分布の違いを測る Measure dissimilarity between distributions KLダイバージェンス KL-divergence

ryamada

2017年3月6日

1 平坦と直線と直交 Flatness, straight lines and perpendicular

1.1 直線 Straigt lines

- θ, η座標系はそれぞれ平坦座標系であるから、θ_i一定、η_j一定の点(上図の４種類の線)は、「直線」である。Because both θ and η coordinate systems are flat, points with same value of θ_i or η_j are in "straight lines", that are shown in the figure above with 4 colors.

- θ_1, θ_2の線形和で表される線も「直線」。η_1, η_2の線形和で表される線も「直線」。Points that are expressed as linear sum of θ_1 and θ_2 are in "straight lines" as well. Same for η_1 and eta_2.

1.2 直交

- θ座標系とη座標系は双直交であるから、θ_1とeta_2は直交し、θ_2とeta_1は直交する。 Because two coordinate systems are in dual orthogonal relataion, θ_1 and η_2 are perpendicular and θ_2 and η_1 are perpendicular.

1.3 例 Example

- 以下の図は、正規分布のθ座標系格子とη座標系格子とを、(m, s)座標系に描いたものである。The figure below is drawn as its x and y axes are m and s and θ-lattice and η-lattice are drawn in it.

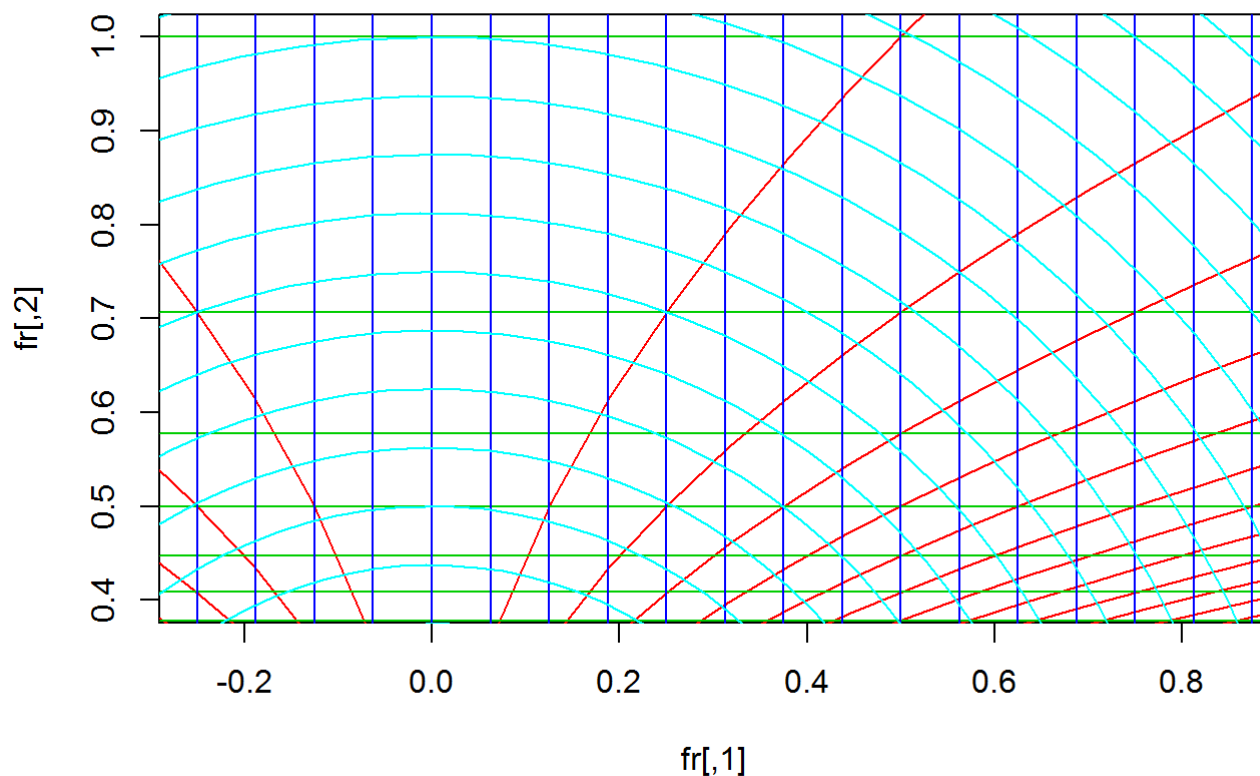- 赤線はθ_1一定の点を描いたもの、同様に、緑はθ_2、青はη_1、水色はη_2。

```
# 正規分布、(m,s)座標に(theta1,theta2)格子
#
theta1 <- seq(from=1/2,to=1000,by=1/2)
theta2 <- -seq(from=1/2,to=1000,by=1/2)
eta1 <- seq(from=-1,to=1,by=1/(2^4))
eta2 <- seq(from=0,to=2,by=1/(2^4))

fr <- matrix(c(-0.5,0,0.5,0.5),byrow=TRUE,2,2)
plot(fr,asp=1,col=0,xlim=c(0,0.6),ylim=c(0.4,1))
t <- seq(from=0,to=1,length=100) * 2*pi

for(i in 1:length(theta1)){
  points(t,sqrt(t/theta1[i]),type="l",col=2)
  points(-t,sqrt(t/theta1[i]),type="l",col=2)
}

for(i in 1:length(theta2)){
    abline(h=sqrt(-1/(2*theta2[i])),col=3)
}
for(i in 1:length(eta1)){
    abline(v=eta1[i],col=4)
}
t <- seq(from=0,to=1,length=100) * 2*pi
for(i in 1:length(eta2)){
    points(eta2[i] * cos(t),eta2[i] * sin(t),type="l",col=5)
}
```


- 描かれた線はすべて「直線」である

- θ_1 一定の線(赤)と η_2 一定の線(水色)は直交し、θ_2 一定の線(緑)と η_1 一定の線(青)も直交している

2 Kullback-Leibler ダイバージェンス Kullback-Leibler divergence

2.1 曲がっているので距離は定まらない Due to non-flatness, no distance is difined

- 双対平坦ではあるが、単純に平坦ではないので、いわゆる距離は定められない。Although dually flat, the space of distributions is not "simply flat". Therefore "distance" in a regular sense can not be defined.

2.2 ダイバージェンス Divergence

- ダイバージェンスと呼ばれる、距離と似た性質を持つものが定義できる。 Divergence that shares some features with distance can be defined.

- 双対平坦座標系にそのようなものを定めると、Kulback-Leibler divergence (KL divergence) となる。 The divergence for dually flat coordinate systems is called Kullback-Leibler divergence (KL divergence).

- 2点 P,Qの確率分布がp,qであるとすると、PからQへのKLdは以下の式で表される。Assume two points P and Q, corresponding to distributions p and q, respectively. KLd from P to Q is expressed as;

$$KLd(P \rightarrow Q) = \int p \frac{\log p}{\log q} dx$$

- 式の非対称性からわかるように、QからPに変えると値が異なる。The formula is asymmetric, that means the exchange of P and Q changes the value of divergence.
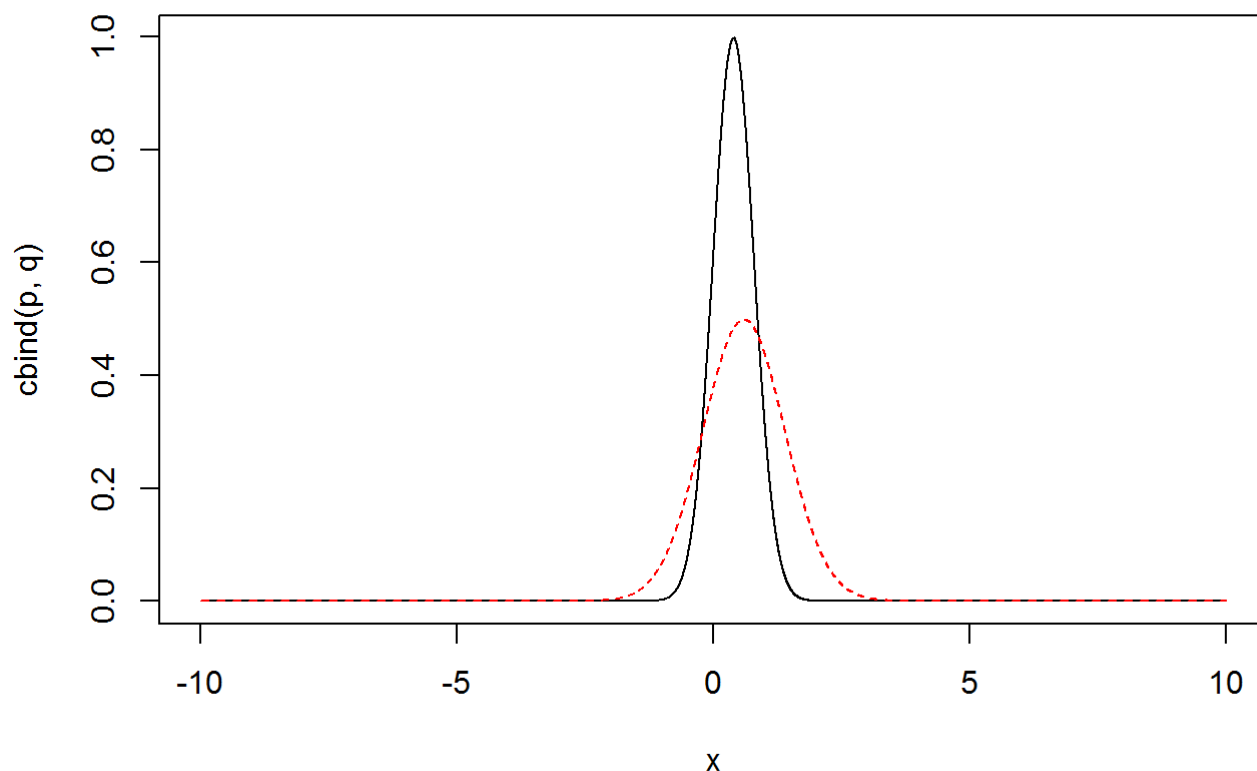
$$KLd(Q \rightarrow P) = \int p \frac{\log p}{\log q} dx \neq KLd(P \rightarrow Q)$$

