演習問題解答

prof.joe.suzuki

May 2025

1　確率と統計の基礎

1. $\Omega \cup \{\} = \Omega$ かつ $\Omega \cap \{\} = \{\}$ より、3. を適用して、$P(\Omega) = P(\Omega) + P(\{\})$ が成立する。これに 2. を適用して、(1.1) が成立する。また、

$$\begin{aligned} A \cup B &= (A - A \cap B) \cup (B - A \cap B) \cup A \cap B \\ A &= (A - A \cap B) \cup A \cap B \\ B &= (B - A \cap B) \cup A \cap B \end{aligned}$$

はいずれも排反な事象の和であるから、各項の確率について、最初の式から最後の 2 式を引くと、

$$P(A \cup B) - P(A) - P(B) = -P(A \cap B)$$

より、(1.2) が成立する。

2. $x = r\cos\theta,\ y = r\sin\theta,\ J = r,\ r \geq 0,\ 0 \leq \theta \leq \pi/2$ を代入して、

$$\begin{aligned} \{\int_{-\infty}^{\infty} \exp(-\frac{x^2}{2})dx\}^2 &= \{\int_{-\infty}^{\infty} \exp(-\frac{x^2}{2})dx\}\{\int_{-\infty}^{\infty} \exp(-\frac{y^2}{2})dy\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\frac{x^2+y^2}{2})dxdy = 4\int_0^{\infty}\int_0^{\infty} \exp(-\frac{x^2+y^2}{2})dxdy \\ &= 4\int_0^{\pi/2}\int_0^{\infty} \exp(-\frac{r^2}{2})rdrd\theta = 4 \cdot \frac{\pi}{2} \cdot \left[-e^{-r^2/2}\right]_0^{\infty} = 2\pi \end{aligned}$$

が得られる。

3. (1.8) を部分積分すると、

$$\Gamma(\alpha+1) = \int_0^{\infty} t^{\alpha}e^{-t}dt = \int_0^{\infty} t^{\alpha}(-e^{-t})'dt = \left[-t^{\alpha}e^{-t}\right]_0^{\infty} + \alpha\int_0^{\infty} t^{\alpha-1}e^{-t}dt = \alpha\Gamma(\alpha)$$

また、$t = x^2/2$ とおくと、

$$\sqrt{\frac{\pi}{2}} = \int_{-\infty}^{\infty} e^{-x^2/2}dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2t}}e^{-t}dt$$

となり、$\Gamma(1/2) = \sqrt{\pi}$ が得られる。

4. $\mu := \mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)dx,\ \sigma^2 := \mathbb{V}[X] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$ とおくと、$\int_{-\infty}^{\infty} f(x)dx = 1$ を用いて、

以下の変形ができる。

$$\mathbb{E}[aX+b] = \int_{-\infty}^{\infty}(ax+b)f(x)dx = a\int_{-\infty}^{\infty}xf(x)dx + b\int_{-\infty}^{\infty}f(x)dx = a\mathbb{E}[X]+b$$
$$\mathbb{V}[aX+b] = \int_{-\infty}^{\infty}\{(ax+b)-(a\mu+b)\}^2 f(x)dx = a^2\int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx = a^2\mathbb{V}[X]$$
$$\mathbb{V}[X] = \int_{-\infty}^{\infty}(x^2-2x\mu+\mu^2)f(x)dx = \mathbb{E}[X^2]-2\mu\mathbb{E}[X]+\mu^2 = \mathbb{E}[X^2]-\mu^2$$

また、

$$\mathbb{E}\left[A\left[\begin{array}{c} X \\ Y \end{array}\right]+B\right] = A\,\mathbb{E}\left[\begin{array}{c} X \\ Y \end{array}\right]+B = A\left[\begin{array}{c} \mathbb{E}_X[X] \\ \mathbb{E}_Y[Y] \end{array}\right]+B$$

より、$\mu_X := \mathbb{E}_X[X]$, $\mu_Y := \mathbb{E}_Y[Y]$ とおくと、

$$\left(A\left[\begin{array}{c} X \\ Y \end{array}\right]+B\right) - \left\{A\left[\begin{array}{c} \mathbb{E}_X[X] \\ \mathbb{E}_Y[Y] \end{array}\right]+B\right\} = A\left[\begin{array}{c} X-\mu_X \\ Y-\mu_Y \end{array}\right]$$

となり、共分散行列は

$$A\left[\begin{array}{c} X-\mu_X \\ Y-\mu_Y \end{array}\right]\left\{A\left[\begin{array}{c} X-\mu_X \\ Y-\mu_Y \end{array}\right]\right\}^{\top} = A\left[\begin{array}{c} X-\mu_X \\ Y-\mu_Y \end{array}\right][X-\mu_X, Y-\mu_Y]A^{\top}$$

の平均をとったものであるから、$A\Sigma A^{\top}$ となる。

5. 任意の $a,b=0,1,\ldots$ で (1.17) が成立することを仮定し、整数 $x,y\geq 0$ を任意とする。(1.17) を $0\leq a\leq x$, $0\leq b\leq y$ で加えると (1.16) が成立する。任意の $x,y=0,1,\ldots$ で (1.16) が成立することを仮定し、$a,b\in\mathbb{N}$ を任意とする。$x=a$ を固定して、$y=b-1,b$ を代入して

$$P(X\leq a, Y\leq b) = P(X\leq a)P(Y\leq b), \qquad P(X\leq a, Y\leq b-1) = P(X\leq a)P(Y\leq b-1)$$

より、$P(X\leq a, Y=b) = P(X\leq a)P(Y=b)$ が得られる。同様に、

$$P(X\leq a-1, Y=b) = P(X\leq a-1)P(Y=b), \qquad P(X\leq a, Y=b) = P(X\leq a)P(Y=b)$$

が成立し、(1.17) が得られる。

6. $cov(Y,Y)=\mathbb{V}[Y]$ であるから、

$$\begin{aligned} cov(Y,Z) &= cov(Y, X-\frac{cov(X,Y)}{\mathbb{V}[Y]}Y) = cov(Y,X) - cov(Y, \frac{cov(X,Y)}{\mathbb{V}[Y]}Y) \\ &= cov(Y,X)-cov(Y,X) = 0 \end{aligned}$$

が成立する。これと、$\mathbb{V}[Z]\geq 0$ より、

$$\mathbb{V}[X] = \mathbb{V}[Z+\frac{cov(X,Y)}{\mathbb{V}[Y]}Y] = \mathbb{V}[Z]+\mathbb{V}[\frac{cov(X,Y)}{\mathbb{V}[Y]}Y] \geq \frac{cov(X,Y)^2}{\mathbb{V}[Y]}$$

が成立する。したがって、相関係数 $\dfrac{cov(X,Y)}{\mathbb{V}[X]\mathbb{V}[Y]}$ は -1 以上 1 以下になる。

7. $f_{XY}(x,y) = f_X(x) f_{Y|X}(x,y)$ の形に変形すると、

$$
\begin{aligned}
&\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\} \\
=&\frac{1}{\sqrt{2\pi\sigma_X^2}}\cdot\frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}}\cdot\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right. \\
&\left.-\frac{1}{2(1-\rho^2)}\left[\rho^2\left(\frac{x-\mu_X}{\sigma_X}\right)^2-2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\} \\
=&\frac{1}{\sqrt{2\pi\sigma_X^2}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right\}\frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_Y}{\sigma_Y}-\rho\frac{x-\mu_X}{\sigma_X}\right)^2\right\}
\end{aligned}
$$

が得られる。実際、この式が

$$
X \sim N(\mu_X, \sigma_X^2), \qquad Y \mid X = x \sim N(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X), (1-\rho^2)\sigma_Y^2)
$$

の確率密度関数の積になっていて、後者を y で積分すると 1 になる。

8. 正規分布では、(1.24) より、$\Psi_X(0) = 1$, $|\Psi_X(t)| = e^{-\sigma^2/t^2}|e^{i\mu t}| \le 1$, $\Psi'_X(0) = i\mu$, $\Psi'_X(t) = (i\mu - \sigma^2 t)\exp(i\mu t - \frac{\sigma^2 t^2}{2})$, $\Psi''_X(t) = -\sigma^2\exp(i\mu t - \frac{\sigma^2t^2}{2}) + (i\mu-\sigma^2 t)^2\exp(i\mu t - \frac{\sigma^2t^2}{2})$, $\Psi''_X(0) = -(\sigma^2+\mu^2)$ が成立する。二項分布では、$\Psi_X(t) = (1-p+pe^{it})^n$ となる。$\Psi_X(0) = (1-p+p)^n = 1$, $|\Psi_X(t)| = |1-p+pe^{it}|^n \le (1-p+p)^n = 1$ (三角不等式を用いた),

$$
\begin{aligned}
\Psi'_X(t) &= n(1-p+pe^{it})^{n-1}(ip)e^{it} = inp(1-p+pe^{it})^{n-1}e^{it} \\
\Psi''_X(t) &= -npe^{it}\left(1-p+pe^{it}\right)^{n-1} + inpe^{it}\cdot(n-1)\left(1-p+pe^{it}\right)^{n-2}\cdot pie^{it} \\
&= -npe^{it}\left(1-p+pe^{it}\right)^{n-1} - np^2(n-1)e^{2it}\left(1-p+pe^{it}\right)^{n-2}
\end{aligned}
$$

$\Psi'_X(0) = inp$, $\Psi''_X(0) = -np - n(n-1)p^2 = -np(1-p) - (np)^2$ より、$\mu = np$, $\sigma^2 = np(1-p)$

9. (1.11) の f_X を用いて、特性関数は次で定義される：

$$
\begin{aligned}
\Psi_X(t) =&\mathbb{E}[e^{itX}] = \int_0^\infty e^{itx} f_X(x)dx = \frac{1}{2^{n/2}\Gamma(n/2)}\int_0^\infty x^{n/2-1}\exp\left(-\frac{x}{2}+itx\right)dx. \\
=&\frac{1}{2^{n/2}\Gamma(n/2)}\int_0^\infty x^{n/2-1}\exp\left(-\frac{1-2it}{2}x\right)dx.
\end{aligned}
$$

$y := (1-2it)x$ とおくと、$dx = (1-2it)^{-1}dy$,

$$
\begin{aligned}
\Psi_X(t) &= (1-2it)^{-n/2}\cdot\frac{1}{2^{n/2}\Gamma(n/2)}\int_0^\infty y^{n/2-1}\exp\left(-\frac{y}{2}\right)dy \\
&= (1-2it)^{-n/2}\int_0^\infty f_X(y)dy = (1-2it)^{-n/2}
\end{aligned}
$$

10. コインを n 回投げ、表を 1, 裏を 0 として加算することで、表の出る回数を S_n とする。その平均と分散はそれぞれ、$\mathbb{E}[S_n] = \dfrac{n}{2}$, $\quad \mathbb{V}[S_n] = \dfrac{n}{4}$ であり、中心極限定理より、

$$
Z_n := \frac{S_n - \frac{n}{2}}{\sqrt{n/4}} = \frac{2}{\sqrt{n}}\left(S_n - \frac{n}{2}\right)
$$

は、n が十分大きければ標準正規分布 $N(0,1)$ に近づく。したがって、1 個の標準正規乱数を生成するには、コインを n 回投げて合計 S_n を計算し、Z_n を得ればよい。

```r
set.seed(123)   # 再現性のため
n <- 50
N <- 1000
Sn <- replicate(N, sum(rbinom(n, 1, 0.5)))
Zn <- (2 / sqrt(n)) * (Sn - n / 2)

# ヒストグラムの描画
hist(Zn, breaks = 30, probability = TRUE,
     main = "標準正規分布に近似されたコイン乱数",
     xlab = "値", col = "lightblue", border = "white")

# 標準正規分布の密度関数を重ねて表示
curve(dnorm(x), col = "red", lwd = 2, add = TRUE)
```

11.
$$\int_0^{\pi} (\sin\theta)^{2(a-1)} d\theta = 2\int_0^{\pi/2} (1-\cos^2\theta)^{2(a-1)} d\theta = 2\int_0^1 (1-t)^{a-1} \frac{dt}{\sqrt{2t(1-t)}}$$
$$= \int_0^1 (1-t)^{a-\frac{3}{2}} t^{-\frac{1}{2}} dt = B(a-\frac{1}{2}, \frac{1}{2}) = \frac{\Gamma(a-\frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(a)}$$

12. 観測データ $\boldsymbol{x} = (x_1, \ldots, x_k)$ に対する尤度関数および対数尤度関数はそれぞれ以下のようになる。

$$L = \frac{n!}{x_1! \cdots x_k!} \prod_{i=1}^{k} p_i^{x_i} \quad (\text{ただし } \sum_{i=1}^{k} p_i = 1,\ p_i > 0)$$

$$l = \log L = \log n! - \sum_{i=1}^{k} \log x_i! + \sum_{i=1}^{k} x_i \log p_i$$

最尤推定量は、制約 $\sum p_i = 1$ のもとで、ℓ を最大化して得られる。

$$\mathcal{L} = \sum_{i=1}^{k} x_i \log p_i + \lambda \left(1 - \sum_{i=1}^{k} p_i\right)$$

を p_i で偏微分して 0 とおくと、$\dfrac{\partial \mathcal{L}}{\partial p_i} = \dfrac{x_i}{p_i} - \lambda = 0$ より、$p_i = \dfrac{x_i}{\lambda}$ が得られる。これを $\sum_{i=1}^{k} p_i = 1$ に代入して $\displaystyle\sum_{i=1}^{k} p_i = \sum_{i=1}^{k} \frac{x_i}{\lambda} = \frac{n}{\lambda} = 1$ より、$\lambda = n$ となり、例 12 の最尤推定量が得られる。

13. $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$（独立）というようなモデルをおく。対数尤度最大は、誤差平方和 $\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$ を最小と等価になる。最初にこの値を β_0 で微分して 0 とおくと、$-\sum_{i=1}^{n}(y_i - \beta_0 - x_i\beta_1) = 0$ より、

$$\beta_0 = \frac{1}{n}\sum_{i=1}^{n} (y_i - x_i \hat{\beta}_1) = \bar{y} - \beta_1 \bar{x}$$

とできる。したがって、仮に $\hat{\beta}_1$ が先に求まれば、$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ の値も求まる。まず、中心化

$$(x_i, y_i) - (\bar{x}, \bar{y}) \to (x_i, y_i)$$

を施す。$\bar{x} = \bar{y} = 0$ が成立する。このとき、回帰直線は原点を通る $(\beta_0 = 0)$ と仮定してよい。そこで、$\sum_{i=1}^{n}(y_i - x_i\beta_1)^2$ を β_1 で微分して 0 とおくと、

$$0 = -2\sum x_i(y_i - \beta_1 x_i) = -2\left(\sum x_i y_i - \beta_1 \sum x_i^2\right) = 0$$

となる。これより、$\hat{\beta}_1 = \dfrac{\sum_i x_i y_i}{\sum_i x_i^2}$ が得られる。中心化前の表記にもどすと、下記が得られる。

$$\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

14. (1.29) は $\dfrac{\partial L}{\partial \sigma^2} = 0$ から得られる。その式を、対数尤度 L の定義式に代入すると (1.30) が得られる。

