

08 関数の凸性と不等式への応用

黒木玄

2018-06-11

- Copyright 2018 Gen Kuroki
- License: MIT <https://opensource.org/licenses/MIT> (<https://opensource.org/licenses/MIT>)
- Repository: <https://github.com/genkuroki/Calculus> (<https://github.com/genkuroki/Calculus>)

このファイルは次の場所できれいに閲覧できる:

- <http://nbviewer.jupyter.org/github/genkuroki/Calculus/blob/master/08%20convexity.ipynb>
(<http://nbviewer.jupyter.org/github/genkuroki/Calculus/blob/master/08%20convexity.ipynb>)
- <https://genkuroki.github.io/documents/Calculus/08%20convexity.pdf>
(<https://genkuroki.github.io/documents/Calculus/08%20convexity.pdf>)

このファイルは [Julia Box](https://juliabox.com) (<https://juliabox.com>) で利用できる.

自分のパソコンに [Julia 言語](https://julialang.org/) (<https://julialang.org/>) をインストールしたい場合には

- [WindowsへのJulia言語のインストール](http://nbviewer.jupyter.org/gist/genkuroki/81de23edcae631a995e19a2ecf946a4f) (<http://nbviewer.jupyter.org/gist/genkuroki/81de23edcae631a995e19a2ecf946a4f>)

を参照せよ.

論理的に完璧な説明をするつもりはない. 細部のいい加減な部分は自分で訂正・修正せよ.

Table of Contents

- 1 [凸関数の定義](#)
- 2 [2階の導関数の符号と関数の凸性](#)
 - 2.1 [2階の導関数の符号と接線の関係](#)
 - 2.2 [2階の導関数の符号と関数の凸性の関係](#)
 - 2.3 [指数・対数関数への応用](#)
- 3 [Jensenの不等式](#)
 - 3.1 [期待値汎関数](#)
 - 3.2 [Jensenの不等式とその証明](#)
 - 3.3 [相加相乗平均の不等式](#)
 - 3.4 [Youngの不等式](#)
 - 3.5 [Hölderの不等式](#)
 - 3.5.1 [\$L^p\$ ノルム](#)
 - 3.5.2 [一般化されたHölderの不等式](#)
 - 3.5.3 [Cauchy-Schwarzの不等式](#)
 - 3.5.4 [分配関数の対数凸性](#)
 - 3.6 [Minkowskiの不等式](#)

```

In [1]: 1 using Base.MathConstants
2 using Base64
3 using Printf
4 using Statistics
5 const e = ℯ
6 endof(a) = lastindex(a)
7 linspace(start, stop, length) = range(start, stop=stop, length=length)
8
9 using Plots
10 gr(); ENV["PLOTS_TEST"] = "true"
11 #clibrary(:colorcet)
12 clibrary(:misc)
13
14 ▼ function pngplot(P...; kwargs...)
15     sleep(0.1)
16     pngfile = tempname() * ".png"
17     savefig(plot(P...; kwargs...), pngfile)
18     showing("image/png", pngfile)
19 end
20 pngplot(; kwargs...) = pngplot(plot!(); kwargs...)
21
22 ▼ showing(mime, fn) = open(fn) do f
23     base64 = base64encode(f)
24     display("text/html", """""")
25 end
26
27 using SymPy
28 #sympy[:init_printing](order="lex") # default
29 #sympy[:init_printing](order="rev-lex")
30
31 using SpecialFunctions
32 using QuadGK

```

1 凸函数の定義

定義: 区間 I 上の実数値関数 f が下に凸な関数であるとは、任意の $a, b \in I$ と実数 t について、

$$0 \leq t \leq 1 \implies f((1-t)a + tb) \leq (1-t)f(a) + tf(b)$$

を満たしていることである。逆向きの不等式で上に凸な関数を定義する。すなわち、区間 I 上の実数値関数 f が上に凸な関数であるとは、任意の $a, b \in I$ と実数 t について、

$$0 \leq t \leq 1 \implies f((1-t)a + tb) \geq (1-t)f(a) + tf(b)$$

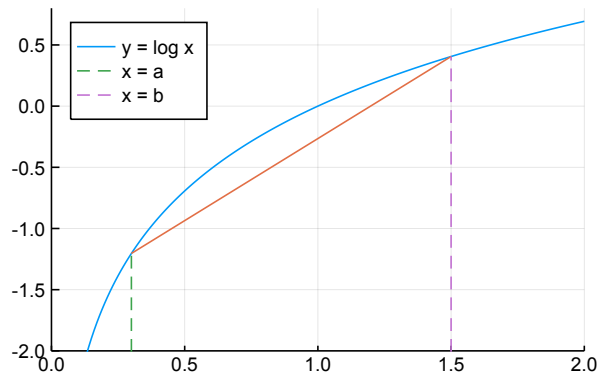
を満たしていることである。

```

In [2]: 1 # log x は上に凸な函数
2
3 x = 0:0.01:2.0
4 a, b = 0.3, 1.5
5 f(x) = log(x)
6 t = 0:0.01:1.0
7 g(a,b,t) = (1-t)*f(a) + t*f(b)
8 X(a,b,t) = (1-t)*a + t*b
9 plot(size=(400,250), legend=:topleft, xlims=(0,2.0), ylims=(-2.0, 0.8))
10 plot!(x, f.(x), label="y = log x")
11 plot!(X.(a,b,t), g.(a,b,t), label="")
12 plot!([a,a], [-10.0, f(a)], label="x = a", ls=:dash)
13 plot!([b,b], [-10.0, f(b)], label="x = b", ls=:dash)

```

Out[2]:

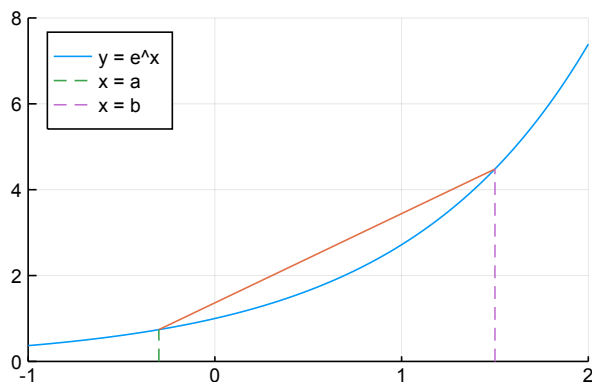


```

In [3]: 1 # e^x は下に凸な函数
2
3 x = -1:0.01:2
4 a, b = -0.3, 1.5
5 f(x) = e^x
6 t = 0:0.01:1.0
7 g(a,b,t) = (1-t)*f(a) + t*f(b)
8 X(a,b,t) = (1-t)*a + t*b
9 plot(size=(400,250), legend=:topleft, xlims=(-1,2), ylims=(0,8))
10 plot!(x, f.(x), label="y = e^x")
11 plot!(X.(a,b,t), g.(a,b,t), label="")
12 plot!([a,a], [-0.0, f(a)], label="x = a", ls=:dash)
13 plot!([b,b], [-0.0, f(b)], label="x = b", ls=:dash)

```

Out[3]:



2 2階の導関数の符号と函数の凸性

2.1 2階の導関数の符号と接線の関係

定理: 区間 I 上の実数値函数 f が C^2 級でかつ I 上で $f'' \geq 0$ ならば, $a \in I$ における $y = f(x)$ のグラフは接線 $y = f'(a)(x - a) + f(a)$ の上側にある. 同様に区間 I 上の実数値函数 f が C^2 級でかつ I 上で $f'' \leq 0$ ならば $a \in I$ における $y = f(x)$ のグラフは接線 $y = f'(a)(x - a) + f(a)$ の下側にある.

