第 4 回　多変量分布と統計的推測に必要な分布（3.3, 3.5）

村澤 康友

2020 年 5 月 19 日

今日のポイント

1. (X,Y) の同時 cdf は $F_{X,Y}(x,y) := \Pr[X \leq x, Y \leq y]$．$X$ または Y のみの cdf を周辺 cdf という．(X,Y) の同時 pmf は $p_{X,Y}(x,y) := \Pr[X = x, Y = y]$．$X$ または Y のみの pmf を周辺 pmf という．多重積分すると同時 cdf が得られる関数（同時 cdf の交差偏導関数）を同時 pdf という．
2. $g(X,Y)$ の期待値は $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$．$X$ と Y の共分散は $\mathrm{cov}(X,Y) := \mathrm{E}((X-\mathrm{E}(X))(Y-\mathrm{E}(Y)))$．標準化した確率変数の共分散を相関係数という．
3. $Y = y$ が与えられたときの X の条件つき pdf は $f_{X|Y}(x|Y=y) := f_{X,Y}(x,y)/f_Y(y)$．$Y = y$ が与えられたときの X の条件つき期待値は $\int_{-\infty}^{\infty} x f_{X|Y}(x|Y=y)\,\mathrm{d}x$．$f_{X|Y}(x|Y=y) = f_X(x)$ なら X と Y は独立という．
4. 測定誤差は正規分布にしたがう．正規分布の線形変換も正規分布であり，標準化した正規分布を標準正規分布という．
5. $Z_1,\dots,Z_n \sim \mathrm{N}(0,1)$ が独立のとき，$Z_1^2 + \dots + Z_n^2 \sim \chi^2(n)$．$Z \sim \mathrm{N}(0,1)$ と $X \sim \chi^2(n)$ が独立のとき，$Z/\sqrt{X/n} \sim \mathrm{t}(n)$．$U \sim \chi^2(m)$ と $V \sim \chi^2(n)$ が独立のとき，$(U/m)/(V/n) \sim \mathrm{F}(m,n)$．

目次

1　同時分布と周辺分布

1.1　累積分布関数

(X,Y) を確率ベクトルとする．

定義 1. (X,Y) の同時（結合）*cdf* は，任意の (x,y) について

$$F_{X,Y}(x,y) := \Pr[X \leq x, Y \leq y]$$

定義 2. X の周辺 *cdf* は，任意の x について

$$F_X(x) := \Pr[X \leq x]$$

注 1. 同時 cdf と周辺 cdf の関係は

$$\begin{aligned} F_X(x) &:= \Pr[X \leq x] \\ &= \Pr[X \leq x, Y < \infty] \\ &= F_{X,Y}(x, \infty) \end{aligned}$$

1.2　確率質量関数（p. 50）

(X, Y) を離散確率ベクトルとする．

定義 3. (X, Y) の同時（結合）*pmf* は，任意の (x, y) について

$$p_{X,Y}(x, y) := \Pr[X = x, Y = y]$$

定義 4. X の周辺 *pmf* は，任意の x について

$$p_X(x) := \Pr[X = x]$$

注 2. 同時 pmf と周辺 pmf の関係は

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

1.3　確率密度関数

(X, Y) を連続確率ベクトルとする．

定義 5. 任意の (x, y) について

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s, t) \,\mathrm{d}s \,\mathrm{d}t$$

となる $f_{X,Y}(.,.)$ を (X, Y) の同時（結合）*pdf* という．

注 3. 任意の a, b, c, d について

$$\begin{aligned} &\Pr[a < X \leq b, c < Y \leq d] \\ &= \int_c^d \int_a^b f_{X,Y}(x, y) \,\mathrm{d}x \,\mathrm{d}y \end{aligned}$$

注 4. $F_{X,Y}(.,.)$ が微分可能なら

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x, y)$$

定義 6. X の周辺 *pdf* は，任意の x について

$$f_X(x) := \int_{-\infty}^{\infty} f_{X,Y}(x, y) \,\mathrm{d}y$$

2　積率

2.1　期待値

定義 7. $g(X, Y)$ の期待値は

$$\begin{aligned} &\mathrm{E}(g(X, Y)) \\ &:= \begin{cases} \sum_x \sum_y g(x, y) p_{X,Y}(x, y) & \text{（離散）} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) \,\mathrm{d}x \,\mathrm{d}y & \text{（連続）} \end{cases} \end{aligned}$$