15. $X_1, \ldots, X_n$ は独立に $N(\mu, \sigma^2)$ に従うとする。標本平均は $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ となる。このとき、正規分布の加法性により、$\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$ となる。よって、その平均を μ、分散を σ^2/n で標準化すると：

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$$

16. $\mu = 5$ のとき（有意水準 $\alpha = 0.05$）に 1 回だけ検定する。

```
set.seed(123)
n <- 10
mu0 <- 5
sigma <- 3
alpha <- 0.05
# データ生成
x <- rnorm(n, mean=5, sd=sigma)
# 検定統計量
z <- sqrt(n)*(mean(x) - mu0)/sigma
pval <- 2*pnorm(-abs(z))  # 両側検定
# 出力
cat("検定統計量 Z =", z, "\n")
cat("P 値 =", pval, "\n")
cat("帰無仮説", ifelse(pval < alpha, "棄却", "採択"), "\n")
```

次に、2 番目の問題で検出力を推定する。$\mu = 7$ のとき、検出率を 1000 回試行すると、以下のようになる。

```
set.seed(456)
B <- 1000
detect_count <- 0
for (i in 1:B) {
  x <- rnorm(n, mean=7, sd=sigma)
  z <- sqrt(n)*(mean(x) - mu0)/sigma
  pval <- 2*pnorm(-abs(z))
  if (pval < alpha) detect_count <- detect_count + 1
}
power_est <- detect_count / B
cat("推定検出率（検出力） =", power_est, "\n")
```

$\alpha = 0.01$ で再度実行する場合は、`alpha <- 0.01` に変更して再試行する。

17. 最初の行は、$\det(B) = \prod_{i=1}^{p} u_{i.i}^2$、$\mathrm{tr}(B) = \sum_{i=1}^{p} \sum_{j=1}^{i} u_{i,j}^2$、$B$ を下三角成分 (対角成分を含む) にするための Jacobian$\prod_{i=1}^{p} u_{i,i}^{p+1-i}$ を代入している。次の行では $\det(B)$ と Jacobian を統合して、$\prod_{i=1}^{p} u_{i,i}^{2a-i}$ というようにまとめている。3 行目では対角成分と非対角成分にわけて、2^p は対角成分の (p 個の) 積の中で 2 倍にしている。4 行目の前半は、i, j によらない定数を $p(p-1)/2$ 個かけている。5 行目はガウス積分と Gamma 関数の定義によった。

18. (a) $X_p = Y_p Y_p^\top$ を考える。下三角行列 Y_p の下三角成分を $\tilde{Y}_p$、対称行列 X_p の下三角成分を $\tilde{X}_p$ と表す。このとき、

$$Y_p = \begin{bmatrix} Y_{p-1} & 0 \\ y_{p-1} & y_{pp} \end{bmatrix} \quad \text{とおくと、} \quad X_p = Y_p Y_p^\top = \begin{bmatrix} X_{p-1} & Y_{p-1} y_{p-1}^\top \\ y_{p-1} Y_{p-1}^\top & y_{p-1} y_{p-1}^\top + y_{pp}^2 \end{bmatrix}$$

したがって、$\tilde{X}_p$ は、$\tilde{X}_{p-1}$（既知)、$Y_{p-1} y_{p-1}^\top$、$y_{p-1} y_{p-1}^\top + y_{pp}^2$ から構成される。これを使って、ヤコビ行列（偏微分行列）をブロック行列として整理すると

$$\begin{aligned} \frac{\partial \tilde{X}_p}{\partial \tilde{Y}_p} &= \begin{bmatrix} \dfrac{\partial \tilde{X}_{p-1}}{\partial \tilde{Y}_{p-1}} & \dfrac{\partial \tilde{X}_{p-1}}{\partial y_{p-1}} & \dfrac{\partial \tilde{X}_{p-1}}{\partial y_{pp}} \\ \dfrac{\partial}{\partial \tilde{Y}_{p-1}} Y_{p-1} y_{p-1}^\top & \dfrac{\partial}{\partial y_{p-1}} Y_{p-1} y_{p-1}^\top & \dfrac{\partial}{\partial y_{pp}} Y_{p-1} y_{p-1}^\top \\ \frac{\partial}{\partial \tilde{Y}_{p-1}} (y_{p-1} y_{p-1}^\top + y_{pp}) & \frac{\partial}{\partial y_{p-1}} (y_{p-1} y_{p-1}^\top + y_{pp}) & \frac{\partial}{\partial y_{pp}} (y_{p-1} y_{p-1}^\top + y_{pp}) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial \tilde{X}_{p-1}}{\partial \tilde{Y}_{p-1}} & * & * \\ 0 & Y_{p-1}^\top & * \\ 0 & 0 & 2y_{pp} \end{bmatrix} \end{aligned}$$

(b) 帰納法で以下の式を示す。

$$\det\left(\frac{\partial \tilde{X}_p}{\partial \tilde{Y}_p}\right) = 2^p \prod_{i=1}^{p} y_{ii}^{p+1-i}$$

まず、$p = 1$ では、$X_1 = y_{11}^2$ であれば $\dfrac{\partial X_1}{\partial y_{11}} = 2y_{11}$ であり、$\det = 2y_{11}$ となり、正しい。次に、

$$\det\left(\frac{\partial \tilde{X}_{p-1}}{\partial \tilde{Y}_{p-1}}\right) = 2^{p-1} \prod_{i=1}^{p-1} y_{ii}^{p-i}$$

を仮定すると、ヤコビアンのブロック構造が**下三角行列**なので、行列式は対角項の積であり、

$$\det\left(\frac{\partial \tilde{X}_p}{\partial \tilde{Y}_p}\right) = \det\left(\frac{\partial \tilde{X}_{p-1}}{\partial \tilde{Y}_{p-1}}\right) \cdot \det(Y_{p-1}^\top) \cdot 2y_{pp} = 2^{p-1} \prod_{i=1}^{p-1} y_{ii}^{p-i} \cdot \prod_{i=1}^{p-1} y_{ii} \cdot 2y_{pp} = 2^p \prod_{i=1}^{p} y_{ii}^{p+1-i}$$

となるので、主張が成立する。

(c) $X_p = Y_p Y_p^\top$ のトレースを直接計算する。

$$\mathrm{tr}(X_{p-1}) = \sum_{i=1}^{p-1} \sum_{j=1}^{i} y_{ij}^2, \qquad X_p = \begin{bmatrix} X_{p-1} & * \\ * & y_{p-1} y_{p-1}^\top + y_{pp}^2 \end{bmatrix} \qquad \text{とおくと、}$$

$$\mathrm{tr}(X_p) = \mathrm{tr}(X_{p-1}) + y_{p-1} y_{p-1}^\top + y_{pp}^2 = \sum_{i=1}^{p-1} \sum_{j=1}^{i} y_{ij}^2 + \sum_{j=1}^{p-1} y_{p,j}^2 + y_{pp}^2 = \sum_{i=1}^{p} \sum_{j=1}^{i} y_{ij}^2$$

より、主張が成立する。

2　グラフィカルモデル

19. (a) 辺がない
    (b) $Y \leftarrow Z, \quad Z \rightarrow Y$
    (c) $Z \leftarrow X, \quad X \rightarrow Z$
    (d) $X \leftarrow Y, \quad Y \rightarrow X$
    (e) $Z \rightarrow X \rightarrow Y, \quad Z \leftarrow X \leftarrow Y, \quad Z \leftarrow X \rightarrow Y$
    (f) $X \rightarrow Y \rightarrow Z, \quad X \leftarrow Y \leftarrow Z, \quad X \leftarrow Y \rightarrow Z$
    (g) $Y \rightarrow Z \rightarrow X, \quad Y \leftarrow Z \leftarrow X, \quad Y \leftarrow Z \rightarrow X$
    (h) $X \rightarrow Y \leftarrow Z$
    (i) $Y \rightarrow Z \leftarrow X$
    (j) $Z \rightarrow X \leftarrow Y$
    (k) $\{Z \leftarrow X \rightarrow Y, Y \rightarrow Z\}$, $\{Z \leftarrow X \rightarrow Y, Z \rightarrow Y\}$
    $\{X \leftarrow Y \rightarrow Z, X \rightarrow Z\}$, $\{X \leftarrow Y \rightarrow Z, Z \rightarrow X\}$
    $\{Y \leftarrow Z \rightarrow X, Y \rightarrow X\}$, $\{Y \leftarrow Z \rightarrow X, X \rightarrow Y\}$
20. (a) $P(X)P(Y)P(Z)$
    (b) $P(X)P(Y,Z)$
    (c) $P(Y)P(X,Z)$
    (d) $P(Z)P(X,Y)$
    (e) $P(Z,X)P(X,Y)/P(X)$
    (f) $P(X,Y)P(Y,Z)/P(Y)$
    (g) $P(Y,Z)P(Z,X)/P(Z)$
21. (d) だけが有向分離される。
22. (a)(b) は真、他は偽
23. (a)(b)(c)(e)(f) は真、他は偽
24. (c) に $Z \rightarrow W$ もしくは $Z \leftarrow W$ を加えても $X \perp\!\!\!\perp_G Y \mid \{Z, W\}$ しか成立しない。$X \rightarrow Y$ もしくは $Y \rightarrow X$ を加えると、自明な分離関係しか残らない。(b)(d) では逆に $X \rightarrow Y$ もしくは $Y \rightarrow X$ を加えても $Z \perp\!\!\!\perp_G W \mid \{X, Y\}$ しか成立しない。$Z \rightarrow W$ もしくは $Z \leftarrow W$ を加えると、自明な分離関係しか残らない。(b) で $X \rightarrow Z$, $X \rightarrow W$, $Y \rightarrow Z$, $Y \rightarrow W$ の 1 個を削除するとそれぞれ、$X \perp\!\!\!\perp_G Z \mid Y$, $X \perp\!\!\!\perp_G W \mid Y$, $Y \perp\!\!\!\perp_G Z \mid X$, $Z \perp\!\!\!\perp_G W \mid Y$ が成立し、意図していない条件付き独立性を表現することになる。(c) で $X \rightarrow Z$, $X \rightarrow W$, $Z \rightarrow Y$, $W \rightarrow Y$ の 1 個を削除するとそれぞれ、$X \perp\!\!\!\perp_G Z \mid W$, $X \perp\!\!\!\perp_G W \mid Z$, $Y \perp\!\!\!\perp_G Z \mid W$, $Y \perp\!\!\!\perp_G W \mid Z$ が成立し、意図していない条件付き独立性を表現することになる。(d) で $X \rightarrow Z$, $X \rightarrow W$, $Z \rightarrow Y$, $Y \rightarrow W$ の 1 個を削除するとそれぞれ、$X \perp\!\!\!\perp_G Z \mid Y$, $X \perp\!\!\!\perp_G W \mid Z$, $Y \perp\!\!\!\perp_G Z \mid X$, $Y \perp\!\!\!\perp_G W \mid Z$ が成立し、意図していない条件付き独立性を表現することになる。
25. (a) $A \perp\!\!\!\perp_G C \mid D$ かつ $B \perp\!\!\!\perp_G C \mid D$ のとき、$a \in A$, $c \in C$ を結ぶすべての経路、$b \in B$, $c \in C$ を結ぶすべての経路が D で分離される。したがって、任意の $x \in A \cup B$, $c \in C$ に対し、x と c を結ぶ経路が D で分離される。すなわち、
    $$(A \cup B) \perp\!\!\!\perp_G C \mid D$$

が成り立つ。逆に、$(A \cup B) \perp\!\!\!\perp_G C \mid D$ であれば、特に $A \subseteq A \cup B,\ B \subseteq A \cup B$ であることから、

$$A \perp\!\!\!\perp_G C \mid D, \quad B \perp\!\!\!\perp_G C \mid D$$

が成り立つ。

(b) 任意の $k \in C$ に対して、

$$A \not\perp\!\!\!\perp_G k \quad \text{かつ} \quad k \not\perp\!\!\!\perp_G B$$

であれば、ある $a \in A,\ b \in B$ を結ぶ経路が 任意の $k \in C$ で分離されないことになる。このことは、$A \perp\!\!\!\perp_G B \mid C$ と矛盾する。

26. 例えば

$$\frac{P(W,X)P(X,Z)}{P(X)} = P(X)P(W \mid X)P(Z \mid X)$$

を X で和をとると、$P(W \mid X)P(Z \mid X)$ の平均になるので、$P(W)P(Z)$ となる。また、(2.12) からと

$$P(X,Y \mid Z,W) = P(X,Y,Z,W)/P(Z,W)$$

$$P(X \mid Z,W) = P(X,Z,W)/P(Z,W), \qquad P(Y \mid Z,W) = P(Y,Z,W)/P(Z,W)$$

から、(2.13) の等号が成立する。

27. BN は $Y \to X \leftarrow Z,\ Z \to Y \leftarrow X,\ X \to Z \leftarrow Y$、MN は完全グラフ。

28. (a) 5 行目から 9 行目で上三角成分を対応する下三角成分にコピーするため、対称性が担保される。

(b) `u[j] = 1` のとき、$j \in U$ を意味する。

(c) `u[j] = 0` のとき、`parent[j] = -1` が設定され、木の根として扱われる。

29. (a) 一般に $x > 0$ で不等式 $\log x \le x - 1$ が成立し、$x = 1$ のとき等号が成立する。したがって、

$$-\log \frac{P'(x_1,\ldots,x_p)}{P(x_1,\ldots,x_p)} \ge 1 - \frac{P'(x_1,\ldots,x_p)}{P(x_1,\ldots,x_p)}$$

が成立し、となり、$P(x_1,\ldots,x_p)$ をかけて $(x_1,\ldots,x_p)$ に関して和をとると (2.12) が非負であるという不等式が得られる。また、等号成立は、

$$P(x_1,\ldots,x_p) = P'(x_1,\ldots,x_p)$$

となるが、(2.12) の各項が非負の値をとることから、すべての $(x_1,\ldots,x_p)$ について一致する必要がある。

(b) 相互情報量 $I(\cdot,\cdot)$ は同時分布 $P(\cdot,\cdot)$ と、独立であると仮定した積分布 $P(\cdot)P(\cdot)$ の Kullback-Leibler 情報量 として表され、常に非負の値をとる。等号が成立するのは、分子と分母が一致するとき、すなわち各 i,j について

$$P(\cdot,\cdot) = P(\cdot)P(\cdot)$$

これは 2 個の確率変数が独立であることを意味する。

(c)

$$\log P'(X_1 = x_1,\ldots,X_p = x_p) = \sum_{k \in V} \log P(X_k = x_k) + \sum_{\{i,j\} \in E} \log \frac{P(X_i = x_i, X_j = x_j)}{P(X_i = x_i)P(X_j = x_j)}$$

