講義4 確率分布と統計的推測

1. 期待値や分散で，確率分布の特徴が分かる！

さァ，これから“**確率分布**”について解説しよう。試行結果が数値で表されるとき，これを**確率変数**とおき，この確率変数の値それぞれに確率が割り当てられるとき，**確率分布**が与えられていると言うんだね。今回は，この確率分布の形を特徴づける **3** つの指標として“**期待値**”，“**分散**”，そして“**標準偏差**”について教えよう。

● 分散で確率分布の広がり具合がわかる!

たとえば，正しいサイコロを **1** 回投げたとき，出る目は **1** から **6** までの数で，それぞれが出る確率は共に等しく $\frac{1}{6}$ になるはずだ。このサイコロの目のように，ある試行結果が数値で表されるとき，これを確率変数 $X = x_1, x_2, \cdots, x_n$ とおくことにする。そして，それぞれの確率変数の値 $x_1, x_2, \cdots, x_n$ に対して，確率 $P = p_1, p_2, \cdots, p_n$ が与えられているとき，「確率変数 X は確率分布が与えられている。」または「確率変数 X は，この確率分布に従う。」と言うんだね。

そして，この確率分布は次のように表やグラフで表すことができる。

確率分布表

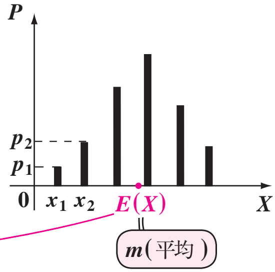
確率変数 X	x_1	x_2	$\cdots$	x_n
確　率 P	p_1	p_2	$\cdots$	p_n

(1)　$\sum_{k=1}^{n} p_k = p_1 + p_2 + \cdots + p_n = \underline{1}$　全確率

(2)　期待値 $E(X) = \sum_{k=1}^{n} \underline{x_k}\,\underline{p_k}$　確率変数　確率

$= x_1p_1 + x_2p_2 + \cdots + x_np_n$

図 1　期待値 $E(X)$ は確率変数 X の平均 m のことだ！



確率変数 $X = x_1, x_2, \cdots, x_n$ に対応する確率をそれぞれ $p_1, p_2, \cdots, p_n$ とおいたとき，この確率の総和は 1 になるんだね。また，$x_k p_k (k=1, 2, \cdots, n)$ の総和が X の**期待値** $E(X)$ と呼ばれるもので，この X の**平均**の値を表しているんだね。したがって，$E(X)$ を $E(X) = m$（**平均**）とおいたりもする。覚えておこう。

"*Expectation*（期待値）"の頭文字　"*mean*（平均）"の頭文字

そして，さらに，**分散**（ぶんさん）と**標準偏差**（ひょうじゅんへんさ）の計算方法を次に示す。

分散と標準偏差

平均からの散らばり具合を表す量

これを $E(X^2) - \{E(X)\}^2$ とも表す！

(1) 分散 $V(X) = \underbrace{\sum_{k=1}^{n} (x_k - m)^2 p_k}_{\text{定義式}} = \underbrace{\sum_{k=1}^{n} x_k^2 p_k - m^2}_{\text{計算式}}$

$\because E(X) = \sum_{k=1}^{n} x_k p_k$
$E(X^2) = \sum_{k=1}^{n} x_k^2 p_k$

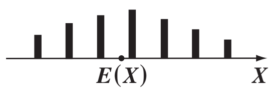
(2) 標準偏差 $D(X) = \sigma = \sqrt{V(X)}$

"シグマ" と読むよ

期待値 $E(X)$ が，確率変数 X の平均を表すのに対して，分散や標準偏差は確率変数の広がり具合を表す。これらの値が大きいと，横に広がった分布となり，逆に小さいとシャープなとんがった分布になるんだね。図 2 を参考にしてくれ。

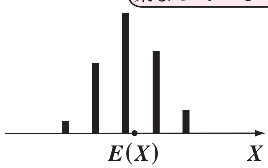
図 2　分散や標準偏差は分布の広がり具合を表す

(i) $V(X)$ や $D(X)$ が大きいと横に広がった分布



(ii) $V(X)$ や $D(X)$ が小さいとたてにシャープな分布

平均のまわりに集まっている



(1) では，分散の定義式と，計算式を示しておいた。実際に問題を解く上では，計算式の方を使うことが多い。この計算式により分散 $V(X)$ を求めるには，まず期待値 $m = E(X)$ を算出し，そして，（**確率変数**）^{2}× **確率** $[= x_k^2 \times p_k]$ の和をとったものから，期待値の 2 乗 $[m^2]$ を引いて分散 $V(X)$ を求める。ここで，期待値 $E(X)$ の記号法についてだけれど，$E(X) = \sum_{k=1}^{n} x_k p_k$ だから，別の確率変数 Y や Z などについても，

