データ解析 (第6回)

静岡大学システム工学科
安藤 和敏

(6) 統計量に関係する確率分布

母集団分布が $N(\mu, \sigma^2)$ の母集団から, n 個のデータ $x_1, \cdots, x_n$ を採取する.
$x_1, \cdots, x_n$ を確率変数と考えると,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.1}$$

は, 確率変数となる.

また,

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} \quad (2.2)$$

も同様に確率変数となる.

これら確率変数の分布を調べてみよう.

平均 $\bar{x}$ の分布

$\bar{x}$ の分布

$$\bar{x} \text{ は正規分布 } N(\mu, \sigma^2/n) \text{ に従う．} \tag{2.57}$$

$\frac{S_{xx}}{\sigma^2}$ の分布

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} \quad (2.2)$$

χ^2 分布

$\chi^2 = \frac{S_{xx}}{\sigma^2}$ は自由度 $\phi = n - 1$ の χ^2 分布 ($\chi^2(\phi)$ と表示) に従う. (2.58)

χ^2 分布の確率密度関数

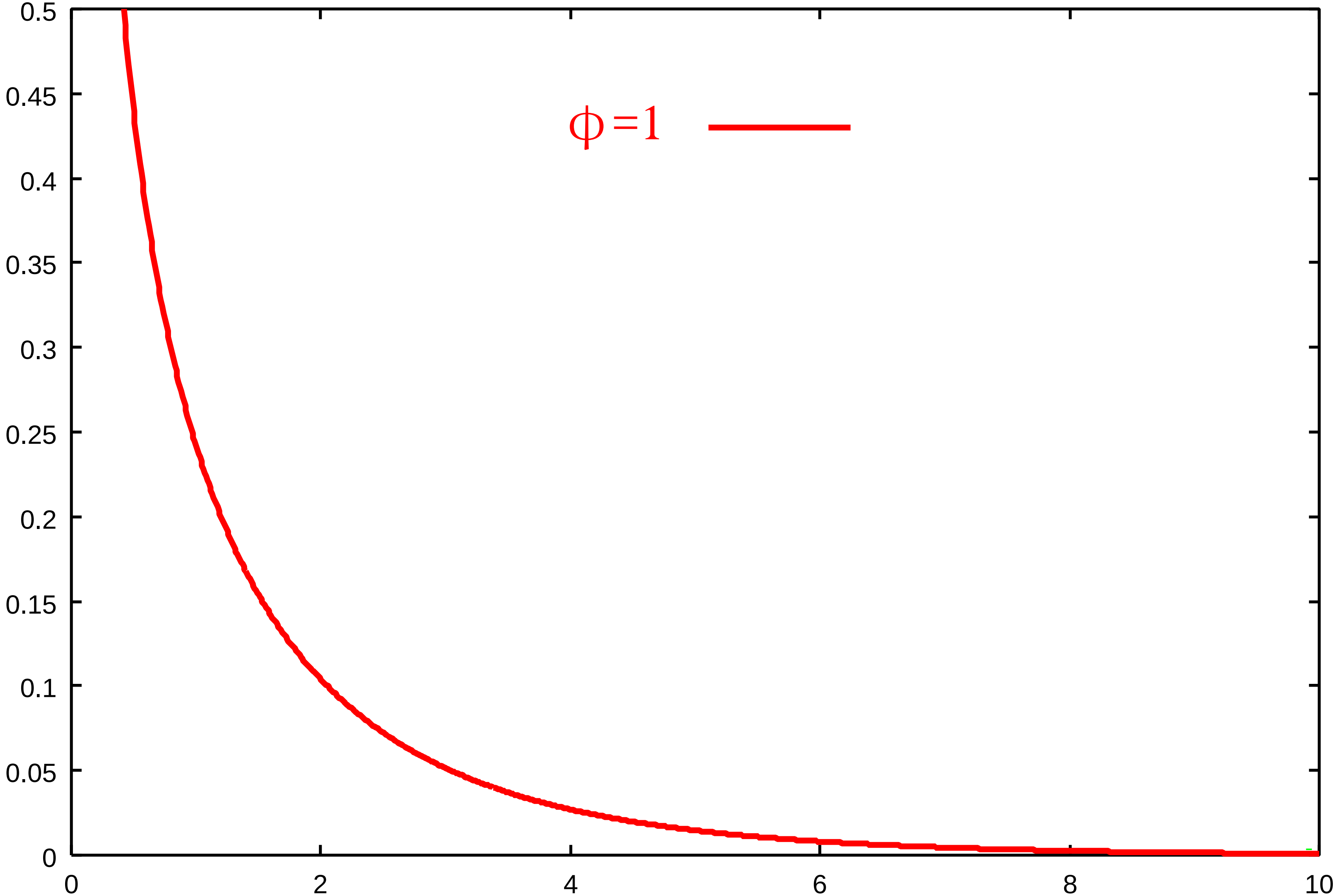
自由度 ϕ の χ^2 分布の密度関数は,

$$\frac{1}{2^{\phi/2}\Gamma\left(\frac{\phi}{2}\right)} \cdot x^{(\phi/2)-1}\mathrm{e}^{-\frac{x}{2}}$$

