情報領域演習第二

第 5 週 K 演習 (確率論) 第 1 回

K 演習

2017 年 5 月

確率論

事象の不確実性を確率で取り扱う学問。

1. 事象の不確実性を確率分布で表す。確率分布の与え方には大きく、表と関数がある。確率表は確率分布を表で与え、確率関数、累積分布関数、あるいは確率密度関数確率分布を関数で与える。
2. 1 次元の確率変数を考えるのは、理論を構築していくために最初に置いた数学的に単純な想定である。
3. 事象が標本空間のある部分集合に含まれる可能性を確率で表す。
4. 事象が標本空間のある部分集合で発生することが予見されるなら、その確率は条件付き確率となる。
5. 現実の不確実な事象は、抽象的な集合では扱いにくい。そのために標本空間を、整数集合や実数空間など少し具体的な集合で記述する。
6. 現実の不確実な事象には、複数の確率変数が関わる。そのために標本空間に次元を定める。
7. 集合で論じた同時確率、周辺確率、条件付き確率、加法法則、乗法法則などは、標本空間を離散集合や連続集合などに具体化しても、次元を導入しても、すべて同様に成り立つ。

演習問題１：確率表とベイズの定理

青・赤・黄の３色の袋がそれぞれ２、２、１個ずつある。３色の袋の中にはそれぞれ白球と黒球が表に書かれた個数入っている。いま、無作為に選ばれた袋から無作為に球を取り出したところ、白球であったという結果を知らされた。この時、次の各問いに答えよ。

各袋の球の内訳	青袋	赤袋	黄袋
白球	２個	１個	４個
黒球	３個	４個	１個

1. 袋の周辺確率表を記せ。
2. 袋の色が判明しているときに、袋から球を１つ取り出す試行の条件付き確率表を記せ。
3. 「無作為に選ばれた袋から無作為に球を取り出したところ、白球であったという結果を知らされた。この白球が黄色の袋から取り出されたもの」である確率をベイズの定理を用いて求めよ。
4. 袋を無作為に１つ選び、その中からさらに球を１つ無作為に選ぶ、という試行の同時確率表を記せ。
5. 同時確率表から、前々問の条件付き確率を求めよ。

演習問題 1 のヒント

1. 教科書の 1 章から 3 章を読み返して同時確率、周辺確率、条件付き確率の意味を確認する。
2. 教科書 p.10-11 のベイズの定理を理解する。
3. まずは $A_1, A_2, \ldots, A_n$ と B をこの問題ではどのように取れば良いかを考えよ。
4. それから定理 1.3 のベイズの定理を適用すると良い。
5. なお、この問題は教科書の p.12 の問題 1.6 である。

確率表

標本空間が 1 次元の離散集合なら、確率表はただの確率表である。

X	1	2	$\cdots$	M	総和
確率	p_1	p_2	$\cdots$	p_M	1

この表の表頭は標本空間であり、各列は互いに排反な標本空間の部分集合である。

確率変数 X	標本空間 $\mathcal{X}$				参考
	$\{1\}$	$\{2\}$	$\cdots$	$\{M\}$	$\{1\} \cup \{2\} \cup \cdots \cup \{1\}$
各部分集合の確率	p_1	p_2	$\cdots$	p_M	1

確率表

標本空間 $\mathcal{X}$ が２つの離散集合 $\mathcal{X}_1, \mathcal{X}_2$ の直積 $\mathcal{X}_1 \otimes \mathcal{X}_2$ である[1]とき、確率表には３の種類がある。

1. 同時確率表：$\mathcal{X}_1 \otimes \mathcal{X}_2$ のすべての組み合わせについて与える、個々の組み合わせの確率の表。
2. 周辺確率表：X_1 のみが観測される時の X_1 が $\mathcal{X}_1$ のそれぞれの値をとる確率、あるいは X_2 のみが観測される時の X_2 が $\mathcal{X}_2$ のそれぞれの値をとる確率の表。
3. 条件付き確率表：X_1 のとる値を $\mathcal{X}_1$ の中の一つに固定したときに X_2 が $\mathcal{X}_2$ のそれぞれの値をとる確率、あるいは X_2 のとる値を $\mathcal{X}_2$ の中の一つに固定したときに X_1 が $\mathcal{X}_1$ のそれぞれの値をとる確率の表。

[1] 二つの集合 A, B の直積 $A \otimes B$ は $\{(a, b)\,;\forall a \in A, \forall b \in B\}$ という、それぞれの集合の要素どうしのすべての組み合わせから成る集合である。組み合わせを記す順序も $a \in A$, $b \in B$ の順と定められていて、ベクトル (a, b) で表すと分かりやすい。

