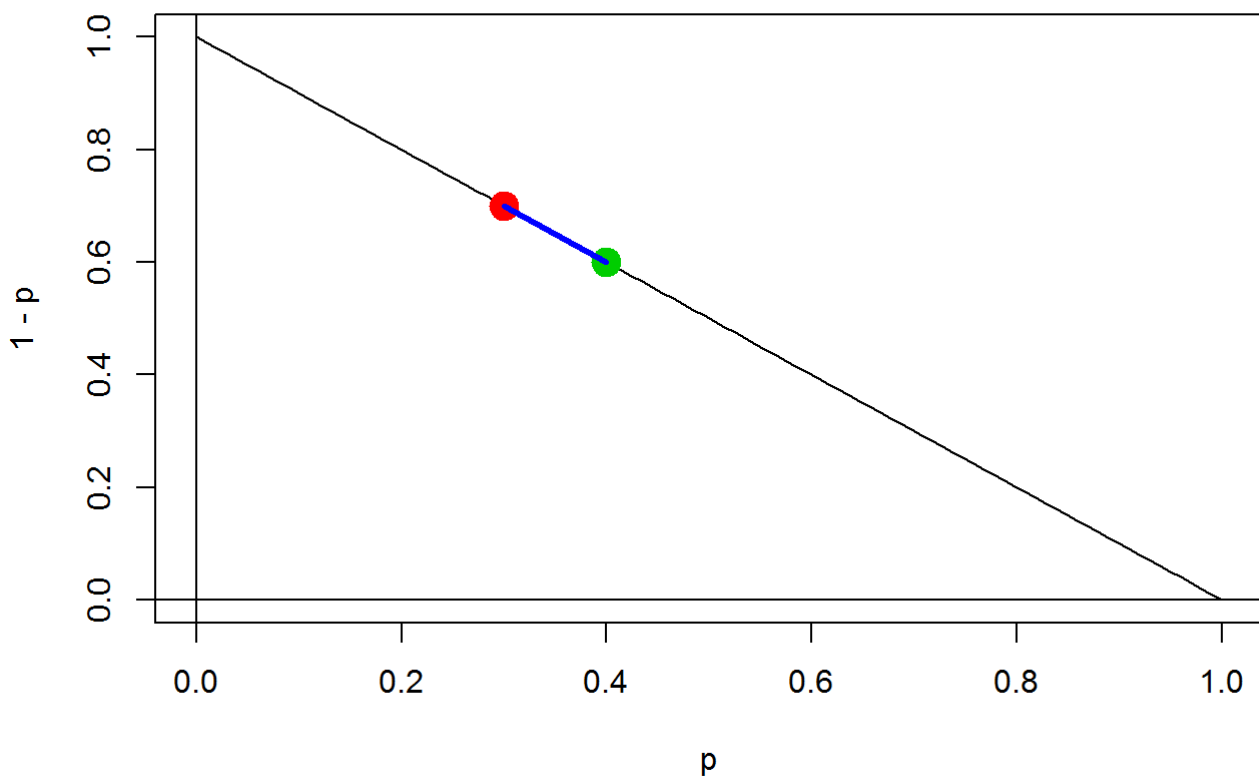
```
Pb <- c(0.4, 0.6)
```

```
sum((Pa-Pb)^2)
```

```
## [1] 0.02
```

- それらしい値になった。It returns non-zero, which might be good.
- これは何かと言えば、 $p_1 + p_2 = 1$ という斜辺に乗った、2つの点の斜辺上の距離の二乗。This square of distance along the line in which distributions are located.

```
p <- seq(from=0, to=1, length=100)
plot(p, 1-p, type="l")
points(Pa[1], Pa[2], pch=20, cex=3, col=2)
points(Pb[1], Pb[2], pch=20, cex=3, col=3)
segments(Pa[1], Pa[2], Pb[1], Pb[2], col=4, lwd=3)
abline(h=0)
abline(v=0)
```