2.3 例 Example

- ２つの正規分布P(m=0.4,s=0.4)とQ(m=0.6,0.8)を考える。 Assume two normal distributions P(m=0.4,s=0.4) and Q(m=0.6,0.8).

```
x <- seq(from=-10, to=10, length=10^4)
p <- dnorm(x, 0.4, 0.4)
q <- dnorm(x, 0.6, 0.8)

matplot(x, cbind(p, q), type="l")
```

- KLd の計算 Calculation of KLd.

```
KLdPQ <- sum(p*(log(p)-log(q))) * (x[2]-x[1])
KLdQP <- sum(q*(log(q)-log(p))) * (x[2]-x[1])
KLdPQ
```

```
## [1] 0.3493972
```

```
KLdQP
```

```
## [1] 0.9318528
```

2.4 正規分布の２次元の広がり Draw normal distributions in 2d space.

$$\theta_1 = \frac{m}{s^2}$$

$$\theta_2 = -\frac{1}{2s^2}$$

$$m = -\frac{\theta_1}{2\theta_2}$$

$$s^2 = -\frac{1}{2\theta_2}$$

$$\eta_1 = m$$
$$\eta_2 = m^2 + s^2$$
$$m = \eta_1$$
$$s^2 = \eta_2 - \eta_1^2$$

- 点$P = (m_p = 0.4, s_p = 0.4)$を取る Point $P = (m_p = 0.4, s_p = 0.4)$.

$$\theta_p 1 = \frac{m_p}{s_p^2} = \frac{0.4}{0.16} = 2.5$$

$$\theta_p 2 = -\frac{1}{2 \times s_p^2} = -\frac{1}{2 \times 0.16}$$

$$\eta_p 1 = m_p = 0.4$$

$$\eta_p 2 = m_p^2 + s_p^2 = 0.32$$

- 点Pが乗っている赤い線は$\theta_1 = 2.5$であり、点の近くの緑の線は$\theta_2 = -6$、2本の青い線($m = \eta_1 = 0.3750, 0.4375$)の間に位置し、水色の円では半径0.3125の付近にある。 A red line in whic P is, is $\theta_1 = 2.5$. The green line near P is $\theta_2 = -6$, two blue lines sandwiching P are $m = \eta_1 = 0.3750$ and 0.4375. The light blue arc close to P is a part of a circle with radius 0.3125.

```
# 正規分布、(m, s)座標に(theta1, theta2)格子
#
theta1 <- seq(from=1/2, to=1000, by=1/2)
theta2 <- -seq(from=1/2, to=1000, by=1/2)
eta1 <- seq(from=-1, to=1, by=1/(2^4))
eta2 <- seq(from=0, to=2, by=1/(2^4))

fr <- matrix(c(-0.5, 0, 0.5, 0.5), byrow=TRUE, 2, 2)
plot(fr, asp=1, col=0, xlim=c(0, 0.6), ylim=c(0.4, 1))
t <- seq(from=0, to=1, length=100) * 2*pi

for(i in 1:length(theta1)){
  points(t, sqrt(t/theta1[i]), type="l", col=2)
  points(-t, sqrt(t/theta1[i]), type="l", col=2)

}

for(i in 1:length(theta2)){
    abline(h=sqrt(-1/(2*theta2[i])), col=3)
}
for(i in 1:length(eta1)){
    abline(v=eta1[i], col=4)
}
t <- seq(from=0, to=1, length=100) * 2*pi
for(i in 1:length(eta2)){
    points(eta2[i] * cos(t), eta2[i] * sin(t), type="l", col=5)
}

P.ms <- c(0.4, 0.4)
P.theta <- my.ms2theta(P.ms[1], P.ms[2])
points(P.ms[1], P.ms[2], pch=20, cex=3)
text(P.ms[1]+0.1, P.ms[2], "P")
```
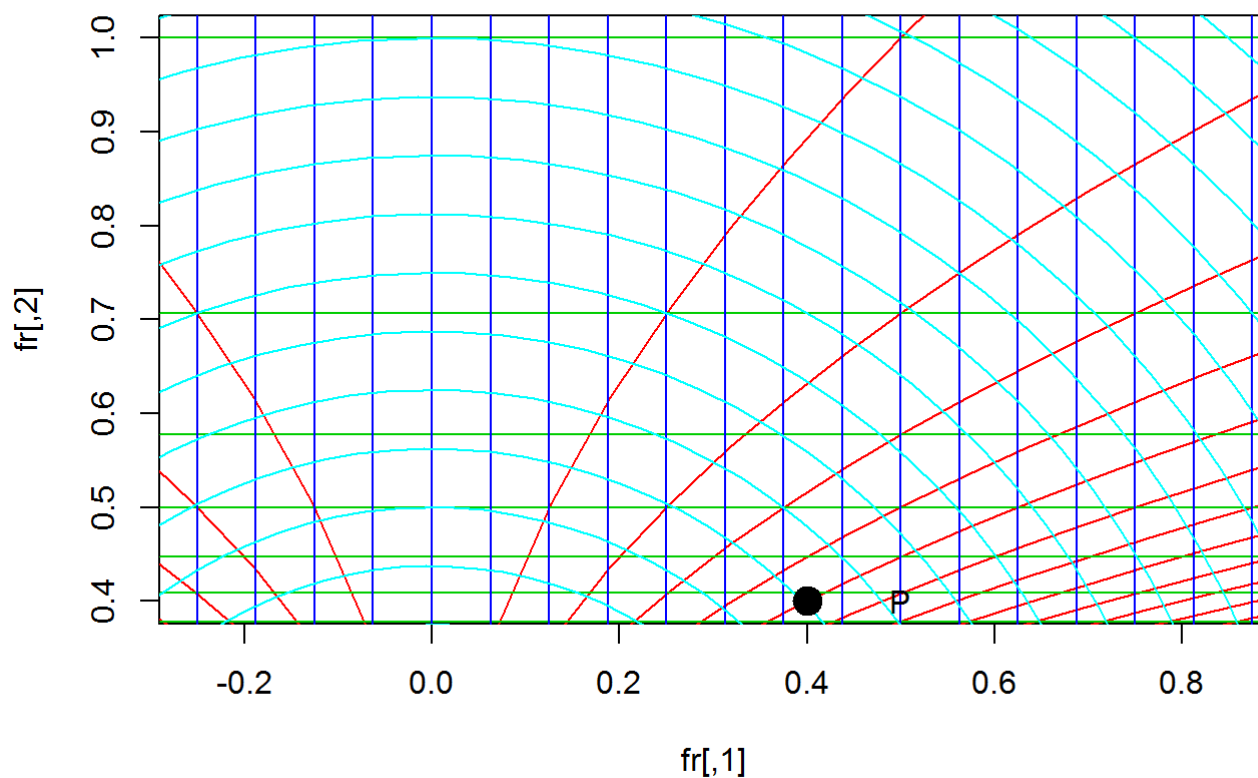
- 第2点$Q = (m_p = 0.6, s_p = 0.8)$を取る The 2nd point $Q = (m_p = 0.6, s_p = 0.8)$