証明: 前半のみを証明する. 区間 I 上の実数値函数 f が C^2 級でかつ I 上で $f'' \geq 0$ であると仮定し, $a \in I$ とする. f は C^2 級なので Taylorの公式より, $x \in I$ について,

$$f(x) = f(a) + f'(a)(x-a) + R, \quad R = f''(\xi) \frac{(x-a)^2}{2}.$$

ここで ξ は a と x のあいだのある実数である。 $f'' \geq 0$ より、 $R \geq 0$ となる。ゆえに、 $f(x) \geq f(a) + f'(a)(x-a)$ となる。これで、 $y = f(x)$ のグラフは接線 $y = f'(a)(x-a) + f(a)$ の上側にあることがわかった。□

注意: 上の定理の C^2 級の仮定は2回微分可能性の仮定に弱めることができる。□

注意: 上の定理の証明中の剰余項は

$$R = \int_a^x dx_1 \int_a^{x_1} dx_2 f''(x_2)$$

とも書ける。 $f'' \geq 0$ のとき、 $x > a$ ならば x_1, x_2 は $a < x_1 < x_2 < x$ の範囲を動くので積分の結果は0以上になり、 $x < a$ ならば x_1, x_2 は $x < x_1 < x_2 < a$ の範囲を動き、 x_2 のみに関する積分の結果は0以下になり、さらに x_1 についても積分すると0以上になる。 $\int_b^a f(x) dx = -\int_a^b f(x) dx$ であることに注意せよ。□

2.2 2階の導関数の符号と関数の凸性の関係

定理: 区間 I 上の実数値関数 f が C^2 級でかつ I 上で $f'' \geq 0$ ならば f は下に凸である。同様に区間 I 上の実数値関数 f が C^2 級でかつ I 上で $f'' \leq 0$ ならば f は上に凸である。

証明: 前半のみを証明する。区間 I 上の実数値関数 f が C^2 級でかつ I 上で $f'' \geq 0$ であると仮定する。 $a, b \in I$ と仮定し、 $c = (1-t)a + tb$ とおく。 $0 \leq t \leq 1$ であると仮定する。 $(1-t)f(a) + tf(b) \geq f(c)$ を示せばよい。 f は C^2 級なので Taylor の公式より、

$$\begin{aligned} f(a) &= f(c) + f'(c)(a-c) + R(a), \\ f(b) &= f(c) + f'(c)(b-c) + R(b), \\ R(x) &= f''(\xi_x) \frac{(x-c)^2}{2}. \end{aligned}$$

ここで、 ξ_x は c と x のあいだのある実数である。 $f'' \geq 0$ という仮定より、 $R(x) \geq 0$ となることがわかる。ゆえに

$$\begin{aligned} f(a) &\geq f(c) + f'(c)(a-c), \\ f(b) &\geq f(c) + f'(c)(b-c). \end{aligned}$$

1つ目の不等式の両辺に $(1-t)$ をかけ、2つ目の不等式の両辺に t をかけると

$$\begin{aligned} (1-t)f(a) &\geq (1-t)f(c) + f'(c)((1-t)a - (1-t)c), \\ tf(b) &\geq tf(c) + f'(c)(tb - tc). \end{aligned}$$

これらをたすと、 $c = (1-t)a + tb$ とおいたことより、

$$(1-t)f(a) + tf(b) \geq f(c) + f'(c)((1-t)a + tb - c) = f(c).$$

これで f が下に凸であることが示された。□

注意: 上の証明の本質はTaylorの公式の剰余項が0以上になることである。 f が C^2 級であるという仮定を2回微分可能性の仮定に弱めても、 c と x のあいだのある実数 ξ_x が存在して、

$$f(x) = f(c) + f'(c)(x-c) + R, \quad R = \frac{1}{2}f''(\xi_x)(x-c)^2$$

となるという形でTaylorの定理が成立している。このことを使えば上の定理における C^2 級の仮定を2回微分可能性の仮定に弱めることができる。□

問題: 上と同様にして $f'' \leq 0$ ならば f が上に凸であることの証明を書き下せ。□

例: $f(x) = e^{ax}$ とおくと $f'(x) = ae^{ax}$, $f''(x) = a^2 e^{ax} \geq 0$ を満たしているので、 f は下に凸な関数である。□

例: $x > 0$ に対して $f(x) = \log x$ とおくと $f'(x) = 1/x$, $f''(x) = -1/x^2 < 0$ なので、 f は上に凸な関数である。□

例題: a は実数であるとし、 $x > 0$ に対して $f(x) = x^a$ とおく。 $f(x)$ は凸性について調べよ。

解答例: $x > 0$ であるとする。 $f'(x) = ax^{a-1}$, $f''(x) = a(a-1)x^{a-2}$ であり、 $x^{a-2} > 0$ である。ゆえに、 $a(a-1) \geq 0$ のとき、すなわち $a \leq 0$ または $1 \leq a$ のとき、 f は下に凸になり、 $a(a-1) \leq 0$ のとき、すなわち $0 \leq a \leq 1$ のとき、 f は上に凸になる。□

2.3 指数・対数関数への応用

問題: すべての $x \in \mathbb{R}$ について $1 + x \leq e^x$ となることを示せ.

解答例: $y = e^x$ の $x = 0$ での接線は $y = 1 + x$ であり, $f''(x) = e^x \geq 0$ であるから, 上の定理より, $1 + x \leq ex$. \square

注意: この問題の結果より, $1 - x \leq e^{-x}$ なので, $x < 1$ のとき $e^x \leq (1 - x)^{-1}$ となる. これで

$$1 + x \leq e^x \leq (1 - x)^{-1} \quad (x < 1)$$

となることがわかった. $a > 0$, $-a \leq t < a$ のとき, この不等式の $x = \frac{t}{a}$ を代入して全体を a 乗すると次が得られる:

$$\left(1 + \frac{t}{a}\right)^a \leq e^t \leq \left(1 - \frac{t}{a}\right)^{-a} \quad (-a \leq t < a). \quad \square$$

注意: 上の問題の結果より, $x > -1$ のとき, $\log(1 + x) \leq x$ となることもわかる. \square

問題: $x > -1$ のとき $\frac{x}{1+x} \leq \log(1+x)$ となることを示せ.