同時確率表

同時確率 $\Pr[X_1 = x_1, X_2 = x_2] = \Pr[(X_1, X_2) = (x_1, x_2)]$ の表のこと。確率変数の組み合わせ (X_1, X_2) が、$\mathcal{X}_1 \otimes \mathcal{X}_2$ の各要素をとる確率

$$\Pr[\{X_1 = 1\} \cap \{X_2 = 1\}] = p_{11},\ \Pr[\{X_1 = 1\} \cap \{X_2 = 2\}] = p_{12},\ \ldots,$$
$$\Pr[\{X_1 = N_1\} \cap \{X_2 = N_2\}] = p_{N_1 N_2}$$

を表に直すと

		X_2			
		1	2	$\cdots$	N_2
	1	p_{11}	p_{12}	$\cdots$	p_{1N_2}
X_1	2	p_{21}	p_{22}	$\cdots$	p_{2N_2}
	$\vdots$	$\vdots$		$\ddots$	$\vdots$
	N_1	$p_{N_1 1}$	$p_{N_1 2}$	$\cdots$	$p_{N_1 N_2}$

と 2 次元の表になる。複数の確率変数の (同時の) 組み合わせの確率を同時確率といい、この表も同時確率表という。同時確率表ですべての確率の総和が 1 となること、またすべての確率が 0 と 1 の間であることは、確率表が確率の公理を満たしている証拠である。

周辺確率表

周辺確率表：X_1 のみの周辺確率 $\Pr[X_1 = x_1]$ は $\mathcal{X}_1$ についてのみ与えられ、X_2 のみの周辺確率 $\Pr[X_2 = x_2]$ は $\mathcal{X}_2$ についてのみ与えられる。同時確率表の最右列の右に列を 1 つ加え、各行の行和をそこに記すと、X_1 の周辺確率表が得られる。同様に最下行の下に行を 1 つ加え、各列の列和をそこに記すと、X_2 の周辺確率表が得られる。

		X_2				
		1	2	$\cdots$	N_2	X_1 の周辺確率
	1	p_{11} $+$	p_{12} $+$	$\cdots +$	p_{1N_2} $=$	$p_{1\cdot}$
		$+$	$+$		$+$	
X_1	2	p_{21} $+$	p_{22} $+$	$\cdots +$	p_{2N_2} $=$	$p_{2\cdot}$
		$+$	$+$		$+$	
	$\vdots$	$\vdots$		$\ddots$	$\vdots$	$\vdots$
	N_1	$p_{N_1 1}$ $+$	$p_{N_1 2}$ $+$	$\cdots +$	$p_{N_1 N_2}$ $=$	$p_{N_1\cdot}$
		$\|$	$\|$		$\|$	
X_2 の周辺確率		$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot N_2}$	1

確率表：周辺確率

組み合わせの確率 (同時確率) が与えられていても、X_1 が観測されず X_2 のみが観測される場合、X_2 の確率分布が必要になる。前のページの X_2 の周辺確率についての計算は

$$
\begin{aligned}
\Pr[\{X_2 = j\}] &= \Pr[\{X_1 = 1, X_2 = j\} \cup \{X_1 = 2, X_2 = j\} \cup \cdots \\
&\qquad\qquad \cup \{X_1 = N_1, X_2 = j\}] \\
&= \Pr[\{X_1 = 1, X_2 = j\}] + \Pr[\{X_1 = 2, X_2 = j\}] + \cdots \\
&\qquad\qquad + \Pr[\{X_1 = N_1, X_2 = j\}]
\end{aligned}
$$

と、加法法則に基づいている。それぞれの事象が互いに排反なことを思い起こすこと。
同様に、X_2 が観測されず X_1 のみが観測される場合には、X_1 の周辺確率を導けばよい。これも加法法則を用いる。

条件付確率

互いに排反とは限らない 2 つの事象 A, B があり、いずれも標本空間の部分集合で

$$\Pr[A], \Pr[B], \Pr[A \cap B], \Pr[A \cup B],$$
$$\Pr\left[\overline{A}\right], \Pr\left[\overline{B}\right], \Pr\left[\overline{A \cap B}\right], \Pr\left[\overline{A \cup B}\right]$$