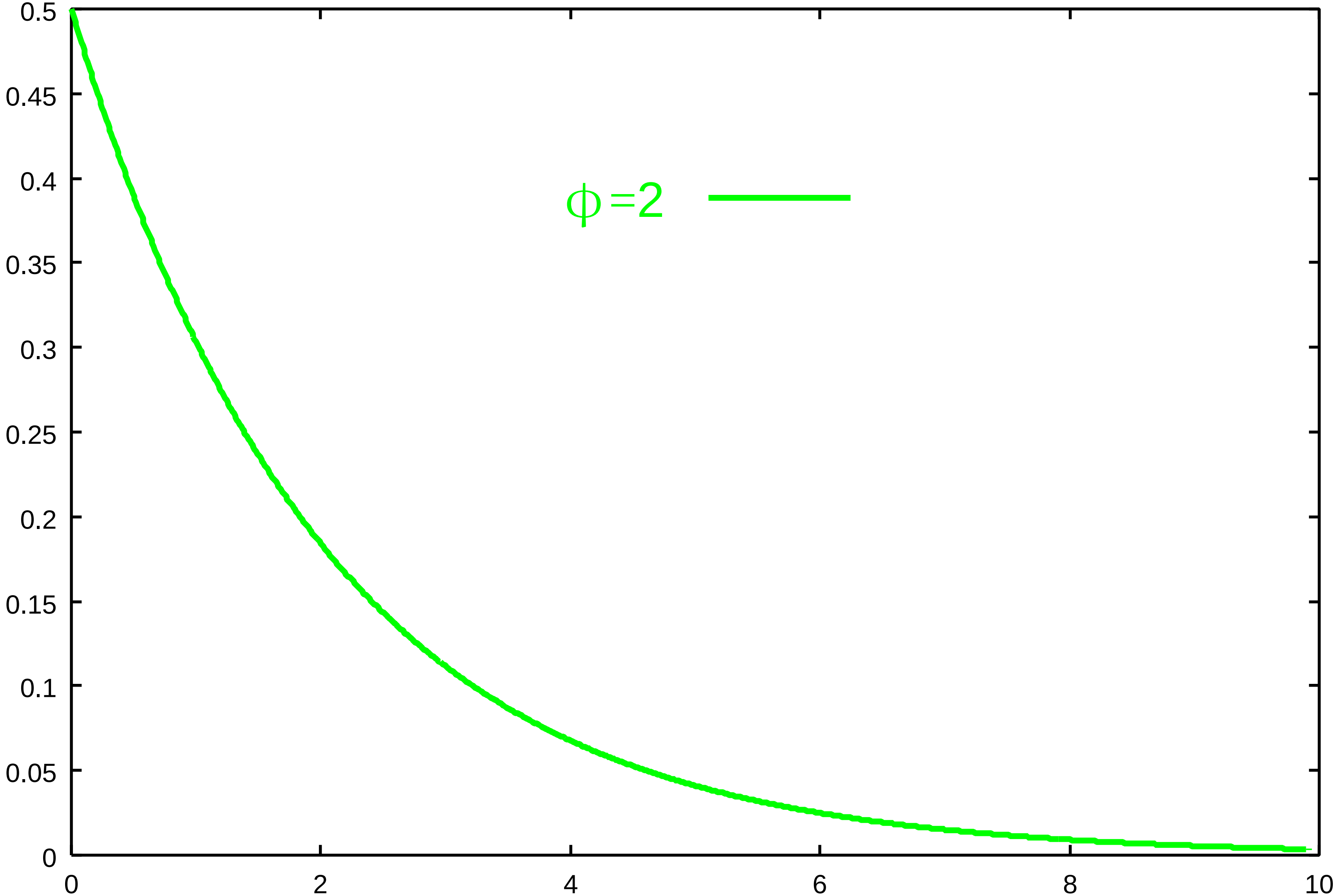
で与えられる. ここで, ガンマ関数 Γ (微積でやったはず) は以下で与えられる.

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}\mathrm{d}t.$$

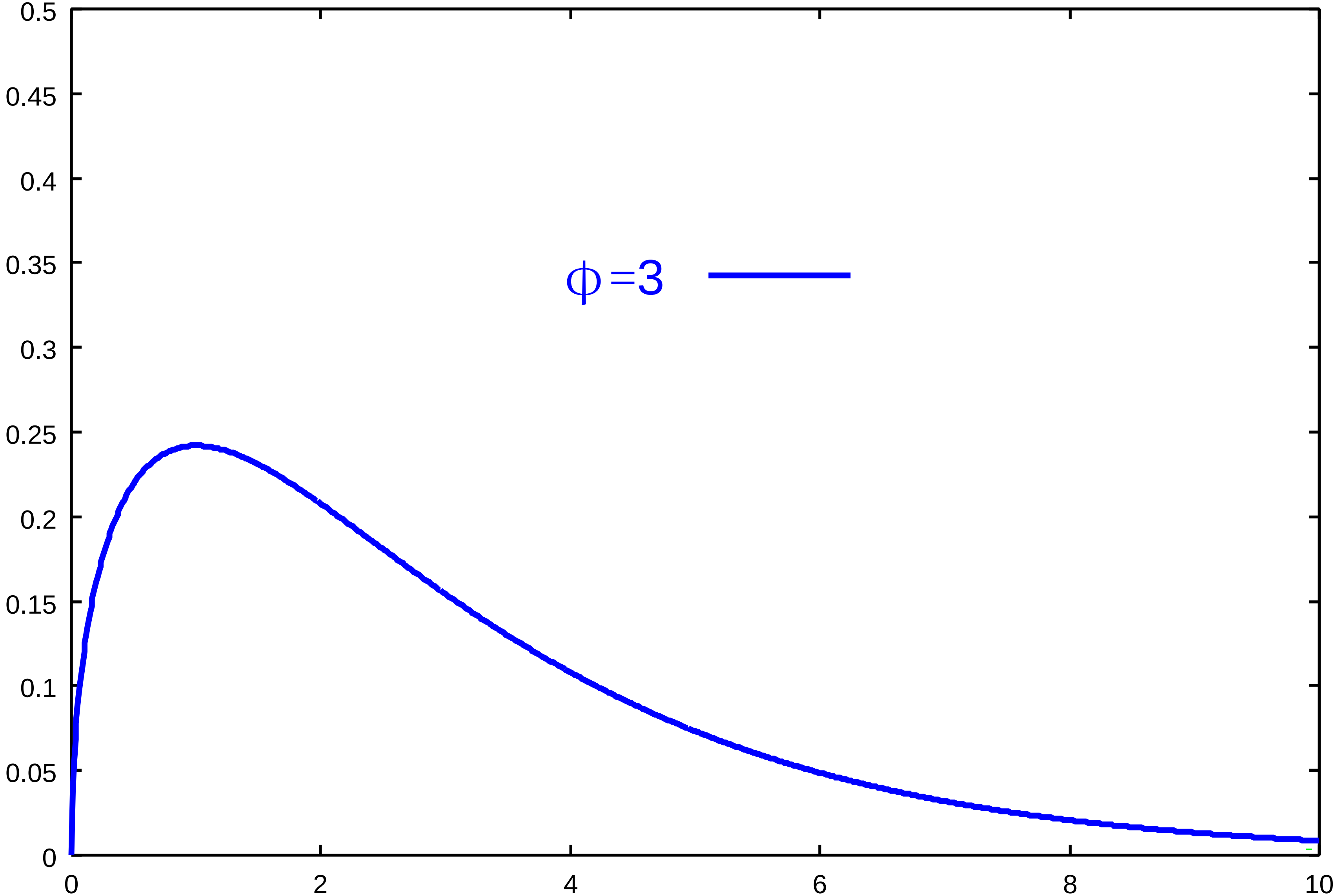
χ^2 分布のグラフ ($\phi = 1$)