$E(Y) = \sum_{k=1}^{n} y_k p_k$，$E(Z) = \sum_{k=1}^{n} z_k p_k$ などと表せるんだよ。だから，$Y = X^2$ とおくと

$y_k = x_k^2$ より $E(X^2) = \sum_{k=1}^{n} x_k^2 p_k$ とも表せる。よって，(1) の分散の計算式も $V(X) = E((X-m)^2) = E(X^2) - \{E(X)\}^2$ と表すこともできる。大丈夫？

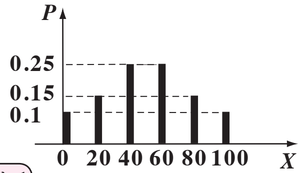
さらに (2) の標準偏差 $D(X)$ (σ とも書く)は，$V(X)$ の正の平方根なんだね。

シグマ

● 例題で，$V(X)$ と $D(X)$ に慣れよう！

次に，A 組の数学のテストの得点分布を例題として示す。得点を確率変数 X として，100 人のクラスのうち 15 人が 80 点をとっている場合，$X = 80$ 点となる確率は，$\frac{15}{100} = 0.15$ となるね。

図 3　A 組の得点の確率分布



A 組の得点の確率分布表

得点 X	0	20	40	60	80	100
確率 P	0.1	0.15	0.25	0.25	0.15	0.1

$\frac{10}{100}$　$\frac{15}{100}$　$\frac{25}{100}$　$\frac{25}{100}$　$\frac{15}{100}$　$\frac{10}{100}$　総和は 1

まず，確率 P の総和が 1 になることを確認しよう。確率分布表を書いたときには，必ず確率 P の総和が 1 (全確率)となることを忘れないでくれ。それでは，A 組の得点 X の期待値，分散，標準偏差を求めてみよう。

期待値 $E(X) = 0\times0.1 + 20\times0.15 + 40\times0.25$

$\quad + 60\times0.25 + 80\times0.15 + 100\times0.1$　←「確率変数 × 確率」の和

$\quad = 0 + 3 + 10 + 15 + 12 + 10 = 50$(点)

分散 $V(X) = (0^2\times0.1 + 20^2\times0.15 + 40^2\times0.25$　(期待値)^2

$\quad + 60^2\times0.25 + 80^2\times0.15 + 100^2\times0.1) - 50^2$　←「(確率変数)^2× 確率」の和から (期待値)^2 を引く！

$\quad = (0 + 60 + 400 + 900 + 960 + 1000) - 2500$

$\quad = \underline{\underline{820}}$

標準偏差 $D(X) = \sqrt{\underline{\underline{820}}} = 2\sqrt{205}$ [≒ 29]　← $\sqrt{分散}$

どう，計算は大変だったけれどやり方はわかった？ それでは，もう 1 つのクラス B 組の得点の確率分布と，期待値，分散，標準偏差を示すよ。

B 組の得点の確率分布表

得点 Y	0	20	40	60	80	100
確率 P	0	0.1	0.4	0.4	0.1	0

確率の総和
0+0.1+0.4+0.4+0.1+0=1
となって OK だね。

では，B 組の得点 Y の期待値 $E(Y)$, 分散 $V(Y)$，標準偏差 $D(Y)$ も同様に求めてみよう。

期待値 $E(Y) = \cancel{0\times 0} + 20\times 0.1 + 40\times 0.4 + 60\times 0.4 + 80\times 0.1 + \cancel{100\times 0}$

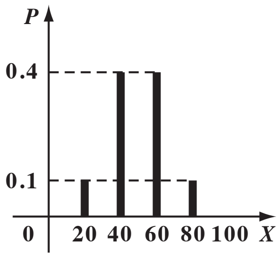
$= 2 + 16 + 24 + 8 = 50$(点) ← これが，平均点だね。

分散 $V(Y) = (\cancel{0^2\times 0} + 20^2\times 0.1 + \cdots + \cancel{100^2\times 0}) - 50^2$