の両辺の平均をとる。すなわち、$P(X_1 = x_1,\cdots,X_p = x_p)$ をかけて和を取ると、

$$\sum_{x_1} \cdots \sum_{x_p} P(X_1 = x_1,\cdots,X_p = x_p) \log P'(X_1 = x_1,\ldots,X_p = x_p) = -\sum_{k \in V} H(k) + \sum_{\{i,j\} \in E} I(i,j)$$

これと、$D(P||P')$, $H(1,\cdots,p)$ の定義から (2.13) 得られる。

3　カーネルによる独立性、条件付き独立性の検定

30.
```r
# Epanechnikov カーネル関数
D <- function(u) {
  u <- abs(u)
  ifelse(u <= 1, 0.75 * (1 - u^2), 0)
}

# RBF (Gaussian) カーネル関数
k <- function(x, y, sigma2) {
  exp(- (x - y)^2 / (2 * sigma2))
}

# Nadaraya-Watson 推定関数
f <- function(x_star, sigma2) {
  weights <- sapply(1:length(x), function(i) k(x_star, x[i], sigma2))
  sum(weights * y) / sum(weights)
}
```

31. 固有値は

$$K_\lambda - tI = \begin{bmatrix} 3/4 & 9/16 & 0 \\ 9/16 & 3/4 & 9/16 \\ 0 & 9/16 & 3/4 \end{bmatrix} - tI$$

の行列式が 0 となる t で、3 次方程式になる。その場合、3 個の解の積が K_λ の行列式に一致する。3 次の行列式の公式から

$$\left(\frac{3}{4}\right)^3 + 2\cdot 0\cdot\left(\frac{9}{16}\right)^2 - 0^2\cdot\frac{3}{4} - 2\cdot\left(\frac{9}{16}\right)^2\cdot\frac{3}{4} = -3^2/2^7$$

となり、負となる。したがって少なくとも 1 個の固有値が負となり、非負定値ではない。

32. 命題 9(Bochner の定理) から、$k(x,y)=\psi(x-y)$ が正定値カーネルとなるためには、ψ がある確率密度関数 $f(x)$ の特性関数 $\phi(t)=\displaystyle\int_{\mathbb{R}^p} e^{it^\top x} f(x)dx$ の形であればよい。

Gauss カーネルの定義は

$$k(x,y) = \exp\left(-\frac{1}{2\sigma^2}\|x-y\|^2\right) = \psi(x-y)$$

この関数 $\psi(\cdot)$ が、正規分布

$$f(x) = \frac{1}{(2\pi\sigma^2)^{p/2}}\exp\left(-\frac{\|x\|^2}{2\sigma^2}\right)$$

の特性関数

$$\phi(t) = \mathbb{E}[e^{it^\top X}] = \exp\left(-\frac{\sigma^2}{2}\|t\|^2\right)$$

の定数倍であるから、$k(x,y)=\psi(x-y)$ は $\psi(x-y)$ の形をしており、正定値カーネルである。

Laplace カーネルの定義は

$$k(x,y) = \frac{1}{2\pi}\cdot\frac{1}{\|x-y\|^2+\beta^2} = \psi(x-y)$$

この関数 $\psi(\cdot)$ が、Laplace 分布

$$f(x) = \frac{\beta}{2}\exp(-\beta|x|), \quad x \in \mathbb{R}$$

の特性関数

$$\phi(t) = \mathbb{E}[e^{itX}] = \frac{\beta^2}{\beta^2 + t^2}$$

の定数倍であるから、$k(x, y) = \frac{1}{2\pi} \cdot \frac{1}{\|x-y\|^2+\beta^2}$ は $\psi(x - y)$ の形をしており、正定値カーネルである。

```r
# Gauss カーネル (σ ^2 = 1)
gauss_kernel <- function(x, y) {
  exp(- (x - y)^2 / 2)
}

# Laplace カーネル (β = 1)
laplace_kernel <- function(x, y) {
  exp(- abs(x - y))
}
```

33. 対称である $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ が正定値カーネルであることは、任意の $N \in \mathbb{N}$、任意の点列 $x_1, \ldots, x_N \in \mathcal{X}$ と任意の係数列 $\alpha_1, \ldots, \alpha_N \in \mathbb{R}$ に対して、

$$\sum_{i=1}^{N}\sum_{j=1}^{n}\alpha_i\alpha_j k(x_i, x_j) \geq 0$$

が成り立つこととして定義される。$k(x_i, x_j) = x_i^\top x_j$ を代入すると、

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j x_i^\top x_j = \left(\sum_{i=1}^{N}\alpha_i x_i\right)^\top \left(\sum_{j=1}^{N}\alpha_j x_j\right) = \left\|\sum_{i=1}^{N}\alpha_i x_i\right\|^2 \geq 0$$

右辺はベクトルのノルムの 2 乗であり、常に非負である。

34.

$$H_0 := \left\{ f \mid f = \sum_{i=1}^{N}\alpha_i k(x_i, \cdot),\ \alpha_i \in \mathbb{R},\ x_i \in \mathcal{X} \right\}$$

とおく。$\alpha, \alpha_i \in \mathbb{R} \Longrightarrow \alpha\alpha_i \in \mathbb{R}$、$\alpha_i, \beta_i \in \mathbb{R} \Longrightarrow \alpha_i + \beta_i \in \mathbb{R}$ より、

$$\begin{aligned}
\alpha f =& \alpha\sum_{i=1}^{N}\alpha_i k(x_i, \cdot) = \sum_{i=1}^{N}\alpha\alpha_i k(x_i, \cdot) \in H_0 \\
f + g =& \sum_{i=1}^{N}\alpha_i k(x_i, \cdot) + \sum_{i=1}^{N}\beta_i k(x_i, \cdot) = \sum_{i=1}^{N}(\alpha_i + \beta_i)k(x_i, \cdot) \in H_0
\end{aligned}$$

H_0 における $\langle\cdot,\cdot\rangle_{H_0}$

$$\langle f, g\rangle_{H_0} := \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\beta_j k(x_i, x_j), \quad f = \sum_{i=1}^{N}\alpha_i k(x_i, \cdot), \quad g = \sum_{j=1}^{N}\beta_j k(x_j, \cdot)$$

について、$k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ の対称性から、対称性 $\langle f, g \rangle_{H_0} = \sum_i \sum_j \alpha_i \beta_j k(x_i, x_j) = \langle g, f \rangle_{H_0}$ が成立する。線形性は

$$
\begin{aligned}
\langle \alpha f + \beta g, h \rangle_{H_0} &= \langle \alpha \sum_i \alpha_i k(x_i, \cdot) + \beta \sum_i \beta_i k(x_i, \cdot), \sum_j \gamma_j \ k(x_j, \cdot) \rangle \\
=& \langle \sum_i (\alpha\alpha_i + \beta\beta_i) k(x_i, \cdot), \sum_j \gamma_j \ k(x_j, \cdot) \rangle = \sum_i \sum_j (\alpha\alpha_i + \beta\beta_i) \gamma_j k(x_i, x_j) \\
=& \alpha \sum_i \sum_j \alpha_i \gamma_j k(x_i, x_j) + \beta \sum_i \sum_j \beta_i \gamma_j k(x_i, x_j) = \alpha \langle f, h \rangle_{H_0} + \beta \langle g, h \rangle_{H_0}
\end{aligned}
$$

から、非負性は

$$
\langle f, f \rangle_{H_0} = \sum_i \sum_j \alpha_i \alpha_j k(x_i, x_j) = \| \sum_i \alpha k(x_i, \cdot) \|_{H_0}^2 \geq 0
$$

から。最後の条件は、

$$
\|f\|_{H_0}^2 = \| \sum_{i=1}^N \alpha_i k(x_i, \cdot) \|_{H_0}^2 = 0 \Longrightarrow \alpha_1 = \ldots = \alpha_N = 0 \Longrightarrow f = \sum_{i=1}^N \alpha_i k(x_i, \cdot) = 0
$$

より成立する。

35.

$$
L^2(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{R} \mid \int_{\mathbb{R}} |f(x)|^2 dx < \infty \right\}
$$

の各要素 $f, g \in L^2(\mathbb{R})$ およびスカラー $\alpha, \beta \in \mathbb{R}$ に対して、線形結合 $\alpha f + \beta g$ を考えると、

$$
\int_{\mathbb{R}} |\alpha f(x) + \beta g(x)|^2 dx \leq 2|\alpha|^2 \int |f(x)|^2 dx + 2|\beta|^2 \int |g(x)|^2 dx < \infty
$$

が成立する（Cauchy-Schwarz および二乗の加法的不等式に基づく）。したがって、$\alpha f + \beta g \in L^2(\mathbb{R})$ が成立する。L^2 空間におけるノルム $\|f\|$ は内積 $\langle f, g \rangle := \int_{\mathbb{R}} f(x) g(x) dx$ から誘発されたもので、具体的には

$$
\|f\| := \left(\int_{\mathbb{R}} |f(x)|^2 dx \right)^{1/2}
$$

で与えられる。このノルムは以下の性質を満たす：$\|f\| \geq 0$(非負性)、$\|\alpha f\| = |\alpha| \cdot \|f\|$(スカラー倍) は明らかである。三角不等式を示すには、ノルムの 2 乗を展開して、

$$
\|f + g\|^2 = \int (f(x) + g(x))^2 dx = \|f\|^2 + 2 \int f(x) g(x) dx + \|g\|^2
$$

これに Cauchy-Schwarz の不等式を適用して、$\left| \int f(x) g(x) dx \right| \leq \|f\| \cdot \|g\|$ より、$2 \int f(x) g(x) dx \leq 2\|f\|\|g\|$ となる。さらに

$$
\|f + g\|^2 \leq \|f\|^2 + 2\|f\|\|g\| + \|g\|^2 = (\|f\| + \|g\|)^2
$$

とでき、両辺の平方根をとると、三角不等式が成り立つ

$$
\|f + g\| \leq \|f\| + \|g\|
$$

Cauchy-Schwarz の不等式は、任意の実数 $t \in \mathbb{R}$ に対して、L^2 関数 $f - tg$ について

$$
0 \leq \int |f(x) - tg(x)|^2 dx = \int \left(f(x)^2 - 2t f(x) g(x) + t^2 g(x)^2 \right) dx = \|f\|^2 - 2t \int f(x) g(x) dx + t^2 \|g\|^2
$$

が成立し、t に関する判別式が非正であることから成立する。

$$\left(\int f(x)g(x)dx\right)^2 \le \|f\|^2\|g\|^2$$

このノルムは、$f \equiv= 0 \Longleftrightarrow$ 至るところすべてで $f=0$ の同値類どうしに対して適用される。

36. 中心化行列 H は

$$H^2 = H, \quad H^\top = H, \quad H\mathbf{1} = \mathbf{0}, \quad \mathbf{1}^\top H = \mathbf{0}^\top$$

を満足する。任意の Gram 行列 K に対し $\tilde{K} := HKH$ とおくと,

$$\tilde{K}\mathbf{1} = HKH\mathbf{1} = HK\mathbf{0} = \mathbf{0}, \qquad \mathbf{1}^\top\tilde{K} = \mathbf{0}^\top KH = \mathbf{0}^\top$$

が成り立つ。したがって $\tilde{K}$ の各列・各行は和が 0 である。特に

$$\tilde{K}_{XZ}, \quad \tilde{K}_{YZ}, \quad \tilde{K}_Z$$

はいずれも列和・行和が 0 である。

次に $\tilde{K}_{XZ}$ の固有分解を

$$\tilde{K}_{XZ} = U\Lambda U^\top$$

とし，$V_{XZ} := U\Lambda^{1/2}$ とおく。このとき

$$\mathrm{Range}(\tilde{K}_{XZ}) = \mathrm{Range}(V_{XZ})$$

である。さらに $\tilde{K}_{XZ}\mathbf{1} = \mathbf{0}$ より，V_{XZ} の各列は $\mathbf{1}$ に直交する。すなわち

$$\mathbf{1}^\top v_j = 0 \qquad (\forall j),$$

ここで v_j は V_{XZ} の第 j 列ベクトルである。同様にして V_{YZ} の各列も $\mathbf{1}$ に直交する。

一方，$R_Z := \lambda(\tilde{K}_Z + \lambda I)^{-1}$ は対称行列であり，

$$(\tilde{K}_Z + \lambda I)\mathbf{1} = \lambda\mathbf{1} \quad \Rightarrow \quad R_Z\mathbf{1} = \lambda(\tilde{K}_Z + \lambda I)^{-1}\mathbf{1} = \mathbf{1}$$

が成立する。したがって任意のベクトル v に対して

$$\mathbf{1}^\top(R_Z v) = (R_Z\mathbf{1})^\top v = \mathbf{1}^\top v$$

が成り立つ。特に $\mathbf{1}^\top v = 0$ なら $\mathbf{1}^\top(R_Z v) = 0$ である。すなわち R_Z は「平均ゼロの部分空間」を保つ。

以上を踏まえると，本文の定義

$$\phi := R_Z V_{XZ}, \qquad \varphi := R_Z V_{YZ}$$

に対して，各列 $\phi_{\cdot j}, \varphi_{\cdot k}$ はいずれも $\mathbf{1}$ に直交する。したがって

$$\sum_{i=1}^n \phi_{i,j} = \mathbf{1}^\top\phi_{\cdot j} = 0, \qquad \sum_{i=1}^n \varphi_{i,k} = \mathbf{1}^\top\varphi_{\cdot k} = 0$$

がすべての j, k に対して成立する。

$\square$

37. (a) $x \sim \mathcal{N}(0,1)$, $y \sim \mathcal{N}(0,1)$（独立）

(b) $y = x + \varepsilon,\ \varepsilon \sim \mathcal{N}(0,1)$（従属）

について、HSIC 統計量と帰無分布を比較する。

```r
## データ生成とHSICの検定
n <- 100; m <- 300
k <- function(x, y) exp(-sum((x - y)^2))  # RBFカーネル

# 独立のケース
x1 <- rnorm(n)
y1 <- rnorm(n)
Kx1 <- K(k, x1)
Ky1 <- K(k, y1)
u1 <- HSIC.1(Kx1, Ky1)

# 従属のケース
x2 <- rnorm(n)
y2 <- x2 + rnorm(n)
Kx2 <- K(k, x2)
Ky2 <- K(k, y2)
u2 <- HSIC.1(Kx2, Ky2)

## 帰無分布の生成関数
null.dist.perm <- function(x, y, m = 300) {
  Ky <- K(k, y)
  w <- numeric(m)
  for (i in 1:m) {
    x_perm <- sample(x)
    Kx_perm <- K(k, x_perm)
    w[i] <- HSIC.1(Kx_perm, Ky)
  }
  return(w)
}

w1 <- null.dist.perm(x1, y1)
w2 <- null.dist.perm(x2, y2)

## グラフ描画
par(mfrow = c(1, 2))
hist(w1, probability = TRUE, main = "独立：x, y ~ N(0,1)",
     xlab = "HSIC値", col = "gray"); abline(v = u1, col = "blue", lwd = 2)
hist(w2, probability = TRUE, main = "従属：y = x + ε",
     xlab = "HSIC値", col = "gray"); abline(v = u2, col = "red", lwd = 2)
```