### 3.3 球面に乗せてみる Put distributions on a sphere surface

- $\sum_i p_i = 1$ という平らな構造(直線、平面三角形、四面体...)を球面(円、いわゆる球面、高次元球面)へと乗せかえる。Points in a flat structure  $\sum_i p_i = 1$  (a straight segment, a triangle, a tetrahedron) can be placed in a sphere surface (a arc of circle, a part of sphere surface, a part of higher dimensional sphere).
- どうして、そのようなことをするかと言えば、後にフィッシャー情報量を扱うときに登場するからである。Why do we want to do that? Because it is used to discuss Fisher's information.

$$\sum_i p_i = 1$$

なので Therefore,

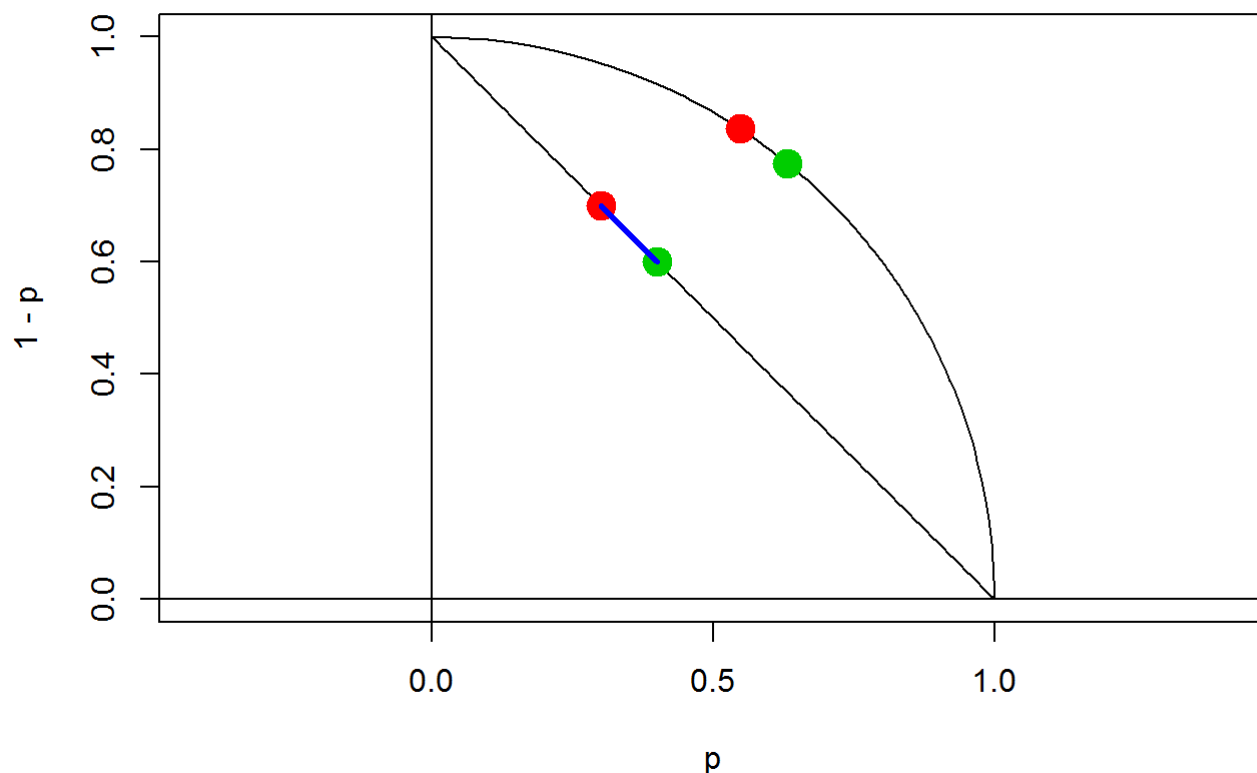
$$\sum_i q_i^2 = 1$$

$$q_i = \sqrt{p_i}$$

であり、これは、 $Q = \sqrt{P} = \{q_1, \dots\}$  が単位球面上の点であることを表す。This is an expression of unit sphere surface.

```
Pa <- c(0.3, 0.7)
Pb <- c(0.4, 0.6)
p <- seq(from=0, to=1, length=100)
plot(p, 1-p, type="l", asp=TRUE)
points(Pa[1], Pa[2], pch=20, cex=3, col=2)
points(Pb[1], Pb[2], pch=20, cex=3, col=3)
segments(Pa[1], Pa[2], Pb[1], Pb[2], col=4, lwd=3)

Qa <- sqrt(Pa)
Qb <- sqrt(Pb)
points(sqrt(p), sqrt(1-p), type="l")
points(Qa[1], Qa[2], pch=20, cex=3, col=2)
points(Qb[1], Qb[2], pch=20, cex=3, col=3)
abline(h=0)
abline(v=0)
```