解答例: $f(x) = \log(1+x) - \frac{x}{1+x}$ とおく. $f(0) = 0$ である. $\frac{x}{1+x} = 1 - \frac{1}{1+x}$ と使って, f を微分すると,

$$f'(x) = \frac{1}{1+x} - \frac{1}{(1+x)^2} = \frac{x}{(1+x)^2}$$

なので, $x < 0$ において $f' < 0$ で f は単調減少し, $x > 0$ において $f' > 0$ で f は単調増加する. さらに $f(0) = 0$ なので, $f(x) \geq 0$ すなわち, $\frac{x}{1+x} \leq \log(1+x)$ である. \square

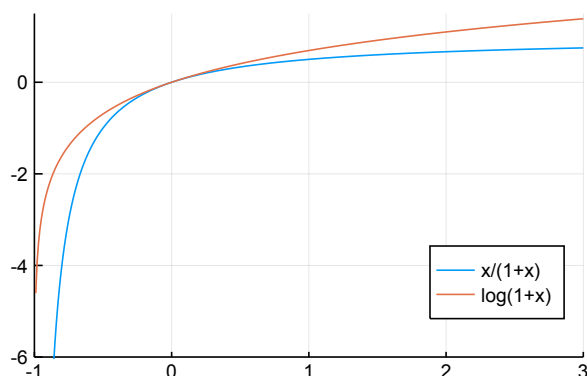
注意: $f''(x) = \frac{1-x}{(x+1)^3}$ なので, $-1 < x \leq 1$ で $f''(x) \geq 0$ となるが, $x > 1$ で $f''(x) < 0$ となってしまうので, 上の問題と同様にこの問題を解くことはできない. \square

以上のまとめ: $x > -1$ のとき

$$\frac{x}{1+x} \leq \log(1+x) \leq x. \quad \square$$

```
In [4]: 1 f(x) = x/(1+x)
2 g(x) = log(1+x)
3 x = -0.99:0.01:3
4 plot(size=(400, 250), legend=:bottomright, xlim=(-1,3), ylims=(-6,1.5))
5 plot!(x, f.(x), label="x/(1+x)")
6 plot!(x, g.(x), label="log(1+x)")
```

Out[4]:



問題: $-\frac{1}{2} \leq x \leq 0$ ならば $\log(1+x) \geq 2x$ となることを示せ.

解答例: $f(x) = \log(1+x) - 2x$ とおく. $f(0) = 0$ である. ゆえに $-\frac{1}{2} \leq x \leq 0$ で単調減少するならば, $-\frac{1}{2} \leq x \leq 0$ で $f(x) \leq 0$ となり, 示したいことが示される. そのことを確認するために $f(x)$ を微分すると,

$$f'(x) = \frac{1}{1+x} - 2 = \frac{-(2x+1)}{1+x}$$

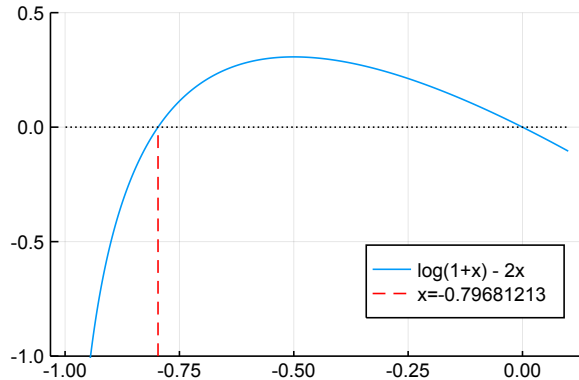
なので $x \geq -\frac{1}{2}$ ならば $f'(x) \leq 0$ となるので, そこで $f(x)$ は単調減少する. \square

注意: $f(x) = \log(1+x) - 2x$ は上に凸な関数なので, $f(a) = 0$ となる $a < 0$ を取れば $f(x) \geq 0$ と $a \leq x \leq 0$ は同値になる. 実際に a を計算すると $a = -0.79681213 \dots$ となる. \square

```
In [5]: 1 f(x) = log(1+x) - 2x
2 @show f(-0.79681213)
3 @show f(-0.79681214)
4
5 a = -0.79681213
6 x = -0.999:0.001:0.1
7 plot(size=(400,250), legend=:bottomright, ylims=(-1,0.5))
8 plot!(x, f.(x), label="log(1+x) - 2x")
9 plot!(x, zero(x), label="", color="black", ls=:dot)
10 plot!([a,a], [-1,0], label="x=$a", color=:red, ls=:dash)
```

```
f(-0.79681213) = 5.848987960632712e-11
f(-0.79681214) = -2.9157047887196086e-8
```

Out[5]:



問題: $a > 0$ であるとし, $-a < t < a$ であるとする. $\left(1 + \frac{t}{a}\right)^a$ は a について単調増加し, $\left(1 - \frac{t}{a}\right)^{-a}$ は a について単調減少することを示せ.

解答例: それぞれの対数を取ってから a で微分する, 上の方の問題で示した $\log(1+x)$ に関する不等式

$$\frac{x}{1+x} \leq \log(1+x) \quad (x > -1)$$

より,

$$\begin{aligned} \frac{\partial}{\partial a} \log \left(1 + \frac{t}{a}\right)^a &= \log \left(1 + \frac{t}{a}\right) - \frac{t/a}{1+t/a} \geq 0, \\ \frac{\partial}{\partial a} \log \left(1 - \frac{t}{a}\right)^{-a} &= -\log \left(1 - \frac{t}{a}\right) + \frac{-t/a}{1-t/a} \leq 0. \end{aligned}$$

ゆえに, $\left(1 + \frac{t}{a}\right)^a$ は a について単調増加し, $\left(1 - \frac{t}{a}\right)^{-a}$ は a について単調減少する. \square

問題: 以上の結果をまとめて用いて,

$$\lim_{a \rightarrow \infty} \left(1 + \frac{t}{a}\right)^a = \lim_{a \rightarrow \infty} \left(1 - \frac{t}{a}\right)^{-a} = e^t$$

を証明せよ.

解答例: 上の方の注意では

$$\left(1 + \frac{t}{a}\right)^a \leq e^t \leq \left(1 - \frac{t}{a}\right)^{-a} \quad (-a < t < a) \quad (*)$$

という不等式が成立していることを指摘している. さらにすぐ上の問題によって, $\left(1 + \frac{t}{a}\right)^a$, $\left(1 - \frac{t}{a}\right)^{-a}$ はそれぞれ a について単調増加, 単調減少するので, どちらも $a \rightarrow \infty$ で収束する. a を十分に大きくして $\frac{t^2}{a^2} \leq \frac{1}{2}$ となるようにすると, 上の方の問題で示した不等式

$$\log(1+x) \geq 2x \quad \left(-\frac{1}{2} \leq x \leq 0\right)$$

を使うことができて、

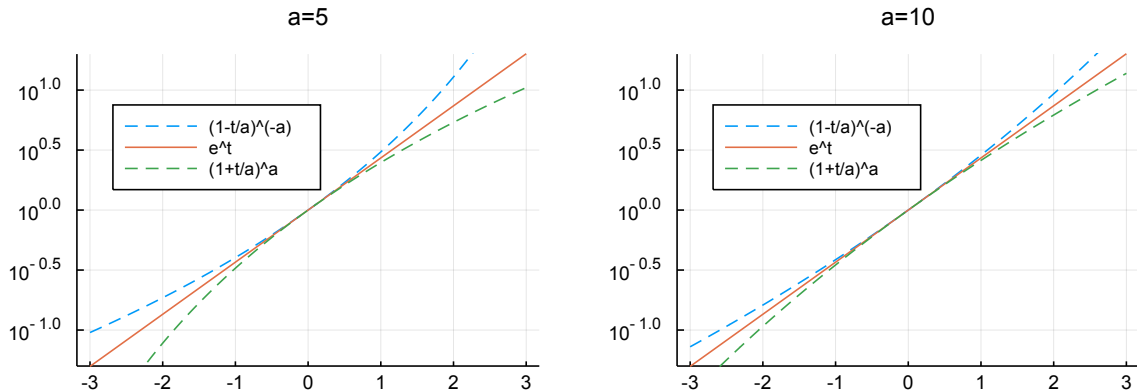
$$0 \geq \log \frac{(1+t/a)^a}{(1-t/a)^{-a}} = a \log \left(1 - \frac{t^2}{a^2} \right) \geq a \cdot 2 \left(-\frac{t^2}{a^2} \right) = -\frac{2t^2}{a} \rightarrow 0 \quad (q \rightarrow \infty).$$