- 独立のケースでは、観測された HSIC 値（青線）は帰無分布内にあり、独立性は棄却されない。
- 従属のケースでは、HSIC 値（赤線）が帰無分布より大きく、独立性が棄却される。

38. (a) ノルムの直交分解性により、以下が成り立つ：

$$\|f\|_H^2 = \|f_0 + f_\perp\|_H^2 = \|f_0\|_H^2 + \|f_\perp\|_H^2$$

ここで、$f_0 = \sum_{i=1}^{n} \alpha_i k(x_i, \cdot)$ である。

(b) 与えられた正則化問題：

$$\arg\min_{f\in H}\left\{\sum_{j=1}^{n}(f(x_j)-y_j)^2+\lambda\|f\|_H^2\right\}$$

に対して、$f=f_0+f_\perp$ と分解し、評価点 x_j において $f(x_j)=f_0(x_j)$ が (a) より成り立つため、目的関数は以下のように書き換えられる：

$$\sum_{j=1}^{n}(f_0(x_j)-y_j)^2+\lambda(\|f_0\|^2+\|f_\perp\|^2)$$

ただし、第 1 項は $f_\perp$ に依存しない一方で、第 2 項は $f_\perp$ を含む。したがって、$f_\perp\neq 0$ であると $\|f\|$ が不必要に大きくなるので、最小化のためには $f_\perp=0$ が最適である。

39. 出力例

```
lambda=0.01, r=3: statistic=0.23451
lambda=0.01, r=20: statistic=0.31042
lambda=1.00,  r=3: statistic=0.20987
lambda=1.00,  r=20: statistic=0.28914
```

- λ を小さくすると、行列 $\tilde{K}_Z$ の逆が大きくなり、R_Z の値が大きくなる傾向があり、統計量も増加傾向にある。
- r の増加は、固有成分のより高次までを利用することを意味し、より多くの情報が統計量に反映されるため、一般に値は大きくなる。
- ただし、$n=4$ のように小規模なデータでは、$r>n$ は意味を持たないため、$r\leq n$ であることが前提である。

ソースコード 1　KCI 統計量のパラメータ依存性の検証

```
##
set.seed(1)
n <- 4
U <- rnorm(n); V <- rnorm(n); W <- rnorm(n)
X <- U; Y <- V; Z <- W

## カーネル定義とカーネル行列の構築
k <- function(x, y) exp(-sum((x - y)^2))
K.x <- K(k, X)
K.y <- K(k, Y)
K.z <- K(k, Z)

## 検定統計量の計算(関数KCI.1は第 2章で定義済みとする)
lambda.list <- c(0.01, 1)
r.list <- c(3, 20)

for (lambda in lambda.list) {
  for (r in r.list) {
    result <- KCI.1(K.x, K.y, K.z, lambda)
    stat <- result$statistics
```

```
    cat(sprintf("lambda=%.2f, r=%d: statistic=%.5f\n", lambda, r, stat))
  }
}
```

40.
- **Case 1**：X, Y が独立 → 統計量は帰無分布と一致
- **Case 2**：W を介して X と Y が依存 → 統計量は大きくなる
- **Case 3**：$Z = U + V$ は X と Y の情報を含む → HSIC が抑えられる
- **Case 4**：$X = U, Y = U + V, Z = V$ → 条件付き独立性が崩れる

```
set.seed(0)
n <- 100
sigma2 <- 1
k <- function(x, y) exp(-sum((x - y)^2) / (2 * sigma2))

par(mfrow = c(2, 2))
for (case in 1:4) {
  U <- rnorm(n)
  V <- rnorm(n)
  W <- rnorm(n)
  if (case == 1) { X <- U; Y <- V; Z <- W }
  if (case == 2) { X <- U + W; Y <- V + W; Z <- W }
  if (case == 3) { X <- U; Y <- V; Z <- U + V }
  if (case == 4) { X <- U; Y <- U + V; Z <- V }

  K.x <- K(k, X)
  K.y <- K(k, Y)

  u <- HSIC.1(K.x, K.y)

  # 帰無分布を生成
  w <- NULL
  m <- 200
  for (i in 1:m) {
    idx <- sample(1:n)
    K.x_perm <- K.x[idx, idx]
    w <- c(w, HSIC.1(K.x_perm, K.y))
  }

  hist(w, prob = TRUE, main = paste("Case", case),
       xlab = "HSIC", col = "lightgray", border = "white")
  abline(v = u, col = "blue", lwd = 2, lty = 2)
  legend("topright", legend = "観測統計量", lty = 2, col = "blue", bty = "n")
}
```

41. 行列の積 AB および BA が定義されるとき $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times m}$ とすると、$AB \in \mathbb{R}^{m\times m}$, $\quad BA \in \mathbb{R}^{n\times n}$ である。$\mathrm{tr}(AB)$ の定義より、

$$\mathrm{tr}(AB) = \sum_{i=1}^{m}(AB)_{ii} = \sum_{i=1}^{m}\sum_{k=1}^{n} A_{ik}B_{ki}, \qquad \mathrm{tr}(BA) = \sum_{k=1}^{n}(BA)_{kk} = \sum_{k=1}^{n}\sum_{i=1}^{m} B_{ki}A_{ik}$$

であり、$\mathrm{tr}(AB) = \mathrm{tr}(BA)$ が示された。次に、行列 $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ に対して、Frobenius ノルムは $\|A\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2$ で定義されるので $(AA^\top)_{ii} = \sum_{k=1}^{n} a_{ik}a_{ik} = \sum_{k=1}^{n} a_{ik}^2$ であり、

$$\mathrm{tr}(AA^\top) = \sum_{i=1}^{m}(AA^\top)_{ii} = \sum_{i=1}^{m}\sum_{k=1}^{n} a_{ik}^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2 = \|A\|_F^2$$

が成り立つ。

42. KCI 統計量は、条件付きカーネル独立性検定において、次の構造を持つ
    - $K_{\mathcal{Z}} = a \cdot J$ とする。ただし J はすべての要素が 1 の行列（$n \times n$）。
    - このとき、$\tilde{K}_{\mathcal{Z}} = HK_{\mathcal{Z}}H = a \cdot HJH = 0$（なぜなら $HJ = 0$）

    したがって、$\tilde{K}_{\mathcal{Z}} = 0$。

    (a) $\tilde{K}_{\mathcal{Z}} = 0$ なので、

    $$R_Z = \lambda\left(\tilde{K}_{\mathcal{Z}} + \lambda I\right)^{-1}\lambda(\lambda I)^{-1} = I$$

    (b) 本文における KCI.1 関数では、

    $$T = \mathrm{tr}(XZ \cdot YZ) \quad \text{ただし } XZ = R_Z\tilde{K}_{XZ}R_Z, \quad YZ = R_Z\tilde{K}_{YZ}R_Z$$

    $R_Z = I$ より、

    $$XZ = \tilde{K}_{XZ} = \tilde{K}_X \circ \tilde{K}_Z, \quad YZ = \tilde{K}_Y \circ \tilde{K}_Z$$

    $K_Z = aJ$ は定数カーネルなので、$\tilde{K}_Z = 0$、しかし元のカーネル $K_Z = a \cdot J$ による Hadamard 積で考えると、

    $$K_{XZ} = K_X \circ K_Z = aK_X, \quad K_{YZ} = K_Y \circ K_Z = aK_Y$$

    よって、$XZ = a\tilde{K}_X, YZ = a\tilde{K}_Y$。したがって、

    $$T = \mathrm{tr}(XZ \cdot YZ) = \mathrm{tr}(a\tilde{K}_X \cdot a\tilde{K}_Y) = a^2 \cdot \mathrm{tr}(\tilde{K}_X \cdot \tilde{K}_Y)$$

    これは、$a = 1$ の場合の統計量を a^2 倍したものである。また、本文の定式化では、近似的に

    $$T \approx \sum_{k=1}^{n^2} \lambda_k Z_k^2 \quad (\lambda_k\text{: 固有値、}Z_k \sim \mathcal{N}(0,1))$$

    前述のように、XZ と YZ の要素が a 倍されているため、W 行列も a 倍、$W^\top W$ は a^2 倍となる。よって、固有値 λ_k はすべて a^2 倍される。

43. プログラムは下記

```
set.seed(42)
n <- 100
alpha <- 0.01

# --------------------------------
# X ~ N(0,1), Y = X^2 の独立性検定(相関 0だが独立でない)
# --------------------------------
X <- rnorm(n)
Y <- X^2
```

```
# RBF カーネル（本書の定義を使う前提）
k <- function(x, y) exp(-sum((x - y)^2))
Kx <- K(k, X)
Ky <- K(k, Y)

# HSIC の計算と固有値の取得
res <- HSIC.2(Kx, Ky)          # 本書定義:return list(statistics=..., eigen_values=...)
hsic_stat <- res$statistics
eigen_vals <- res$eigen_values

# 帰無分布の臨界値(固有値を使って漸近分布をシミュレート)
crit <- null.dist(eigen_vals, alpha)$critical

# 結果出力
cat("HSIC 値（観測):", hsic_stat, "\n")
cat("臨界値（1%）:", crit, "\n")
cat("→ 独立性検定の結果:", ifelse(hsic_stat > crit, "独立でない", "独立"), "\n")
# ------------------------------
# (2) Z=0 なら Y=X^2, Z=1 なら X=Y^2 の KCI 検定(条件付き独立性)
# ------------------------------
set.seed(42)
n <- 100
alpha <- 0.01

# データ生成
Z <- sample(0:1, n, replace = TRUE)
X <- numeric(n)
Y <- numeric(n)

for (i in 1:n) {
  if (Z[i] == 0) {
    X[i] <- rnorm(1)
    Y[i] <- X[i]^2
  } else {
    Y[i] <- rnorm(1)
    X[i] <- Y[i]^2
  }
}

# RBF カーネルの定義
k <- function(x, y) exp(-sum((x - y)^2))

# カーネル行列の構成(本書定義の K を使用)
Kx <- K(k, X)
Ky <- K(k, Y)
Kz <- K(k, Z)

# KCI 統計量と固有値の取得
```

```
result <- KCI.1(Kx, Ky, Kz, lambda = 1e-3)
stat <- result$statistics
eigen <- result$eigen_values

# 帰無分布の臨界値(有意水準α)
crit <- null.dist(eigen, alpha)$critical

# 検定結果出力
cat("KCI 値 (観測):", stat, "\n")
cat("臨界値 (1%) :", crit, "\n")
cat("→ 条件付き独立性検定の結果:", ifelse(stat > crit, "条件付きで独立でない", "条件付きで独立"), "\n")
```

4 PC アルゴリズム

44. $\mathcal{X} \perp\!\!\!\perp \mathcal{Z}$ および $\mathcal{Y} \perp\!\!\!\perp \mathcal{Z}$ なので、X_i と Z_j、Y_k と Z_j の共分散が 0 になる。$X_i \in \mathcal{X}, Y_k \in \mathcal{Y}, Z_j \in \mathcal{Z}$ とすると、$s \cdot u$ 個（$\mathcal{X} \times \mathcal{Z}$）$t \cdot u$ 個（$\mathcal{Y} \times \mathcal{Z}$）合計 $2u(s+t) = 2u(p-u)$ 個が 0。
45.
```
library(pcalg)
p <- 10
n <- 1000
vars <- c("1", 2:p)
gGtrue <- randomDAG(p, prob = 0.5, V = vars)
gmG <- list(x = rmvDAG(n, gGtrue), g = gGtrue)

suffStat <- list(C = cor(gmG$x), n = nrow(gmG$x))
skeleton.fit <- skeleton(suffStat, indepTest = gaussCItest, p = p, alpha = 0.01)
pc.fit <- pc(suffStat, indepTest = gaussCItest, p = p, alpha = 0.01)

par(mfrow = c(1, 3))
plot(gmG$g, main = "True DAG")
plot(skeleton.fit, main = "Skeleton")
plot(pc.fit, main = "CPDAG")
```
46. 空欄 (1)：rep(2, p)（全変数が 2 値）, 空欄 (2)：suffStat
47. `K <- matrix(1, n, n)` としなければ、空集合に対する条件付き独立性（$S = \emptyset$）でエラーになる。単位行列だとアダマール積の際にゼロになる部分が出て、誤判定や行列サイズ不整合が起こる。
48. `length(S) >= m` でなければ、サイズ m の部分集合が作れないので combn(S, m) に失敗する。また、s[i, j, k] = TRUE となるのは、$X_i \perp\!\!\!\perp X_j \mid X_k$ が成立したとき、k が分離集合に含まれていたことを記録するため。
49. ルール 0, 1(a) ともに i と k が隣接していると、$i - j - k$ が V-構造にならないため、j を媒介としたコライダー構造の証明に使えない。つまり、$i \perp\!\!\!\perp k$ を j 経由で示せない。

50.

DAG	$(1,2,3)$	$(1,3,2)$	$(2,1,3)$	$(2,3,1)$	$(3,1,2)$	$(3,2,1)$
(a)	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE
(b)	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE
(c)	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
(d)	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
(e)	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE
(f)	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE
(g)	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
(h)	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE
(i)	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE
(j)	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE
(k)	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE

51. $1 \to 3 \leftarrow 2,\quad 2 \leftrightarrow 4,\quad 4 \leftrightarrow 1$ 。$1 \to 3 \leftarrow 2$ は、$S[1,2,3] =$ FALSE に基づくルール 0（V 構造）による。それ以外のルールは適用できない。

52. プログラムを実行して確認する。

53.

```
set.seed(1)
# --- 条件付き独立性検定を呼び出す簡易関数(本書定義のCI.PCを使用) ---
check_CI <- function(x, y, z, alpha = 0.05, lambda = 1e-3) {
  n <- length(x)
  p <- ncol(z)
  # 全体を結合
  X <- cbind(x, y, z)
  KK <- list()
  gaussian_kernel <- function(x, y) exp(-sum((x - y)^2) / 2)
  for (j in 1:(p + 2)) {
    KK[[j]] <- K(gaussian_kernel, X[, j])
  }
  # x → col 1, y → col 2, z → col 3:(p+2)
  return(CI.PC(1, 2, 3:(p + 2), KK, alpha, lambda))
}
# --- 条件付き独立が成立する例 ---
n <- 100; p <- 2
z <- matrix(rnorm(n * p), n, p)
x <- rnorm(n)
y <- rnorm(n)
cat(check_CI(x, y, z), "\n")  # TRUE が期待される
# --- 条件付き独立が成立しない例 ---
z <- matrix(rnorm(n * p), n, p)
x <- rnorm(n)
y <- x + z[,1] + rnorm(n, sd = 0.1)
cat(check_CI(x, y, z), "\n")  # FALSE が期待される
```

5 LiNGAM

54. 対称性がある

$$c(x^n, y^n) = c(y^n, x^n), \quad v(x^n) = v(x^n), \quad v(y^n) = v(y^n)$$

$$c(x^n, z^n) \leftrightarrow c(y^n, z^n) \quad (x^n \leftrightarrow y^n \text{ の入れ替えにより})$$