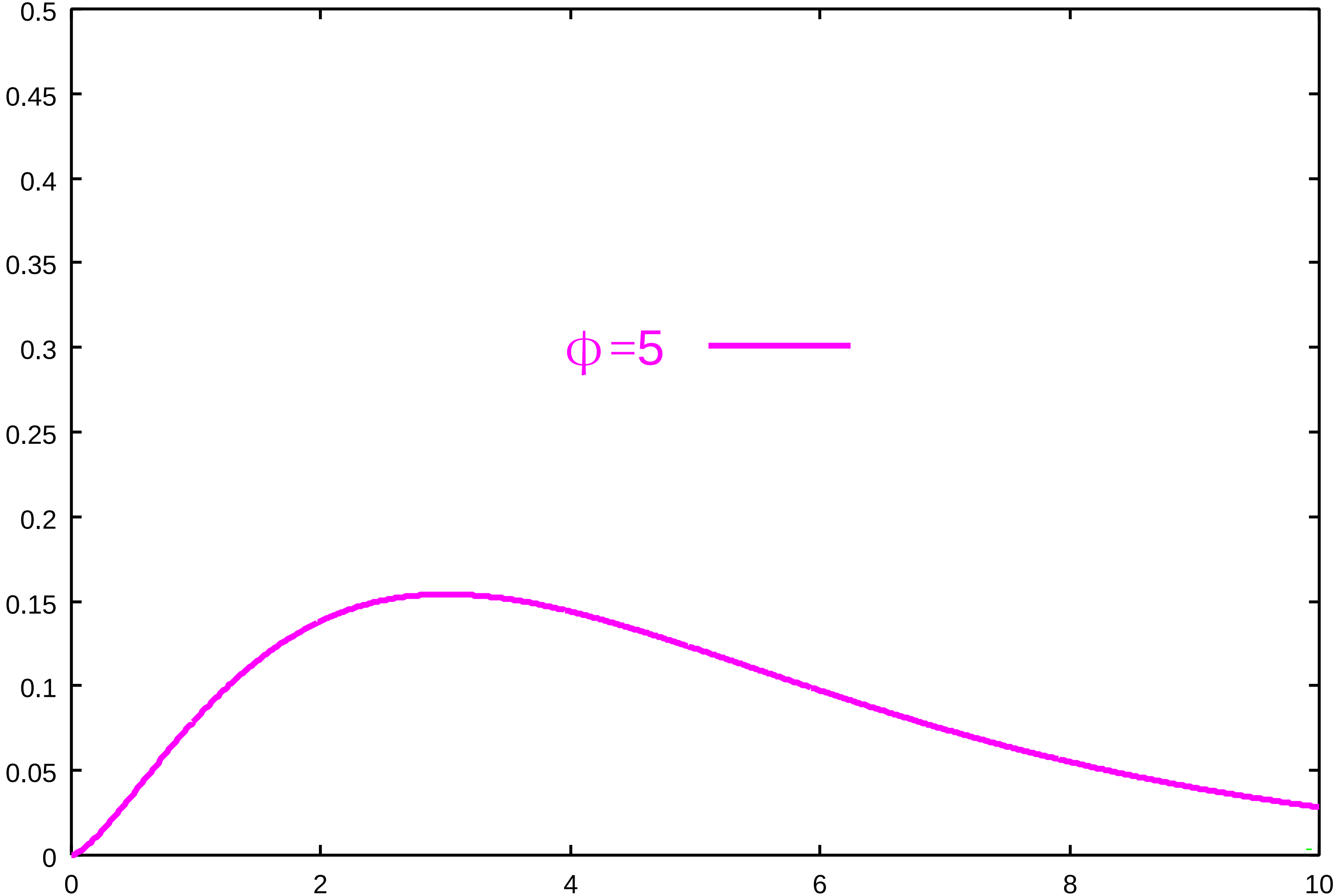
χ^2 分布のグラフ ($\phi = 2$)



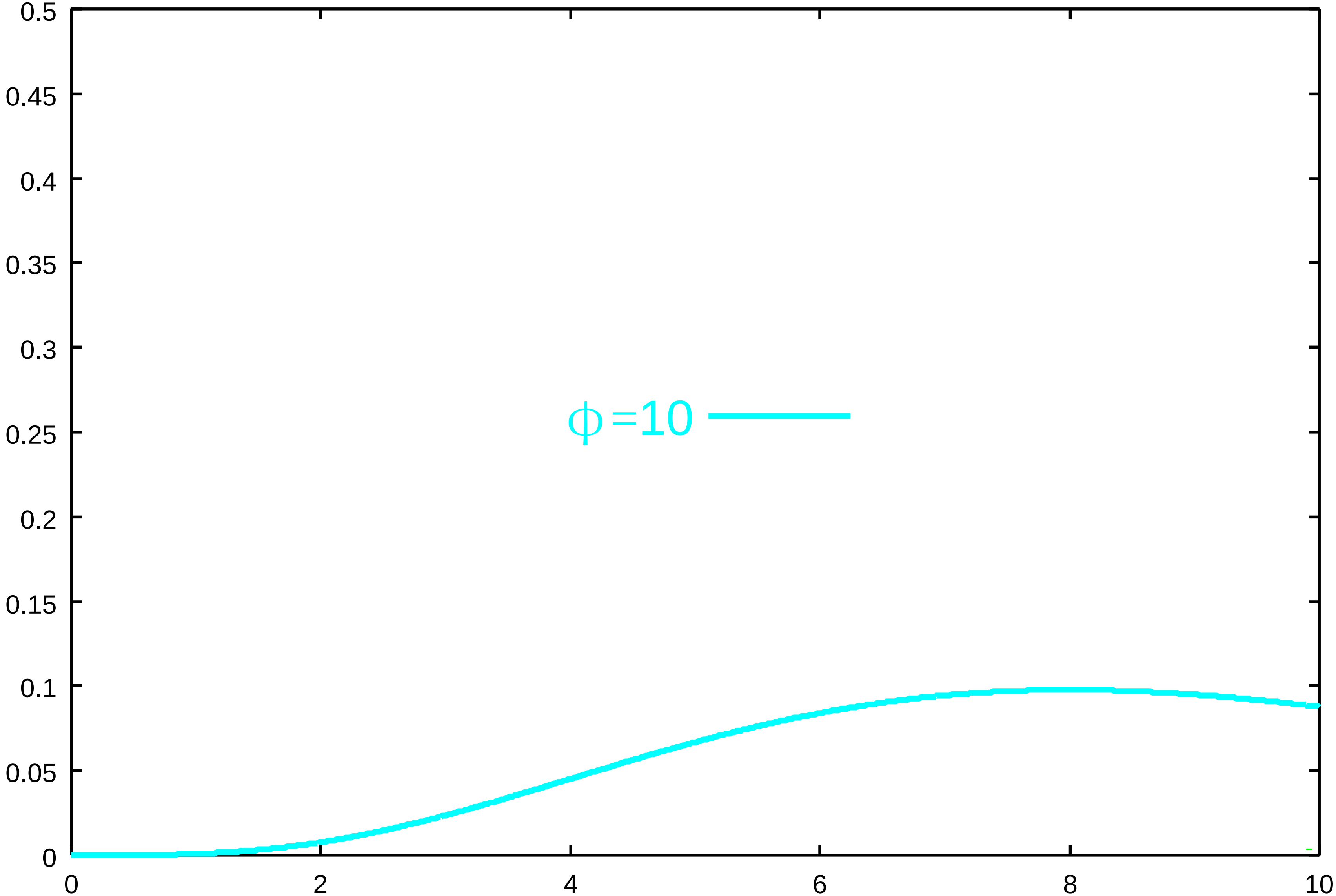
χ^2分布のグラフ ($\phi = 3$)