となるので、 $a \rightarrow \infty$ のとき $\log \frac{(1+t/a)^a}{(1-t/a)^{-a}} \rightarrow 0$ すなわち $\frac{(1+t/a)^a}{(1-t/a)^{-a}} \rightarrow 1$ となる。これで $\left(1 + \frac{t}{a}\right)^a, \left(1 - \frac{t}{a}\right)^{-a}$ が同じ値に収束することがわかった。したがって不等式(*)より、欲しい結果が得られる。□

注意: 不等式(*)は極限で指数関数が現われる結果の極限操作を初等的に実行したいときに便利な不等式である。□

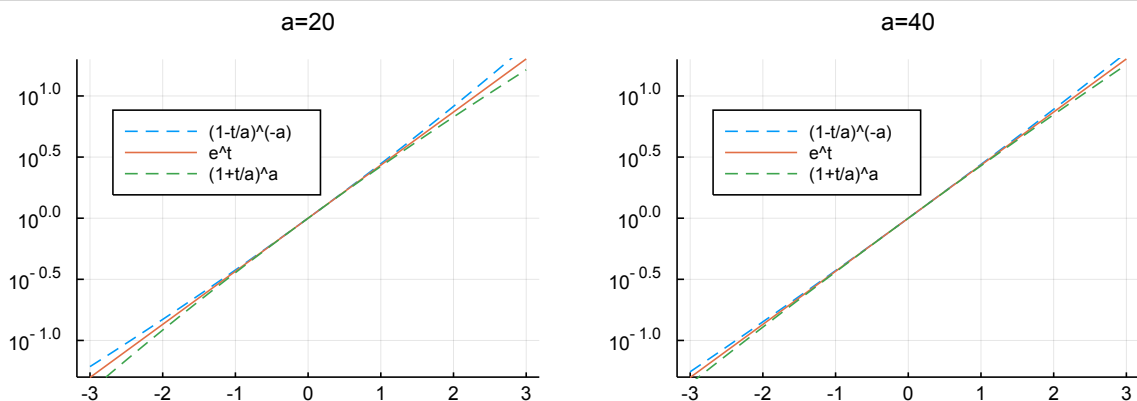
```
In [6]: 1 f(t) = e^t
2 g(a,t) = (1+t/a)^a
3 h(a,t) = (1-t/a)^(-a)
4 t = -3:0.01:3
5
6 PP = []
7 for a in [5, 10, 20, 40]
8     P = plot(title="a=%a", titlefontsize=10, legendfontsize=7)
9     plot!(legend=:topleft, yscale=:log, ylims=(10^(-1.3), 10^1.3))
10    plot!(t, h.(a,t), label="(1-t/a)^(-a)", ls=:dash)
11    plot!(t, f.(t), label="e^t")
12    plot!(t, g.(a,t), label="(1+t/a)^a", ls=:dash)
13    push!(PP, P)
14 end
15 plot(PP[1:2]..., size=(750,250), layout=@layout([a b]))
```

Out[6]:



```
In [7]: 1 plot(PP[3:4]..., size=(750,250), layout=@layout([a b]))
```

Out[7]:



3 Jensenの不等式

3.1 期待値汎函数

函数を数に対応させる函数を**汎函数** (functional)と呼ぶことがある。

函数の例: x を x に対応させる函数を x と書く。

すべての x を一定の値 α に対応させる関数も同じ記号の α で書き、定数関数と呼ぶ。 $f(x) = \alpha$ において記号 f を利用すると煩雑になるので、単に α と書いて代用するということである。数としての α と定数関数としての α を同じ記号で書くので混乱しないように注意して欲しい。

例えば、すべての x を 1 に対応させる関数も単に 1 と書く。□

定義: 関数 f を数 $E[f]$ に対応させる関数 $E[\]$ (汎関数)が以下の条件を満たしているとき、 $E[\]$ は**期待値汎関数** (expectation functional)であると言うことにする:

- (1) **線形性:** 関数 f, g と数 α, β に対して、 $E[\alpha f + \beta g] = \alpha E[f] + \beta E[g]$.
- (2) **単調性:** 関数 f, g のあいだで常に $f \leq g$ が成立しているならば $E[f] \leq E[g]$.
- (3) **規格化条件:** 定数関数 α について、 $E[\alpha] = \alpha$. □

例: 区間 I 上の関数を数 $E[f]$ に対応させる関数 $E[\]$ を以下のように定めると、 $E[\]$ は期待値汎関数になる。まず $a_1, \dots, a_n \in I$ を任意に取る。 w_1, \dots, w_n は 0 以上の実数でかつ $w_1 + \dots + w_n = 1$ を満たしていると仮定する。そして区間 I 上の関数 f に対して

$$E[f] = \sum_{i=1}^n w_i f(a_i) = w_1 f(a_1) + \dots + w_n f(a_n)$$

と定める。このようにして定められた $E[\]$ が実際に期待値汎関数の性質を満たしていることは以下のようにして確認できる。

(1) 区間 I 上の関数 f, g と数 α, β について

$$\begin{aligned} E[\alpha f + \beta g] &= \sum_{i=1}^n w_i (\alpha f(a_i) + \beta g(a_i)) = \sum_{i=1}^n (\alpha w_i f(a_i) + \beta w_i g(a_i)) \\ &= \alpha \sum_{i=1}^n w_i f(a_i) + \beta \sum_{i=1}^n w_i g(a_i) = \alpha E[f] + \beta E[g]. \end{aligned}$$

(2) 区間 I 上で $f \leq g$ が成立していると仮定すると、 $w_i \geq 0$ より $w_i f(a_i) \leq w_i g(a_i)$ も成立するので

$$E[f] = \sum_{i=1}^n w_i f(a_i) \leq \sum_{i=1}^n w_i g(a_i) = E[g].$$

(3) 定数関数 $f(x) = \alpha$ について $f(a_i) = \alpha$ が成立しているので、 $w_1 + \dots + w_n = 1$ という仮定より、

$$E[\alpha] = E[f] = \sum_{i=1}^n w_i f(a_i) = \sum_{i=1}^n w_i \alpha = \alpha.$$

これで示すべきことがすべて示された。

特に $w_i = 1/n$ であるとき、 $E[f]$ は

$$E[f] = \frac{f(a_1) + \dots + f(a_n)}{n}$$

と $f(a_i)$ たちの加法平均になる。□

抽象的な期待値汎関数の概念が難しいと感じる人は、一般的な期待値汎関数の定義を忘れてこの例の $E[f]$ またはさらにその特別な場合である加法平均の場合のみを考えれば十分である。