などの確率が適切に定義されているとする。$\overline{A}$ は標本空間における A の補集合 (あるいは余事象)。A が起こることが事前に分かっているときに B も起こる確率は

$$\frac{\Pr[A \cap B]}{\Pr[A]} \tag{1}$$

で求められる。これを条件付き確率といい、$\Pr[B|A]$ と書く。
久保木先生は $P(B|A)$ と書いている。

条件付確率表

条件付き確率表は、$\mathcal{X}_1$ の方を一つに固定したときの $\mathcal{X}_2$ の条件付き確率

$$\Pr\left[X_1 = x_1 \,|X_2 = x_2\right] = \Pr\left[(X_1, X_2) = (x_1, x_2)\,|X_2 = x_2\right]$$

あるいは $\mathcal{X}_2$ の方を一つに固定したときの $\mathcal{X}_1$ の条件付き確率

$$\Pr\left[X_2 = x_2 \,|X_1 = x_1\right] = \Pr\left[(X_1, X_2) = (x_1, x_2)\,|X_1 = x_1\right]$$

の表である。

確率表：条件付き確率 $(X_2|X_1)$

条件付き確率は条件を与えたときの、条件づけされていない部分の確率で、

$$\Pr\left[\text{残り}\,\middle|\,\text{条件}\right] = \frac{\text{すべての確率変数の同時確率}}{\text{条件の周辺確率}} \tag{2}$$

と憶えると良い。複数の確率変数の一部が先に観測できたときの、残りの確率変数に関する確率も条件付き確率である。
先の同時確率表と周辺確率表から、X_2 が先に観測されたときの X_1 の条件付き確率は次の計算で与えられる。

		X_2			
		1	2	$\cdots$	N_2
X_1	1	$p_{11}/p_{\cdot 1}$	$p_{12}/p_{\cdot 2}$	$\cdots$	$p_{1N_2}/p_{\cdot N_2}$
	2	$p_{21}/p_{\cdot 1}$	$p_{22}/p_{\cdot 2}$	$\cdots$	$p_{2N_2}/p_{\cdot N_2}$
	$\vdots$	$\vdots$		$\ddots$	$\vdots$
	N_1	$p_{N_1 1}/p_{\cdot 1}$	$p_{N_1 2}/p_{\cdot 2}$	$\cdots$	$p_{N_1 N_2}/p_{\cdot N_2}$

この表は列ごと (縦) の総和が 1 になる。

確率表：条件付き確率 $(X_1|X_2)$

先の同時確率表と周辺確率表から、X_2 が先に観測されたときの X_1 の条件付き確率は前のページと同様に、次の計算で与えられる。

		X_2			
		1	2	$\cdots$	N
	1	$p_{11}/p_{1\cdot}$	$p_{12}/p_{1\cdot}$	$\cdots$	$p_{1N_2}/p_{1\cdot}$
X_1	2	$p_{21}/p_{2\cdot}$	$p_{22}/p_{2\cdot}$	$\cdots$	$p_{2N_2}/p_{2\cdot}$
	$\vdots$	$\vdots$		$\ddots$	$\vdots$
	N_1	$p_{N_11}/p_{N_1\cdot}$	$p_{N_12}/p_{N_1\cdot}$	$\cdots$	$p_{N_1N_2}/p_{N_1\cdot}$

この表は行ごと (横) の総和が 1 になる。

確率表：周辺確率

降水確率は常に、周辺確率が報じられていることに注意すると良い。

		12時から15時 降水なし	 降水あり	周辺確率
15時から18時	降水あり	?	?	10%
	降水なし	?	?	90%
	周辺確率	0%	100%	100%

ベイズの定理

条件付き確率は条件と結果の確率関係の表現である。次は、条件付確率に関する有用な定理である。

定理

ベイズの定理標本空間が互いに排反な部分集合 $A_1, A_2, \ldots, A_n$ に分割されている。このとき任意の事象 B に対して、

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{l=1}^{n} P(A_l)P(B|A_l)} \tag{3}$$

が成り立つ。

この定理は「条件が与えられたときの各種の結果が発生する確率」から「結果が与えられたときに条件がどれであったかを考える確率」を与え、条件と結果の関係の反転をもたらす。確率表で言えば $P(B|A_k)$ の表から、$P(A_k|B)$ の表を得る計算に用いる。

宿題１ー１：ベイズの定理

ある病気の感染率は 0.1% である。この病気の診断方法は、感染している患者を診断した時には 90% の確率で「感染あり」との結果を得る。また感染していない患者を診断した時には 10% の確率で「感染あり」という結果を得る。

1. 病気の感染の有無に関する周辺確率表を記せ。
2. 患者の感染の有無が判明しているときの、診断結果の「感染あり」「感染なし」の条件付き確率表を記せ。
3. 「診断結果が感染ありであった」という結果を知らされた。この患者が実際に感染している確率をベイズの定理を用いて求めよ。