から、x^n と y^n を入れ替えると、第 2 項と第 3 項が入れ替わるだけで、式全体としての値は変化しない。

55.

```
search.p.val <- function(index, Z) {
  # index の長さが 2 未満であれば、そのまま index を返す
  if (length(index) < 2) return(index)
  # 初期化:最大の p 値を設定
  max <- -1
  # index の各要素についてループ
  for (i in index) {
    # i を先頭に移動し、残りを順に並べる
    index.2 <- c(i, setdiff(index, i))
    # LiNGAM.2関数で p 値を実行
    result <- LiNGAM.2(1, Z[index.2], proc = "p.val")  # 空欄 (1)
    # 最大の p 値を更新し、その際の Z と i の値を記録
    if (result$p.val > max) {  # 空欄 (2)
      max <- result$p.val      # 空欄 (3)
      W <- Z
      W[index.2] <- result$Z  # p 値を適用後の Z を保存
      k <- i
    }
  }
  # 更新された W で再帰呼び出しを行い、結果をまとめる
  index.3 <- search.p.val(setdiff(index, k), W)  # 空欄 (4)
  return(c(k, index.3))
}
```

56. 下記の関数を用いる

`K, tilde, HSIC.2, null.dist, LiNGAM.2, search.HSIC`

```
# crime.txt の読み込み
crime <- read.table(
  system.file("misc", "crime.txt", package = "glmnet"),
  header = TRUE
)
# 必要な変数のみを抽出(X1〜X7)
data <- as.matrix(crime[, 1:7])
colnames(data) <- paste0("X", 1:7)
# 各変数を中心化&スケーリング(重要)
data <- scale(data)
# リスト形式に変換(search.HSIC 用)
Z <- lapply(1:ncol(data), function(j) data[, j])
```

```
# 因果順序の同定
order <- search.HSIC(Z)
cat("推定された因果順序:\n")
print(order)
```

57. まず、X と Y の関係として以下の 2 通りの加法雑音モデルを考える：

$$\begin{cases} X = e_1 \\ Y = aX + e_2 \end{cases} \qquad \text{または} \qquad \begin{cases} Y = e_1' \\ X = a'Y + e_2' \end{cases}$$

ここで、$e_1 \perp\!\!\!\perp e_2$ および $e_1' \perp\!\!\!\perp e_2'$ と仮定する。e_1 と e_2 が独立であることから、$X \perp\!\!\!\perp Y$ が成り立つのは $a = 0$ のときである。実際、$Y = aX + e_2$ において $a = 0$ ならば $Y = e_2$ となり、$X = e_1$ とは独立である。したがって、

$$a = 0 \implies X \perp\!\!\!\perp Y$$

同様に、$Y = e_1'$ および $X = a'Y + e_2'$ において、$a' = 0$ であれば $X = e_2'$ となり、$Y = e_1'$ とは独立になるため、

$$a' = 0 \implies X \perp\!\!\!\perp Y$$

逆に、$X \perp\!\!\!\perp Y$ であれば、$cov(X, Y) = 0$ が成り立ち、線形回帰係数の定義より、

$$a = \frac{cov(X,Y)}{var(X)} = 0, \quad a' = \frac{cov(X,Y)}{var(Y)} = 0$$

が導かれる。したがって、

$$X \perp\!\!\!\perp Y \implies a = a' = 0$$

以上より、

$$a = 0 \iff a' = 0 \iff X \perp\!\!\!\perp Y$$

が成立する。

次に、$aa' = 1$ が生じ得ない理由を示す。命題 19 直後の関係式として、

$$a' = \frac{a\sigma_1^2}{a^2\sigma_1^2 + \sigma_2^2} \tag{1}$$

が与えられている。このとき、

$$aa' = a \cdot \frac{a\sigma_1^2}{a^2\sigma_1^2 + \sigma_2^2} = \frac{a^2\sigma_1^2}{a^2\sigma_1^2 + \sigma_2^2}$$

ここで、$\sigma_2^2 > 0$ であるため、分母は分子より大きい：

$$a^2\sigma_1^2 + \sigma_2^2 > a^2\sigma_1^2 \implies \frac{a^2\sigma_1^2}{a^2\sigma_1^2 + \sigma_2^2} < 1$$

したがって、$aa' < 1$ が常に成立し、$aa' = 1$ となることはありえない。

58. $e_1 \perp\!\!\!\perp e_2$ および e_1, e_2 が正規分布に従うと仮定する。また、$cov(e_1, e_2) = 0$ であるから、正規分布の独立性の特徴から、

$$e_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \quad e_2 \sim \mathcal{N}(\mu_2, \sigma_2^2), \quad e_1 \perp\!\!\!\perp e_2 \implies (e_1, e_2) \sim \mathcal{N}_2$$

という 2 変数の独立な正規分布に従う。このとき、e'_1 および e'_2 は e_1, e_2 の線形結合であるから、それぞれ正規分布に従う。さらに、e'_1 と e'_2 の共分散を計算すると：

$$\begin{aligned}
cov(e'_1, e'_2) &= cov(ae_1 + e_2, (1 - aa')e_1 - a'e_2) \\
&= a(1 - aa')cov(e_1, e_1) - aa'cov(e_2, e_2) + (1 - aa')cov(e_2, e_1) - a'cov(e_2, e_2) \\
&= a(1 - aa')\sigma_1^2 - aa' \cdot 0 + 0 - a'\sigma_2^2 \quad (\text{since } cov(e_1, e_2) = 0) \\
&= a(1 - aa')\sigma_1^2 - a'\sigma_2^2
\end{aligned}$$

ここで、a' を次のように選べば $cov(e'_1, e'_2) = 0$ が成立する：

$$a' = \frac{a\sigma_1^2}{a^2\sigma_1^2 + \sigma_2^2}$$

このとき、

$$cov(e'_1, e'_2) = 0$$

さらに、e'_1 および e'_2 は e_1, e_2 の独立な正規分布の線形結合であり、かつ共分散が 0 なので、再び**独立な正規分布**になる：

$$e'_1, e'_2 \text{ は正規分布に従い、かつ } cov(e'_1, e'_2) = 0 \quad \Longrightarrow \quad e'_1 \perp\!\!\!\perp e'_2$$

59. (a) 真の順序を

$$\begin{cases} X = e_1 \\ Y = aX + e_2 \\ Z = bX + cY + e_3 \end{cases}$$

とし、誤った順序を

$$\begin{cases} Y = e'_1 \\ Z = a'Y + e'_2 \\ X = b'Y + c'Z + e'_3 \end{cases}$$

とする。このとき、$Y = aX + e_2 = ae_1 + e_2$ より

$$e'_1 = ae_1 + e_2$$

さらに、

$$Z = bX + cY + e_3 = be_1 + c(ae_1 + e_2) + e_3 = (b + ac)e_1 + ce_2 + e_3$$

したがって、

$$e'_2 = Z - a'e'_1 = (b + ac)e_1 + ce_2 + e_3 - a'(ae_1 + e_2) = \{b + a(c - a')\}e_1 + (c - a')e_2 + e_3$$

最後に、

$$X = e_1 = b'Y + c'Z + e'_3 = b'(ae_1 + e_2) + c'\{(b + ac)e_1 + ce_2 + e_3\} + e'_3$$

移項すると、

$$e'_3 = e_1 - b'(ae_1 + e_2) - c'\{(b + ac)e_1 + ce_2 + e_3\} = \{1 - ab' - c'(b + ac)\}e_1 - (b' + c'c)e_2 - c'e_3$$

よって、

$$\begin{aligned}
e'_1 &= ae_1 + e_2 \\
e'_2 &= \{b + a(c - a')\}e_1 + (c - a')e_2 + e_3 \\
e'_3 &= \{1 - ab' - c'(b + ac)\}e_1 - (b' + c'c)e_2 - c'e_3
\end{aligned}$$

(b) e'_1, e'_2, e'_3 の間の共分散が全てゼロであると仮定し、それぞれの係数を a, b, c と $\sigma_1^2, \sigma_2^2, \sigma_3^2$ によって表すと、以下のようになる：

$$e'_1 = ae_1 + e_2$$
$$e'_2 = \frac{b\sigma_2^2}{a^2\sigma_1^2 + \sigma_2^2}e_1 - \frac{ab\sigma_1^2}{a^2\sigma_1^2 + \sigma_2^2}e_2 + e_3$$
$$e'_3 = \frac{\sigma_1^2\sigma_2^2}{b^2\sigma_1^2\sigma_2^2 + (a^2\sigma_1^2 + \sigma_2^2)\sigma_3^2}e_1 - \frac{a\sigma_1^2\sigma_3^2}{b^2\sigma_1^2\sigma_2^2 + (a^2\sigma_1^2 + \sigma_2^2)\sigma_3^2}e_2 - \frac{b\sigma_1^2\sigma_2^2}{b^2\sigma_1^2\sigma_2^2 + (a^2\sigma_1^2 + \sigma_2^2)\sigma_3^2}e_3$$

(c) ここで、$a, b, c \neq 0$ かつ e_1, e_2, e_3 のいずれかが正規分布に従わないとする。e'_1, e'_2, e'_3 は e_1, e_2, e_3 の線形結合である。e'_1, e'_2, e'_3 が相互に独立であるためには、Darmois – Skitovich の定理により、e_1, e_2, e_3 がすべて正規分布でなければならない。したがって、1 つでも非正規であれば e'_1, e'_2, e'_3 の間に独立性は成り立たない。

(d) この設問は前問の部分的な確認である：

- $X = e_1$ が非正規 $\Rightarrow$ e_1 が非正規
- $Y = ae_1 + e_2$ が非正規 $\Rightarrow$ e_1 または e_2 が非正規
- $Z = be_1 + cY + e_3 = (b + ac)e_1 + ce_2 + e_3$ が非正規 $\Rightarrow$ e_1, e_2, e_3 のいずれかが非正規

よって、いずれの場合も構成要素のどれかが非正規になる。

(e) まとめると、X, Y, Z のいずれか 1 つが非正規分布に従うならば、e_1, e_2, e_3 の少なくとも 1 つは非正規である。そのため、Darmois – Skitovich の定理により、e'_1, e'_2, e'_3 の間には独立性が成立しない。したがって、$X \to Y \to Z$ を $Y \to Z \to X$ に誤って識別することはない。

60. (a) 真の順序が

$$\begin{cases} X = e_1 \\ Y = aX + e_2 \\ Z = cY + e_3 \end{cases}$$

であるとき、別の順序として以下のような 2 通りを仮定する：

- 誤識別順序 $Y \to X \to Z$：

$$\begin{cases} Y = e'_1 \\ X = a'Y + e'_2 \\ Z = b'Y + c'X + e'_3 \end{cases}$$

- 誤識別順序 $X \to Z \to Y$：

$$\begin{cases} X = e''_1 \\ Z = a''X + e''_2 \\ Y = b''X + c''Z + e''_3 \end{cases}$$

これは、元のモデルに対して別の構造方程式モデルがあてはめられることを意味する。したがって、順序の再定義としてこれらのモデルは正しく記述可能である。

(b) $e_1 = X$, $e_2 = Y - aX$, $e_3 = Z - cY$ とおく。まず、$Y = aX + e_2 = ae_1 + e_2$ より

$$e'_1 = Y = ae_1 + e_2$$

また、$X = e_1 = a'e'_1 + e'_2$ より

$$e'_2 = e_1 - a'e'_1 = e_1 - a'(ae_1 + e_2) = (1 - aa')e_1 - a'e_2$$

このとき、$cov(e'_1, e'_2) = 0$ となるような a' は次の式で与えられる：

$$a' = \frac{a\sigma_1^2}{a^2\sigma_1^2 + \sigma_2^2}$$

これを代入すると

$$e_2' = \frac{\sigma_2^2}{a^2\sigma_1^2+\sigma_2^2}e_1 - \frac{a\sigma_1^2}{a^2\sigma_1^2+\sigma_2^2}e_2$$

また、$Z = cY + e_3 = c(ae_1 + e_2) + e_3$ を Y, X で表す $Z = b'Y + c'X + e_3'$ に代入すると、$e_3 = e_3'$ となる。したがって、

$$\begin{cases} e_1' = ae_1 + e_2 \\ e_2' = \dfrac{\sigma_2^2}{a^2\sigma_1^2+\sigma_2^2}e_1 - \dfrac{a\sigma_1^2}{a^2\sigma_1^2+\sigma_2^2}e_2 \\ e_3' = e_3 \end{cases}$$

次に、誤順序 $X \to Z \to Y$ に対して：

$$Z = cY + e_3 = c(ae_1 + e_2) + e_3 = ace_1 + ce_2 + e_3 \Rightarrow e_2'' = Z - a''X = Z - a''e_1$$

とおく。$cov(X, Z) = cov(e_1, ace_1 + ce_2 + e_3) = ac\sigma_1^2$ $var(X) = \sigma_1^2$, $var(Z) = a^2c^2\sigma_1^2 + c^2\sigma_2^2 + \sigma_3^2$ したがって、

$$a'' = \frac{cov(X, Z)}{var(X)} = ac \Rightarrow e_2'' = (Z - acX) = ce_2 + e_3$$

これを $e_2'' = ce_2 + e_3$ と書く。また、$Y = aX + e_2 = ae_1 + e_2 = b''X + c''Z + e_3''$ として e_3'' を導出すると：

$$e_3'' = Y - b''X - c''Z = ae_1 + e_2 - b''e_1 - c''(ace_1 + ce_2 + e_3)$$

$cov(Y, Z)$, $cov(X, Z)$ などから c'' を選ぶと、e_3'' は以下のように表せる：

$$e_3'' = \frac{\sigma_3^2}{c^2\sigma_2^2+\sigma_3^2}e_2 - \frac{c\sigma_2^2}{c^2\sigma_2^2+\sigma_3^2}e_3$$

したがって、

$$\begin{cases} e_1'' = e_1 \\ e_2'' = ce_2 + e_3 \\ e_3'' = \dfrac{\sigma_3^2}{c^2\sigma_2^2+\sigma_3^2}e_2 - \dfrac{c\sigma_2^2}{c^2\sigma_2^2+\sigma_3^2}e_3 \end{cases}$$

(c) e_3 が非正規、e_1, e_2 が正規であるとする。このとき、e_3' は e_3 に等しいため非正規である。e_1', e_2' は e_1, e_2 の線形結合なので正規分布に従う。Darmois – Skitovich の定理により、e_1', e_2', e_3' の独立性は成立しない。したがって、誤識別が発生する可能性がある。

(d) e_1 が非ガウス、e_2, e_3 がガウスのとき、$e_1'' = e_1$ は非ガウス、$e_2'' = ce_2 + e_3$ はガウスの線形結合なのでガウス、e_3'' はガウスの線形結合なのでガウス。e_1'', e_2'', e_3'' のうち e_1'' が非ガウスであることから、独立性は成立しない。よって $X \to Z \to Y$ に誤識別することがある。