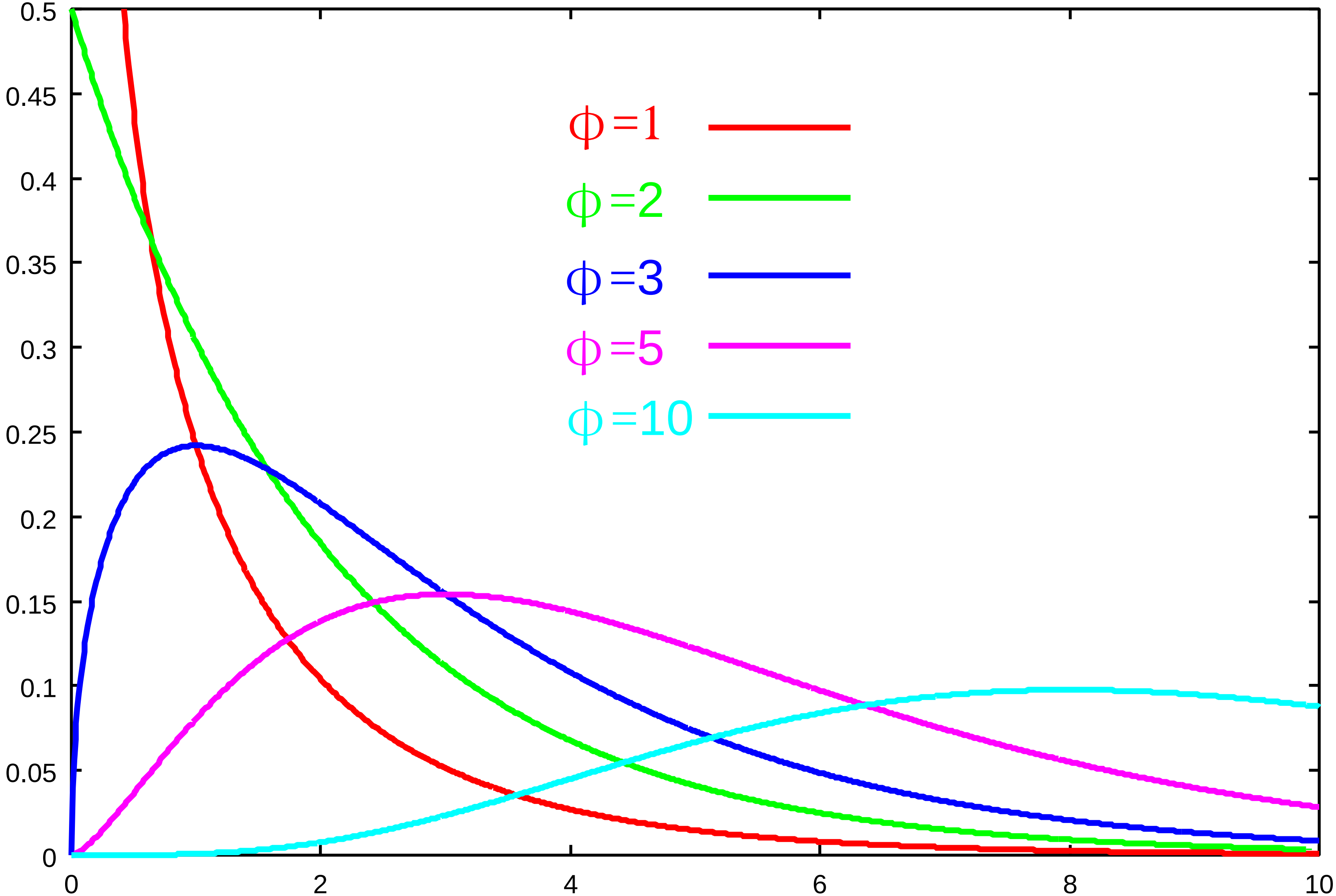
χ^2分布のグラフ ($\phi = 5$)