$= 2760 - 2500 = 260$

標準偏差 $D(Y) = \sqrt{260} = 2\sqrt{65}$ [$\fallingdotseq 16$]

図 4　B 組の得点の確率分布



以上より，A，B 2 つの組の得点の平均 (期待値) は共に 50 点で等しかったんだけれど，分散 $V(X)$ は，A 組より B 組の方が小さいので，B 組の方がより多くの生徒が平均点あたりにいる，まとまりのいい (?) クラスといえるんだね。以上で，期待値（平均），分散，標準偏差の求め方と，それぞれの値の意味もマスターできたと思う。

● 新たな確率変数の期待値，分散，標準偏差を求めよう！

ある確率変数 X を使って新たな確率変数 Y を，$Y = aX + b$ (a, b : 定数) で定義したとき，この Y の期待値，分散，標準偏差は，次のように書けることも覚えておこう！

確率変数 $aX + b$

$Y = aX + b$ のとき，(a, b は定数)

(1)　期待値　$E(Y) = E(aX + b) = aE(X) + b$

(2)　分散　$V(Y) = V(aX + b) = a^2V(X)$ ← b は無関係！

(3)　標準偏差　$D(Y) = \sqrt{V(Y)} = \sqrt{a^2V(X)} = |a|D(X)$

$Y(=aX+b)$ の期待値 $E(Y)$ は，

$E(Y)=E(aX+b)=aE(X)+b$ と変形できるので，元の X の期待値 $E(X)$ をそのまま a 倍して，b をたせばいいんだね。

また，分散 $V(Y)$ の公式：

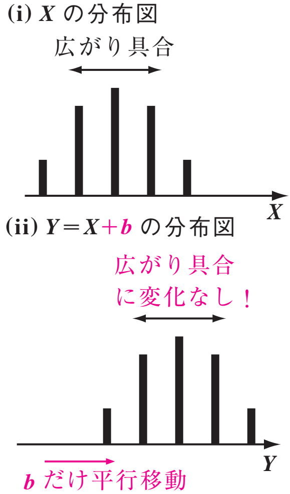
$$V(Y)=V(aX+b)=\underline{a^2V(X)}$$

b の影響がない！

から，分散 $V(Y)$ には b の影響がないことに注意しよう。$a=1$ の特殊な場合，すなわち，$Y=X+b$ のとき，図 5 に示すように，X に b をたすってことは，確率変数 X を b だけ平行移動することなんだね。だから，b が確率分布の広がり具合，つまり分散に影響しないのは当り前だったんだね。大丈夫？

図 5　$a=1$ のとき

$Y=X+b$ は，X を b だけ平行移動したもの。



◆例題 12◆

確率変数 X の期待値 $E(X)=20$，分散 $V(X)=16$，標準偏差 $D(X)=4$ のとき，次の確率変数の期待値，分散，標準偏差を求めよ。

(1) $Y=2X-10$　　　　(2) $Z=40-X$

(1) Y の期待値 $E(Y)$，分散 $V(Y)$，標準偏差 $D(Y)$ を求めよう。

・$E(Y)=E(2X-10)=2E(X)-10$ ← $E(aX+b)=aE(X)+b$

$\quad=2\times20-10=30$

・$V(Y)=V(2X\underline{-10})=2^2\cdot V(X)$ ← $V(aX+b)=a^2V(X)$

これは影響しない！

$\quad=4\times16=64$

・$D(Y)=\sqrt{V(Y)}=\sqrt{64}=8$ ← $D(aX+b)=|a|D(X)$ を使うと $2\times4=8$ と同じ結果だ。

(2) Z の期待値 $E(Z)$，分散 $V(Z)$，標準偏差 $D(Z)$ も同様に，

・$E(Z) = E(-1 \cdot X + 40) = -E(X) + 40 = -20 + 40 = 20$

・$V(Z) = V(-1 \cdot X \underline{+40}) = (-1)^2 \cdot V(X) = 1 \times 16 = 16$

これは影響しない！

・$D(Z) = \sqrt{V(Z)} = \sqrt{16} = 4$

さらに，期待値に関しては，次の重要な公式があることも要注意だ。

$X+Y$ の期待値 $E(X+Y)$

$E(X+Y) = E(X) + E(Y)$ ……$(*)$

これは，2 つの確率変数 X と Y の和の期待値は，それぞれ別々に計算した期待値の和 $E(X)+E(Y)$ になると言ってるんだね。したがって，たとえば，2 つの変数 X と Y の期待値がそれぞれ $E(X)=32$，$E(Y)=18$ であったとすると，$X+Y$ の期待値は新たに $Z=X+Y$ とおいて，Z の期待値を計算しなくても，公式（$*$）を使えば $E(X+Y)=E(X)+E(Y)=32+18=50$ と，アッという間に求まるんだね。簡単だろう？