定理 1 (期待値の線形性)**.**

$$\mathrm{E}(aX + bY) = a \,\mathrm{E}(X) + b \,\mathrm{E}(Y)$$

証明. 復習テスト． □

2.2　共分散（p. 50）

定義 8. X と Y の共分散は

$$\mathrm{cov}(X, Y) := \mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y)))$$

注 5. σ_{XY} と表す．

注 6. X が大きいと Y も大きいなら共分散は正，X が大きいと Y は小さいなら共分散は負．

定理 2.

$$\mathrm{cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X) \,\mathrm{E}(Y)$$

証明. 復習テスト． □

定理 3.

$$\begin{aligned} \mathrm{var}(aX + bY) = a^2 \,\mathrm{var}(X) + 2ab \,\mathrm{cov}(X, Y) \\ + b^2 \,\mathrm{var}(Y) \end{aligned}$$

証明. 復習テスト． □

2.3　相関係数（p. 51）

定義 9. 確率変数から平均を引き標準偏差で割る変換を**標準化**という．

注 7. 式で表すと

$$Z := \frac{X - \mu_X}{\sigma_X}$$

$\mathrm{E}(Z) = 0$，$\mathrm{var}(Z) = 1$ となる．

定義 10. 標準化した確率変数の共分散を**相関係数**という．

注 8. X と Y の関係の強さを表す．

注 9. ρ_{XY} と表す．すなわち

$$\begin{aligned}\rho_{XY} &:= \operatorname{cov}\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right) \\ &= \mathrm{E}\left(\frac{X-\mu_X}{\sigma_X}\frac{Y-\mu_Y}{\sigma_Y}\right) \\ &= \frac{\mathrm{E}((X-\mu_X)(Y-\mu_Y))}{\sigma_X\sigma_Y} \\ &= \frac{\sigma_{XY}}{\sigma_X\sigma_Y}\end{aligned}$$

定義 11. $\rho_{XY} = 0$ なら X と Y は無相関という．

定理 4 (コーシー＝シュワルツの不等式)**.**

$$|\operatorname{cov}(X,Y)| \le \operatorname{var}(X)^{1/2}\operatorname{var}(Y)^{1/2}$$

証明. 省略. □

系 1.

$$|\rho_{XY}| \le 1$$

3　条件つき分布と確率変数の独立性

3.1　条件つき分布（p. 54）

定義 12. $Y \le y$ が与えられたときの X の条件つき *cdf* は，任意の x について

$$F_{X|Y}(x|Y \le y) := \frac{F_{X,Y}(x,y)}{F_Y(y)}$$

注 10. 条件つき確率で定義する．

定義 13. $Y = y$ が与えられたときの X の条件つき *pmf* は，任意の x について

$$p_{X|Y}(x|Y = y) := \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

定義 14. $Y = y$ が与えられたときの X の条件つき *pdf* は，任意の x について

$$f_{X|Y}(x|Y = y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

注 11. 条件つき確率と同様に定義する．

定義 15. $Y = y$ が与えられたときの X の条件つき期待値は

$$\begin{aligned}&\mathrm{E}(X|Y=y) \\ &:= \begin{cases}\sum_x x p_{X|Y}(x|Y=y) & \text{（離散）} \\ \int_{-\infty}^{\infty} x f_{X|Y}(x|Y=y)\,\mathrm{d}x & \text{（連続）}\end{cases}\end{aligned}$$

定義 16. $Y = y$ が与えられたときの X の条件つき分散は

$$\operatorname{var}(X|Y=y) := \mathrm{E}\left((X - \mathrm{E}(X|Y=y))^2|Y=y\right)$$

定理 5 (繰り返し期待値の法則)**.**

$$\mathrm{E}(\mathrm{E}(X|Y)) = \mathrm{E}(X)$$

証明.