$$\theta_q 1 = \frac{m_q}{s_q^2} = \frac{0.6}{0.64} = 0.9375$$

$$\theta_q 2 = -\frac{1}{2 \times s_q^2} = -\frac{1}{2 \times 0.64}$$

$$\eta_q 1 = m_p = 0.6$$

$$\eta_q 2 = m_p^2 + s_p^2 = 1$$

```
fr <- matrix(c(-0.5,0,0.5,0.5),byrow=TRUE,2,2)
plot(fr,asp=1,col=0,xlim=c(0,0.6),ylim=c(0.4,1))
t <- seq(from=0,to=1,length=100) * 2*pi

for(i in 1:length(theta1)){
  points(t,sqrt(t/theta1[i]),type="l",col=2)
  points(-t,sqrt(t/theta1[i]),type="l",col=2)

}

for(i in 1:length(theta2)){
    abline(h=sqrt(-1/(2*theta2[i])),col=3)
}
for(i in 1:length(eta1)){
    abline(v=eta1[i],col=4)
}
t <- seq(from=0,to=1,length=100) * 2*pi
for(i in 1:length(eta2)){
    points(eta2[i] * cos(t),eta2[i] * sin(t),type="l",col=5)
}

P.ms <- c(0.4,0.4)
points(P.ms[1],P.ms[2],pch=20,cex=3)
text(P.ms[1]+0.1,P.ms[2],"P")
Q.ms <- c(0.6,0.8)
points(Q.ms[1],Q.ms[2],pch=20,cex=3,col=2)
text(Q.ms[1]+0.1,Q.ms[2],"Q")
```
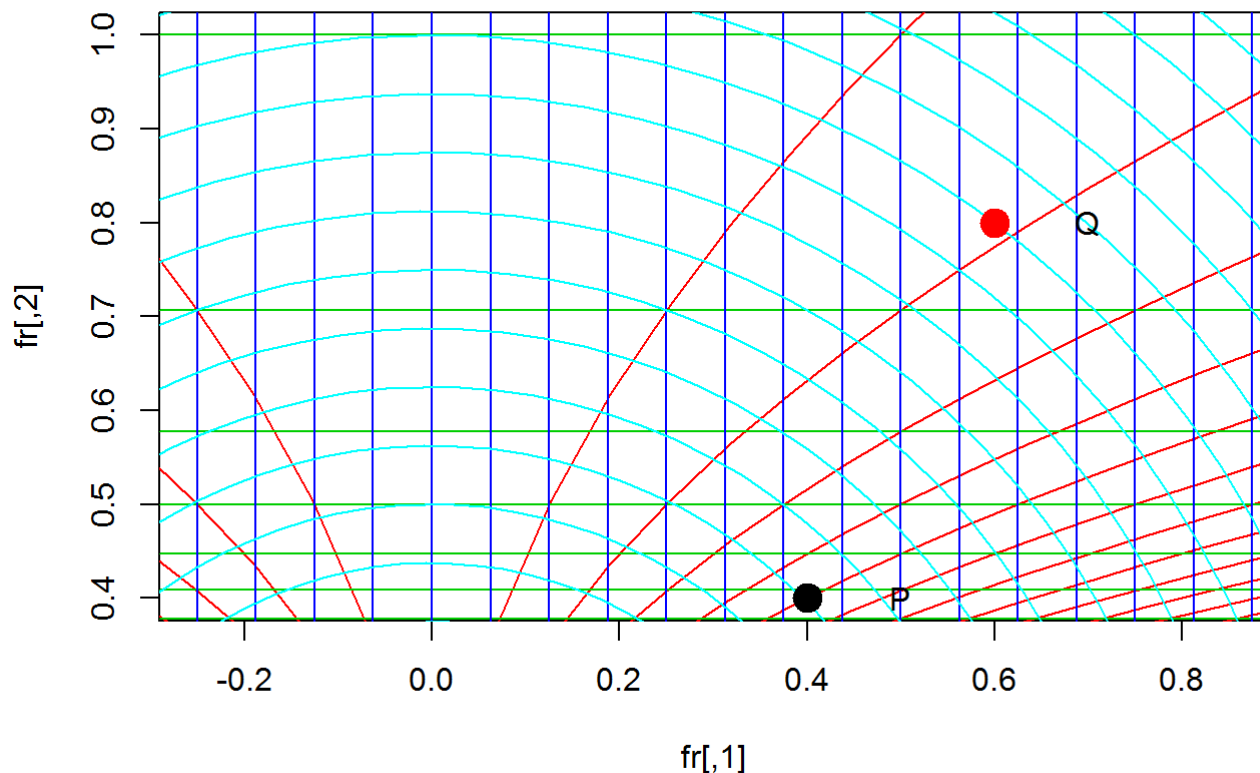
- 点Pから点Qへと向かうにあたり、θ_1が一定(赤)かθ_2(緑)が一定か、η_1(青)が一定かη_2(水色)が一定な『直線』に沿って進むことにする。Assume you move from P to Q. You take a path on which θ_1 or θ_2 or η_1 or η_2 is constant, which is red, green, blue or light blue. This means you move "straight".

- このとき、θ_1一定の線(赤)とη_2一定の線(水色)は直交しているから、そのどちらかが一定になるように進む方法と Because red lines and light blue lines (θ_1 constant and η_2 constant) are perpendicular, you can move only using these.

- もう片方の直交ペアθ_2一定の線(緑)とη_1一定の線(青)のみを使って進む方法の２つを考える。Another way is to use only greens and blues, that are θ_2 constant or η_1 constant.

- この２通りの進み方だと、２点間を結ぶ線を斜辺とする「直角三角形」の直角を挟む２辺となるからこのようにする。そして、そのようにするのは、斜辺の長さの取り方として好都合であるからである。どのように好都合かは後述する。The reason why we take these combinations is because we want to have right-angled triangles that are used later.

- θ_1, η_2 の場合 Case of θ_1 or η_2

- 太い黒い線に沿って「長方形」の２辺を取ると、片方はθ座標系の直線、もう片方はη座標系の直線。Thick black lines make "rectangle", the edges of which are straight in θ or η coordinate systems.

```
fr <- matrix(c(-0.5,0,0.5,0.5),byrow=TRUE,2,2)
plot(fr,asp=1,col=0,xlim=c(0,0.6),ylim=c(0.4,1))
t <- seq(from=0,to=1,length=100) * 2*pi

for(i in 1:length(theta1)){
  points(t,sqrt(t/theta1[i]),type="l",col=2)
  points(-t,sqrt(t/theta1[i]),type="l",col=2)

}


for(i in 1:length(theta2)){
    abline(h=sqrt(-1/(2*theta2[i])),col=3)
}
for(i in 1:length(eta1)){
    abline(v=eta1[i],col=4)
}
t <- seq(from=0,to=1,length=100) * 2*pi
for(i in 1:length(eta2)){
    points(eta2[i] * cos(t),eta2[i] * sin(t),type="l",col=5)
}

P.ms <- c(0.4,0.4)
points(P.ms[1],P.ms[2],pch=20,cex=5)
text(P.ms[1]+0.1,P.ms[2],"P")
Q.ms <- c(0.6,0.8)
points(Q.ms[1],Q.ms[2],pch=20,cex=5,col=2)
text(Q.ms[1]+0.1,Q.ms[2],"Q")
P.thetas <- my.ms2theta(P.ms[1],P.ms[2])
P.etas <- my.ms2eta(P.ms[1],P.ms[2])
Q.thetas <- my.ms2theta(Q.ms[1],Q.ms[2])
Q.etas <- my.ms2eta(Q.ms[1],Q.ms[2])
points(t,sqrt(t/P.thetas[1]),type="l",col=1,lwd = 3)
points(t,sqrt(t/Q.thetas[1]),type="l",col=1,lwd = 3)

points(sqrt(P.etas[2]) * cos(t),sqrt(P.etas[2])* sin(t),type="l",col=1,lwd = 3)
points(sqrt(Q.etas[2]) * cos(t),sqrt(Q.etas[2])* sin(t),type="l",col=1,lwd = 3)
```
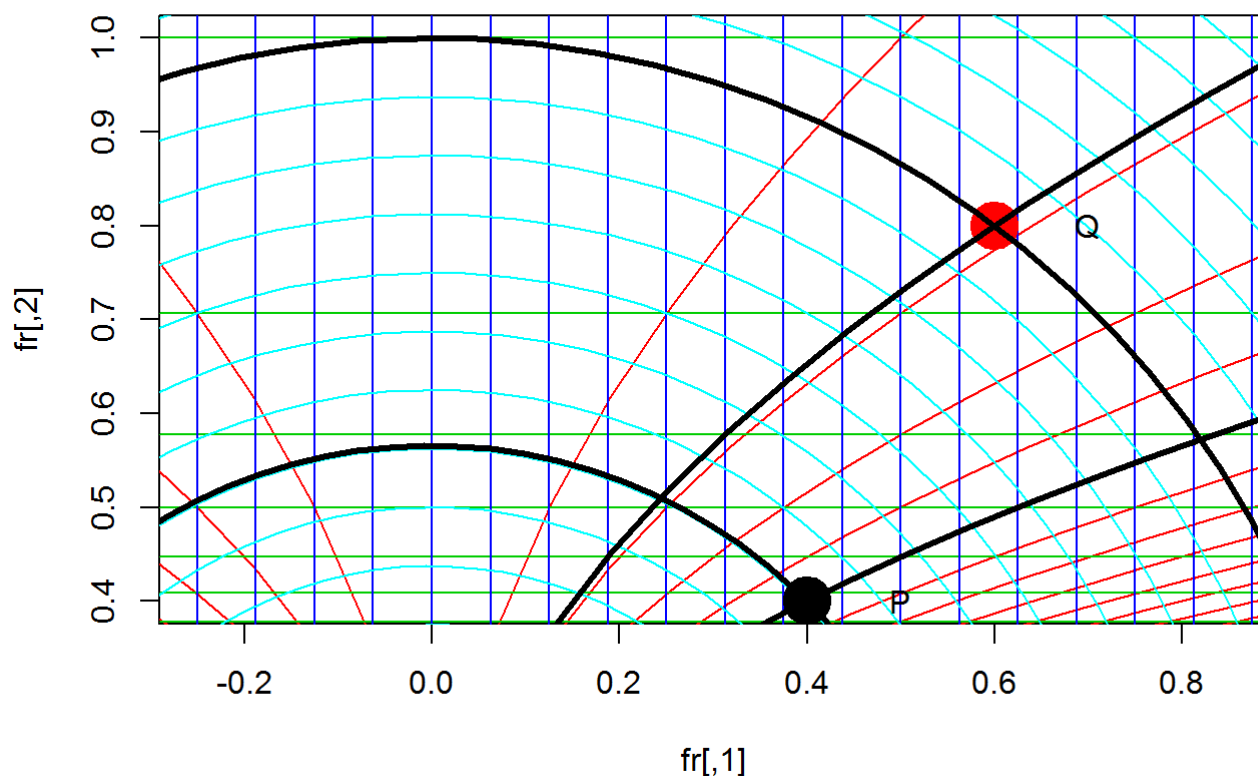
- 2本の直線の交点が2つある。R_1, R_2とすれば、その座標は Let R_1 and R_2 denote two crossing points.