問題: $a < b$ であるとする。閉区間 $I = [a, b]$ 上の連続関数 f を

$$E[f] = \frac{1}{b-a} \int_a^b f(x) dx$$

に対応させる関数 $E[\]$ は期待値汎関数であることを示せ。□

問題: \mathbb{R} 上の関数 $p(x)$ は $p(x) \geq 0$ ($x \in \mathbb{R}$) と $\int_{-\infty}^{\infty} p(x) dx = 1$ を満たしていると仮定する。(このような $p(x)$ は**確率密度関数**と呼ばれる。) このとき、 \mathbb{R} 上の(適当により条件を仮定した)関数 f を

$$E[f] = \int_{-\infty}^{\infty} f(x) p(x) dx$$

に対応させる関数 $E[\]$ が期待値汎関数であることを示せ。($E[f]$ は確率密度関数 $p(x)$ が定める確率分布に関する**確率変数** f の期待値と呼ばれる。)

注意: 上の問題はこの問題で

$$p(x) = \begin{cases} \frac{1}{b-a} & (a \leq x \leq b) \\ 0 & (\text{それ以外}) \end{cases}$$

とした場合になっている. \square

解答例: 以下では $\int_{-\infty}^{\infty}$ を \int と略記する.

(1) 数 α, β と関数 f, g について

$$E[\alpha f + \beta g] = \int (\alpha f(x) + \beta g(x))p(x) dx = \alpha \int f(x)p(x) dx + \beta \int g(x)p(x) dx = \alpha E[f] + \beta E[g].$$

(2) \mathbb{R} 上で $f \leq g$ が成立していると仮定すると, $p(x) \geq 0$ より $f(x)p(x) \leq g(x)p(x)$ も成立するので

$$E[f] = \int f(x)p(x) dx \leq \int g(x)p(x) dx = E[g].$$

(3) 定数関数 $f(x) = \alpha$ について, $\int p(x) dx = 1$ という仮定より,

$$E[\alpha] = \int \alpha p(x) dx = \alpha.$$

これで示すべきことがすべて示された. \square

3.2 Jensenの不等式とその証明

Jensenの不等式: $E[\cdot]$ は区間 I 上の関数の期待値汎関数であるとし, f は区間 I 上の上に凸(もしくは下に凸)な関数であるとする. このとき

$$E[f(x)] \leq f(E[x]) \quad (\text{もしくは } E[f(x)] \geq f(E[x])).$$

例えば

$$E[f(x)] = \sum_{i=1}^n w_i f(a_i), \quad w_i \geq 0, \quad \sum_{i=1}^n w_i = 1$$

のとき,

$$\begin{aligned} w_1 f(a_1) + \cdots + w_n f(a_n) &\leq f(w_1 a_1 + \cdots + w_n a_n) \\ (\text{もしくは } w_1 f(a_1) + \cdots + w_n f(a_n) &\geq f(w_1 a_1 + \cdots + w_n a_n)). \end{aligned}$$

特にこれの $n = 2$ の場合には f が上に凸(もしくは下に凸)であることの定義とこれらは同値であることに注意せよ. すなわち, 関数 f がJensenの不等式を満たすことと上(もしくは下)に凸なことは同値になる. この意味でJensenの不等式は関数の凸性を言い換えた結果になっている.

証明: 簡単のため f は C^1 級であると仮定する. (C^1 級でない場合にも同様の方法で証明できるが, 「接線」の存在を別に証明する必要がある.)

f は上に凸であると仮定する.

$\mu = E[x]$ とおく. $E[f(x)] \leq f(\mu)$ を示せばよい. $x = \mu$ における $y = f(x)$ の接線を $y = \alpha(x - \mu) + f(\mu)$ と書く. f は上に凸であると仮定したので

$$f(x) \leq \alpha(x - \mu) + f(\mu)$$

が成立している(図を描いてみよ). ゆえに $E[\cdot]$ の性質より

$$E[f(x)] \leq E[\alpha(x - \mu) + f(\mu)] = \alpha(E[x] - E[\mu]) + E[f(\mu)] = f(\mu).$$

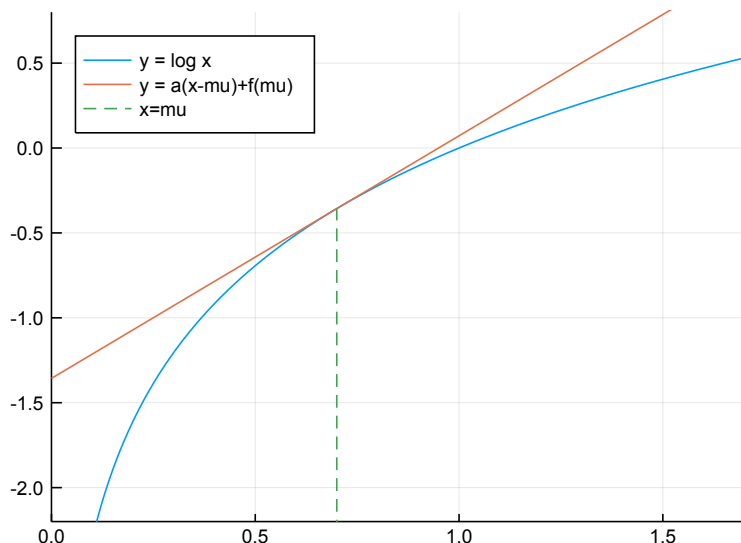
最初の不等号は期待値汎関数の単調性より, 1つ目の等号は線形性より, 2つ目の等号は規格化条件より. これで示すべきことがすべて示された. \square

```

In [8]: 1 x = 0:0.01:2.0
2 mu = 0.7
3 f(x) = log(x)
4 g(mu,x) = (1/mu)*(x-mu) + f(mu)
5 plot(size=(500,350), legend=:topleft, xlims=(0,1.7), ylims=(-2.2, 0.8))
6 plot!(x, f.(x), label="y = log x")
7 plot!(x, g.(mu,x), label="y = a(x-mu)+f(mu)")
8 plot!([mu, mu], [-3.0, f(mu)], label="x=mu", ls=:dash)

```

Out[8]:



問題: 上に凸な函数 $f(x)$ と

$$E[g(x)] = w_1 g(a_1) + \cdots + w_n g(a_n), \quad w_i \geq 0, \quad w_1 + \cdots + w_n = 1$$

に関するJensenの不等式

$$w_1 f(a_1) + \cdots + w_n f(a_n) \leq f(w_1 a_1 + \cdots + w_n a_n) \quad (*)$$

を n に関する数学的帰納法を用いて直接証明せよ.

解答例: $n = 1$ のとき, $(*)$ は $f(a_1) \leq f(a_1)$ を意味するので, 自明に正しい.

