9. 連続的確率変数と正規分布

1. 連続的確率変数

(1) 確率変数 X が，ある区間内のすべての実数値をとり得るとき，「X は連続的である」または「X は連続的確率変数である」という。これは，試行の標本点から定まる X の実現値がすべての実数値をとるという意味ではなく，X はすべての実数値をとり得る変数という意味である。

(2) X が試験の点数の場合，試行を何度行っても，X のとり得る値は 0～100 の整数値であるが，X が身長や体重の場合は，ある範囲内のすべての実数値をとり得る。

(3) 連続的確率変数の例は，以下のように多数ある。

① 自動車を時速 50km で走らせ，地点（P）で急ブレーキをかけてから停止するという試行を行い，地点（P）から停止するまので距離を X m とする。およそ 17m で停止することは分かっているが，実際には，X はいろいろな値をとる。X は，17.12345678m のような値もとり得る。従って，X は連続的である。

② ビール工場でつぎつぎに生産される，内容量が 500mL のビン・ビール。内容量をチェックするために，無作為に 1 本選ぶという試行を行う。選んだビールのフタを開けて，内容量を正確に測定し，測定結果を X mL とする。このとき，X の値がちょうど 500mL に一致することはほとんどなく，500mL の前後のいろいろな値をとる。従って，X は連続的である。この試行は，何度も繰り返すことはできない。製品がムダになるからである。しかし，頭の中で繰り返すことはできる。これを繰り繰り返せば，理論上，X は 500mL 前後の様々な実数値をとり得るのである。

③ 多数の人からなる母集団がある。母集団から 1 人を無作為に抽出するという試行を行い，抽出した人の身長（特性値）を X する。X は確率変数になるが，とり得る値という意味で，X は連続的である。

(4) X が身長の場合，1 つの実数 x に対して，確率 $P(X=x)$ を定めることができない。例えば，$170 \leq x \leq 171$ を満たす実数 x は無限に多くあるので，

$$P(X=170.001),\quad P(X=170.002),\quad P(X=170.003),\ \cdots$$

のような確率を定めていっても，すべての確率 $P(X=x)$ を定めたことにはならない。

(5) そこで，連続的確率変数 X では，$P(X=x)$ の値は常に 0 と考え，X の値が区間 $a \leq x \leq b$ に入る確率 $P(a \leq X \leq b)$ を定めていく。この確率は，離散的な場合と同様に，a 以上 b 以下の値をとる標本点の個数の割合と考えればよい。

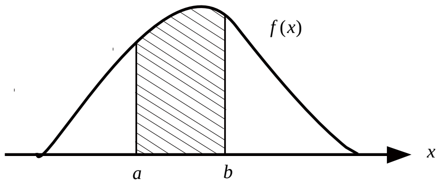
(6) 一般に，連続的確率変数 X に対して，次の①～③を満たす関数 $f(x)$ を，X の「確率密度関数」または単に「密度関数」という。また，関数 $y = f(x)$ のグラフを「X の分布曲線」

という。

① 常に $f(x) \geq 0$

② 任意の a, b（$a \leq b$）に対して，$P(a \leq X \leq b) = \int_a^b f(x)dx$

③ $\int_{-\infty}^{+\infty} f(x)dx = 1$



● 注意

② $\int_a^b f(x)dx$ は $f(x)$ の a から b までの定積分であり，

その値は，$f(x)$ のグラフと 2 直線 $x = a$，$x = b$ および x 軸で囲まれた部分（斜線部）の面積を表す。

③ $\int_{-\infty}^{+\infty} f(x)dx$ は $f(x)$ の $-\infty$ から $+\infty$ までの広義積分といい，その値は $f(x)$ のグラフと x 軸で囲まれた部分の全面積を表す。従って，密度関数といえば，そのグラフと x 軸で囲まれた部分の全面積は常に 1 である。