$$R_1(\theta_1 = \theta_1^P = 2.5, \eta_2 = \eta_2^Q = 1)$$

$$R_2(\theta_1 = \theta_1^Q = \frac{0.6}{0.64}, \eta_2 = \eta_2^P = 0.32)$$

- θ_1とη_2とでできた座標なので、「混合座標」と呼ばれる。この2点のm, sを出しておこう。 Their coordinates can be expressed as mixture of θ and η; this is why these coordinates are called "mixture coordinates".

```
my.mix2ms12 <- function(theta1,eta2){
  s.sq <- (-1 + sqrt(1+4*theta1^2*eta2))/(2*theta1^2)
  s <- sqrt(s.sq)
  m <- theta1*s.sq
  return(c(m,s))
}
my.mix2ms21 <- function(eta1,theta2){
  m <- eta1
  s.sq <- -1/(2*theta2)
  s <- sqrt(s.sq)
  return(c(m,s))
}
R1.ms <- my.mix2ms12(P.thetas[1],Q.etas[2])
R2.ms <- my.mix2ms12(Q.thetas[1],P.etas[2])
R1.ms
```

```
## [1] 0.8198039 0.5726444
```

```
R2.ms
```

```
## [1] 0.2441269 0.5102960
```

```
fr <- matrix(c(-0.5,0,0.5,0.5),byrow=TRUE,2,2)
plot(fr,asp=1,col=0,xlim=c(0,0.6),ylim=c(0.4,1))
t <- seq(from=0,to=1,length=100) * 2*pi

for(i in 1:length(theta1)){
  points(t,sqrt(t/theta1[i]),type="l",col=2)
  points(-t,sqrt(t/theta1[i]),type="l",col=2)

}

for(i in 1:length(theta2)){
    abline(h=sqrt(-1/(2*theta2[i])),col=3)
}
for(i in 1:length(eta1)){
    abline(v=eta1[i],col=4)
}
t <- seq(from=0,to=1,length=100) * 2*pi
for(i in 1:length(eta2)){
    points(eta2[i] * cos(t),eta2[i] * sin(t),type="l",col=5)
}

P.ms <- c(0.4,0.4)
points(P.ms[1],P.ms[2],pch=20,cex=5)
text(P.ms[1]+0.1,P.ms[2],"P")
Q.ms <- c(0.6,0.8)
points(Q.ms[1],Q.ms[2],pch=20,cex=5,col=2)
text(Q.ms[1]+0.1,Q.ms[2],"Q")
P.thetas <- my.ms2theta(P.ms[1],P.ms[2])
P.etas <- my.ms2eta(P.ms[1],P.ms[2])
Q.thetas <- my.ms2theta(Q.ms[1],Q.ms[2])