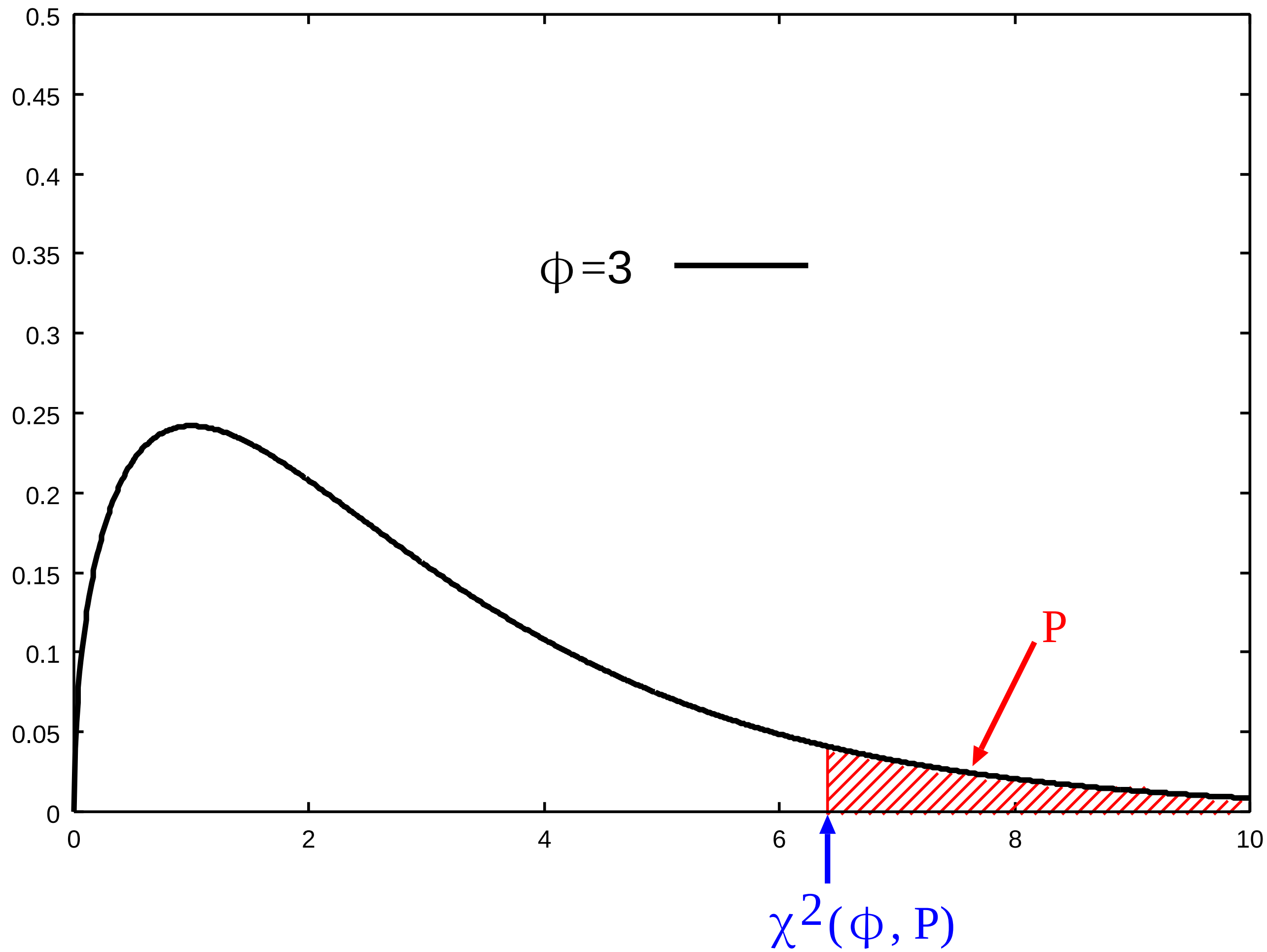
χ^2 分布のグラフ ($\phi = 10$)



χ^2 分布のグラフ ($\phi = 1, 2, 3, 5, 10$)



χ^2 表 (p. 236, 付表3) の読み方



$$P = \Pr(\chi^2(\phi, P) \leq x) \tag{2.59}$$

ϕ と P から $\chi^2(\phi, P)$ を求める (例)

ϕ \ P	···	.50	.25	.10	.05	.025
1						
2						
3				6.25		
4						
5						

$$\phi = 3, P = 0.10 \Longrightarrow \chi^2(3, 0.10) = 6.25$$

$\chi^2(\phi, P)$ を自由度 ϕ の χ^2 分布の上側 100P% 点と呼ぶ.

t 分布

(2.57) より, $\bar{x} \sim N(\mu, \sigma^2/n)$ であるから,

$$\frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1) \qquad (2.60)$$

である. (2.60) の σ^2 に, $\hat{\sigma} = V_x = S_{xx}/(n-1)$ を代入して得られる確率変数を t と置く:

$$t = \frac{\bar{x}-\mu}{\sqrt{V_x/n}}.$$

t 分布

すると, 以下が成り立つ.

t 分布

$t = \dfrac{\bar{x} - \mu}{\sqrt{V_x / n}}$ は自由度 $\phi = n - 1$ の t 分布 ($t(\phi)$ と表示) に従う. (2.61)