61. (a) 確率変数 Y が正規分布 $Y \sim \mathcal{N}(\mu, \sigma^2)$ に従うとする。正規分布は左右対称な分布であるため、奇数次の中心化モーメント（特に 3 次）は 0 になる。よって $\gamma_1 = 0$ となる、標準正規分布 $Z \sim \mathcal{N}(0, 1)$ に対しては、以下の既知の事実がある：

$$\mathbb{E}[Z^4] = 3 \quad \Rightarrow \quad \gamma_2 = 3 - 3 = 0$$

(b) 本章では、非ガウス性を測るために非線形関数 $G(s)$ を用い、$\mathbb{E}[G(s)]$ の最大化を行っている。これに基づき、歪度および尖度はそれぞれ次のような $G(s)$ によって書ける：

$$G_{\text{skew}}(s) := s^3 \quad \Rightarrow \quad \mathbb{E}[G_{\text{skew}}(s)] = \mathbb{E}[s^3]$$

$$G_{\text{kurt}}(s) := s^4 - 3 \quad \Rightarrow \quad \mathbb{E}[G_{\text{kurt}}(s)] = \mathbb{E}[s^4] - 3$$

ここで、$s = (Y - \mu)/\sigma$ を標準化変数とした。

62. (a) 確率変数 $X \sim N(0, \sigma^2)$ の確率密度関数 $f(x)$ は、

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

である。このときの微分エントロピー $h(f)$ は以下のように計算できる：

$$\begin{aligned} h(f) &= -\int_{-\infty}^{\infty} f(x) \log f(x)\, dx = -\int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)\right) \left[\log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)\right)\right] dx \\ &= -\int_{-\infty}^{\infty} f(x) \left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{x^2}{2\sigma^2}\right] dx = \frac{1}{2}\log(2\pi\sigma^2) \int f(x)dx + \frac{1}{2\sigma^2} \int x^2 f(x)dx \\ &= \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \cdot \mathbb{E}[X^2] = \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \cdot \sigma^2 = \frac{1}{2}\log(2\pi e\sigma^2) \end{aligned}$$

(b)

$$\int_{-\infty}^{\infty} f_{gauss}(x)\{-\frac{1}{2}\log(2\pi\sigma^2)\}dx = \int_{-\infty}^{\infty} f(x)\{-\frac{1}{2}\log(2\pi\sigma^2)\}dx$$

$$\int_{-\infty}^{\infty} f_{gauss}(x)(x-\mu)^2 dx = \sigma^2 = f(x)(x-\mu)^2 dx$$

より、

$$\begin{aligned} h(f_{gauss}) &= -\int_{-\infty}^{\infty} f_{gauss}(x) \log\{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)}{2\sigma^2}\}\}dx \\ &= -\int_{-\infty}^{\infty} f(x) \log\{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)}{2\sigma^2}\}\}dx \\ &= -\int_{-\infty}^{\infty} f(x) \log\{f_{gauss}(x)\}dx \end{aligned}$$

したがって、

$$h(f_{gauss}) - h(f) = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{f_{gauss}(x)} dx \geq 0$$

63. 空欄 (1): `S[,1]`、空欄 (2): `S[,2]`、空欄 (3): `X[,1]`、空欄 (4): ”混合信号 2”、空欄 (5): `a$S[,1]`、空欄 (6): `a$S[,2]`

64. それぞれ実行して結果を確認する

65. (a) $p = 3$ のとき、3×3 の置換行列は以下の 6 通り（$3! = 6$ 個）存在する：

$$P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad P_3 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

$$P_4 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad P_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad P_6 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

これらはそれぞれ、元の行列 $A = (a_{ij})$ に左から作用して **行の順番を入れ替える**。たとえば、

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \text{に対して} \quad P_4 A = \begin{bmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{11} & a_{12} & a_{13} \end{bmatrix}$$

となる。

(b) 行列 A の**列を入れ替える**には、**右から置換行列をかける**。たとえば、列 1 と列 2 を入れ替えるには、以下のような置換行列 Q を右から掛ければよい：

$$Q = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \Rightarrow \quad AQ = \begin{bmatrix} a_{12} & a_{11} & a_{13} \\ a_{22} & a_{21} & a_{23} \\ a_{32} & a_{31} & a_{33} \end{bmatrix}$$

つまり、左から掛けると行が並び替わり、右から掛けると列が並び替わる。また、置換行列は、p 個の行（または列）の順番を並べ替える操作を行う行列である。

(c) 順列の定義により、p 個の要素を並び替える方法は $p!$ 通りあるため、$p \times p$ の置換行列も $p!$ 個存在する。各置換行列 P は、ある順列 $\pi : \{1, 2, \ldots, p\} \to \{1, 2, \ldots, p\}$ に対応し、$P_{ij} = 1$ となるのは $j = \pi(i)$ のときのみである。

66. 最短経路探索における「CLOSE に追加された頂点」は、OPEN にあるノードの中で、その時点で最も短い距離 $d(v)$ をもつノード v を確定したことを意味する。すなわち、アルゴリズムの不変条件として：

$$\text{任意の OPEN ノード } u \text{ について } d(v) \le d(u)$$

が成立しており、v を経由した他の経路が v に到達しても、それは必ず $d(v)$ より大きくなる。なぜなら、最短経路の距離関数 $d(\cdot)$ は非負のコスト（例：相互情報量や HSIC）に基づいて構成されており、追加的な遷移（辺）を含む経路は v への直接経路よりも距離が長くなるからである。よって、v が CLOSE に移動された後に、他の経路から再び v に短い距離で到達することは不可能である。

67. z^n_{xy} を x^n, y^n, z^n を用いて展開すると、以下のようになる。

$$\begin{aligned} z^n_{xy} &= z^n - \frac{c(x^n, z^n)}{v(x^n)} x^n - \left\{ \frac{c(y^n_x, z^n)}{v(y^n_x)} \cdot \left(y^n - \frac{c(x^n, y^n)}{v(x^n)} x^n \right) \right\} \\ &= z^n - \frac{c(x^n, z^n)}{v(x^n)} x^n - \left(\frac{c(y^n_x, z^n)}{v(y^n_x)} y^n - \frac{c(y^n_x, z^n)}{v(y^n_x)} \cdot \frac{c(x^n, y^n)}{v(x^n)} x^n \right) \\ &= z^n - \left(\frac{c(x^n, z^n)}{v(x^n)} + \frac{c(y^n_x, z^n)}{v(y^n_x)} \cdot \frac{c(x^n, y^n)}{v(x^n)} \right) x^n - \frac{c(y^n_x, z^n)}{v(y^n_x)} y^n \end{aligned}$$

次に、$c(y^n_x, z^n)$ を x^n, y^n, z^n の共分散の関数として書き直す。

$$c(y^n_x, z^n) = c\left(y^n - \frac{c(x^n, y^n)}{v(x^n)} x^n, z^n \right) = c(y^n, z^n) - \frac{c(x^n, y^n)}{v(x^n)} c(x^n, z^n)$$

また、

$$v(y^n_x) = v\left(y^n - \frac{c(x^n, y^n)}{v(x^n)} x^n \right) = v(y^n) - \frac{c(x^n, y^n)^2}{v(x^n)}$$

したがって、

$$\frac{c(y^n_x, z^n)}{v(y^n_x)} = \frac{v(x^n)c(y^n, z^n) - c(x^n, y^n)c(x^n, z^n)}{v(x^n)v(y^n) - c(x^n, y^n)^2}$$

これらを代入すると、主張が得られる。

$$z^n_{xy} = z^n - \left\{ \frac{v(y^n)c(x^n, z^n) - c(x^n, y^n)c(y^n, z^n)}{v(x^n)v(y^n) - c(x^n, y^n)^2} \right\} x^n - \left\{ \frac{v(x^n)c(y^n, z^n) - c(x^n, y^n)c(x^n, z^n)}{v(x^n)v(y^n) - c(x^n, y^n)^2} \right\} y^n$$

68. 次式を ϵ について解く。

$$(r + \epsilon)\gamma^q - r + \epsilon < 0$$

これを整理すると：

$$r(\gamma^q - 1) + \epsilon(\gamma^q + 1) < 0$$

したがって、この不等式が逆転するために必要な条件は以下のようになる。

$$\epsilon < \frac{r(1-\gamma^q)}{1+\gamma^q}$$

特に $\gamma \in (0,1)$ のときは、$1-\gamma^q > 0$ かつ $1+\gamma^q > 1$ なので、右辺は正であり、十分小さい ϵ を取ることで不等式は逆転する。また、

$$Q(t_s) = a_q t_s^q + O(A^{q-1}) = rA^q + O(A^{q-1})$$

は、t_s を以下のように定義していたことに基づく：

$$t_s = A \cdot e^{i\phi}, \quad \text{ただし } \phi = \frac{2\pi s - \theta}{q} \quad \text{より } t_s^q = A^q \cdot e^{i(2\pi s - \theta)}$$

したがって、主張が得られる。

$$a_q t_s^q = a_q A^q e^{i(2\pi s - \theta)} = rA^q$$

69. $X_1 \to X_2 \to X_3$ という構造をもつデータを非ガウス雑音付きで生成し、fastICA によって独立成分を抽出した後に `lingam()` を適用した（ICA-LiNGAM）。また、元のデータに `lingam()` をそのまま適用することで Direct LiNGAM を実行した。500 サンプルかつ非ガウスノイズ（例：$e_1 \sim$ 非対称分布、$e_2 \sim$ 一様分布、$e_3 \sim \chi^2$）のもとでは、両手法とも正しい因果順序 $X_1 \to X_2 \to X_3$ を再現した。しかし、サンプル数が小さい場合や、ノイズが正規分布である場合には識別性が失われたり、推定結果に差異が生じることを確認した。Direct LiNGAM は条件付き独立性検定に依存するため小サンプルで不安定になりやすい。他方、ICA-LiNGAM は成分分離に成功すれば、順序決定は比較的頑健である。

```r
library(pcalg)
library(fastICA)
set.seed(123)
n <- 500  # サンプルサイズ(変更可能)
# 非ガウスなノイズ
e1 <- sign(rnorm(n)) * abs(rnorm(n))^(1/3)
e2 <- runif(n) - 0.5
e3 <- rnorm(n)^2
# 真の因果順序: X1 -> X2 -> X3
x1 <- e1
x2 <- 0.8 * x1 + e2
x3 <- -0.5 * x2 + e3
X <- cbind(x1, x2, x3)
# ICA の実行
ica_res <- fastICA(X, 3, alg.typ = "parallel", fun = "logcosh", method = "R")
# ICA で得られた成分（独立成分）に LiNGAM を適用
lingam_ica <- lingam(ica_res$S)
# 推定されたDAG（隣接行列）を確認
cat("ICA-LiNGAM の推定因果構造:\n")
print(as(lingam_ica, "amat"))
lingam_direct <- lingam(X)
cat("Direct LiNGAM の推定因果構造:\n")
print(as(lingam_direct, "amat"))
```

6　周辺尤度と情報量規準

70. 変数変換 $\theta = t^2$ を導入し、$t \in (0,1)$ における新たなパラメータ t に対する Jeffreys の事前分布 $\varphi_T(t)$ を導出する。ヤコビアンを用いて、以下が成立する：

$$\varphi_T(t)\,dt = \varphi_\Theta(t^2)\cdot\left|\frac{d\theta}{dt}\right|dt = \frac{C_\Theta}{\sqrt{t^2(1-t^2)}}\cdot 2t\,dt = \frac{2C_\Theta t}{\sqrt{t^2(1-t^2)}}\,dt = \frac{2C_\Theta}{\sqrt{1-t^2}}dt$$

より、$\varphi_T(t) = \dfrac{C_T}{\sqrt{1-t^2}}$、ただし $C_T = \left[\displaystyle\int_0^1 \frac{1}{\sqrt{1-t^2}}dt\right]^{-1}$ とおいた。

71. X_1, X_2, X_3 がベルヌーイ分布（パラメータ θ）に従うとする。事前分布として $a = b = 1$ を用いると、ベータ分布 Beta$(1,1)$ に従い、これは一様分布に相当する。このとき、$q(x_1, x_2, x_3)$ は (6.13) で定義される：$x = (0,0,1)$ のときは $k = 1$ なので、

$$q(0,0,1) = B(2,3) = \frac{\Gamma(2)\Gamma(3)}{\Gamma(5)} = \frac{1!\cdot 2!}{4!} = \frac{2}{24} = \frac{1}{12}$$

$x = (0,0,0)$ のときは $k = 0$ なので、

$$q(0,0,0) = B(1,4) = \frac{\Gamma(1)\Gamma(4)}{\Gamma(5)} = \frac{1\cdot 6}{24} = \frac{1}{4}$$

となる。他の 6 個の場合も、対称性から $k = 0, 1$ のいずれかの周辺尤度になる。一方、（6.15）では、$n = 3$ のとき、$k = 0$ の場合：

$$q = \frac{0!\cdot 3!}{4!} = \frac{6}{24} = \frac{1}{4}$$

$k = 1$ の場合：

$$q = \frac{1!\cdot 2!}{4!} = \frac{2}{24} = \frac{1}{12}$$

となる。

72. $a = b = 1$ は、安定性を重視する設定であり、事前の影響が強く、学習の進行は緩やかだが堅実である。$a = b = 0.5$ は、柔軟性を重視する設定であり、初期データに対して敏感に反応し、より高速に学習が進む反面、ばらつきが大きくなる。

73. $a = b = 0.5$ の場合におけるベルヌーイ系列 $x_1, \ldots, x_n$ の周辺尤度は

$$q(x_1, \ldots, x_n) = \frac{\Gamma(k+\frac{1}{2})\Gamma(n-k+\frac{1}{2})\Gamma(1)}{\Gamma(n+1)\Gamma(\frac{1}{2})^2}$$

と表される。ただし、k は 1 の出現回数である。周辺尤度のマイナス対数は以下のように書ける：

$$-\log q(x_1, \ldots, x_n) = -\log\Gamma\left(k+\frac{1}{2}\right) - \log\Gamma\left(n-k+\frac{1}{2}\right) + \log\Gamma(n+1) + 2\log\Gamma\left(\frac{1}{2}\right)$$

Stirling の公式（$\Gamma(z) \approx \sqrt{2\pi}z^{z-\frac{1}{2}}e^{-z}$）を用いて、各項を近似する。大数の法則のもとで $k \approx np_j$ として（ここでは多項分布の α 項一般化に基づいて議論する）：

$$-\log\Gamma\left(k+\frac{1}{2}\right) \sim -\left(k+\frac{1}{2}\right)\log\left(k+\frac{1}{2}\right)+\left(k+\frac{1}{2}\right)+\frac{1}{2}\log(2\pi)$$

$$-\log\Gamma\left(n-k+\frac{1}{2}\right) \sim -\left(n-k+\frac{1}{2}\right)\log\left(n-k+\frac{1}{2}\right)+\left(n-k+\frac{1}{2}\right)+\frac{1}{2}\log(2\pi)$$

$$\log\Gamma(n+1) \sim n\log n - n + \frac{1}{2}\log(2\pi n)$$

以上をまとめて計算すると、$-\log q(x_1,\ldots,x_n)$ の主要項は

$$\sum_{j=1}^{\alpha} -k_j \log\left(\frac{k_j+1/2}{n+\alpha/2}\right) + \frac{\alpha-1}{2}\log(n+1) + \frac{\alpha-1}{2}\log(2\pi) - \log\frac{\Gamma(\alpha/2)}{\Gamma(1/2)^\alpha}$$