Q.etas <- my.ms2eta(Q.ms[1],Q.ms[2])
points(t,sqrt(t/P.thetas[1]),type="l",col=1,lwd = 3)
points(t,sqrt(t/Q.thetas[1]),type="l",col=1,lwd = 3)

points(sqrt(P.etas[2]) * cos(t),sqrt(P.etas[2])* sin(t),type="l",col=1,lwd = 3)
points(sqrt(Q.etas[2]) * cos(t),sqrt(Q.etas[2])* sin(t),type="l",col=1,lwd = 3)

points(R1.ms[1],R1.ms[2],pch=20,cex=5,col=3)
points(R2.ms[1],R2.ms[2],pch=20,cex=5,col=4)
text(R1.ms[1]+0.1,R1.ms[2],"R1")
text(R2.ms[1]-0.1,R2.ms[2],"R2")
```
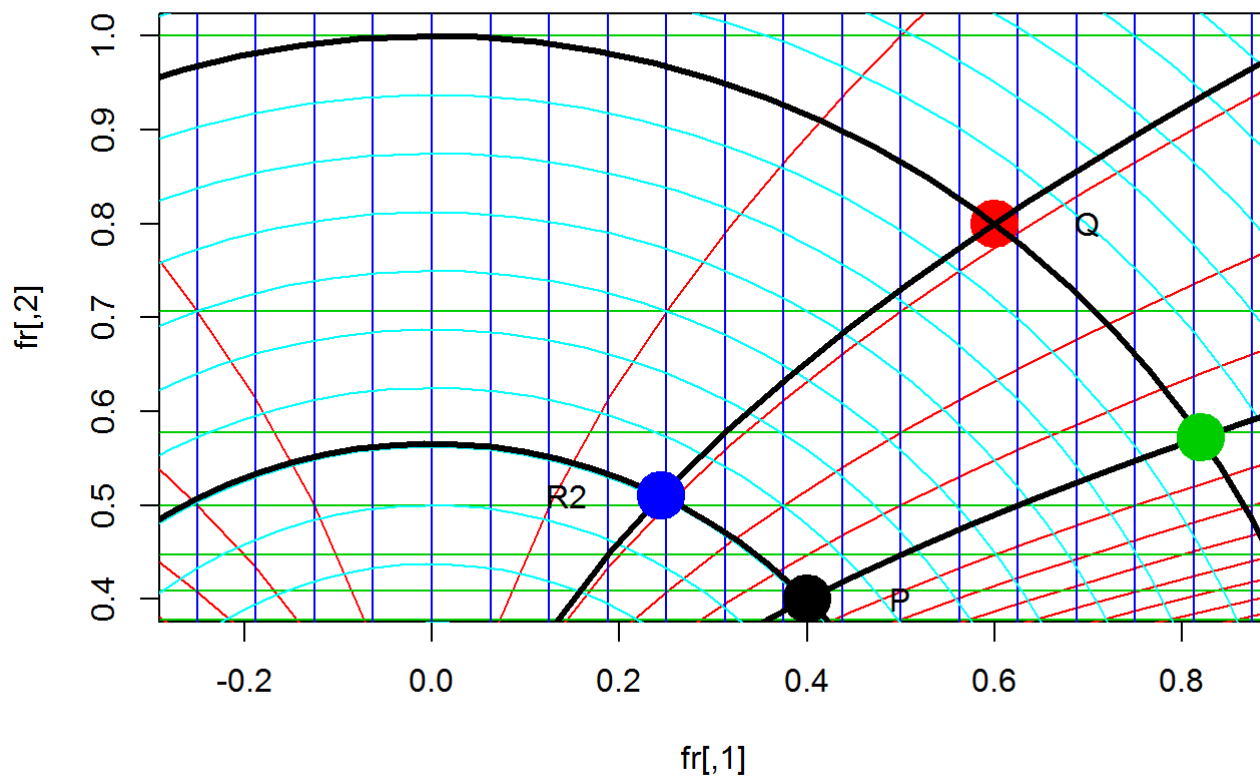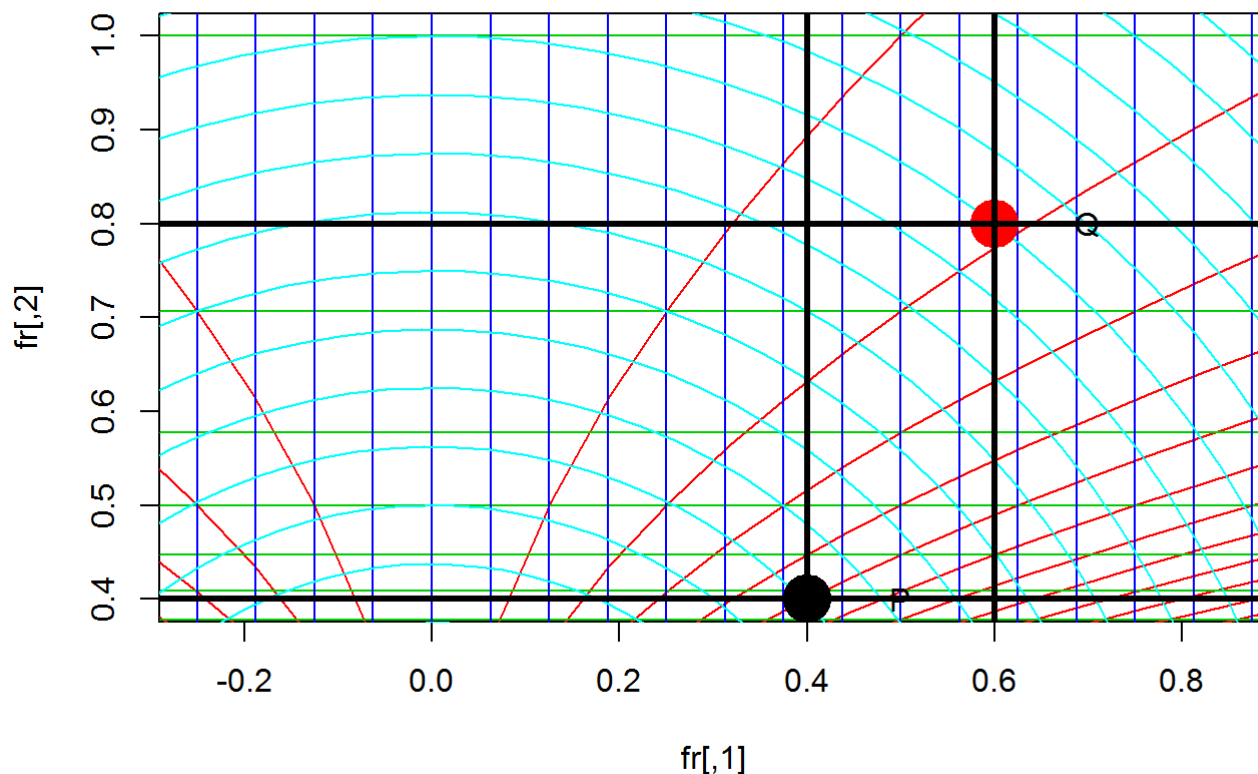
- θ_2, η_1 の場合 Another case with θ_2 or η_1 .

```r
fr <- matrix(c(-0.5,0,0.5,0.5),byrow=TRUE,2,2)
plot(fr,asp=1,col=0,xlim=c(0,0.6),ylim=c(0.4,1))
t <- seq(from=0,to=1,length=100) * 2*pi

for(i in 1:length(theta1)){
  points(t,sqrt(t/theta1[i]),type="l",col=2)
  points(-t,sqrt(t/theta1[i]),type="l",col=2)

}

for(i in 1:length(theta2)){
    abline(h=sqrt(-1/(2*theta2[i])),col=3)
}
for(i in 1:length(eta1)){
    abline(v=eta1[i],col=4)
}
t <- seq(from=0,to=1,length=100) * 2*pi
for(i in 1:length(eta2)){
    points(eta2[i] * cos(t),eta2[i] * sin(t),type="l",col=5)
}

P.ms <- c(0.4,0.4)
points(P.ms[1],P.ms[2],pch=20,cex=5)
text(P.ms[1]+0.1,P.ms[2],"P")
Q.ms <- c(0.6,0.8)
points(Q.ms[1],Q.ms[2],pch=20,cex=5,col=2)
text(Q.ms[1]+0.1,Q.ms[2],"Q")
P.thetas <- my.ms2theta(P.ms[1],P.ms[2])
P.etas <- my.ms2eta(P.ms[1],P.ms[2])
Q.thetas <- my.ms2theta(Q.ms[1],Q.ms[2])
Q.etas <- my.ms2eta(Q.ms[1],Q.ms[2])
abline(h=sqrt(-1/(2*P.thetas[2])),col=1,lwd=3)
abline(h=sqrt(-1/(2*Q.thetas[2])),col=1,lwd=3)

abline(v=P.etas[1],col=1,lwd=3)
abline(v=Q.etas[1],col=1,lwd=3)
```