(7) (6)の②においては，$a = b$ であれば，定積分の値は 0 である。従って，任意の a に対して，

$$P(X = a) = P(a \leq X \leq a) = \int_a^a f(x)dx = 0$$

また，確率 $P(a < X < b)$ も斜線部の面積を表すので，次が成り立つ。（つまり，連続的確率変数の確率計算では，等号は無意味になる。）

$$P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b)$$

(8) 連続的確率変数 X では，X の確率関数 $p(x) = P(X = x)$ の値は常に 0 になるので，この関数は無意味になる。一方，X の分布関数 $F(x) = P(X \leq x)$ は重要な関数になり，

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)\,dx$$

となる。$F(a)$は，区間 $x \leq a$ における，曲線と x 軸，および直線 $x = a$ で囲まれる部分の面積（上の図の斜線部の左側の部分の面積）を表す。

(9) 離散的確率変数 X の場合，X の確率分布とは，X の確率関数 $p(x) = P(X = x)$ や，確率分布表を意味する。連続的確率変数 X の場合は，厳密な定義はあるが，X の分布関数や密度関数などを，「X の確率分布」または単に「X の分布」という。

(10) (6)の確率 $P(a \leq X \leq b)$ は，試行を行ったとき，X の実現値が区間 $a \leq x \leq b$ に入る確率である。この確率は，直感的に理解してよい。つまり，試行を n 回行ったとき，X の実現値 x_i において，$a \leq x_i \leq b$ となる x_i の個数の相対度数は

$$\frac{a \leq x_i \leq b \text{となる } x_i \text{ の個数}}{n}$$

であるが，n を限りなく大きくしたときの，この相対度数の極限値が $P(a \leq X \leq b)$ であ

る。

(11) ただし，X が母集団の個体の特性値の場合は，$a \leq x_i \leq b$ となる特性値 x_i をもつ個体の個数の割合が，$P(a \leq X \leq b)$ であると考えてよい。

■ 例題

X の確率密度関数 $f(x)$ が，次のように与えられている。c と L は定数で，$L > 0$ である。

$$f(x) = \begin{cases} c & (0 \leq x \leq L) \\ 0 & (\text{それ以外}) \end{cases}$$

(1) c を L で表わせ。

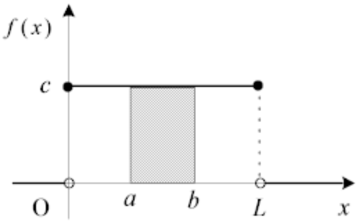
(2) $0 \leq a < b \leq L$ を満たす実数 a, b に対して，確率 $P(a \leq X \leq b)$ を求めよ。

＜解説＞

密度関数のグラフは，直線の場合もある。

(1) 常に，$f(x) \geq 0$ であるから，$c \geq 0$ である。

また，$y = f(x)$ のグラフと，x 軸で囲まれた部分の面積は，右の図の長方形の面積であるから，



長方形の面積 ＝ 縦 × 横 ＝ $c \times L$

一方，$f(x)$ は密度関数であるから，

$$c \times L = 1 \qquad \therefore\ c = \frac{1}{L} \quad \text{（答）}$$

(2) 求める確率は，斜線部の長方形の面積であるから，

$$P(a \leq X \leq b) = c \times (b - a) = \frac{b - a}{L} \quad \text{（答）}$$

上記のような確率分布，すなわち，密度関数 $f(x)$ の値が，ある区間内では一定値，それ以外の範囲では常に 0 になる場合，この確率分布を「一様分布」という。

2. 自然対数の底 e

(1) 無限数列 a_1，a_2，…，a_n，… において，項の番号 n を限りなく大きくするとき，a_n が一定の値 b に限りなく近づくならば，

$$\lim_{n\to\infty} a_n = b \qquad \text{または} \qquad n \to \infty \text{ のとき } a_n \to b$$

と表し，b を数列 $\{a_n\}$ の極限値という。

（例）$\displaystyle \lim_{n\to\infty} \frac{1}{n} = 0$，$\displaystyle \lim_{n\to\infty} \frac{n+1}{n} = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right) = 1$

(2) n を自然数とするとき，無限数列

$$\left(1+\frac{1}{n}\right)^n \quad (n=1,\ 2,\ \cdots)$$

の値は，$n \to \infty$ のとき，一定の値 2.7182818 … に限りなく近づくことが知られている。この極限値を e で表し，「自然対数の底」または「ネイピアの数」と呼ぶ。