宿題１－２：ベイズの定理

ある病気の感染率は 0.1% である。この病気の診断方法 A は、感染している患者を診断した時の正答率が 80%、感染していない患者を診断したときの正答率は 60% である。なお正答率は、患者の真の状態を正しく当てる確率とする。

1. 病気の感染の有無に関する周辺確率表を記せ。
2. 患者の感染の有無が分かっているという条件の下で、方法 A の診断結果が「感染あり」、「感染なし」それぞれとなる条件付き確率表を記せ。
3. 「診断結果が感染ありであった」という結果を知らされた。この患者が実際に感染している確率をベイズの定理を用いて求めよ。
4. 別の診断方法 B は、感染している患者を診断した時の正答率が 70%、感染していない患者を診断したときの正答率も 70% である。診断方法 B で「診断結果が感染ありであった」という結果を知らされた患者が、実際に感染している確率をベイズの定理を用いて求めよ。

演習問題２：累積分布関数と期待値の計算

累積分布関数が

$$F(x) = \begin{cases} 0 & x < 0 \\ 1/4 & x = 0 \\ 1/4 + x/2 & 0 < x < 1 \\ 1 & x \geq 1 \end{cases} \tag{4}$$

と与えられている確率分布がある。この分布につき、以下の問いに順に答えよ。

1. この確率分布の標本空間を記せ。
2. 累積分布関数のグラフを描け。(グラフの横軸の範囲は-1 から 2 までにとり、不連続点は、端点を含む場合に ○、端点を含まない場合には ● で示すこと)
3. この確率分布に従う確率変数を X とするとき、X の関数 $h(X)$ の期待値を求める式を導出せよ。
4. 次の X の関数の期待値を求めよ。(a) $h(X) = X$。 (b) $h(X) = X^2$。 (c) $h(X) = (X - \mu)^2$。 (d) $h(X) = \exp(tX)$ 。ただし、t は任意の実数とする。

演習問題 2 のヒント

1. 確率分布を与えるのは、確率や期待値を計算して、分布の特徴を調べたり、将来の予測のために分布の中心 (平均) やばらつきの大きさ (標準偏差) を調べるため。
2. 確率分布の表現は主に確率表（標本空間が順序の定められていない離散集合の場合)、確率関数（標本空間が順序の定められた離散集合の場合)、累積分布関数（標本空間が順序の定められた集合の場合)、そして確率密度関数（標本空間がユークリッド空間もしくはそのアフィン部分空間の場合）などにより定式化される。
3. どの表現を与えたときにも確率変数の期待値の計算式の定式化をまず考える。
4. これは確率分布の累積分布関数が与えられている問題であることを認識し、2.2 節と 2.3 節を読み返す。
5. 期待値の定式化さえできれば、あとは計算問題となる。

期待値の例

確率分布に関する計算するのは、それがたとえどのように書いてあったとしても、期待値の計算を指示している。

- 確率分布の平均は確率変数の期待値

$$\mu = \mathrm{E}[X]$$

- 分散も期待値

$$\sigma^2 = \mathrm{E}\left[(X-\mu)^2\right]$$

- 分散は確率変数の二乗の期待値と確率変数の期待値の 2 乗の差

$$\sigma^2 = \mathrm{E}\left[(X-\mu)^2\right] = \mathrm{E}\left[X^2\right] - \mu^2$$

期待値の例 (続き)

- 確率も期待値

$$\Pr[X \leq x] = \mathrm{E}[I(X \leq x)]$$

- モーメント母関数も期待値

$$M_X(t) = \mathrm{E}[\exp(tX)] \tag{5}$$

- 原点モーメントは期待値のテイラー展開の係数

期待値の計算

期待値を計算するときには、次のように状況を整理すると良い。

- 何を確率変数 X とするか、を決める。
- 何の期待値を計算するか、を見分ける。

$$X,\ X^2,\ aX+b,\ (X-\mu)^2,\ (X-a)^2,\ \exp(tX),\ \ldots \tag{6}$$

- 標本空間を確認する。

$$\{x; a \le x \le b, x \in \mathcal{R}\},\ \{x; a \le x \le b, x \in \mathcal{Z}\},\ \{\mathrm{True}, \mathrm{False}\},\ \ldots \tag{7}$$

- どんな分布に関する期待値を計算するか、を識別する。

$$p(x) = ?,\ f(x) = ?,\ \ldots \tag{8}$$

- 確率変数の定め方と確率分布の表現に、他の可能性がないか検討する。

期待値の計算 (続き)