- 球面に乗せたので、この2点間の距離(2つの分布間の距離)は、弧の長さを測ればよい。Because distributions are now on a sphere surface, we may measure the distance between two points and the distance in the sphere surface world is an arc.



### 3.3.1 弧の長さを測る Measuring arc length

#### 3.3.1.1 角度で測る Angle

$$\begin{aligned} L(B \rightarrow A) &= \theta \\ &= \int_{\theta_b}^{\theta_a} d\theta \\ &= \theta_a - \theta_b \end{aligned}$$

```
theta.a = acos(Qa[1])
theta.b = acos(Qb[1])
theta.a - theta.b
```

```
## [1] 0.1050795
```

#### 3.3.1.2 内積を使う Inner product is also useful

$$\begin{aligned} L(B \rightarrow A) &= \theta \\ \cos \theta &= \frac{(v_1, v_2)}{|v_1||v_2|} = (v_1, v_2) \end{aligned}$$

ただし、 $(v_1, v_2)$  は内積。wehre  $(v_1, v_2)$  stands for inner product.

```
cos.theta = sum(Qa*Qb)
acos(cos.theta)
```

```
## [1] 0.1050795
```

#### 3.3.1.3 積分する Integral

$$\begin{aligned} \sqrt{p} &= \cos \theta \\ \sqrt{(1-p)} &= \sin \theta \\ p &= \cos^2 \theta \\ \frac{dp}{d\theta} &= 2 \cos \theta \times (-\sin \theta) \\ d\theta &= -\frac{1}{2 \cos \theta \sin \theta} dp \\ L(B \rightarrow A) &= \theta \\ &= \int_{\theta_b}^{\theta_a} d\theta \\ &= \int_{Pa}^{Pb} -\frac{1}{2 \cos \theta \sin \theta} dp \\ &= \int_{Pa}^{Pb} -\frac{1}{2\sqrt{p}\sqrt{(1-p)}} dp \end{aligned}$$