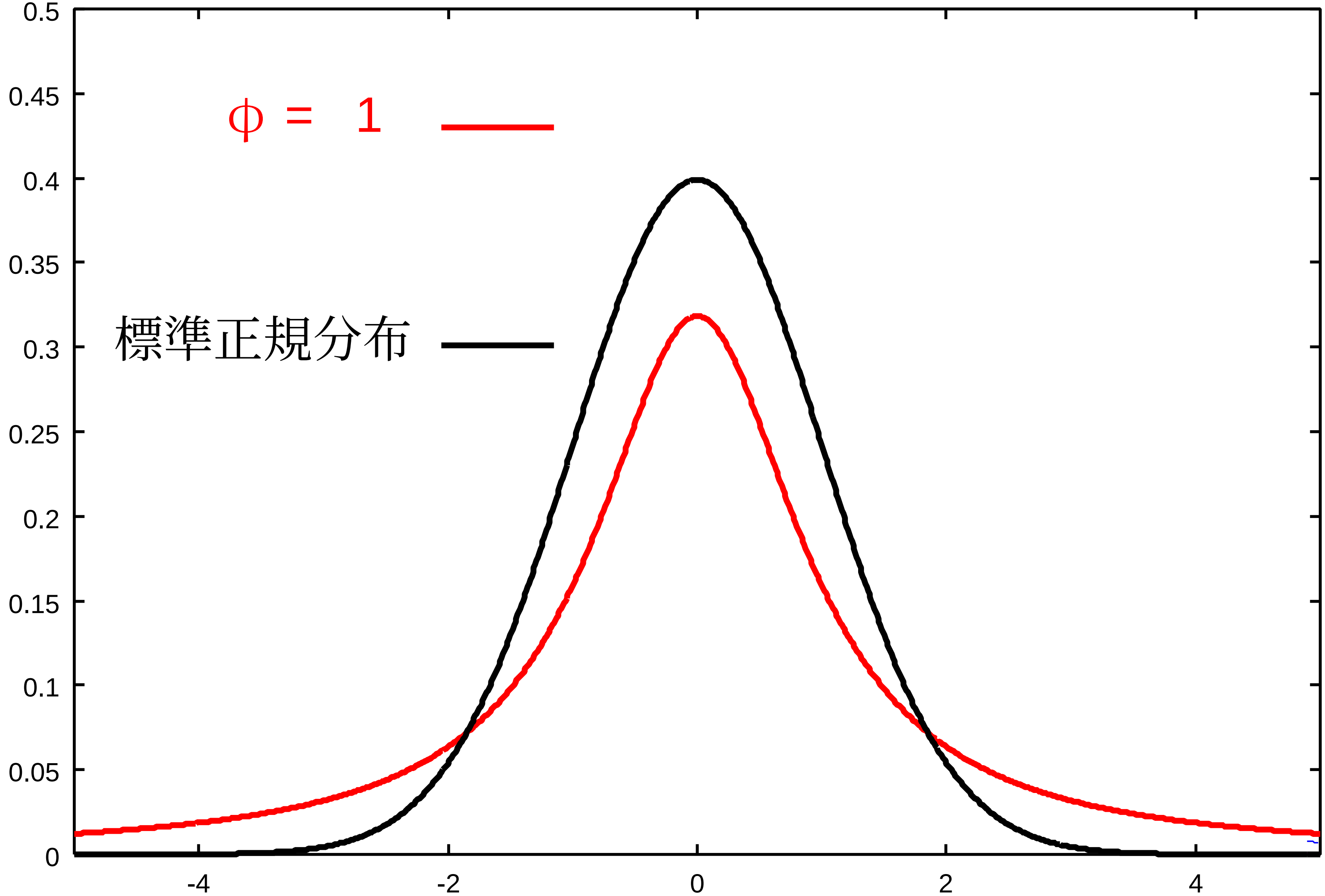
t 分布の確率密度関数

自由度 ϕ の t 分布の密度関数は,

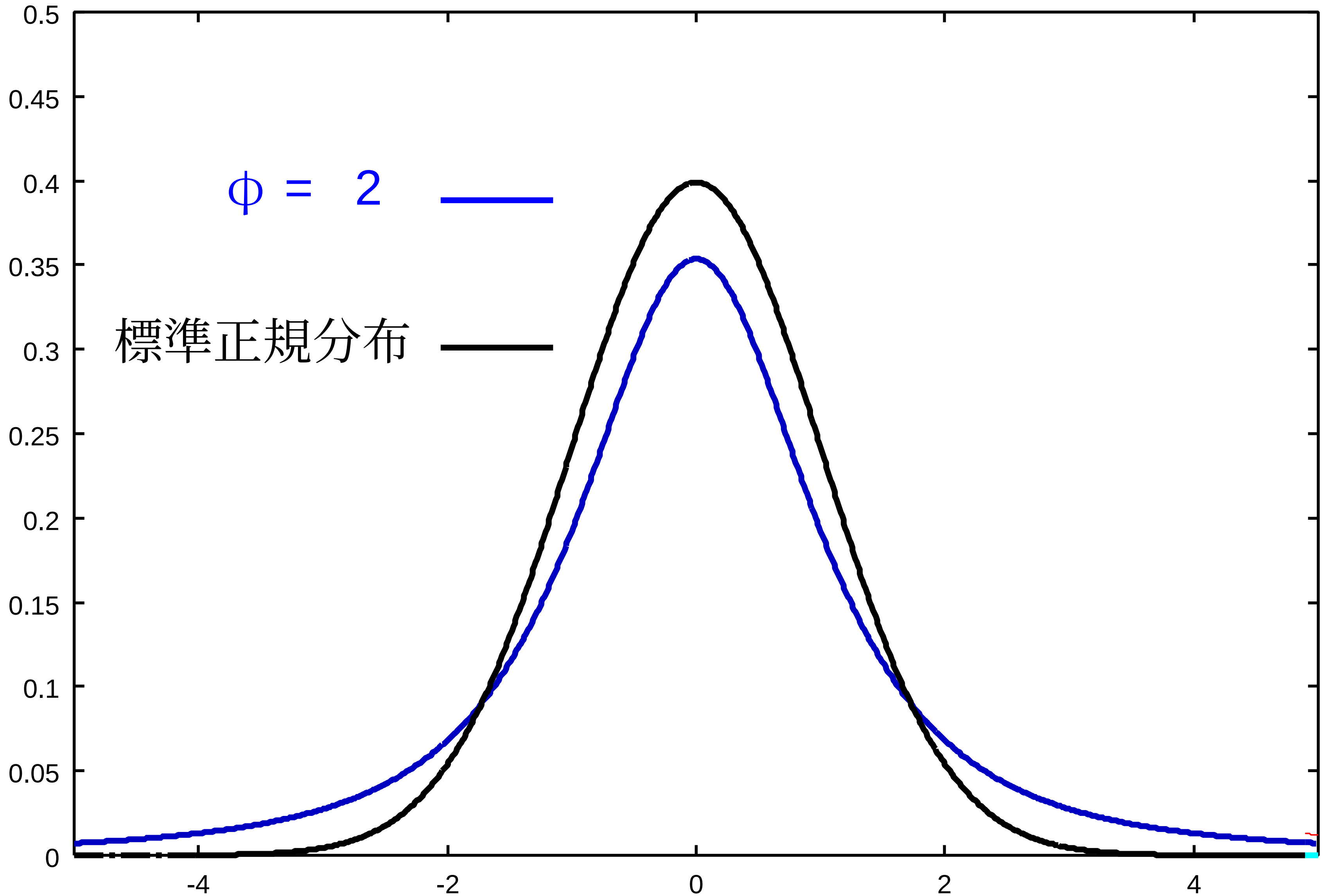
$$\frac{\Gamma\left(\frac{\phi+1}{2}\right)}{\sqrt{\phi\pi}\,\Gamma\left(\frac{\phi}{2}\right)\left(1+\frac{x^2}{\phi}\right)^{\frac{\phi+1}{2}}}$$

で与えられる. (ここに書いてあることは覚えなくても良い. 参考として書いている.)