そして，（$*$）は公式 $E(aX+b)=aE(X)+b$ と組み合わせれば，さらに

$E(aX+bY+c) = aE(X)+bE(Y)+c$ や

$E(aX+bY+cZ) = aE(X)+bE(Y)+cE(Z)$ （ただし，a, b, c：実数定数）

など…の公式も導けることが分かると思う。

ン？ では，分散についても $X+Y$ の分散は，

$V(X+Y) = V(X) + V(Y)$ ……$(**)$ となるのかって!?

いい質問だ！実はこの $(**)$ は常に成り立つとは限らない。2 つの確率変数 X と Y が独立な確率変数のときのみ成り立つと言えるんだ。これについては，次回の講義でも詳しく解説しよう。もう少し待ってくれ！

$V(X)$, $E(aX+b)$, $V(aX+b)$ の公式

絶対暗記問題 49	難易度 ★★	CHECK1	CHECK2	CHECK3

$m = E(X) = \sum_{k=1}^{n} x_k p_k$ $\left(\sum_{k=1}^{n} p_k = 1\right)$ と定義する。次の式を証明せよ。

(1) X の分散 $V(X) = E((X-m)^2) = E(X^2) - \{E(X)\}^2$

(2) $Y = aX + b$ $(a, b$：実数定数$)$ のとき，

　(ⅰ) $E(Y) = aE(X) + b$　　(ⅱ) $V(Y) = a^2 V(X)$

ヒント！ (1)(2) 共に公式の証明だね。$E(X) = \sum_{k=1}^{n} x_k p_k$ から，$V(X) = E((X-m)^2) = \sum_{k=1}^{n}(x_k - m)^2 p_k$，$E(X^2) = \sum_{k=1}^{n} x_k{}^2 p_k$ などとなるんだよ。大丈夫？

解答＆解説

(1) $V(X) = E((X-m)^2) = \sum_{k=1}^{n}(x_k - m)^2 \cdot p_k$ ← 分散 $V(X)$ の定義式

$= \sum_{k=1}^{n}(x_k{}^2 - 2mx_k + m^2)p_k = \underbrace{\sum_{k=1}^{n} x_k{}^2 p_k}_{E(X^2)} - 2m\underbrace{\sum_{k=1}^{n} x_k p_k}_{m=E(X)} + m^2 \underbrace{\sum_{k=1}^{n} p_k}_{1(\text{全確率})}$

$= E(X^2) - 2m^2 + m^2 = E(X^2) - m^2$

$= E(X^2) - \{E(X)\}^2$ ……………………(終)

(2) $Y = aX + b$ $(a, b$：実数定数$)$ のとき，

(ⅰ) $E(Y) = E(aX+b) = \sum_{k=1}^{n}(ax_k + b)p_k$

$= a\underbrace{\sum_{k=1}^{n} x_k p_k}_{E(X)} + b\underbrace{\sum_{k=1}^{n} p_k}_{1} = aE(X) + b$ ……① ……………(終)

(ⅱ) Y の平均を $m' = E(Y)$ とおくと，Y の分散 $V(Y)$ は，

$V(Y) = E((\underbrace{Y}_{aX+b} - \underbrace{m'}_{(am+b)\,(\text{①より})})^2) = E(\{aX + \cancel{b} - (am + \cancel{b})\}^2)$

$= E((aX - am)^2) = E(\underbrace{a^2}_{\text{定数}}(X-m)^2)$

$= \underline{a^2 E((X-m)^2)} = a^2 V(X)$ ……………………(終)

↑ $E(aX) = aE(X)$ (①より)

期待値，分散，標準偏差の計算

絶対暗記問題 50	難易度 ★	CHECK1	CHECK2	CHECK3

赤球 **6** 個，白球 **4** 個が入った袋の中から **2** 個の球を同時に取り出す。取り出された赤球の個数を X とするとき，X の期待値 $E(X)$，分散 $V(X)$，標準偏差 $D(X)$ の値を求めよ。