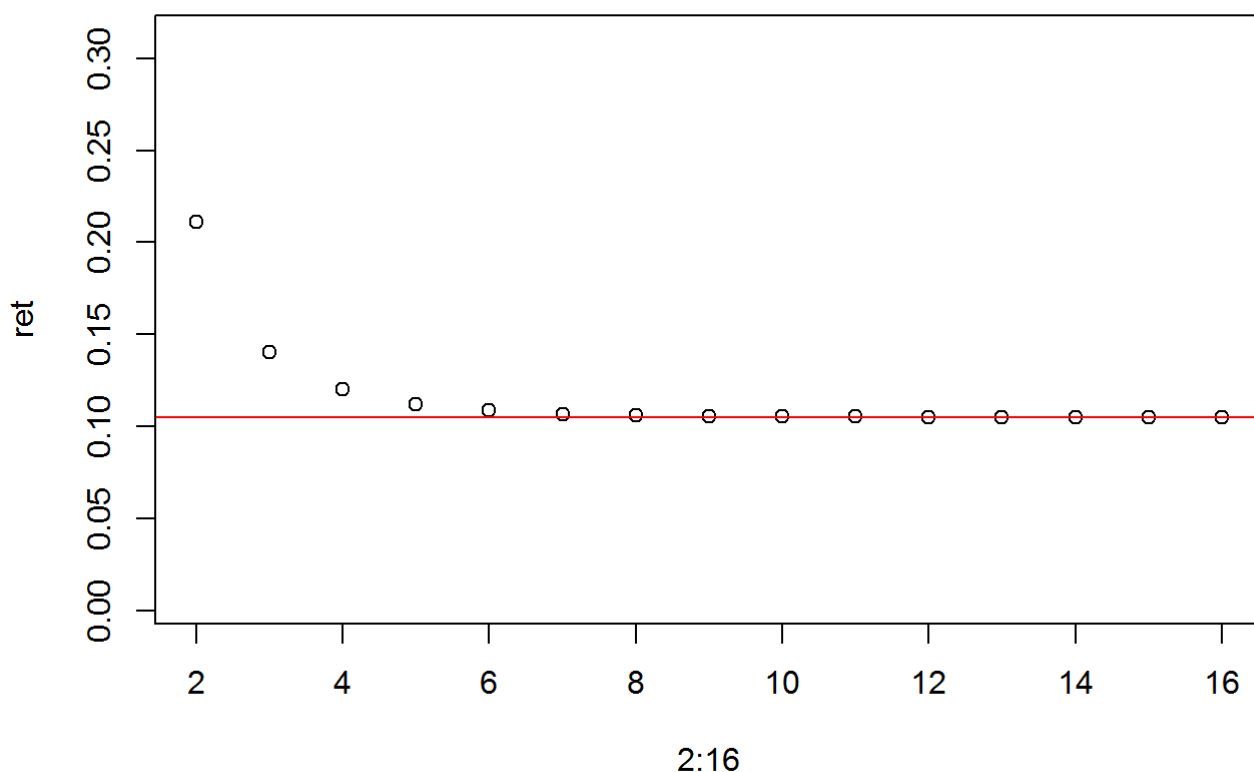
$p$  の刻み幅をだんだん小さくして、上記積分を計算すると、確かに、角度に収束する。

Numerical calculation with difference of sequence values of  $p$  makes the integral converging to the angle.

```

dps <- (1/2)^seq(2:16)
ret <- rep(0, length(dps))
for(i in 1:length(dps)) {
  ps <- seq(from=Pb[1], to=Pa[1], length=1/dps[i])
  dtheta <- -(ps[2]-ps[1])*(1/(2*sqrt(ps)*sqrt(1-ps)))
  ret[i] <- sum(dtheta)
}
plot(2:16, ret, ylim=c(min(ret)-0.1, max(ret)+0.1))
abline(h=acos(cos.theta), col=2)