となる。また、その 2 個の不等式を第 1 項に代入して (6.23) が得られる。

74.

```
library(MASS)
df <- Boston
X <- as.matrix(df[,c(1,3,5,6,7,8,10,11,12,13)])
y <- df[[14]]
n <- nrow(X)
p <- ncol(X)
X <- cbind(rep(1, n), X)  # 切片

# 関数定義: 自由エネルギー (負の対数周辺尤度)
Q.2 <- function(X, y, mu.0, Lambda.0, kappa.0, m.0) {
  n <- nrow(X)
  d <- ncol(X)
  bar.x <- colMeans(X)
  beta.hat <- solve(t(X) %*% X) %*% t(X) %*% y
  S <- sum((y - X %*% beta.hat)^2)
  kappa.n <- kappa.0 + n
  m.n <- m.0 + n
  Lambda.n <- Lambda.0 + t(X) %*% X + (kappa.0 * n) / (kappa.0 + n) * (bar.x - mu.0) %*% t(bar.x - mu.0)
  value <- n * d / 2 * log(pi) +
    sum(lgamma((m.n + 1 - 1:d)/2)) - sum(lgamma((m.0 + 1 - 1:d)/2)) +
    m.0 / 2 * log(det(Lambda.0)) - m.n / 2 * log(det(Lambda.n)) +
    d / 2 * log(kappa.0 / kappa.n)
  return(value)
}
# モデル選択ループ
Q.seq <- c()
for (k in 1:p) {
  T <- combn(1:p, k)
  m <- ncol(T)
  Q.min <- Inf
  for (j in 1:m) {
    q <- c(1, T[, j] + 1)
    Q <- Q.2(X[, q], y, rep(0, k + 1), diag(k + 1), 1, 0.5)
    if (Q < Q.min) Q.min <- Q
```

```
  }
  Q.seq <- c(Q.seq, Q.min)
}
# 図の作成
plot(1:p, Q.seq, type = "b", col = "purple",
     xlab = "変数の個数", ylab = "-log 周辺尤度",
     main = "自由エネルギーによるモデル選択")
grid()
```

75.

$$B(x,y)=\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

を用いると、各項は以下のように展開できる：

$$\frac{\Gamma(k_{j-1}+a_{j-1})\Gamma(\sum_{h=j}^{\alpha}(k_h+a_h))}{\Gamma(\sum_{h=j-1}^{\alpha}(k_h+a_h))}\cdot\frac{\Gamma(\sum_{h=j-1}^{\alpha}a_h)}{\Gamma(a_{j-1})\Gamma(\sum_{h=j}^{\alpha}a_h)}$$

この積の形を $\prod_{j=2}^{\alpha}$ で繰り返すことで、分子には $\Gamma(k_j+a_j)$ の積が構成され、分母には $\Gamma(a_j)$、および全体の和に対する Γ 関数が現れる。整理すると、全体として以下の形になる：

$$\frac{\prod_{j=1}^{\alpha}\Gamma(k_j+a_j)}{\Gamma(\sum_{j=1}^{\alpha}(k_j+a_j))}\cdot\frac{\Gamma(\sum_{j=1}^{\alpha}a_j)}{\prod_{j=1}^{\alpha}\Gamma(a_j)}$$

76.

```
f <- function(x) log(x^(x - 0.5) * exp(-x) * C)                # 下限
g <- function(x) log(gamma(x))                                 # 真の log Γ (x)
h <- function(x) log(x^(x - 0.5) * exp(-x + 1 / (12 * x)) * C)   # 上限
```

77. `n * log(S) + 2 * k`

78. まず、$\sum_{i=1}^{n}(x_i-\mu)^\top\Sigma^{-1}(x_i-\mu)$ を展開する：

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\mu)^\top\Sigma^{-1}(x_i-\mu)&=\sum_{i=1}^{n}(x_i^\top\Sigma^{-1}x_i-2x_i^\top\Sigma^{-1}\mu+\mu^\top\Sigma^{-1}\mu)\\&=\sum_{i=1}^{n}x_i^\top\Sigma^{-1}x_i-2n\bar{x}^\top\Sigma^{-1}\mu+n\mu^\top\Sigma^{-1}\mu\end{aligned}$$

また、

$$(\mu-\mu_0)^\top\Sigma^{-1}(\mu-\mu_0)=\mu^\top\Sigma^{-1}\mu-2\mu_0^\top\Sigma^{-1}\mu+\mu_0^\top\Sigma^{-1}\mu_0$$

これらを用いて、全体の指数部をまとめる：

$$\begin{aligned}&-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}\Lambda_0)-\frac{1}{2}\sum_{i=1}^{n}x_i^\top\Sigma^{-1}x_i+n\bar{x}^\top\Sigma^{-1}\mu-\frac{n}{2}\mu^\top\Sigma^{-1}\mu\\&\quad-\frac{\kappa_0}{2}\mu^\top\Sigma^{-1}\mu+\kappa_0\mu_0^\top\Sigma^{-1}\mu-\frac{\kappa_0}{2}\mu_0^\top\Sigma^{-1}\mu_0\end{aligned}$$

この中で、μ に関する 2 次形式は以下のように平方完成される：

$$-\frac{\kappa_0+n}{2}\mu^\top\Sigma^{-1}\mu+(\kappa_0\mu_0+n\bar{x})^\top\Sigma^{-1}\mu=-\frac{\kappa_0+n}{2}(\mu-\mu_n)^\top\Sigma^{-1}(\mu-\mu_n)+\frac{1}{2(\kappa_0+n)}(\kappa_0\mu_0+n\bar{x})^\top\Sigma^{-1}(\kappa_0\mu_0+n\bar{x})$$

さらに、

$$\begin{aligned}&\frac{1}{2(\kappa_0+n)}(\kappa_0\mu_0+n\bar{x})^\top\Sigma^{-1}(\kappa_0\mu_0+n\bar{x})-\frac{\kappa_0}{2}\mu_0^\top\Sigma^{-1}\mu_0\\&=-\frac{n\kappa_0}{2(\kappa_0+n)}(\mu_0-\bar{x})^\top\Sigma^{-1}(\mu_0-\bar{x})+\frac{n}{2}\bar{x}^\top\Sigma^{-1}\bar{x}\end{aligned}$$

最後に、

$$\frac{n}{2}\bar{x}^\top\Sigma^{-1}\bar{x}-\frac{1}{2}\sum_{i=1}^n x_i^\top\Sigma^{-1}x_i=-\frac{1}{2}\sum_{i=1}^n(x_i-\bar{x})^\top\Sigma^{-1}(x_i-\bar{x})$$

以上をすべてまとめると、指数部は以下のように整理される：

$$\begin{aligned}&-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}\Lambda_0)-\frac{\kappa_0+n}{2}(\mu-\mu_n)^\top\Sigma^{-1}(\mu-\mu_n)\\&-\frac{n\kappa_0}{2(\kappa_0+n)}(\mu_0-\bar{x})^\top\Sigma^{-1}(\mu_0-\bar{x})-\frac{1}{2}\sum_{i=1}^n(x_i-\bar{x})^\top\Sigma^{-1}(x_i-\bar{x})\end{aligned}$$

7　スコアベースの構造学習

79. (a) asia データセットを用いて、BIC に基づく構造学習を行い、ネットワーク構造を igraph によって可視化する。

```r
library(bnlearn)
data("asia")

# BIC による構造学習(hill-climbing 法)
bn.bic <- hc(asia, score = "bic")
edges.bic <- amat(bn.bic)

# igraph 形式に変換して描画
library(igraph)
g.bic <- graph_from_adjacency_matrix(edges.bic, mode = "directed")
plot(g.bic, vertex.label = colnames(edges.bic))
```

(b) 次に BIC を AIC に変更し、同様に構造を学習する。

```r
bn.aic <- hc(asia, score = "aic")
edges.aic <- amat(bn.aic)
g.aic <- graph_from_adjacency_matrix(edges.aic, mode = "directed")
plot(g.aic, vertex.label = colnames(edges.aic))
```

BIC はペナルティ項として $\frac{d}{2}\log n$ を含み、複雑なモデルに対して厳格である。一方、AIC のペナルティは d のみであり、比較的自由度の高い構造を許す。その結果、AIC によるネットワークは BIC よりもエッジ数が多くなる傾向がある。これは過学習のリスクと引き換えに柔軟な表現力を得る。

80. 変数 `X.1`, `X.2`, `X.3` に対して、以下の R コマンドを実行した。

```r
table(X.1, X.2, X.3)
ftable(X.1, X.2, X.3)
ftable(X.2, X.3, X.1)
```

(a) table(X.1, X.2, X.3) は、多次元配列（3 次元配列）として出力される。行、列、スライス（3 番目の変数）という形で表示され、直感的には分かりやすいが、視覚的に把握しにくいことがある。
(b) ftable(X.1, X.2, X.3) は、table の出力を「フラット」に整形して表示する関数である。階層的に表示されるが、行方向と列方向に変数が展開され、コンパクトで表形式に近い。
(c) ftable(X.2, X.3, X.1) のように、変数の順序を入れ替えることで、行・列の配置が異なる出力になる。これは表形式における可読性や解析の観点で重要である。たとえば、条件付き分布の確認などにおいて柔軟な出力が得られる。

このように、table は配列的な出力を、ftable は人間にとって読みやすい表形式の出力を提供する。また、ftable は変数の順序変更に敏感であり、表示構造が大きく変わる。

81.

$$H + \frac{d}{2}\log n \le H' + \frac{d'}{2}\log n \tag{1}$$
$$H + d \ge H' + d' \tag{2}$$

と書ける。(1) より、両辺から H' を引くと、

$$H - H' \le \frac{d' - d}{2}\log n \tag{3}$$

(2) より、同様に整理すると、

$$H - H' \ge d' - d \tag{4}$$

(3), (4) を組み合わせると、

$$d' - d \le H - H' \le \frac{d' - d}{2}\log n$$

ここで、$d' - d < 0$（すなわち $d > d'$）と仮定すると、左辺は負の値となり、右辺はその半分に $\log n$ を掛けた値になる。通常、$n \ge 8$ 程度のデータ数があれば $\log n > 2$ となり、

$$\frac{d' - d}{2}\log n < d' - d$$

が成立するため、上の不等式は矛盾する。よって、(1) と (2) が同時に成り立つためには、$d \le d'$ が必要である。

82.

```r
library(bnlearn)
IC.min.discrete <- function(df, score = "bic") {
  hc(df, score = score)  # hill-climbing 法を用いる
}
# データの読み込みと構造学習
data("asia")
model_bic <- IC.min.discrete(asia, score = "bic")
model_aic <- IC.min.discrete(asia, score = "aic")
# グラフ描画
library(igraph)
# エッジリストに変換して描画
plot(graph_from_edgelist(arcs(model_bic), directed = TRUE),
     main = "BIC による構造学習",
     vertex.label = colnames(asia))
plot(graph_from_edgelist(arcs(model_aic), directed = TRUE),
```

```
    main = "AIC による構造学習",
    vertex.label = colnames(asia))
```

BIC による構造は、スパース（辺が少ない）になりやすい。過学習を抑制。AIC による構造は、複雑（辺が多め）になる傾向がある。モデル適合を優先。`hc(df, score = "...")` は両者に対応しており、`asia` データにおいては辺の数と方向にわずかな違いが見られる。

83. 11 個の Markov 同値な構造（DAG）は、次の形式に分類される。
    (a) $X \to Y \to Z$
    (b) $X \leftarrow Y \leftarrow Z$
    (c) $X \leftarrow Y \to Z$
    (d) $X \to Z \leftarrow Y$
    (e) $X \to Y \leftarrow Z$
    (f) $X \leftarrow Z \to Y$
    (g) $X \to Y, X \to Z$
    (h) $X \leftarrow Y, Z \leftarrow Y$
    (i) $X \leftarrow Z, Y \leftarrow Z$
    (j) $X \to Y, Y \to Z$
    (k) $X \to Z, Y \to Z$

    これらに対して、関数 Q を用いてスコアを計算する。構造が DAG として有効である限り、条件付き独立性の分解に基づいてスコアは次のように表せる。

```
Q1 <- Q(cbind(x)) + Q(cbind(y, x)) + Q(cbind(z, y))  # x→y→z
Q2 <- Q(cbind(z)) + Q(cbind(y, z)) + Q(cbind(x, y))  # x←y←z
Q3 <- Q(cbind(y)) + Q(cbind(x, y)) + Q(cbind(z, y))  # x←y→z
Q4 <- Q(cbind(x)) + Q(cbind(y)) + Q(cbind(z, x, y))  # x→z←y
Q5 <- Q(cbind(x, y)) + Q(cbind(z, y)) - Q(cbind(y))  # x→y←z
Q6 <- Q(cbind(x, z)) + Q(cbind(y, z)) - Q(cbind(z))  # x←z→y
Q7 <- Q(cbind(x)) + Q(cbind(y, x)) + Q(cbind(z, x))  # x→y, x→z
Q8 <- Q(cbind(y)) + Q(cbind(x, y)) + Q(cbind(z, y))  # x←y, z←y
Q9 <- Q(cbind(z)) + Q(cbind(x, z)) + Q(cbind(y, z))  # x←z, y←z
Q10 <- Q(cbind(x)) + Q(cbind(y, x)) + Q(cbind(z, y)) # x→y, y→z
Q11 <- Q(cbind(x)) + Q(cbind(y)) + Q(cbind(z, x, y)) # x→z, y→z
```

    これらをすべて実行し、最も値が小さいもの（尤度が最大の構造）を選択すればよい。

84. X と Y が独立であると仮定する。すなわち、任意の (i, j) に対して

$$p_{XY}(i, j) = p_X(i) \cdot p_Y(j)$$

が成立する。また、$k_{XY}(i, j)$ は $(X, Y) = (i, j)$ の観測回数であり、$k_X(i) = \sum_j k_{XY}(i, j)$、$k_Y(j) = \sum_i k_{XY}(i, j)$ である。このとき、K_n を展開すると

$$\begin{aligned} K_n &= \sum_i -k_X(i) \log p_X(i) + \sum_j -k_Y(j) \log p_Y(j) - \sum_{i,j} -k_{XY}(i, j) \log p_{XY}(i, j) \\ &= -\sum_i k_X(i) \log p_X(i) - \sum_j k_Y(j) \log p_Y(j) + \sum_{i,j} k_{XY}(i, j) \log p_{XY}(i, j) \end{aligned}$$

独立性より、$p_{XY}(i,j)=p_X(i)p_Y(j)$ を代入すると：

$$\sum_{i,j} k_{XY}(i,j)\log p_{XY}(i,j)=\sum_{i,j} k_{XY}(i,j)(\log p_X(i)+\log p_Y(j))$$

$$=\sum_{i,j} k_{XY}(i,j)\log p_X(i)+\sum_{i,j} k_{XY}(i,j)\log p_Y(j)=\sum_i k_X(i)\log p_X(i)+\sum_j k_Y(j)\log p_Y(j)$$

$$K_n=-\sum_i k_X(i)\log p_X(i)-\sum_j k_Y(j)\log p_Y(j)+\sum_i k_X(i)\log p_X(i)+\sum_j k_Y(j)\log p_Y(j)=0$$