$$\begin{aligned}\mathrm{E}(\mathrm{E}(X|Y)) &:= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,\mathrm{d}x f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\frac{f_{X,Y}(x,y)}{f_Y(y)}\,\mathrm{d}x f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \mathrm{E}(X)\end{aligned}$$

□

3.2　確率変数の独立性（p. 52）

定義 17. 任意の (x,y) について

$$f_{X|Y}(x|Y=y) = f_X(x)$$

なら X と Y は独立という．

注 12. 条件つき pdf の定義より

$$\begin{aligned}&f_{X|Y}(x|Y=y) = f_X(x) \\ &\iff f_{X,Y}(x,y) = f_X(x)f_Y(y)\end{aligned}$$

定義 18. 任意の $(x_1,\dots,x_n)$ について

$$f_{X_1,\dots,X_n}(x_1,\dots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n)$$

なら $X_1,\dots,X_n$ は独立という．

注 13. cdf で定義してもよい．

定理 6. X と Y が独立なら，任意の $f(.)$ と $g(.)$ について

$$\mathrm{E}(f(X)g(Y)) = \mathrm{E}(f(X))\,\mathrm{E}(g(Y))$$

証明．(X,Y) が連続なら

$$\begin{aligned}
&\mathrm{E}(f(X)g(Y)) \\
&:= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)g(y)f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)g(y)f_X(x)f_Y(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} f(x)f_X(x)\,\mathrm{d}x \int_{-\infty}^{\infty} g(y)f_Y(y)\,\mathrm{d}y \\
&= \mathrm{E}(f(X))\,\mathrm{E}(g(Y))
\end{aligned}$$

離散の場合も同様． □

系 2. X と Y が独立なら

$$\mathrm{cov}(X,Y) = 0$$

証明．復習テスト． □

注 14．すなわち独立なら無相関．逆は必ずしも成立しない．

4 統計的推測に必要な分布

4.1 正規分布（p. 64）

定義 19. **正規分布**の pdf は

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

注 15．$\mathrm{N}\left(\mu,\sigma^2\right)$ と書く．

例 1. 測定誤差，標本平均（中心極限定理）．

定義 20. $\mathrm{N}(0,1)$ を**標準正規分布**という．

注 16．$\mathrm{N}(0,1)$ の cdf を $\Phi(.)$，pdf を $\phi(.)$ で表す．すなわち

$$\begin{aligned}
\phi(x) &:= \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2} \\
\Phi(x) &:= \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\,\mathrm{d}z
\end{aligned}$$

例 2. $\mathrm{N}(0,1)$ の cdf と pdf は図 1 の通り．

定理 7. $X \sim \mathrm{N}\left(\mu,\sigma^2\right)$ なら

$$\begin{aligned}
\mathrm{E}(X) &= \mu \\
\mathrm{var}(X) &= \sigma^2
\end{aligned}$$

証明．省略． □

定理 8. $X \sim \mathrm{N}\left(\mu,\sigma^2\right)$ なら

$$aX + b \sim \mathrm{N}\left(a\mu + b, a^2\sigma^2\right)$$

証明．省略． □

系 3. $X \sim \mathrm{N}\left(\mu,\sigma^2\right)$ なら

$$\frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$$

証明．前の定理で $a := 1/\sigma$，$b := -\mu/\sigma$ とする． □

注 17．したがって $X \sim \mathrm{N}\left(\mu,\sigma^2\right)$ の累積確率は標準正規分布表から求まる．すなわち

$$\begin{aligned}
F_X(x) &:= \Pr[X \le x] \\
&= \Pr\left[\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right] \\
&= \Phi\left(\frac{x-\mu}{\sigma}\right)
\end{aligned}$$

ただし $\Phi(.)$ でなく $Q(.) := 1 - \Phi(.)$ の表の場合も多い．

例 3. $X \sim \mathrm{N}(1,9)$ について $\Pr[X \le 2]$ を求める．$(X-1)/3 \sim \mathrm{N}(0,1)$ より

$$\begin{aligned}
\Pr[X \le 2] &= \Pr\left[\frac{X-1}{3} \le \frac{2-1}{3}\right] \\
&= \Phi\left(\frac{1}{3}\right) \\
&= 1 - Q\left(\frac{1}{3}\right) \\
&= 1 - .3707 \\
&= .6293
\end{aligned}$$