```



### 3.3.2 球面上の座標とその伸び縮み Coordinates on the sphere and local elongation/shrinkage

- $P = \{p_1, p_2\}$  という分布を、パラメタ  $t = p_1$  を使って表すことにする。Parameterize  $P = \{p_1, p_2\}$  with  $t = p_1$ .
- $\sum p_i = 1$  直線上で等間隔に並ぶ点が、 $\sum q_i^2 = 1; q_i = \sqrt{p_i}$  という円弧上では等間隔ではなく並ぶ様子がわかる。They are evenly spaced in the line  $\sum p_i = 1$ . However they are not evenly spaced in the arc of sphere  $\sum q_i^2 = 1; q_i = \sqrt{p_i}$

```

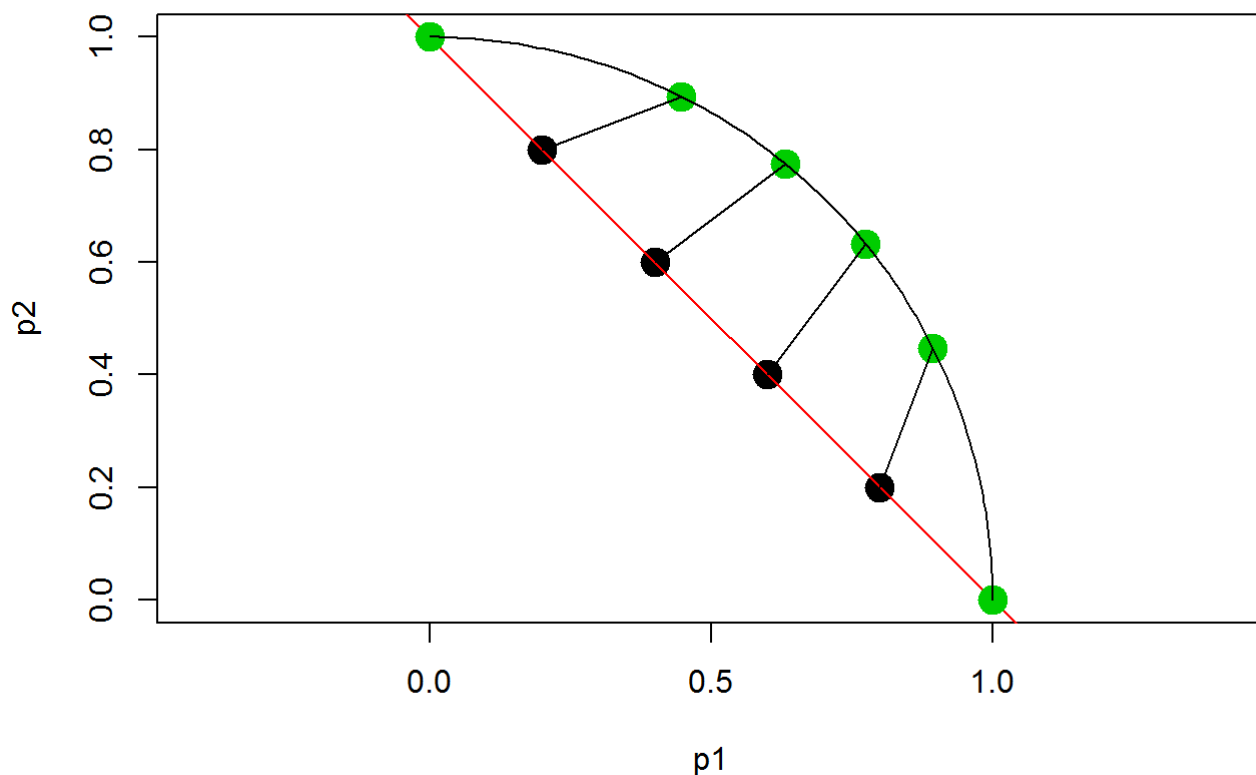
t <- seq(from=0, to=1, length=6)

p1 <- t
p2 <- 1-t

q1 <- sqrt(p1)
q2 <- sqrt(p2)

plot(p1, p2, pch=20, cex=3, asp=TRUE)
abline(1, -1, col=2)
points(q1, q2, pch=20, cex=3, col=3)
tt <- seq(from=0, to=pi/2, length=100)
points(cos(tt), sin(tt), type="l")
segments(p1, p2, q1, q2)

```



- $t$ が $t + dt$ まで少しだけ増えたとき、 $(\sqrt{t}, \sqrt{1-t})$ と $(\sqrt{t+dt}, \sqrt{1-t-dt})$ とを結ぶ弧 $dl$ の二乗は、  
When the parameter is increased a little bit from  $t$  to  $t + dt$ , the small arc's length  $dl$  between  $(\sqrt{t}, \sqrt{1-t})$  and  $(\sqrt{t+dt}, \sqrt{1-t-dt})$  is expressed as;

$$\begin{aligned}
 dl^2 &= (\sqrt{t+dt} - \sqrt{t})^2 + (\sqrt{1-t} - \sqrt{1-t-dt})^2 \\
 &= ((\frac{d\sqrt{t}}{dt})^2 + (\frac{d\sqrt{1-t}}{dt})^2) dt^2 \\
 &= ((\frac{1}{2\sqrt{t}})^2 + (\frac{-1}{2\sqrt{1-t}})^2) dt^2 \\
 &= \frac{1}{2} (\frac{1}{t} + \frac{1}{1-t}) dt^2 \\
 &= \frac{1}{2t(1-t)} dt^2
 \end{aligned}$$

なので therefore,

$$dl = \frac{1}{2t(1-t)} dt$$

これは、積分で弧の長さを計算したときに出てきた項。This term appeared in the integral to calculate the length of arc in the previous section.