- 2本の直線の交点が2つある。R_3, R_4とすれば、その座標は Let R_3, R_4 denote the crossing points.

$$R_3(\eta_1 = \eta_1^P = 0.4, \theta_2 = \theta_2^Q = -\frac{1}{2 \times 0.64})$$

$$R_4(\eta_1 = \eta_1^Q = 0.6, \theta_2 = \theta_2^P = -\frac{1}{2 \times 0.16})$$

- θ_2とη_1とでできた座標なので、これも「混合座標」と呼ばれる。この2点のm, sを出しておこう。Again they are mixture.

```
R3.ms <- my.mix2ms21(P.etas[1],Q.thetas[2])
R4.ms <- my.mix2ms21(Q.etas[1],P.thetas[2])
R3.ms
```

```
## [1] 0.4 0.8
```

```
R4.ms
```

```
## [1] 0.6 0.4
```

```r
fr <- matrix(c(-0.5,0,0.5,0.5),byrow=TRUE,2,2)
plot(fr,asp=1,col=0,xlim=c(0,0.6),ylim=c(0.4,1))
t <- seq(from=0,to=1,length=100) * 2*pi

for(i in 1:length(theta1)){
  points(t,sqrt(t/theta1[i]),type="l",col=2)
  points(-t,sqrt(t/theta1[i]),type="l",col=2)

}

for(i in 1:length(theta2)){
    abline(h=sqrt(-1/(2*theta2[i])),col=3)
}
for(i in 1:length(eta1)){
    abline(v=eta1[i],col=4)
}
t <- seq(from=0,to=1,length=100) * 2*pi
for(i in 1:length(eta2)){
    points(eta2[i] * cos(t),eta2[i] * sin(t),type="l",col=5)
}

P.ms <- c(0.4,0.4)
points(P.ms[1],P.ms[2],pch=20,cex=5)
text(P.ms[1]+0.1,P.ms[2],"P")
Q.ms <- c(0.6,0.8)
points(Q.ms[1],Q.ms[2],pch=20,cex=5,col=2)
text(Q.ms[1]+0.1,Q.ms[2],"Q")
P.thetas <- my.ms2theta(P.ms[1],P.ms[2])
P.etas <- my.ms2eta(P.ms[1],P.ms[2])
Q.thetas <- my.ms2theta(Q.ms[1],Q.ms[2])
Q.etas <- my.ms2eta(Q.ms[1],Q.ms[2])
abline(h=sqrt(-1/(2*P.thetas[2])),col=1,lwd=3)
abline(h=sqrt(-1/(2*Q.thetas[2])),col=1,lwd=3)

abline(v=P.etas[1],col=1,lwd=3)
abline(v=Q.etas[1],col=1,lwd=3)

points(R3.ms[1],R3.ms[2],pch=20,cex=5,col=5)
points(R4.ms[1],R4.ms[2],pch=20,cex=5,col=6)
text(R3.ms[1]-0.1,R3.ms[2],"R3")
text(R4.ms[1]+0.1,P.ms[2],"R4")
```
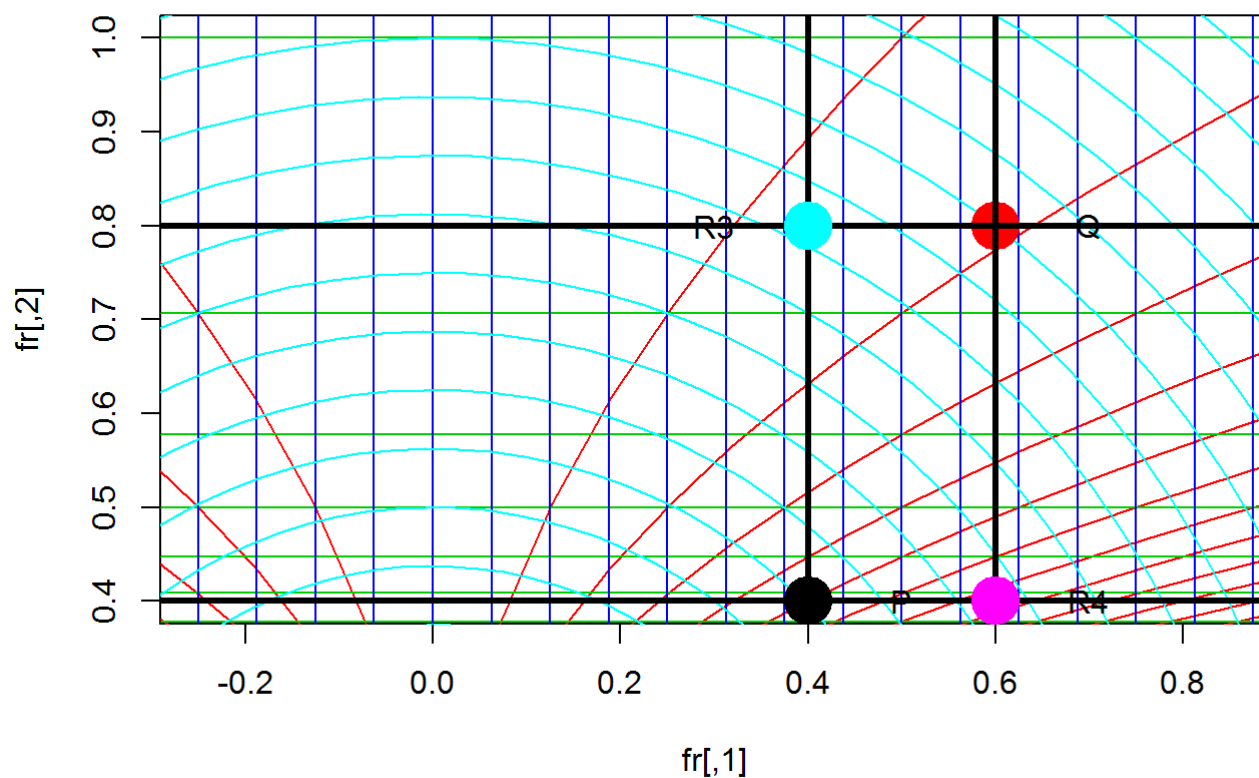
- θ_1, η_2 のペアでの「長方形」も θ_2, η_1 のペアでの「長方形」も、次の意味で「同じ」である。The 1st pair θ_1, η_2 generated a rectangle and θ_2, η_1 pair also generated a rectangle.

```
fr <- matrix(c(-0.5,0,0.5,0.5),byrow=TRUE,2,2)
plot(fr,asp=1,col=0,xlim=c(0,0.6),ylim=c(0.4,1))
t <- seq(from=0,to=1,length=100) * 2*pi

for(i in 1:length(theta1)){
  points(t,sqrt(t/theta1[i]),type="l",col=2)
  points(-t,sqrt(t/theta1[i]),type="l",col=2)

}

for(i in 1:length(theta2)){
    abline(h=sqrt(-1/(2*theta2[i])),col=3)
}
for(i in 1:length(eta1)){
    abline(v=eta1[i],col=4)
}
t <- seq(from=0,to=1,length=100) * 2*pi
for(i in 1:length(eta2)){
    points(eta2[i] * cos(t),eta2[i] * sin(t),type="l",col=5)
}

P.ms <- c(0.4,0.4)
points(P.ms[1],P.ms[2],pch=20,cex=5)
text(P.ms[1]+0.1,P.ms[2],"P")
Q.ms <- c(0.6,0.8)
points(Q.ms[1],Q.ms[2],pch=20,cex=5,col=2)
text(Q.ms[1]+0.1,Q.ms[2],"Q")
P.thetas <- my.ms2theta(P.ms[1],P.ms[2])
P.etas <- my.ms2eta(P.ms[1],P.ms[2])
Q.thetas <- my.ms2theta(Q.ms[1],Q.ms[2])
Q.etas <- my.ms2eta(Q.ms[1],Q.ms[2])
points(t,sqrt(t/P.thetas[1]),type="l",col=4,lwd = 3)
points(t,sqrt(t/Q.thetas[1]),type="l",col=4,lwd = 3)

points(sqrt(P.etas[2]) * cos(t),sqrt(P.etas[2])* sin(t),type="l",col=1,lwd = 3)
points(sqrt(Q.etas[2]) * cos(t),sqrt(Q.etas[2])* sin(t),type="l",col=1,lwd = 3)


abline(h=sqrt(-1/(2*P.thetas[2])),col=4,lwd=3)
abline(h=sqrt(-1/(2*Q.thetas[2])),col=4,lwd=3)

abline(v=P.etas[1],col=1,lwd=3)
abline(v=Q.etas[1],col=1,lwd=3)

points(R1.ms[1],R1.ms[2],pch=20,cex=5,col=3)
points(R2.ms[1],R2.ms[2],pch=20,cex=5,col=4)
points(R3.ms[1],R3.ms[2],pch=20,cex=5,col=5)
points(R4.ms[1],R4.ms[2],pch=20,cex=5,col=6)
text(R1.ms[1]+0.1,R1.ms[2],"R1")
text(R2.ms[1]-0.1,R2.ms[2],"R2")
text(R3.ms[1]-0.1,R3.ms[2],"R3")
text(R4.ms[1]+0.1,R4.ms[2],"R4")
```

2.5 ２点間のダイバージェンス　Divergence between two points.

- ２点間の距離のようなものであるダイバージェンスを考える。２点間に直線のようなものを取って、その長さが知りたい。Divergence that something like asquared distance is introduced as a measure between two points. Distance is a measure along straight lines and divergence should be also defined along "straight lines".

- 距離の二乗に相当する量である。Divergence is like square of distance.

- この空間には、θ座標系での直線も引けるし、η座標系での直線も引ける。We have two kinds of straight lines in θ and η coordinates systems.

- それぞれの座標系の直線は、線形に表せるから、２点$P = (\theta_1^P, \theta_2^P)$と$Q = (\theta_1^Q, \theta_2^Q)$との間の$\theta$座標系な直線は以下のように表せる。Straignt lines are expressed as linear sum of each coordinates and the coordinates of lines in θ system are:

$$(\theta_1, \theta_2) = (\theta_1^P, \theta_2^P) + t(\theta_1^Q - \theta_1^P, \theta_2^Q - \theta_2^P)$$

- また、η座標系の直線は次のように表せる。Same for straight lines in η system.

$$(\eta_1, \eta_2) = (\eta_1^P, \eta_2^P) + t(\eta_1^Q - \eta_1^P, \eta_2^Q - \eta_2^P)$$