4.2 χ^2 分布（p. 67）

定義 21. $Z_1,\dots,Z_n \sim \mathrm{N}(0,1)$ が独立のとき $Z_1^2 + \dots + Z_n^2$ の分布を**自由度** n **の** χ^2 **分布**という．

注 18．$\chi^2(n)$ と書く．

注 19．累積確率は χ^2 分布表を参照．

例 4. $\chi^2(n)$ の pdf の例は図 2 の通り．

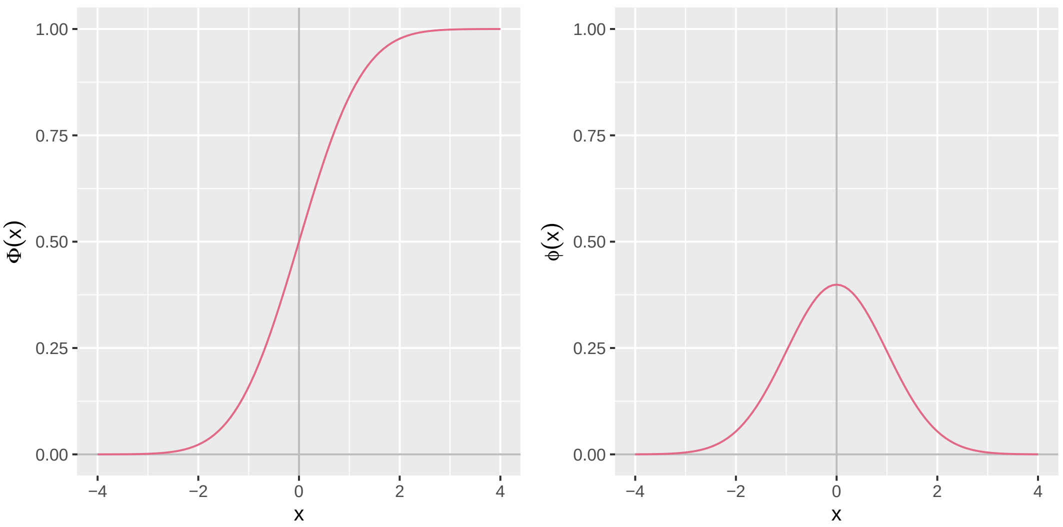

図 1　$\mathrm{N}(0,1)$ の cdf と pdf

定理 9. $X \sim \chi^2(n)$ なら

$$\mathrm{E}(X) = n$$

証明. $X = Z_1^2 + \cdots + Z_n^2$ とすると

$$\begin{aligned}\mathrm{E}(X) &= \mathrm{E}\left(Z_1^2 + \cdots + Z_n^2\right) \\ &= \mathrm{E}\left(Z_1^2\right) + \cdots + \mathrm{E}\left(Z_n^2\right) \\ &= \mathrm{var}(Z_1) + \cdots + \mathrm{var}(Z_n) \\ &= n\end{aligned}$$

□

4.3　t 分布（p. 69）

定義 22. $Z \sim \mathrm{N}(0,1)$ と $X \sim \chi^2(n)$ が独立のとき $Z/\sqrt{X/n}$ の分布を**自由度** n **の** t **分布**という.

注 20. $\mathrm{t}(n)$ と書く.

注 21. 累積確率は t 分布表を参照.

注 22. $\mathrm{t}(1)$ はコーシー分布，$\mathrm{t}(\infty)$ は $\mathrm{N}(0,1)$.

例 5. $\mathrm{t}(n)$ の pdf の例は図 3 の通り.

4.4　F 分布（p. 70）

定義 23. $U \sim \chi^2(m)$ と $V \sim \chi^2(n)$ が独立のとき $(U/m)/(V/n)$ の分布を**自由度** (m,n) **の** F **分布**という.