$n = N \geq 1$ に関する $(*)$ が常に成立していると仮定し(帰納法の仮定),

$$w_i \geq 0, \quad w_1 + \cdots + w_N + w_{N+1} = 1$$

と仮定する. もしも $w_{N+1} = 1$ ならば $n = N + 1$ の場合の $(*)$ は自明に成立している. 以下では $w_{N+1} < 1$ と仮定し,

$$t_i = \frac{w_i}{1 - w_{N+1}}$$

とおく. このとき, $w_i \geq 0, w_1 + \cdots + w_N = 1 - w_{N+1}$ より,

$$t_i \geq 0, \quad t_1 + \cdots + t_N = \frac{w_1 + \cdots + w_N}{1 - w_{N+1}} = 1.$$

したがって, $n = N$ の場合の $(*)$ より,

$$t_1 f(a_1) + \cdots + t_N f(a_N) \leq f(t_1 a_1 + \cdots + t_N a_N).$$

f は上に凸と仮定してあったので, $w_i = (1 - w_{N+1})t_i$ を使うと,

$$\begin{aligned} w_1 f(a_1) + \cdots + w_N f(a_N) + w_{N+1} f(a_{N+1}) &= (1 - w_{N+1})(t_1 f(a_1) + \cdots + t_N f(a_N)) + w_{N+1} f(a_{N+1}) \\ &\leq (1 - w_{N+1})f(t_1 a_1 + \cdots + t_N a_N) + w_{N+1} f(a_{N+1}) \\ &\leq f((1 - w_{N+1})(t_1 a_1 + \cdots + t_N a_N) + w_{N+1} a_{N+1}) \\ &= f(w_1 a_1 + \cdots + w_N a_N + w_{N+1} a_{N+1}). \end{aligned}$$

ここで, 2行目で $n = N$ の場合の $(*)$ を使い, 3行目で f が上に凸であることを使った.

ゆえに数学的帰納法によってすべての n について $(*)$ が成立する. \square

3.3 相加相乗平均の不等式

例: $a_i > 0$ であるとし,

$$E[f(x)] = \frac{f(a_1) + \cdots + f(a_n)}{n}$$

とおく. $x > 0$ に対して $f(x) = \log x$ とおく. このとき, $f(x) = \log x$ は下に凸なので, Jensenの不等式より $E[f(x)] \leq f(E[x])$. そして,

$$E[f(x)] = \frac{\log a_1 + \cdots + \log a_n}{n} = \log(a_1 \cdots a_n)^{1/n}, \quad f(E[x]) = \log \frac{a_1 + \cdots + a_n}{n}$$

なので, \log の単調増加性より,

$$(a_1 \cdots a_n)^{1/n} \leq \frac{a_1 + \cdots + a_n}{n}.$$

これでJensenの不等式が相加相乗平均の不等式を含んでいることがわかった. \square

例題: 下に凸な函数 $f(x) = e^x$ に関するJensenの不等式を用いて相加相乗平均の不等式を証明し直せ.

解答例: $a_i > 0$ であるとし, $A_i = \log a_i$, $E[f(x)] = \frac{f(A_1) + \cdots + f(A_n)}{n}$ とおくと, $f(x) = e^x$ のとき

$$E[f(x)] = E[e^x] = \frac{e^{A_1} + \cdots + e^{A_n}}{n} = \frac{a_1 + \cdots + a_n}{n},$$

$$f(E[x]) = e^{(A_1 + \cdots + A_n)/n} = (e^{A_1} \cdots e^{A_n})^{1/n} = (a_1 \cdots a_n)^{1/n}.$$

下に凸な函数 $f(x) = e^x$ に関するJensenの不等式 $E[f(x)] \geq f(E[x])$ より,

$$\frac{a_1 + \cdots + a_n}{n} \geq (a_1 \cdots a_n)^{1/n}.$$

これで示すべきことが示された. \square

3.4 Youngの不等式

Youngの不等式の一般化: $p_i > 0$, $\frac{1}{p_1} + \cdots + \frac{1}{p_n} = 1$, $a_i > 0$ のとき

$$a_1 \cdots a_n \leq \frac{a_1^{p_1}}{p_1} + \cdots + \frac{a_n^{p_n}}{p_n}$$

$n = 2$ の場合をYoungの不等式と呼ぶ. すなわち $p > 0$, $q > 0$, $\frac{1}{p} + \frac{1}{q} = 1$, $a > 0$, $b > 0$ のときの

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

という不等式を**Youngの不等式**と呼ぶ.

証明: 上に凸な函数 $f(x) = \log x$ に関するJensenの不等式より

$$\log(a_1 \cdots a_n) = \frac{\log a_1^{p_1}}{p_1} + \cdots + \frac{\log a_n^{p_n}}{p_n} \leq \log \left(\frac{a_1^{p_1}}{p_1} + \cdots + \frac{a_n^{p_n}}{p_n} \right)$$

ゆえに

$$a_1 \cdots a_n \leq \frac{a_1^{p_1}}{p_1} + \cdots + \frac{a_n^{p_n}}{p_n} \quad \square$$

問題: $a, b > 0$, $p, q > 0$, $\frac{1}{p} + \frac{1}{q} = 1$ のとき

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

となること(つまりYoungの不等式)を直接的に証明してみよ.

解答例1: $f(x) = \log x$ が上に凸であることより, $A, B > 0$ について

$$\frac{1}{p} \log A + \frac{1}{q} \log B \leq \log \left(\frac{A}{p} + \frac{B}{q} \right).$$

ゆえに $A = a^p, B = b^q$ とおくと,

$$\log(ab) = \frac{1}{p} \log a^p + \frac{1}{q} \log b^q \leq \log \left(\frac{a^p}{p} + \frac{b^q}{q} \right).$$

したがって,

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}. \quad \square$$

解答例2: $f(x) = e^x$ が下に凸であることより, 実数 A, B について

$$\frac{1}{p} e^A + \frac{1}{q} e^B \geq e^{A/p+B/q} = e^{A/p} e^{B/q}.$$

ゆえに $A = \log a^p, B = \log b^q$ とおくと,

$$\frac{a^p}{p} + \frac{b^q}{q} \geq ab. \quad \square$$

解答例3: $f(x)$ は $x \geq 0$ の狭義単調増加連続関数で $f(0) = 0$ を満たすものであるとし,

$$X = \{ (x, y) \mid 0 \leq x \leq a, 0 \leq y \leq f(x) \}, \quad Y = \{ (x, y) \mid 0 \leq x \leq f^{-1}(y), 0 \leq y \leq b \}$$

とおくと, $[0, a] \times [0, b] \subset X \cup Y$ となるので(図を描いてみよ), 面積を比較することによって

$$ab \leq \int_0^a f(x) dx + \int_0^b f^{-1}(y) dy. \quad (\star)$$

$f(x) = x^{p-1}$ とおくと,

$$\frac{1}{p} + \frac{1}{q} = 1 \iff (p-1)(q-1) = 1$$

より, $f^{-1}(y) = y^{1/(p-1)} = y^{q-1}$ となることがわかる. これに上の不等式を適用すると

$$ab \leq \frac{a^p}{p} + \frac{b^p}{q}$$

が得られる. \square

注意: $a^{p-1} = b$ ならば (\star) で等号が成立する.