85. 表の 2 行目以降の ✓ の配置は、それぞれの行に示された因果順序（トポロジカル順序）と整合的な DAG（有向非巡回グラフ）の構造に対して印が付けられている。たとえば、$X\to Y\to Z$ という因果順序は、X が最も早く、次に Y、最後に Z の順に原因となることを意味する。このとき、すべての辺がこの順序に沿っていれば、すなわち、$X\to Y$、$Y\to Z$ といった向きであれば、✓ が付く。逆に、$Z\to Y$ のように因果順序に逆らう向きの辺が含まれていれば、その構造には ✓ は付かない。以下は具体例である。
    - 構造 (a): $X\to Y\to Z$ は、因果順序 $X\to Y\to Z$ と完全に一致するので、すべての行において ✓ が付いている。
    - 構造 (f): $X\to Y$, $Z\to Y$ は、$Z\to Y$ は Z が Y よりも後に来る順序では整合しないため、$X\to Z\to Y$ のような順序では ✓ は付かない。
    - 構造 (g): $X\to Z\leftarrow Y$（V 字構造）は、Z に集まる構造は、X と Y がともに Z に向かっているため、Z が最後に来る順序（たとえば $X\to Z\to Y$）では整合的である。したがって、そのような行には ✓ が付く。

    このように、因果順序（変数のトポロジカル順）が与えられたとき、その順序に反しない向きの辺だけから構成される DAG に対してのみ ✓ が付いている。これは、探索空間を制限するうえで重要な情報であり、因果順序に基づいた構造学習の際に利用される。
86. 3 変数（X,Y,Z）から成る 8 個のマルコフネットワークに対して、スコア（例えば、最小記述長や負の対数周辺尤度）を計算するには、それぞれの無向グラフ構造に対応する分布クラスの分解形を考慮して尤度（またはその近似）を求める。各グラフに対応する尤度の計算には、以下のようにクリーク分解（clique decomposition）を用いて分布を表現し、事後分布の積の形に変換してスコアを評価する。
    (a) X,Y,Z 間に辺が存在しない完全独立構造: $Q(x^n)Q(y^n)Q(z^n)$
    (b) Y–Z のみ接続: $Q(x^n)Q(y^n,z^n)$
    (c) Z–X のみ接続: $Q(y^n)Q(z^n,x^n)$
    (d) X–Y のみ接続: $Q(z^n)Q(x^n,y^n)$
    (e) Y–Z, Z–X: $Q(x^n,z^n)Q(y^n,z^n)/Q(z^n)$
    (f) X–Y, Y–Z: $Q(x^n,y^n)Q(y^n,z^n)/Q(y^n)$
    (g) X–Z, Y–Z: $Q(x^n,z^n)Q(y^n,z^n)/Q(z^n)$
    (h) 完全グラフ: $Q(x^n,y^n,z^n)$ は完全な 3 次元分布
87. (a)(b) ともに、実行すれば良い。

88.

$$\begin{aligned}
R(X \mid \emptyset) &= Q(X) \\
R(X \mid \{Y\}) &= \max\{R(X \mid \emptyset), Q(X \mid Y)\} \\
R(X \mid \{Z\}) &= \max\{R(X \mid \emptyset), Q(X \mid Z)\} \\
R(X \mid \{Y, Z\}) &= \max\{R(X \mid \{Y\}), R(X \mid \{Z\}), Q(X \mid Y, Z)\} \\
R(Y \mid \emptyset) &= Q(Y) \\
R(Y \mid \{X\}) &= \max\{R(Y \mid \emptyset), Q(Y \mid X)\} \\
R(Y \mid \{Z\}) &= \max\{R(Y \mid \emptyset), Q(Y \mid Z)\} \\
R(Y \mid \{X, Z\}) &= \max\{R(Y \mid \{X\}), R(Y \mid \{Z\}), Q(Y \mid X, Z)\} \\
R(Z \mid \emptyset) &= Q(Z) \\
R(Z \mid \{X\}) &= \max\{R(Z \mid \emptyset), Q(Z \mid X)\} \\
R(Z \mid \{Y\}) &= \max\{R(Z \mid \emptyset), Q(Z \mid Y)\} \\
R(Z \mid \{X, Y\}) &= \max\{R(Z \mid \{X\}), R(Z \mid \{Y\}), Q(Z \mid X, Y)\}
\end{aligned}$$

$$\begin{aligned}
T(\emptyset) &= 1 \\
T(\{X\}) &= R(X \mid \emptyset) \\
T(\{Y\}) &= R(Y \mid \emptyset) \\
T(\{Z\}) &= R(Z \mid \emptyset) \\
T(\{X, Y\}) &= \max\{T(\{X\}) \cdot R(Y \mid X),\ T(\{Y\}) \cdot R(X \mid Y)\} \\
T(\{X, Z\}) &= \max\{T(\{X\}) \cdot R(Z \mid X),\ T(\{Z\}) \cdot R(X \mid Z)\} \\
T(\{Y, Z\}) &= \max\{T(\{Y\}) \cdot R(Z \mid Y),\ T(\{Z\}) \cdot R(Y \mid Z)\} \\
T(\{X, Y, Z\}) &= \max \left\{ \begin{array}{l} T(\{Y, Z\}) \cdot R(X \mid Y, Z), \\ T(\{X, Z\}) \cdot R(Y \mid X, Z), \\ T(\{X, Y\}) \cdot R(Z \mid X, Y) \end{array} \right\}
\end{aligned}$$

89. Jeffreys の事前分布では、すべての $a_{i,j} = 1/2$ と設定する。$X \in \{1, \ldots, \alpha\}$、$Y \in \{1, \ldots, \beta\}$ とし、$n_{i,j}$ を $(X, Y) = (i, j)$ の頻度、$n_{\cdot,j} = \sum_i n_{i,j}$ を $Y = j$ の頻度とする。周辺尤度の一般式を用いると、条件付き周辺尤度は

$$Q_{X|Y}(x^n, y^n) = \frac{\prod_{i=1}^{\alpha} \prod_{j=1}^{\beta} \Gamma(n_{i,j} + 1/2)}{\Gamma(n + \alpha\beta/2)} \cdot \frac{\Gamma(\alpha\beta/2)}{\prod_{i=1}^{\alpha} \prod_{j=1}^{\beta} \Gamma(1/2)} \cdot \frac{\Gamma(\beta/2)}{\prod_{j=1}^{\beta} \Gamma(n_{\cdot,j} + 1/2)} \cdot \frac{\prod_{j=1}^{\beta} \Gamma(1/2)}{\Gamma(n + \beta/2)}$$

これを整理してまとめると、

$$Q_{X|Y}(x^n, y^n) = \frac{\Gamma(\alpha\beta/2)}{\Gamma(n + \alpha\beta/2)} \cdot \prod_{i=1}^{\alpha} \prod_{j=1}^{\beta} \frac{\Gamma(n_{i,j} + 1/2)}{\Gamma(1/2)} \cdot \left[\frac{\Gamma(\beta/2)}{\Gamma(n + \beta/2)} \cdot \prod_{j=1}^{\beta} \frac{\Gamma(n_{\cdot,j} + 1/2)}{\Gamma(1/2)} \right]^{-1}$$

が成立する。

90. AIC, BIC の定義において、$S \subset S'$ のときは、$H(X \mid S') \leq H(X \mid S)$ が常に成立する（条件付きエントロピーの性質による）。また、$d(X, S') \geq d(X, S)$ も成立する（パラメータ数が増えるため）。

(1) AIC に関して：

$$\begin{aligned}
AIC(X, S) &= H(X \mid S) + d(X, S) \\
AIC(X, S') &= H(X \mid S') + d(X, S') \\
AIC(X, S) > AIC(X, S') &\Rightarrow H(X \mid S) + d(X, S) > H(X \mid S') + d(X, S')
\end{aligned}$$

ここで、$d(X,S') > d(X,S)$ であることを用いると、$H(X \mid S) > H(X \mid S')$ が必要条件である。

(2) BIC に関して：

$$BIC(X,S) = H(X \mid S) + \frac{d(X,S)}{2}\log n$$
$$BIC(X,S') = H(X \mid S') + \frac{d(X,S')}{2}\log n$$
$$BIC(X,S) > BIC(X,S') \Rightarrow H(X \mid S) + \frac{d(X,S)}{2}\log n > H(X \mid S') + \frac{d(X,S')}{2}\log n$$

このときも同様に、$d(X,S') > d(X,S)$ より、$H(X \mid S) > H(X \mid S')$ が成立する。

91.
$$\frac{a_{n+1}}{a_n} = \frac{(n+\frac{\alpha\beta}{2})(n+\frac{1}{2})}{(n+\frac{\alpha}{2})(n+\frac{\beta}{2})} \geq 1 \Longleftrightarrow (\alpha-1)(\beta-1)n \geq 0$$
$$\frac{b_{n+1}}{b_n} = \frac{(n+\delta)(n+\frac{\delta}{\alpha\beta})}{(n+\frac{\delta}{\alpha})(n+\frac{\delta}{\beta})} \geq 1 \Longleftrightarrow (1-\frac{1}{\alpha})(1-\frac{1}{\beta})n\delta \geq 0$$

より、ともに単調非減少で、$n=0$ 以外では不等式の等号が成立しない。

92. 実行すればよい。

93. (a) について、例えば
$$\frac{\Gamma(\frac{9}{4})}{\Gamma(\frac{1}{4})} = \frac{\frac{5}{4}\cdot\frac{1}{4}\cdot\Gamma(\frac{1}{4})}{\Gamma(\frac{1}{4})} = \frac{5}{4}\cdot\frac{1}{4}$$

のように有理数に変形してから比較する。そうすると、Jeffreys の事前分布では左辺、右辺がそれぞれが $\frac{1}{5}, \frac{1}{7}$ となるので、不等式が成立する。BDeu では左辺、右辺がそれぞれ $(\frac{1}{3})^2$, $(\frac{9}{20})^2$ となのるで、不等式が成立する。前者に関しては (7.9) の値を左辺に、j を (j,k) に置き換えた式をそれぞれ左辺、右辺においている。後に関しては (7.8) について同様の比較を行っている。(a) と同様の不等式は (b) ではそれぞれ、以下のようになる。

$$\frac{\Gamma(8/2)}{\Gamma(8+8/2)}\left\{\frac{\Gamma(2+1/2)}{\Gamma(1/2)}\right\}^4\left[\frac{\Gamma(4/2)}{\Gamma(8+4/2)}\left\{\frac{\Gamma(2+1/2)}{\Gamma(1/2)}\right\}^4\right]^{-1} = \frac{3\cdot 2}{11\cdot 10}$$
$$> \frac{5}{13\cdot 15} = \frac{\Gamma(16/2)}{\Gamma(8+16/2)}\left\{\frac{\Gamma(2+1/2)}{\Gamma(1/2)}\right\}^4\left[\frac{\Gamma(8/2)}{\Gamma(8+8/2)}\left\{\frac{\Gamma(2+1/2)}{\Gamma(1/2)}\right\}^2\right]^{-1}$$

$$\left\{\left[\frac{\Gamma(2+1/4)}{\Gamma(1/4)}\right]^{-1}\frac{\Gamma(2+1/8)}{\Gamma(1/8)}\right\}^4 = \left(\frac{9}{20}\right)^4 < \left(\frac{17}{36}\right)^4 = \left\{\left[\frac{\Gamma(2+1/8)}{\Gamma(1/8)}\right]^{-1}\frac{\Gamma(2+1/16)}{\Gamma(1/16)}\right\}^4$$

94. `x[i] <- x[i] - prob[j]`

95. 実行すれば良い。

96.
$$\log\frac{Q_{XY}(x^n,y^n)}{Q_X(x^n)\cdot Q_Y(y^n)} = \log p(x^n,y^n \mid \hat{\theta}_{XY}) - \log p(x^n \mid \hat{\theta}_X) - \log p(y^n \mid \hat{\theta}_Y)$$
$$- \left[\frac{(\alpha\beta-1)}{2} - \frac{(\alpha-1)}{2} - \frac{(\beta-1)}{2}\right]\log n + o(\log n)$$
$$= I_n(x^n,y^n) - \frac{(\alpha-1)(\beta-1)}{2}\log n + o(\log n)$$

実行すれば良い。

97. 実行すれば良い。

98.

```
J.n <- function(x, y) {
  n <- length(x)
  return(I.n(x, y) - log(n) / n)
}
```

を実行すれば良い。

99. 仮定より以下が成り立つ：

$$X_k \perp\!\!\!\perp X_{\overline{\pi}_k} \mid X_{\pi_k}$$

ここで、π_k は X_k の親集合、$\overline{\pi}_k$ は $\{1,\dots,k-1\} \setminus \pi_k$ を表す。次に、集合

$$\pi^* := \overline{\pi}_k \cap \overline{S \cup T \cup U}$$

を定義する。これは、S, T, U, π_k のいずれにも含まれない変数の集合である。また、$T \cap \pi_k = \emptyset$ であることに注意する。なぜなら、T に属する変数から k に矢印が存在すれば、$k \perp\!\!\!\perp_G T \mid U$ という仮定と矛盾するためである。よって、$T \subseteq \overline{\pi}_k$ が成立し、$T \cup \pi^* \subseteq \overline{\pi}_k$ が成り立つ。

これにより、縮小性（graphoid axioms）を用いて、

$$X_k \perp\!\!\!\perp X_{T \cup \pi^*} \mid X_{\pi_k}$$

が導かれる。

さらに、条件付き独立の弱結合性（weak union）により、条件の変数集合に $X_{S \cup U}$ を追加しても条件付き独立性は保たれるので、

$$X_k \perp\!\!\!\perp X_{T \cup \pi^*} \mid X_{S \cup U \cup \pi_k}$$

が得られる。

最後に、縮小性を再度用いることで、$X_{T \cup \pi^*}$ の部分集合である X_T に関しても、

$$X_k \perp\!\!\!\perp X_T \mid X_{S \cup U \cup \pi_k}$$

が成り立つ。

100. I_n に対して定数項を引いた形である J_n は、項 $\frac{(\alpha-1)(\beta-1)}{2n}$ が変数 X, Y の状態数に依存して異なることから、辺の重みの順位が I_n と異なる場合がある。したがって、結合する辺の順序（すなわち MST の構造）も異なる可能性がある。しかし、すべての確率変数が 2 値（すなわち $\alpha = \beta = 2$）である場合には、補正項はすべて

$$\frac{(2-1)(2-1)}{2n} = \frac{1}{2n}$$

と等しい。よって、すべての変数対 (X, Y) について $J_n(X,Y) = I_n(X,Y) - \frac{1}{2n}$ となり、すべての辺で等しく同じ定数を引いたことになる。これは順位には影響を与えないため、J_n による重み付けと I_n による重み付けは順序が完全に一致する。