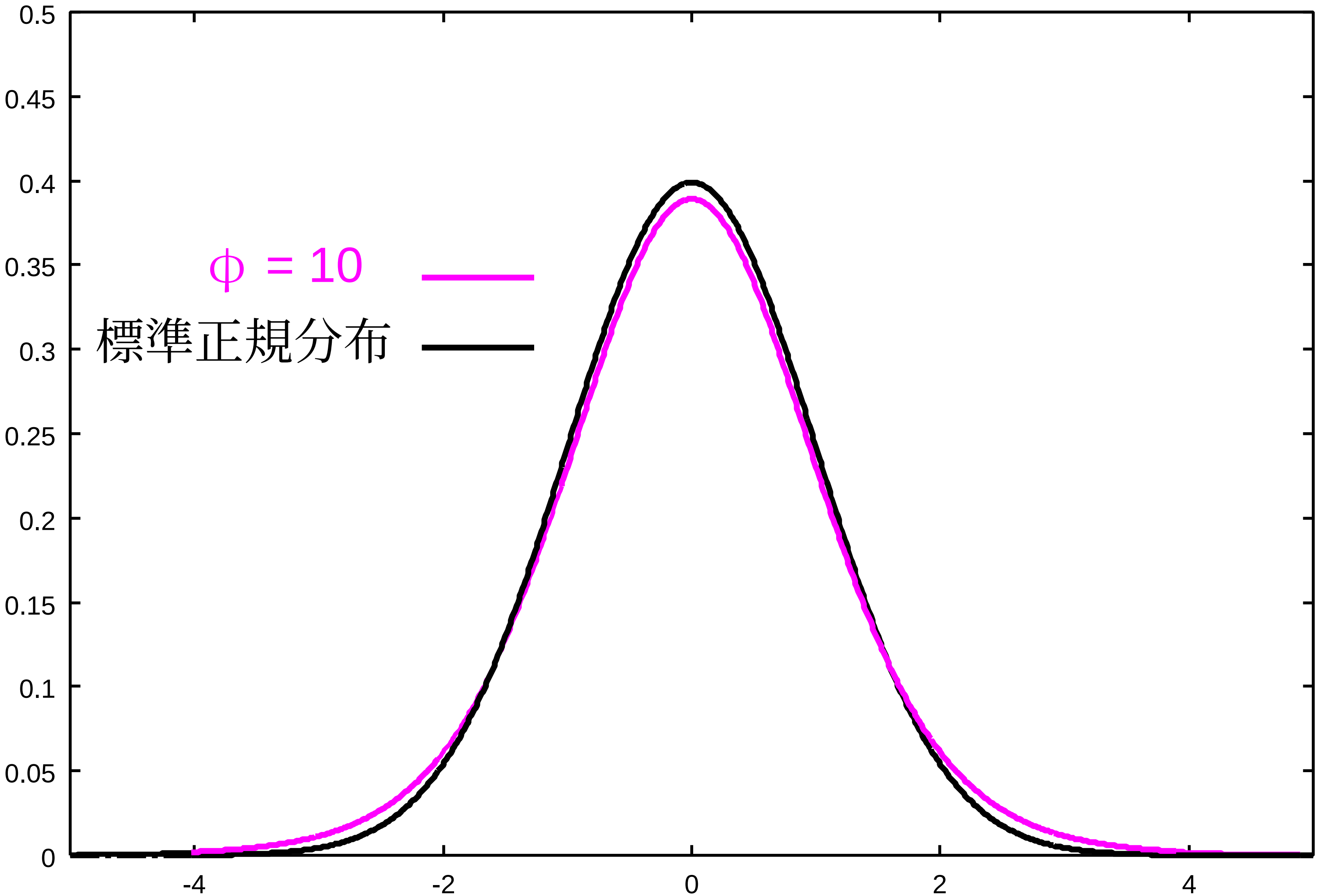
t 分布のグラフ ($\phi = 1$)



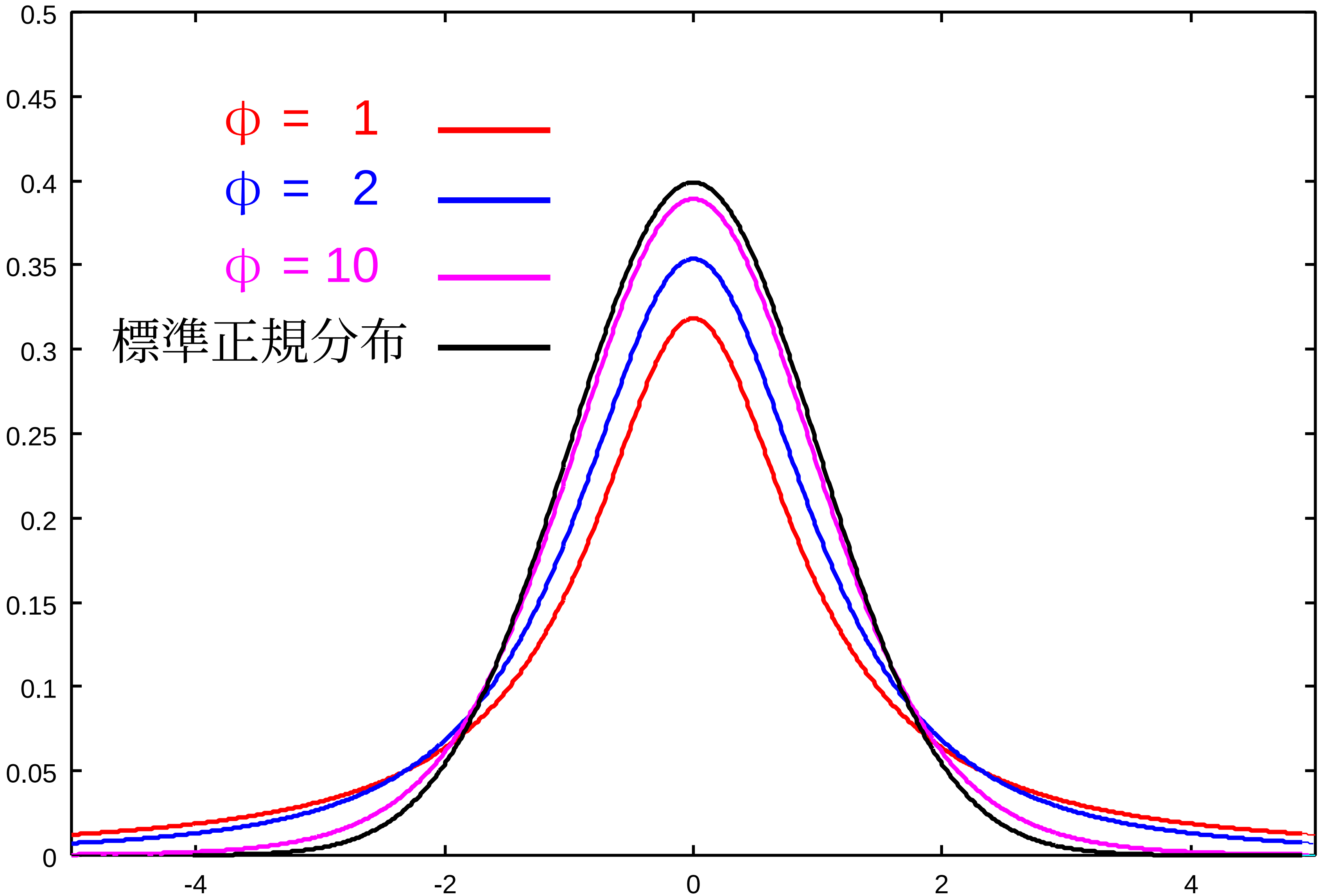
t 分布のグラフ ($\phi = 2$)