結局、弧の長さは The length of arc is when the dimension is finite;

$$\begin{aligned}
 \theta &= \int_{ta}^{tb} dl \\
 \frac{dl}{dt} &= \sqrt{\sum_i (\frac{d\sqrt{p_i}}{dt})^2} \\
 \frac{dl}{dt} \frac{dl}{dt} &= \sum_i \frac{d\sqrt{p_i}}{dt} \frac{d\sqrt{p_i}}{dt}
 \end{aligned}$$

もし無限次元なら when infinite,

$$\frac{dl}{dt} \frac{dl}{dt} = \int \frac{d\sqrt{p_i}}{dt} \frac{d\sqrt{p_i}}{dt} dx$$

- 球面局所で、変数tをわずかに動かしたときに球面上の微小直線がどのくらいの長さの二乗は、

$\frac{dl}{dt} \frac{dl}{dt} = \sum_i \frac{d\sqrt{p_i}}{dt} \frac{d\sqrt{p_i}}{dt} \frac{dl}{dt} \frac{dl}{dt} = \int \frac{d\sqrt{p_i}}{dt} \frac{d\sqrt{p_i}}{dt} dx$  The square of length of small arc when  $t$  is increased by  $dt$  is given in this formula.

## 4 まとめ：球面上の座標 球面上の距離の二乗 座標変換 微分・積分 Summary: Coordinates on sphere, square of distance on the sphere, coordinate transformation, differentiation/integration.

- 分布は  $\sum_i^k p_i = 1$  を満足する Distributions meet  $\sum_i^k p_i = 1$ .
- $k$  は有限かもしれないし無限かもしれない  $k$  can be infinite.
- $\sum_i^k q_i^2 = 1$ ;  $q_i = \sqrt{p_i}$  と変換すると、分布は  $k$  次元空間の単位球面上の点とみなせる They are points on the  $k$  dimensional sphere surface with transformation  $\sum_i^k q_i^2 = 1$ ;  $q_i = \sqrt{p_i}$ .

- 弧の長さは  $\int_{\theta_a}^{\theta_b} d\theta = \int_{t_a}^{t_b} \frac{d\theta}{dt} dt$  となるから、局所の微分  $\frac{d\theta}{dt}$  が大事 local distance of small arc is  $\int_{\theta_a}^{\theta_b} d\theta = \int_{t_a}^{t_b} \frac{d\theta}{dt} dt$  and  $\frac{d\theta}{dt}$  is essential.
- 分布の違いを  $\sum_i (p_i^a - p_i^b)^2$  で測るのは、「直線距離の二乗」を使うことで可能 Dissimilarity among distributions can be measured with square of straight line as  $\sum_i (p_i^a - p_i^b)^2$ .
- 分布の違いを球面上で測ることにすれば  $\theta = \int_{t_a}^{t_b} dl; \sum_i \frac{d\sqrt{p_i}}{dt} \frac{d\sqrt{p_i}}{dt}$  When measuring dissimilarity on a surface of sphere,
- 分布関数の平方根がパラメタ表示されているとき、パラメタでの微分  $\frac{d\sqrt{p_i}}{dt}$  を全  $\sqrt{p_i}$  について足し合わせたものが、局所の微小距離の二乗になっているから、その平方根を積分すると、球面配置における分布間の距離となる Sum of  $\frac{d\sqrt{p_i}}{dt}$  is square of local infinitesimal distance and summing up its square-root for all  $p_i$  (or integration of  $p$ ) is distance among distributions
- 最後の積分はフィッシャー情報量で再度登場する The last integral will appear in the section for Fisher information.