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e \qquad (e = 2.7182818\ \cdots)$$

(3)　実数全体で定義された指数関数

$$y = e^x \quad (-\infty < x < \infty)$$

は，常に正の値をとる増加関数である。e^x を $\exp(x)$ と表すこともある。

（例） $\exp(3x) = e^{3x}$,　　$\exp\left\{\frac{(x-1)^2}{2}\right\} = e^{\frac{(x-1)^2}{2}}$

3.　正規分布（normal distribution）の定義

(1)　連続的確率変数 X の密度関数 $f(x)$ が

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

（ただし，μ, σ は定数で，$\sigma > 0$）

であるとき，X は正規分布 $N(\mu,\ \sigma^2)$ に従うといい，$X \sim N(\mu,\ \sigma^2)$ で表す。(式の中の π は円周率 $\pi = 3.14159\cdots$ である。)

(2)　次が成立する。

> **● 定理（正規分布の平均・分散）**
>
> $X \sim N(\mu,\ \sigma^2)$ のとき，次が成り立つ。
>
> ①　X の平均は μ，　　すなわち　$E(X) = \mu$
>
> ②　X の分散は σ^2，　　すなわち　$V(X) = \sigma^2$
>
> ③　X の標準偏差は σ，すなわち　$\sigma(X) = \sqrt{V(X)} = \sigma$

この証明には，連続的確率変数である X の平均や分散の定義，および，数学的な知識が必要になるので，証明なしにそのまま認めてよい。ただし，X の密度関数 $f(x)$ が μ と σ のみで決まる関数であること，従って，正規分布 $N(\mu,\ \sigma^2)$ は，平均 μ と標準偏差 σ の 2 つで一意に決まる分布であることに注意する。

(3)　$\mu = 0$，$\sigma = 1$ の特別な正規分布 $N(0,\ 1)$ を，標準正規分布という。その密度関数は，次のようになる。

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

(4)　X が正規分布 $N(\mu,\ \sigma^2)$ に従うとき，その密度関数 $f(x)$ のグラフを，正規分布曲線または正規曲線という。この曲線は，次の性質をもつ。（曲線の形はテキスト p.144 を参照）

①　関数 $f(x)$ は，$x = \mu$ で最大になる。

②　曲線 $y = f(x)$ は，直線 $x = \mu$ に関して対象である。

③　密度関数であるから，曲線 $y = f(x)$ と x 軸で囲まれる部分の面積は 1 である。

④　平均 μ から標準偏差 σ の 1 倍，2 倍，3 倍離れている区間における面積は，次のようになる。

(a)　$P(\mu - \sigma < X < \mu + \sigma) = 0.683$ （約 70%）
(b)　$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.954$ （約 95%）
(c)　$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.997$ （約 100%）

(5)　確率変数 Z が標準正規分布 $N(0,\ 1)$ に従うときは，その密度関数 $f(z)$ のグラフは，次の性質をもつ。

①　関数 $f(z)$ は，$z = 0$ で最大になる。

②　曲線 $y = f(z)$ は，y 軸に関して対称である。

③　密度関数であるから，曲線 $y = f(z)$ と z 軸で囲まれる部分の面積は 1 である。

④　平均 0 から標準偏差 1 の 1 倍，2 倍，3 倍離れている区間における面積は，次のようになる。

(a)　$P(-1 < Z < 1) = 0.683$ （約 70%）
(b)　$P(-2 < Z < 2) = 0.954$ （約 95%）
(c)　$P(-3 < Z < 3) = 0.997$ （約 100%）

■ 例 1

確率変数 X が正規分布 $N(10,\ 20)$ に従うとき，X の平均，分散，標準偏差を求めよ。

（解）X の平均は $E(X) = 10$
X の分散は $V(X) = 20$
X の標準偏差は $\sigma(X) = \sqrt{V(X)} = \sqrt{20} = 2\sqrt{5}$