$$\frac{1}{p} + \frac{1}{q} = 1 \iff (p-1)(q-1) = 1 \iff (p-1)q = p$$

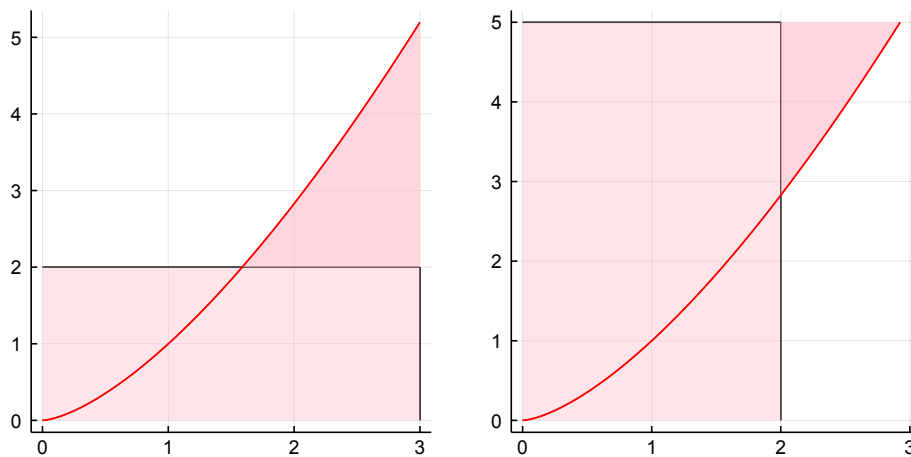
なので $a^{p-1} = b$ と $a^p = b^q$ は同値である(前者の両辺を q 乗すると後者になる). \square

```

In [9]: 1 function plot_Young(a, b; p=2.5, kwarg...)
2         q = p/(p-1)
3         f(x) = x^(p-1)
4         g(y) = y^(q-1)
5         A = max(a, g(b))
6         x = 0:a/200:a
7         y = 0:b/200:b
8         x1 = 0:A/200:A
9         P = plot(legend=false; kwarg...)
10        plot!([a,a], [0,b], color=:black)
11        plot!([0,a], [b,b], color=:black)
12        plot!(x, f.(x), color=:red, fill=(0, 0.4, :pink))
13        plot!(x1, f.(x1), color=:red, fill=(b, 0.4, :pink))
14        if A > a
15            x2 = a:(A-a)/200:A
16            plot!(x2, f.(x2), color=:red, fill=(b, 0.4, :pink))
17        end
18        P
19    end
20
21    P1 = plot_Young(3,2)
22    P2 = plot_Young(2,5)
23    plot(P1, P2, size=(600, 300))

```

Out[9]:



3.5 Hölderの不等式

3.5.1 L^p ノルム

区間 $[a, b]$ 上の連続関数 f に対して $\|f\|_p$ を

$$\|f\|_p = \left(\int_a^b |f(x)|^p dx \right)^{1/p}$$

と定める。これを関数 f の L^p ノルムと呼ぶ。

問題: 定数 α について $\|\alpha f\|_p = |\alpha| \|f\|_p$ が成立していることを示せ。

解答例:

$$\|\alpha f\|_p = \left(\int_a^b |\alpha f(x)|^p dx \right)^{1/p} = \left(|\alpha|^p \int_a^b |f(x)|^p dx \right)^{1/p} = |\alpha| \left(\int_a^b |f(x)|^p dx \right)^{1/p} = |\alpha| \|f\|_p. \quad \square$$

注意: 上の問題の結果より、特に $g = \frac{f}{\|f\|_p}$ とおくと $\|g\|_p = 1$ となることがわかる。 \square

3.5.2 一般化されたHölderの不等式

$p > 0, p_i > 0$ であるとし、

$$\frac{1}{p_1} + \cdots + \frac{1}{p_n} = \frac{1}{p}$$

と仮定する. このとき, 区間 $[a, b]$ 上の連続関数 f_i 達について

$$\|f_1 \cdots f_n\|_p \leq \|f_1\|_{p_1} \cdots \|f_n\|_{p_n} \quad (*)$$

が成立する. これを**一般化されたHölderの不等式**と呼ぶことにする. $n = 2, p = 1$ の場合はHölderの不等式と呼ばれる. すなわち, $p > 0, q > 0$ かつ

$$\frac{1}{p} + \frac{1}{q} = 1$$

のときの

$$\|fg\|_1 \leq \|f\|_p \|g\|_q$$

という不等式は**Hölderの不等式**と呼ばれる. さらにこれの $p = q = 2$ の場合は**Cauchy-Schwarzの不等式**の特別な場合である.

一般化されたHölderの不等式の証明の概略: $g_i(x) = \frac{f_i(x)}{\|f_i\|_{p_i}}$ とおく. そのとき $\|g_i\|_{p_i} = 1$ となる.

$\frac{1}{p_1} + \cdots + \frac{1}{p_n} = \frac{1}{p}$ は $\frac{p}{p_1} + \cdots + \frac{p}{p_n} = 1$ と同値なので, $q_i = p_i/p$ とおくと,

$$q_i > 0, \quad pq_i = p_i, \quad \frac{1}{q_1} + \cdots + \frac{1}{q_n} = 1.$$

ゆえに $a_i = |g_i(x)|^p$ にYoungの不等式の一般化を適用することによって,

$$|g_1(x) \cdots g_n(x)|^p = |g_1(x)|^p \cdots |g_n(x)|^p \leq \frac{|g_1(x)|^{pq_1}}{q_1} + \cdots + \frac{|g_n(x)|^{pq_n}}{q_n} = \frac{|g_1(x)|^{p_1}}{q_1} + \cdots + \frac{|g_n(x)|^{p_n}}{q_n}.$$

これを x で積分すると,

$$(\|g_1 \cdots g_n\|_p)^p \leq \frac{(\|g_1(x)\|_{p_1})^{p_1}}{q_1} + \cdots + \frac{(\|g_n(x)\|_{p_n})^{p_n}}{q_n} = \frac{1}{q_1} + \cdots + \frac{1}{q_n} = 1.$$

ゆえに,

$$\frac{\|f_1 \cdots f_n\|_p}{\|f_1\|_{p_1} \cdots \|f_n\|_{p_n}} = \|g_1 \cdots g_n\|_p \leq 1.$$

したがって,

$$\|f_1 \cdots f_n\|_p \leq \|f_1\|_{p_1} \cdots \|f_n\|_{p_n}. \quad \square$$

問題: (一般化されていない)Youngの不等式から(一般化されていない)Hölderの不等式を導け.

解答例の概略: $p, q > 0, \frac{1}{p} + \frac{1}{q} = 1$ と仮定する.

$$\|fg\|_1 \leq \|f\|_p \|g\|_q$$

を証明したい. 左辺を右辺で割ったものが 1 以下になることを示せばよい.

$$a(x) = \frac{|f(x)|}{\|f\|_p}, \quad b(x) = \frac{|g(x)|}{\|g\|_q}$$

とおくと, $\|a\|_p = \|b\|_q = 1$ となる. $a(x), b(x)$ にYoungの不等式を適用すると,

$$a(x)b(x) \leq \frac{a(x)^p}{p} + \frac{b(x)^q}{q}.$$

両辺を x について積分すると,

$$\int_a^b a(x)b(x) dx = \|ab\|_1 = \frac{\|fg\|_1}{\|f\|_p \|g\|_q},$$

$$\int_a^b \left(\frac{a(x)^p}{p} + \frac{b(x)^q}{q} \right) dx = \frac{(\|a\|_p)^p}{p} + \frac{(\|b\|_q)^q}{q} = \frac{1}{p} + \frac{1}{q} = 1.$$

ゆえに

$$\frac{\|fg\|_1}{\|f\|_p \|g\|_q} \leq 1.$$

これより, Hölderの不等式が成立することがわかる. \square

3.5.3 Cauchy-Schwarzの不等式

$p, q > 0, \frac{1}{p} + \frac{1}{q} = 1$ と仮定する. このとき, $\|f\|_p < \infty, \|g\|_q < \infty$ を満たす函数 f, g の内積を

$$\langle f, g \rangle = \int_a^b \overline{f(x)} g(x) dx$$

と定義できる. ここで $\overline{f(x)}$ は $f(x)$ の複素共役を意味している. $f(x)$ が実数ならば $\overline{f(x)} = f(x)$ である. このとき, Hölderの不等式より,

$$|\langle f, g \rangle| \leq \int_a^b |\overline{f(x)}| |g(x)| dx = \int_a^b |f(x)| |g(x)| dx = \|fg\|_1 \leq \|f\|_p \|g\|_q.$$

すなわち,

$$|\langle f, g \rangle| \leq \|fg\|_1 \leq \|f\|_p \|g\|_q.$$

この不等式の $p = q = 1/2$ の場合は**Cauchy-Schwarzの不等式**と呼ばれている.