t 分布のグラフ ($\phi = 10$)



t 分布のグラフ ($\phi = 1, 2, 10$)



t分布の極限 $(\phi \to \infty)$

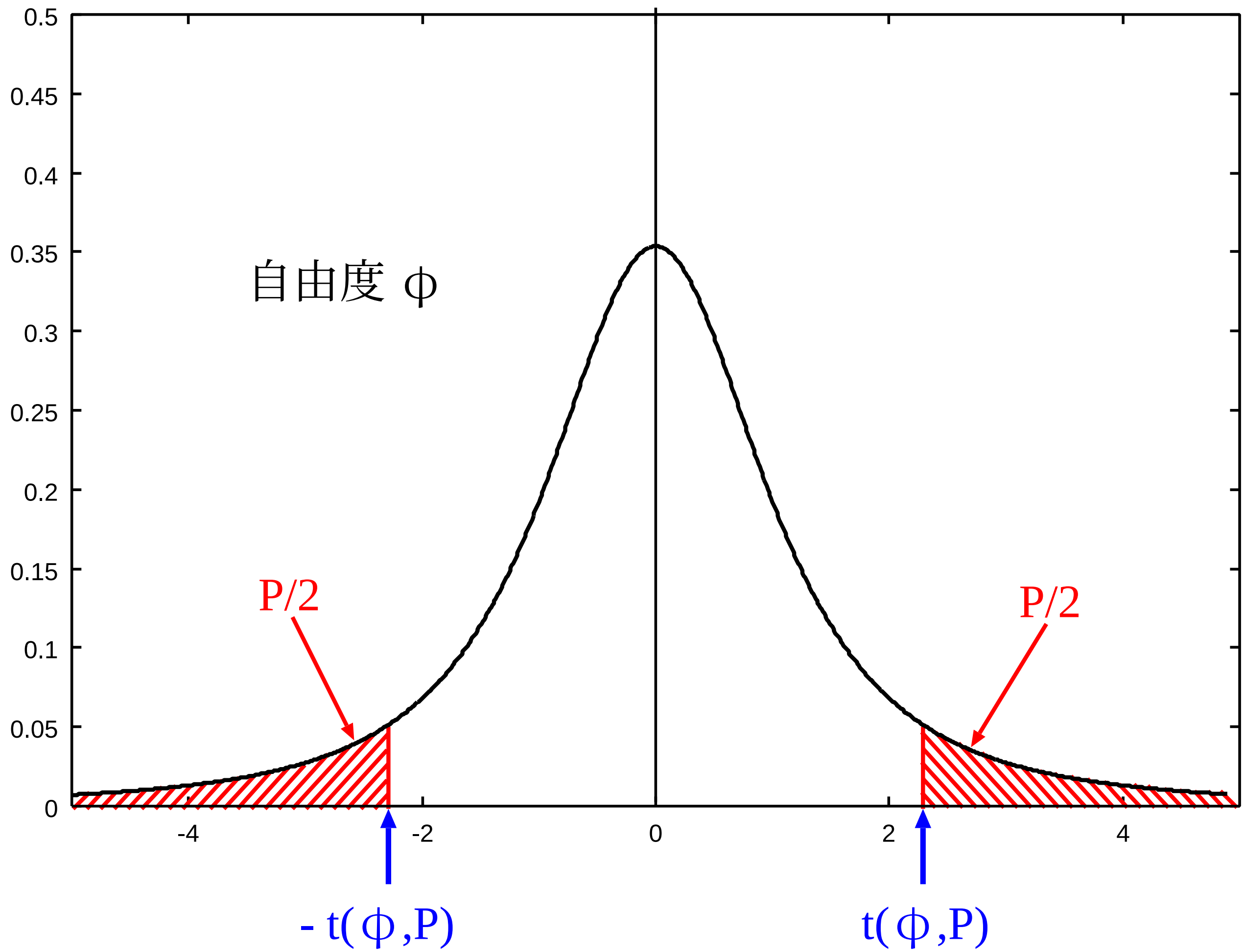
グラフの形から推測できるように,

$$\text{自由度}\,\phi\,\text{の}\,t\,\text{分布} \to \text{正規分布} \quad (\phi \to \infty)$$

であることが知られている.

(ていうか, $\phi = 10$ くらいの t 分布でさえ, 正規分布かなり近い.)

t 表 (p. 237, 付表 4) の読み方



$$P = \Pr(|t| \geq t(\phi, P)) \qquad (2.62)$$

ϕ と P から $t(\phi, P)$ を求める

ϕ \ P	···	.20	.10	.05	.02	.001
1						
2						
3				3.182		
4						
5						

$$\phi = 3, P = 0.05 \Longrightarrow t(3, 0.05) = 3.182$$

$t(\phi, P)$ を自由度 ϕ の t 分布の両側 100P%点と呼ぶ.