■ 例 2

母集団の個体の特性値を X とする。$X \sim N(8,\ 4)$ のとき，母平均，母分散，母標準偏差を求めよ。

（解）母平均は $E(X) = 8$，母分散は $V(X) = 4$
母標準偏差は $\sigma(X) = \sqrt{V(X)} = \sqrt{4} = 2$

4. 正規分布の１次変換

● **定理（正規分布の１次変換）**

確率変数 X が正規分布 $N(\mu,\ \sigma^2)$ に従うとき，

$$Y = aX + b \quad (a, b \text{ は定数})$$

とおくと，確率変数 Y も正規分布に従う。従って，

$$X \sim N(\mu, \sigma^2) \text{ ならば，} aX + b \sim N(a\mu + b,\ a^2\sigma^2)$$

(1)　正規分布に従う確率変数 X を１次変換しても，正規分布に従うことが知られている。これは，データ x_i の１次変換で説明したが，直感的に認めてよい。

(2)　従って，$X \sim N(\mu,\ \sigma^2)$ のとき，

$$Y = aX + b \quad (a, b \text{ は定数})$$

とおくと，

$$E(Y) = E(aX + b) = aE(X) + b = a\mu + b$$
$$V(Y) = V(aX + b) = a^2V(X) = a^2\sigma^2$$

であるから，次が成り立つことが分かる。

$$X \sim N(\mu,\ \sigma^2) \text{ ならば，} aX + b \sim N(a\mu + b, a^2\sigma^2)$$

5. 正規分布の標準化

● **定理（正規分布の標準化）**

$X \sim N(\mu,\ \sigma^2)$ のとき，

$$Z = \frac{X - E(X)}{\sigma(X)} = \frac{X - \mu}{\sigma} \quad (Z \text{ は } X \text{ の標準測度})$$

とおくと，$Z \sim N(0,\ 1)$ となる。

さらに，$N(\mu,\ \sigma^2)$ と $N(0,\ 1)$ の対応する部分の面積は等しい。

すなわち，以下が成り立つ。

(a)　$P(a \leq X \leq b) = P\left(\dfrac{a - \mu}{\sigma} \leq Z \leq \dfrac{b - \mu}{\sigma}\right)$

(b)　$P(X \leq a) = P\left(Z \leq \dfrac{a - \mu}{\sigma}\right)$

(c)　$P(X \geq b) = P\left(Z \geq \dfrac{b - \mu}{\sigma}\right)$

(1)　上記の前半は，もはや自明である。確率変数 X を標準化して，その標準測度 Z を考えれば，

$$E(Z) = 0, \quad \sigma(Z) = 1$$

になった。一方，標準化は 1 次変換であるから， $X \sim N(\mu,\ \sigma^2)$ であれば， Z は正規分布に従うので， $Z \sim N(0,\ 1)$ となる。

(2)　後半は，対応する部分の面積は変わらないことを意味しているが，これも直感的に認めてよい。従って，正規分布での確率（面積）の計算は，標準化して標準正規分布で計算できる。

6.　標準正規分布の確率計算

標準正規分布 $N(0,\ 1)$ における確率（面積）は，正規分布表に示されている。ただし，正規分布表には，次の 2 種類がある。

① 実数 $z \geq 0$ に対して， $P(Z \leq z)$ の値が示されているもの

② 実数 $z \geq 0$ に対して， $P(0 \leq Z \leq z)$ の値が示されているもの

①は，分布関数 $F(z) = P(Z \leq z)$ の $z \geq 0$ の範囲での値が示された形である。両者は本質的に同じであり，①の表の各数値から 0.5 を引けば，②の表が得られる。テキスト p.264 の標準正規分布表は，①の方である。

$Z \sim N(0,\ 1)$ のとき， Z の密度関数は

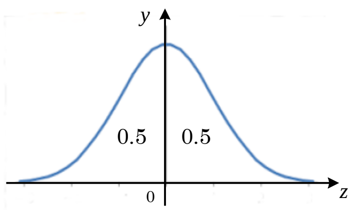
$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}$$