注 23. $\mathrm{F}(m,n)$ と書く.

注 24. 累積確率は F 分布表を参照.

注 25. $X \sim \mathrm{F}(m,n)$ なら $1/X \sim \mathrm{F}(n,m)$.

注 26. $t \sim \mathrm{t}(n)$ なら $t^2 \sim \mathrm{F}(1,n)$.

例 6. F 分布の pdf の例は図 4 の通り.

5　今日のキーワード

同時（結合）cdf，周辺 cdf，同時（結合）pmf，周辺 pmf，同時（結合）pdf，周辺 pdf，期待値，共分散，標準化，相関係数，無相関，条件つき cdf，条件つき pmf，条件つき pdf，条件つき期待値，条件つき分散，繰り返し期待値の法則，独立，正規分布，標準正規分布，χ^2 分布，t 分布，F 分布

6　次回までの準備

提出　宿題 2
復習　教科書第 3 章 3, 5 節，復習テスト 4
予習　教科書第 4 章

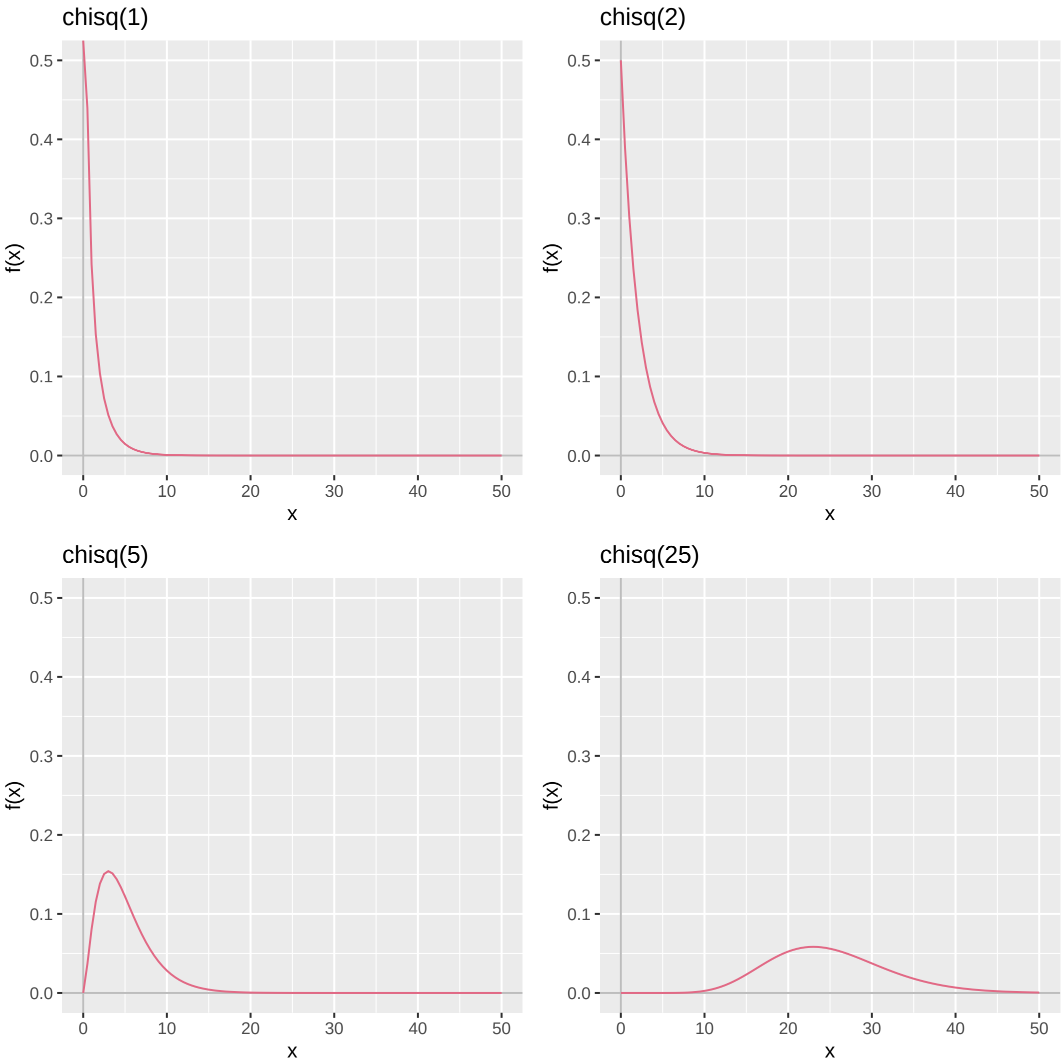

図 2　$\chi^2(n)$ の pdf の例

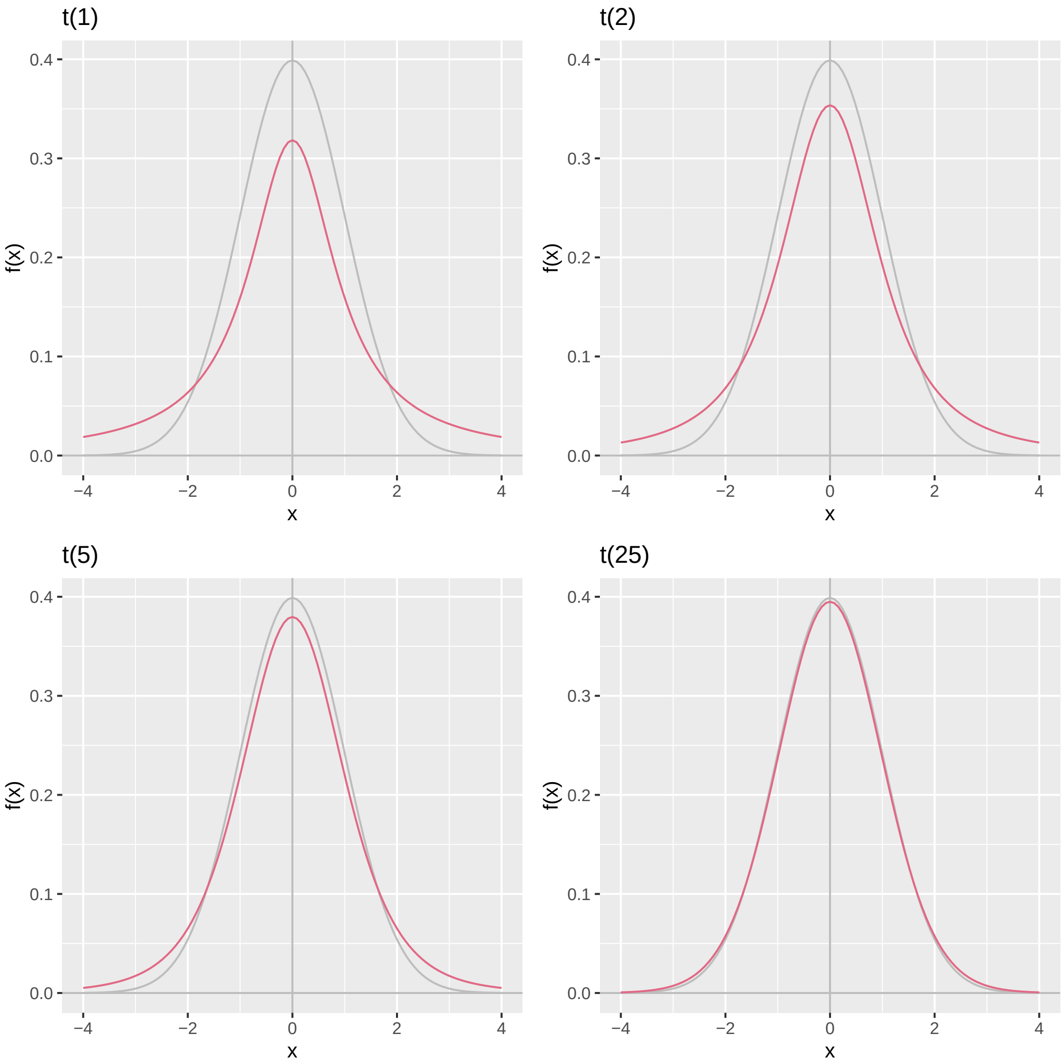

図 3　t(n) の pdf の例

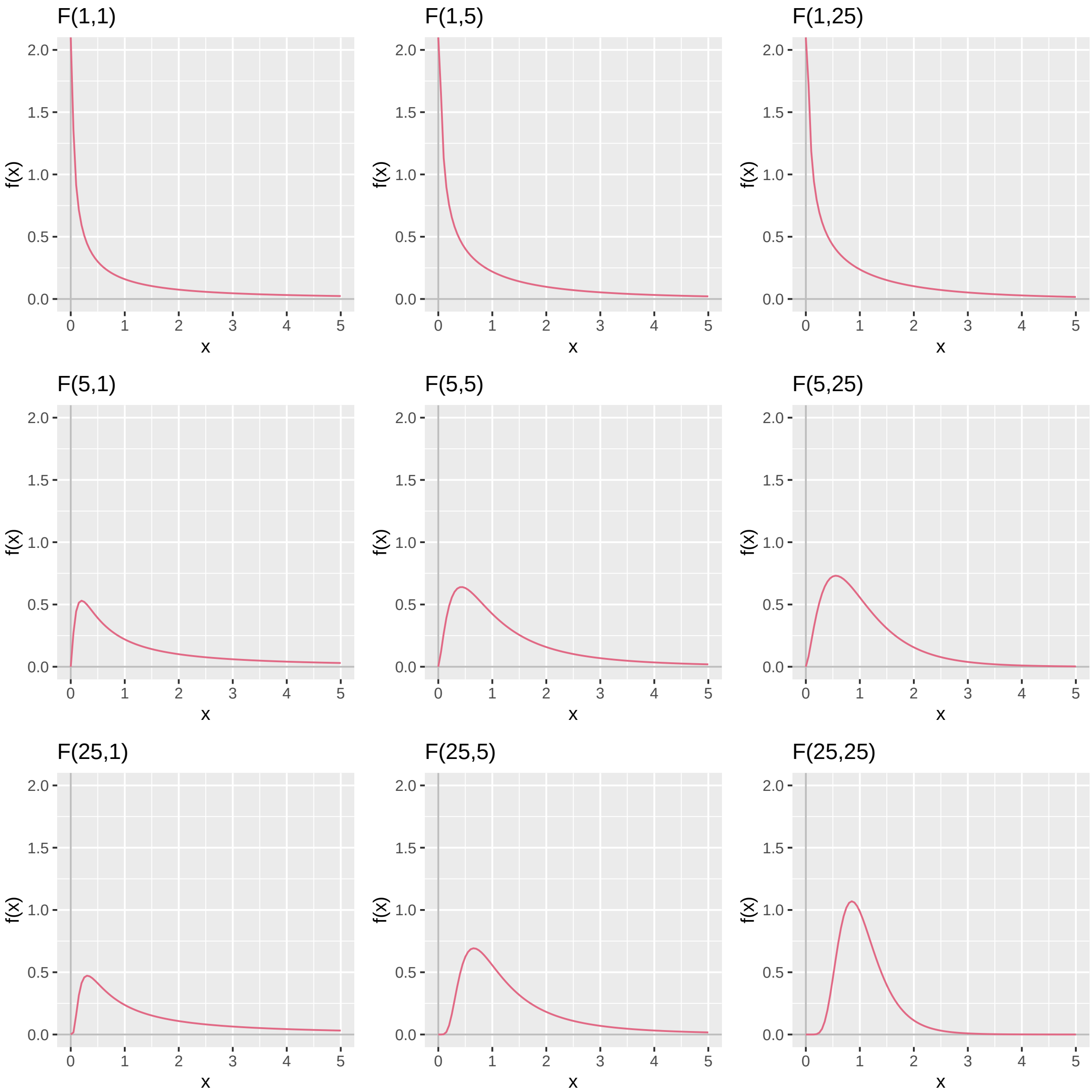

図 4　F 分布の pdf の例