ヒント！ まず，$X=0, 1, 2$ のそれぞれがとる確率を求めて確率分布表を作るんだね。それから，期待値 $E(X)$，分散 $V(X)$，標準偏差 $D(X)$ の値を計算式通りに求めていけばいいんだよ。今回は，ウォーミングアップ問題だ。

解答＆解説

赤 **6** 個，白 **4** 個，計 **10** 個の球の入った袋から **2** 個を取り出す全場合の数は

$n(U) = {}_{10}C_2 = 45$ 通り　←　${}_{10}C_2 = \frac{10!}{2!8!} = \frac{10\times9}{2} = 45$

取り出された **2** 個の球のうち赤球の個数を X とおくと，$X=0, 1, 2$ であり，それぞれがとる確率を求める。

○○

赤球 **6** 個
白球 **4** 個

$P(X=0) = \frac{\overbrace{{}_4C_2}^{\text{白4コから2コ}}}{45} = \frac{6}{45} = \frac{2}{15}$

$P(X=1) = \frac{\overbrace{{}_6C_1}^{\text{赤6コから1コ}}\cdot\overbrace{{}_4C_1}^{\text{白4コから1コ}}}{45} = \frac{6\times4}{45} = \frac{8}{15}$

$P(X=2) = \frac{\overbrace{{}_6C_2}^{\text{赤6コから2コ}}}{45} = \frac{15}{45} = \frac{5}{15}$

確率分布表

X	0	1	2
P	$\frac{2}{15}$	$\frac{8}{15}$	$\frac{5}{15}$

以上より，X の確率分布表を右に示す。これから，

$E(X) = \sum_{k=1}^{n} x_k p_k$

期待値 $E(X) = \cancel{0\times\frac{2}{15}} + 1\times\frac{8}{15} + 2\times\frac{5}{15} = \frac{8+10}{15} = \underline{\underline{\frac{6}{5}}}$ ……(答)

$V(X) = \sum_{k=1}^{n} x_k^2 p_k - m^2$

分散 $V(X) = \cancel{0^2\times\frac{2}{15}} + 1^2\times\frac{8}{15} + 2^2\times\frac{5}{15} - \left(\underline{\underline{\frac{6}{5}}}\right)^2 = \frac{32}{75}$ ……(答)

$D(X) = \sqrt{V(X)}$

標準偏差 $D(X) = \sqrt{\frac{32}{75}} = \frac{4\sqrt{2}}{5\sqrt{3}} = \frac{4\sqrt{6}}{15}$ ……………………(答)

確率変数 $aX+b$ の期待値と標準偏差

絶対暗記問題 51　難易度 ★★　CHECK1　CHECK2　CHECK3

(1) 右の確率変数 X の確率分布に対して，期待値 $E(X)$，標準偏差 $D(X)$ の値を求めよ。

変数 X	60	70	80	90	100
確率 P	0.1	0.2	0.4	0.2	0.1

(2) (1)の X に対して，$Y=50+\dfrac{10\{X-E(X)\}}{D(X)}$ とおくとき，この新たな確率変数 Y の期待値 $E(Y)$，標準偏差 $D(Y)$ を求めよ。　(日本大)

ヒント！ 変数 X に対して，新たな変数 Y を $Y=aX+b$ で定義すると，Y の期待値，分散，標準偏差はそれぞれ $E(Y)=aE(X)+b$，$V(Y)=a^2V(X)$，$D(Y)=|a|\sqrt{V(X)}=|a|D(X)$ となるんだったね。覚えて使いこなしてくれ。

解答＆解説

(1) 与えられた X の確率分布表より，

期待値　$E(X)=60\times0.1+70\times0.2+80\times0.4+90\times0.2+100\times0.1=\underline{80}$　…(答)

分散　$V(X)=60^2\times0.1+70^2\times0.2+80^2\times0.4+90^2\times0.2+100^2\times0.1-\underline{80}^2$

$=6520-6400=\underline{120}$

標準偏差　$D(X)=\sqrt{V(X)}=\sqrt{120}=\underline{\underline{2\sqrt{30}}}$　……(答)

X の期待値 $E(X)$ と標準偏差 $D(X)$ は，計算式通り求める。

(2) 新たな確率変数 Y は，X を使って，

$$Y=50+\frac{10\{X-E(X)\}}{D(X)}=50+\frac{10(X-80)}{2\sqrt{30}}$$

この変換式は実は，得点 X から偏差値 Y に変換する式だ！(P167)

$$\therefore Y=\underbrace{\boxed{\frac{5}{\sqrt{30}}}}_{a}X+\underbrace{\boxed{50-\frac{400}{\sqrt{30}}}}_{b}$$