であり，関数 $y = f(z)$ のグラフは y 軸に関して対称である。従って，以下のようなことは自明である。（グラフを書いて考えればよい。）

(1)　$P(Z \geq 0)$

全体の面積は 1 で，グラフは y 軸に関して対称なので

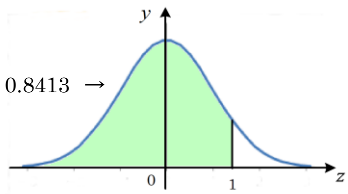
$P(Z \geq 0) = P(Z \leq 0) = 0.5$



(2)　$P(Z \leq 1)$

$1 = 1.00 > 0$ なので，正規分布表から

$P(Z \leq 1) = P(Z \leq 1.00) = 0.8413$

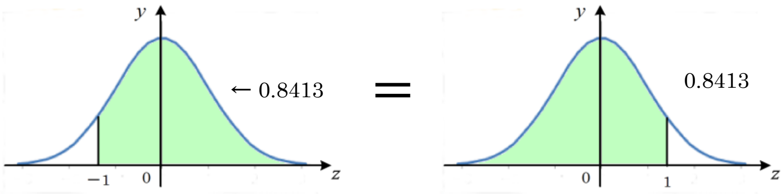


$P(Z \leq 1)$

(3)　P(Z ≥ −1)

グラフは y 軸に関して対称なので

$$P(Z \geq -1) = P(Z \leq 1) = 0.8413$$

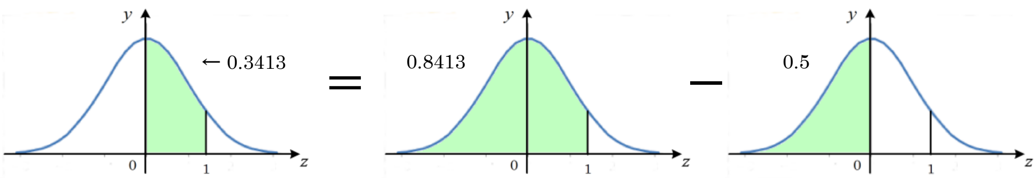


$P(Z \geq -1)$　　$P(Z \leq 1)$

(4)　P(0 ≤ Z ≤ 1)

$P(Z \leq 1)$から 0.5 を引けばよいので

$$P(0 \leq Z \leq 1) = P(Z \leq 1) - 0.5 = 0.8413 - 0.5 = 0.3413$$

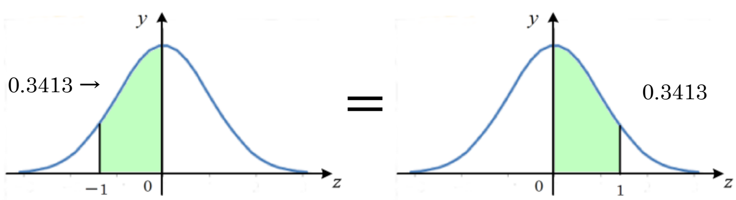


$P(0 \leq Z \leq 1)$　　$P(Z \leq 1)$　　$P(Z \leq 0)$

(5)　P(−1 ≤ Z ≤ 0)

グラフは y 軸に関して対称なので

$$P(-1 \leq Z \leq 0) = P(0 \leq Z \leq 1) = P(Z \leq 1) - 0.5 = 0.8413 - 0.5 = 0.3413$$



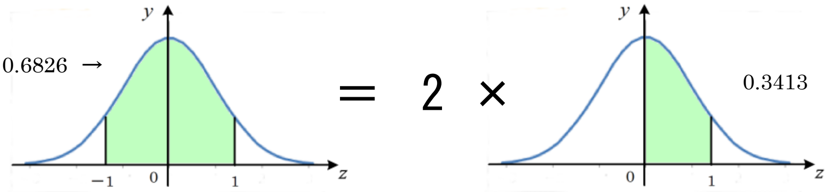
$P(-1 \leq Z \leq 0)$　　$P(0 \leq Z \leq 1)$

(6)　P(−1 ≤ Z ≤ 1)