- 二通りの「直線」を描く(青太線、赤破線)。Two straight lines are drawn with thick blue line and red broken line, which are ovelapped.

```
fr <- matrix(c(-0.5,0,0.5,0.5),byrow=TRUE,2,2)
plot(fr,asp=1,col=0,xlim=c(0,0.6),ylim=c(0.4,1))
t <- seq(from=0,to=1,length=100) * 2*pi

for(i in 1:length(theta1)){
  points(t,sqrt(t/theta1[i]),type="l",col=2)
  points(-t,sqrt(t/theta1[i]),type="l",col=2)

}

for(i in 1:length(theta2)){
    abline(h=sqrt(-1/(2*theta2[i])),col=3)
}
for(i in 1:length(eta1)){
    abline(v=eta1[i],col=4)
}
t <- seq(from=0,to=1,length=100) * 2*pi
for(i in 1:length(eta2)){
    points(eta2[i] * cos(t),eta2[i] * sin(t),type="l",col=5)
}

P.ms <- c(0.4,0.4)
points(P.ms[1],P.ms[2],pch=20,cex=5)
text(P.ms[1]+0.1,P.ms[2],"P")
Q.ms <- c(0.6,0.8)
points(Q.ms[1],Q.ms[2],pch=20,cex=5,col=2)
text(Q.ms[1]+0.1,Q.ms[2],"Q")
P.thetas <- my.ms2theta(P.ms[1],P.ms[2])
P.etas <- my.ms2eta(P.ms[1],P.ms[2])
Q.thetas <- my.ms2theta(Q.ms[1],Q.ms[2])
Q.etas <- my.ms2eta(Q.ms[1],Q.ms[2])
points(t,sqrt(t/P.thetas[1]),type="l",col=4,lwd = 3)
points(t,sqrt(t/Q.thetas[1]),type="l",col=4,lwd = 3)

points(sqrt(P.etas[2]) * cos(t),sqrt(P.etas[2])* sin(t),type="l",col=1,lwd = 3)
points(sqrt(Q.etas[2]) * cos(t),sqrt(Q.etas[2])* sin(t),type="l",col=1,lwd = 3)


abline(h=sqrt(-1/(2*P.thetas[2])),col=4,lwd=3)
abline(h=sqrt(-1/(2*Q.thetas[2])),col=4,lwd=3)

abline(v=P.etas[1],col=1,lwd=3)
abline(v=Q.etas[1],col=1,lwd=3)

points(R1.ms[1],R1.ms[2],pch=20,cex=5,col=3)
points(R2.ms[1],R2.ms[2],pch=20,cex=5,col=4)
points(R3.ms[1],R3.ms[2],pch=20,cex=5,col=5)
points(R4.ms[1],R4.ms[2],pch=20,cex=5,col=6)
text(R1.ms[1]+0.1,R1.ms[2],"R1")
text(R2.ms[1]-0.1,R2.ms[2],"R2")
text(R3.ms[1]-0.1,R3.ms[2],"R3")
text(R4.ms[1]+0.1,R4.ms[2],"R4")

t <- seq(from=0,to=1,length=100)
line.theta1 <- P.thetas[1] + t * (Q.thetas[1]-P.thetas[1])
line.theta2 <- P.thetas[2] + t * (Q.thetas[2]-P.thetas[2])
line.eta1 <- P.etas[1] + t * (Q.etas[1]-P.etas[1])
```

```
line.eta2 <- P.etas[2] + t * (Q.etas[2]-P.etas[2])
line.theta.ms <- matrix(0, length(t), 2)
line.eta.ms <- matrix(0, length(t), 2)
for(i in 1:length(t)){
  line.theta.ms[i,] <- my.theta2ms(line.theta1[i], line.theta2[i])
  line.eta.ms[i,] <- my.eta2ms(line.eta1[i], line.eta2[i])
}
points(line.theta.ms[,1], line.theta.ms[,2], type="l", col=4, lwd=5)
points(line.eta.ms[,1], line.eta.ms[,2], type="l", col=2, lwd=2, lty=2)
```


2.6 二通りのダイバージェンスとKL divergence, Relation between KL divergence and two types of divergences

- ２点P,Qの間のダイバージェンスには、次のような２通りの値がある。PとQとを交換した形になっていて、"$P||Q$"の、もしくは"$Q||P$"のKL divergenceと呼ばれる。Between two points P and Q, two divergence values are defined. KL divergence with "$P||Q$" and "$Q||P$".

$$KLd(P||Q) = \int p(\log(p) - \log(q))dx$$

$$KLd(Q||P) = \int q(\log(q) - \log(p))dx$$

- 情報幾何では、これに次のような４つの意味を持たせる。Information geometry interprets these two in four ways.
- $D^\theta(P \to Q)$:PからQへ、θ的に($\alpha = 1$的に)取ったダイバージェンス Divergence from P to Q in terms of $, or $=1.

- $D^\eta(P \to Q)$:PからQへ、η的に($\alpha = -1$的に)取ったダイバージェンス Divergence from P to Q in terms of $, or $=-1.
- $D^\theta(Q \to P)$:QからPへ、θ的に($\alpha = 1$的に)取ったダイバージェンス Divergence from Q to P in terms of $, or $=1.
- $D^\eta(Q \to P)$:QからPへ、η的に($\alpha = -1$的に)取ったダイバージェンス Divergence from Q to P in terms of $, or $=-1.

- これが、以下のような対応関係を持っている。See following relation among them.

$$D^\eta(P \to Q) = D^\theta(Q \to P) = KLd(P||Q)$$
$$D^\theta(P \to Q) = D^\eta(Q \to P) = KLd(Q||P)$$

- ダイバージェンスを計算してみる。Calculation of divergence values.

```
my.divergence <- function(dP,dQ,dx,alpha=-1){
  if(alpha==-1){
    ret <- sum(dP * (log(dP)-log(dQ))) * dx
  }else{
    ret <- sum(dQ * (log(dQ)-log(dP))) * dx
  }
  return(ret)
}
x <- seq(from=-10,to=10,length=10^4)

dP <- dnorm(x,P.ms[1],P.ms[2])
dQ <- dnorm(x,Q.ms[1],Q.ms[2])

dx <- x[2]-x[1]
# P->Q alpha=1
my.divergence(dP,dQ,dx,alpha=1)
```

```
## [1] 0.9318528
```

```
# P->Q alpha=-1
my.divergence(dP,dQ,dx,alpha=-1)
```

```
## [1] 0.3493972
```

```
# Q->P alpha=1
my.divergence(dQ,dP,dx,alpha=1)
```

```
## [1] 0.3493972
```

```
# Q->P alpha=-1
my.divergence(dQ,dP,dx,alpha=-1)
```

```
## [1] 0.9318528
```

2.7 拡張ピタゴラスの定理 Extended Pythagorean theorem

- θ, ηは双直交なので、ピタゴラスの定理のような性質が、上記のダイバージェンスには存在する。Because θ and η are dual orthogonal, divergeces for them have a feature related with Pythagorean theorem.

- ピタゴラスの定理は、直角三角形の３辺の長さに$a^2 + b^2 = c^2$という関係があることを言う。Pythagorean theorem shows relation among length of three edges of right-angled triangle, $a^2 + b^2 = c^2$.

- 双対座標系での拡張ピタゴラスの定理は次のような関係のことを言う。The extended Pythagorean theorem in dual coordinate systems is as follow.

- PとR_iとはθでの「直線(測地線)」で結ばれ、R_iとQとはηでの「直線(測地線)」で結ばれているとき、以下の関係にある。具体的には、R_1, R_4がこれに相当する。Assume a point R_i for P and Q, where R_i should meet; the "straight"" line between P and R_i is θ' geodesics and the "straight" line between Q and R_i is η's geodesics, for example R_1 and R_4, then;

$$D^\theta(P \to Q) = D^\theta(P \to R_i) + D^\theta(R_i \to Q)$$

```
dR1 <- dnorm(x,R1.ms[1],R1.ms[2])
dR2 <- dnorm(x,R2.ms[1],R2.ms[2])
dR3 <- dnorm(x,R3.ms[1],R3.ms[2])
dR4 <- dnorm(x,R4.ms[1],R4.ms[2])

my.divergence(dP,dQ,dx,alpha=1)
```

```
## [1] 0.9318528
```

```
my.divergence(dP,dR1,dx,alpha=1) + my.divergence(dR1,dQ,dx,alpha=1)
```

```
## [1] 0.9318528
```

```
my.divergence(dP,dR4,dx,alpha=1) + my.divergence(dR4,dQ,dx,alpha=1)
```

```
## [1] 0.9318528
```

- PとR_iとはηでの「直線(測地線)」で結ばれ、R_iとQとはθでの「直線(測地線)」で結ばれているとき、以下の関係にある。具体的には、$R2, R3$がこれに相当する。Assume a point R_i for P and Q, where R_i should meet; the "straight"" line between P and R_i is η' geodesics and the "straight" line between Q and R_i is θ's geodesics, for example R_2 and R_3, then;

$$D^\eta(P \to Q) = D^\eta(P \to R_i) + D^\eta(R_1 \to Q)$$

```
my.divergence(dP,dQ,dx,alpha=-1)
```

```
## [1] 0.3493972
```

```
my.divergence(dP,dR2,dx,alpha=-1) + my.divergence(dR2,dQ,dx,alpha=-1)
```

```
## [1] 0.3493972
```

```
my.divergence(dP,dR3,dx,alpha=-1) + my.divergence(dR3,dQ,dx,alpha=-1)
```

```
## [1] 0.3493972
```