$Y=aX+b$ の形だ！

$Y=aX+b$ の期待値 $E(Y)$ と標準偏差 $D(Y)$ は(1)の $E(X)$ と $D(X)$ を利用する。

以上より，Y の期待値 $E(Y)$ と標準偏差 $D(Y)$ は，

$E(Y)=aE(X)+b$

$$E(Y)=\underbrace{\boxed{\frac{5}{\sqrt{30}}}}_{a}\underbrace{E(X)}_{80}+\underbrace{\boxed{50-\frac{400}{\sqrt{30}}}}_{b}=50\quad\cdots\cdots(\text{答})$$

$D(Y)=|a|D(X)$

$$D(Y)=\underbrace{\boxed{\frac{5}{\sqrt{30}}}}_{|a|}\underbrace{D(X)}_{2\sqrt{30}}=\frac{10\sqrt{30}}{\sqrt{30}}=10\quad\cdots\cdots\cdots\cdots(\text{答})$$

公式 $E(X+Y)=E(X)+E(Y)$ の活用

絶対暗記問題 52 難易度 ★★ CHECK1 CHECK2 CHECK3

125 から 324 までの番号札が 1 枚ずつ，計 200 枚ある。この中から無作為に 1 枚取り出して，その番号の百の位の数字を X，十の位の数字を Y とおく。確率変数 X の期待値 $E(X)$ と，確率変数 $X+Y$ の期待値 $E(X+Y)$ を求めよ。

ヒント！ 百の位の数 X について，$X=1, 2, 3$ の札がそれぞれ 75 枚，100 枚，25 枚あるが，十の位の数 Y は，$Y=0\sim9$ までの10個の数字の札が等しく20枚ずつある。ここでは，期待値の重要公式 $E(X+Y)=E(X)+E(Y)$ を使う。

解答＆解説

125 から 324 までの番号の札のうち，$X=1$ の札が 75 枚，$X=2$ は 100 枚，$X=3$ は 25 枚現われるから，確率変数 X の確率分布は右のようになる。

X の確率分布

X	1	2	3
P	$\frac{3}{8}$	$\frac{4}{8}$	$\frac{1}{8}$

$P(X=1)=\frac{75}{200}=\frac{3}{8}$ （125 ～ 199）

$P(X=2)=\frac{100}{200}=\frac{4}{8}$ （200 ～ 299）

$P(X=3)=\frac{25}{200}=\frac{1}{8}$ だね。（300 ～ 324）

$$\therefore E(X)=1\times\frac{3}{8}+2\times\frac{4}{8}+3\times\frac{1}{8}$$
$$=\frac{3+8+3}{8}=\frac{14}{8}=\frac{7}{4}\quad\cdots\cdots\cdots(\text{答})$$

次に，十の位の数 Y は，$0, 1, \cdots, 9$ までの 10 個の数字が 20 枚ずつ等しく現われるので，期待値 $E(Y)$ は

$P(Y=0)=P(Y=1)=\cdots=P(Y=9)=\frac{20}{200}=\frac{1}{10}$ だ。

$\sum_{k=0}^{9}k=0+1+2+\cdots+9=1+2+\cdots+9=\sum_{k=1}^{9}k$ だね。

$$E(Y)=\sum_{k=0}^{9}k\cdot\frac{1}{10}=\frac{1}{10}\sum_{k=1}^{9}k=\frac{1}{10}\cdot\frac{1}{2}\cdot9\cdot10=\frac{9}{2}$$

期待値の重要公式

$$\therefore \boxed{E(X+Y)=E(X)+E(Y)}=\frac{7}{4}+\frac{9}{2}=\frac{25}{4}\quad\cdots\cdots\cdots\cdots\cdots\cdots(\text{答})$$

頻出問題にトライ・14 難易度 ★★ CHECK1 CHECK2 CHECK3

1 から 4 までの数字を一つずつ書いた4 個の玉が袋に入っている。この袋の中から玉を同時に 2 個取り出すとき，番号の大きい数を X，小さい数を Y とする。

(1) X と Y のそれぞれの期待値と分散を求めよ。

(2) $Z=aX-Y$ の期待値が 5 のとき，定数 a の値を求めよ。　（帯広畜産大＊）

解答は P193