3.5.4 分配函数の対数凸性

問題: $a < b$ と仮定する. $Z(\beta) = \int_a^b e^{-\beta f(x)+g(x)} dx$ の形の函数 $Z(\beta)$ について, $\log Z(\beta)$ が β の下に凸な函数になることをHölderの不等式を使って示せ.

解答例: $0 < t < 1$ であるとし, $p = 1/(1-t), q = 1/t$ とおく. そのとき $\frac{1}{p} + \frac{1}{q} = 1$ となる. ゆえにHölderの不等式より,

$$\begin{aligned} Z((1-t)\alpha + t\beta) &= Z\left(\frac{\alpha}{p} + \frac{\beta}{q}\right) = \int_a^b e^{-(\alpha/p + \beta/q)f(x) + (1/p + 1/q)g(x)} dx = \int_a^b (e^{-\alpha f(x)+g(x)})^{1/p} (e^{-\beta f(x)+g(x)})^{1/q} dx \\ &\leq \left(\int_a^b e^{-\alpha f(x)+g(x)} dx \right)^{1/p} \left(\int_a^b e^{-\beta f(x)+g(x)} dx \right)^{1/q} \\ &= Z(\alpha)^{1/p} Z(\beta)^{1/q} = Z(\alpha)^{1-t} Z(\beta)^t. \end{aligned}$$

ゆえに

$$\log Z((1-t)\alpha + t\beta) \leq (1-t) \log Z(\alpha) + t \log Z(\beta).$$

これは $\log Z(\beta)$ が β に関する下に凸な函数であることを意味している. \square

注意: 上の問題の $Z(\beta)$ の形の函数を**分配函数**と呼ぶことがある(統計力学). 上の問題より, Hölderの不等式は分配函数の対数が下に凸な函数になることを意味していることがわかる. \square

例(ガンマ函数の対数凸性): $s > 0$ に対して $\Gamma(s) = \int_0^\infty e^{-x} x^{s-1} dx = \int_0^\infty e^{(s-1) \log x - x} dx$ とおくと, その対数 $\log \Gamma(s)$ は $s > 0$ の下に凸な函数になる. \square

注意: 上の問題の記号のもとで, 函数 $h(x)$ に対する $\langle h(x) \rangle = \langle h(y) \rangle$ を次のように定める:

$$\langle h(x) \rangle = \frac{1}{Z(\beta)} \int_a^b h(x) e^{-\beta f(x)+g(x)} dx, \quad \langle h(y) \rangle = \frac{1}{Z(\beta)} \int_a^b h(y) e^{-\beta f(y)+g(y)} dy.$$

さらに函数 $h(x, y)$ に対する $\langle h(x, y) \rangle$ を次のように定める:

$$\langle h(x, y) \rangle = \frac{1}{Z(\beta)^2} \int_a^b \int_a^b h(x, y) e^{-\beta f(x)+g(x)} e^{-\beta f(y)+g(y)} dx dy.$$

このとき $\langle f(x) \rangle^2 = \langle f(x)f(y) \rangle$ などが成立している。

もしも上の問題の $Z(\beta)$ について β に関する微分と積分の順序交換を自由にできるならば

$$\begin{aligned} -\frac{d}{d\beta} \log Z(\beta) &= \frac{\int_a^b e^{-\beta f(x)+g(x)} f(x) dx}{Z(\beta)} = \langle f(x) \rangle, \\ \frac{d^2}{d\beta^2} \log Z(\beta) &= \langle f(x)^2 \rangle - \langle f(x) \rangle^2 = \langle f(x)f(x) \rangle - \langle f(x)f(y) \rangle \\ &= \frac{1}{2} (\langle f(x)f(x) \rangle - 2\langle f(x)f(y) \rangle + \langle f(y)f(y) \rangle) = \frac{1}{2} \langle (f(x) + f(y))^2 \rangle \geq 0. \end{aligned}$$

これからも, $\log Z(\beta)$ が下に凸な関数であることがわかる。□

3.6 Minkowskiの不等式

L^1 ノルムについては, 三角不等式 $|f(x) + g(x)| \leq |f(x)| + |g(x)|$ より,

$$\|f + g\|_1 = \int_a^b |f(x) + g(x)| dx \leq \int_a^b (|f(x)| + |g(x)|) dx = \|f\|_1 + \|g\|_1$$

が成立している。この L^1 ノルムに関する三角不等式はそのまま L^p ノルム ($p > 1$) に一般化される。

$p > 1$ であると仮定する。 $q = \frac{p}{p-1}$ とおくと, $\frac{1}{p} + \frac{1}{q} = 1$ 。このとき,

$$\|\varphi\|_q^{p-1} = \|\varphi\|_{p/(p-1)}^{p-1} = \left(\int_a^b |\varphi(x)|^p dx \right)^{(p-1)/p} = (\|\varphi\|_p)^{p-1}.$$

これを $\varphi = f + g$ に適用することによって,

$$\begin{aligned} (\|f + g\|_p)^p &= \int_a^b |f(x) + g(x)|^p dx = \int_a^b |f(x) + g(x)| |f(x) + g(x)|^{p-1} dx \\ &\leq \int_a^b |f(x)| |f(x) + g(x)|^{p-1} dx + \int_a^b |g(x)| |f(x) + g(x)|^{p-1} dx \\ &\leq \|f\|_p \| |f(x) + g(x)|^{p-1} \|_q + \|g\|_p \| |f(x) + g(x)|^{p-1} \|_q \\ &= \|f\|_p (\|f(x) + g(x)\|_p)^{p-1} + \|g\|_p (\|f(x) + g(x)\|_p)^{p-1} \\ &= (\|f\|_p + \|g\|_p) (\|f(x) + g(x)\|_p)^{p-1} \end{aligned}$$

ここで, 2行目の不等号で三角不等式 $|f(x) + g(x)| \leq |f(x)| + |g(x)|$ を使い, 3行目の不等号でHölderの不等式を使った。したがって,

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

これを**Minkowskiの不等式**と呼ぶ。

定義: $\|f\|_p < \infty$ すなわち $\int_a^b |f(x)|^p dx < \infty$ を満たす関数を L^p 関数 もしくは p 乗可積分関数 と呼ぶ。特に L^1 関数を単に可積分関数と呼ぶ。□

Minkowskiの不等式より, L^p 関数の和も L^p 関数になる。