$P(0 \leq Z \leq 1)$の 2 倍なので

$$P(-1 \leq Z \leq 1) = 2 \times P(0 \leq Z \leq 1) = 2 \times \{ P(Z \leq 1) - 0.5 \}$$

$$= 2 \times (0.8413 - 0.5) = 2 \times 0.3413 = 0.6826$$

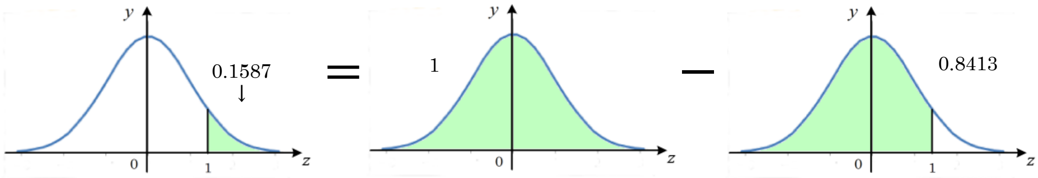


$P(-1 \leq Z \leq 1)$ $P(0 \leq Z \leq 1)$

(7) $P(Z \geq 1)$

全面積 1 から $P(Z \leq 1)$ を引けばよいので

$$P(Z \geq 1) = 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587$$

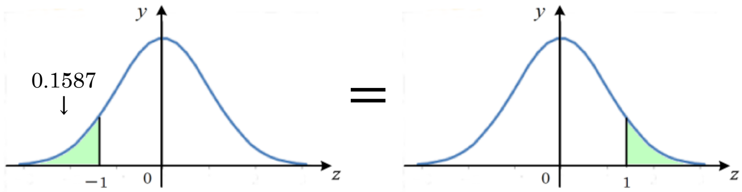


$P(Z \geq 1)$ 全面積は 1 $P(Z \leq 1)$

(8) $P(Z \leq -1)$

y 軸に関して対称なので， $P(Z \geq 1)$ と同じ

$$P(Z \leq -1) = P(Z \geq 1) = 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587$$

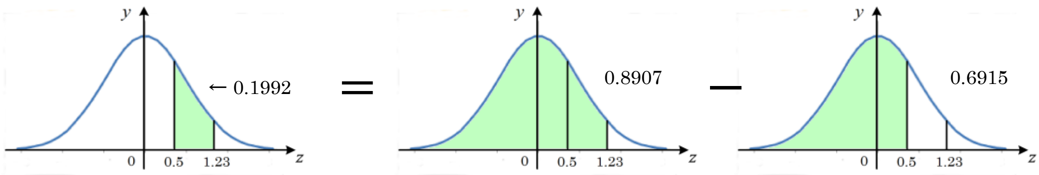


$P(Z \leq -1)$ $P(Z \geq 1)$

(9) $P(0.5 \leq Z \leq 1.23)$

$P(Z \leq 1.23)$ から $P(Z \leq 0.5)$ を引けばよいので

$$P(0.5 \leq Z \leq 1.23) = P(Z \leq 1.23) - P(Z \leq 0.5) = 0.8907 - 0.6915 = 0.1992$$

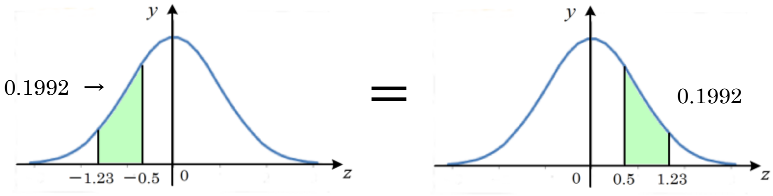


$P(0.5 \leq Z \leq 1.23)$ $P(Z \leq 1.23)$ $P(Z \leq 0.5)$

（10） $P(-1.23 \leq Z \leq 0.5)$

$P(0.5 \leq Z \leq 1.23)$ と同じなので

$$P(-1.23 \leq Z \leq -0.5) = P(0.5 \leq Z \leq 1.23) = P(Z \leq 1.23) - P(Z \leq 0.5)$$
$$= 0.8907 - 0.6915 = 0.1992$$



$P(-1.23 \leq Z \leq -0.5)$ $P(0.5 \leq Z \leq 1.23)$