- 斜辺を共有する直角三角形が円周(球面)をなすことを利用すると、双対平坦空間では、2つの円周(球面)が作られる。Right-angled triangles sharing a oblique segment make a circle(sphere surface). Therefore in the dual flat space, two circles(spheres) appear.

```
fr <- matrix(c(-0.5,0,0.5,0.5),byrow=TRUE,2,2)
plot(fr,asp=1,col=0,xlim=c(0,0.6),ylim=c(0.4,1))
t <- seq(from=0,to=1,length=100) * 2*pi

for(i in 1:length(theta1)){
  points(t,sqrt(t/theta1[i]),type="l",col=2)
  points(-t,sqrt(t/theta1[i]),type="l",col=2)


}

for(i in 1:length(theta2)){
    abline(h=sqrt(-1/(2*theta2[i])),col=3)
}
for(i in 1:length(eta1)){
    abline(v=eta1[i],col=4)
}
t <- seq(from=0,to=1,length=100) * 2*pi
for(i in 1:length(eta2)){
    points(eta2[i] * cos(t),eta2[i] * sin(t),type="l",col=5)
}


P.ms <- c(0.4,0.4)
points(P.ms[1],P.ms[2],pch=20,cex=5)
text(P.ms[1]+0.1,P.ms[2],"P")
Q.ms <- c(0.6,0.8)
points(Q.ms[1],Q.ms[2],pch=20,cex=5,col=2)
text(Q.ms[1]+0.1,Q.ms[2],"Q")
P.thetas <- my.ms2theta(P.ms[1],P.ms[2])
P.etas <- my.ms2eta(P.ms[1],P.ms[2])
Q.thetas <- my.ms2theta(Q.ms[1],Q.ms[2])
Q.etas <- my.ms2eta(Q.ms[1],Q.ms[2])
points(t,sqrt(t/P.thetas[1]),type="l",col=4,lwd = 3)
points(t,sqrt(t/Q.thetas[1]),type="l",col=4,lwd = 3)

points(sqrt(P.etas[2]) * cos(t),sqrt(P.etas[2])* sin(t),type="l",col=1,lwd = 3)
points(sqrt(Q.etas[2]) * cos(t),sqrt(Q.etas[2])* sin(t),type="l",col=1,lwd = 3)



abline(h=sqrt(-1/(2*P.thetas[2])),col=4,lwd=3)
abline(h=sqrt(-1/(2*Q.thetas[2])),col=4,lwd=3)

abline(v=P.etas[1],col=1,lwd=3)
abline(v=Q.etas[1],col=1,lwd=3)

points(R1.ms[1],R1.ms[2],pch=20,cex=5,col=3)
points(R2.ms[1],R2.ms[2],pch=20,cex=5,col=4)
points(R3.ms[1],R3.ms[2],pch=20,cex=5,col=5)
points(R4.ms[1],R4.ms[2],pch=20,cex=5,col=6)
text(R1.ms[1]+0.1,R1.ms[2],"R1")
text(R2.ms[1]-0.1,R2.ms[2],"R2")
text(R3.ms[1]-0.1,R3.ms[2],"R3")
text(R4.ms[1]+0.1,R4.ms[2],"R4")

t <- seq(from=0,to=1,length=100)
line.theta1 <- P.thetas[1] + t * (Q.thetas[1]-P.thetas[1])
line.theta2 <- P.thetas[2] + t * (Q.thetas[2]-P.thetas[2])
line.eta1 <- P.etas[1] + t * (Q.etas[1]-P.etas[1])
```

```
line.eta2 <- P.etas[2] + t * (Q.etas[2]-P.etas[2])
line.theta.ms <- matrix(0,length(t),2)
line.eta.ms <- matrix(0,length(t),2)
for(i in 1:length(t)){
  line.theta.ms[i,] <- my.theta2ms(line.theta1[i],line.theta2[i])
  line.eta.ms[i,] <- my.eta2ms(line.eta1[i],line.eta2[i])
}
points(line.theta.ms[,1],line.theta.ms[,2],type="l",col=4,lwd=5)
points(line.eta.ms[,1],line.eta.ms[,2],type="l",col=2,lwd=2,lty=2)

KLD1 <- my.divergence(dP,dQ,dx,alpha=1)
KLD2 <- my.divergence(dP,dQ,dx,alpha=-1)

dx <- x[2]-x[1]
n.r <- 10000
r.ms <- cbind(runif(n.r)*1,runif(n.r)*0.7+0.3)

kld1 <- kld2 <- rep(0,length(r.ms[,1]))

for(i in 1:length(r.ms[,1])){
    tmp.d <- dnorm(x,r.ms[i,1],r.ms[i,2])
    kld1[i] <- my.divergence(dP,tmp.d,dx,alpha=1) + my.divergence(tmp.d,dQ,dx,alpha=1)

    kld2[i] <- my.divergence(dP,tmp.d,dx,alpha=-1) + my.divergence(tmp.d,dQ,dx,alpha=-1)
    col <- 1
    #if(col>1)col<-1
    #if(col<0)col<-0
    if(abs(kld1[i]-KLD1)<0.01)col <- 2
    if(abs(kld2[i]-KLD2)<0.01)col <- 3
    if(col>1)points(r.ms[i,1],r.ms[i,2],pch=20,col=col)
}
```
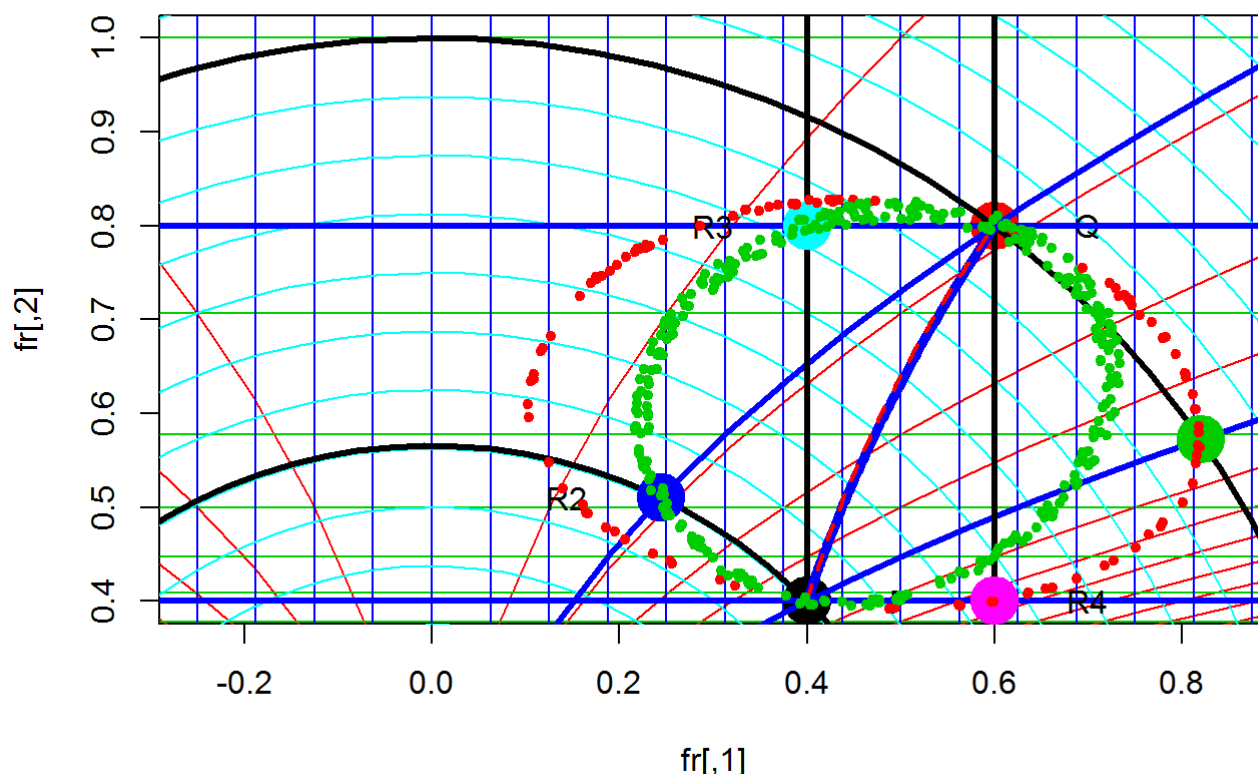
2.8 練習問題 Exercise

- PとQとを入れ替えて、拡張ピタゴラスの定理を確かめてみよ。Exchange P and Q and check the theorem.

- 上記で選ばれたR_i, αの組み合わせと異なる組み合わせでは、拡張ピタゴラスの定理が成立しないことを確かめよ。Check the sum does not fit if you apply combinations of R_i and α not mentioned above.

3 推定 Estimation

- 今、観測データが指し示す確率分布が点Pに相当しているとする。Assume your data set indicates a distribution corresponding to a point P.

- モデルを考えていてそのモデルでは、$\eta_2 = a$という固定値だとする。そのときに点Pから、$\eta_2 = a$という部分空間への最短ダイバージェンスをとるとする。You want to apply a model where η_2 value is fixed at a value a. You want to fit your data set or P to a subspace where $\eta_2 = a$. You can draw a shortest path from P to the subspace. The shortest path should have "shortest" divergence.

- そのような点は、R_1である。そのような点R1を探すにあたって、θ_1が一定の直線に沿って探せばよい。In case of the expample shown above, the point to be estimated is R_1. You can reach to R_1 by taking a straight line of θ_1 being constant.

- これは推定の情報幾何的解釈(θ, η座標系を用いた解釈)の一例である。This is an interpretation of estimation taski in information geometry world with θ and η coordinate systems.