- 期待値の式を準備する。(式を立てる、という)

$$\mathrm{E}\left[h\left(X\right)\right] = \begin{cases} \sum_{x=a}^{b} h\left(x\right) p\left(x\right) \\ \int_a^b h\left(x\right) f\left(x\right) dx \\ \cdots \end{cases} \tag{9}$$

- 立てた式に確信が持てたら、計算を実行する。(計算する、という)

確率関数

確率分布の確率関数 $p(x)$ が与えられたら、期待値は確率関数の値が正になる k の範囲での $p(k)$ による $h(k)$ の重み付き平均を総和で求める。

$$
\begin{aligned}
\sum_{k=-\infty}^{\infty} h(k)\,p(k) &= \sum_{k:p(k)>0} h(k)\,p(k) \\
&= h(a)\,p(a) + h(a+1)\,p(a+1) + \cdots + h(b)\,p(b)
\end{aligned}
$$

確率密度関数

確率分布の確率密度関数 $f(x)$ が与えられたら期待値は確率関数の値が正になる x の範囲での $f(x)$ による $h(x)$ の重み付き平均を定積分で求める。

$$\int_{-\infty}^{\infty} h(x) f(x)\, dx = \int_{x:f(x)>0} h(x) f(x)\, dx$$

識別関数

識別関数 I(条件文) は

$$I(\text{条件文}) = \begin{cases} 1 & \text{条件文が真 (TRUE)} \\ 0 & \text{条件文が偽 (FALSE)} \end{cases} \tag{10}$$

は条件文が真 (TRUE) ならば 1、偽 (FALSE) ならば 0 を返す関数である。

累積分布関数

確率分布の累積分布関数 $F(x)$ が与えられたら、まず連続かつ微分可能か調べる。

- $F(x)$ が標本空間のすべての点で微分可能なら、微分して確率密度関数を求める。

$$f(x) = \frac{d}{dx} F(x) \tag{11}$$

- $F(x)$ が不連続な点を一つでも見つけたら、少し見つめてみる。
  1. 不連続な点以外のすべての点で傾きが 0 なら、その分布は離散分布。不連続点の場所を x_k、高さを p_k と置くと、その累積分布関数は

$$F(x) = \sum_{k=1}^{\infty} p_k I(x \geq x_k) \tag{12}$$

     と表せる。このとき

$$p(x_k) = p_k = \lim_{\delta \to +0} \{F(x_k + \delta) - F(x_k - \delta)\} \tag{13}$$

     である。
  2. 不連続な点以外に、正の傾きを持つ点があるときは離散分布と連続分布が混ざっている。（続く）

累積分布関数

不連続な点以外に正の傾きを持つ点や範囲があるとき、標本空間を不連続な点、傾きが 0 の範囲、傾きが正の範囲に分けていく。

1. 不連続点を小さい順に並べる。

$$x_0 < x_1 < x_2 < \cdots < x_n \tag{14}$$

   x_0 は標本空間の最小値、x_n は最大値となることが多い。希に n が ∞ のこともある。

2. 不連続点 x_k での差の高さを p_k 、隣り合った不連続点の間の区間 $x_{k-1} < x < x_k$ での累積分布関数の微分を $f_k(x)$ と置く。この確率分布に関する $h(X)$ の期待値は、次の式で計算できる。

$$\mathrm{E}\left[h(X)\right] = \sum_k p_k h(x_k) + \sum_k \int_{x_{k-1}}^{x_k} h(x) f_k(x)\, dx \tag{15}$$

宿題２：期待値の計算

1. 標本空間が $\{1, 2, 3\}$ の確率変数 X の確率関数が
$$p(1) = 1/3,\ p(2) = 1/6,\ p(3) = 1/2 \tag{16}$$
と与えられている。(a) まず累積分布関数 $F(x)$ を描け。 (b) 次に X の平均と分散を求めよ。
2. 練習問題２が一回に 0.5 の費用がかかる投資のリターン（回収額）を表しているとする。この投資の平均リターンを求めよ。またこの投資のリスク（標準偏差）を求めよ。
3. 標本空間が $\{x; 0 \leq x \leq 2\}$ の確率変数 X の確率密度関数が
$$f(x) = \left\{ \begin{array}{ll} x & 0 \leq x < 1 \\ -x + 2 & 1 \leq x \leq 2 \end{array} \right\} \tag{17}$$
と与えられている。(a) この確率分布の累積分布関数を描け。 (b) 確率変数 X の平均と分散を求めよ。

演習問題３：確率不等式

1. 非負の確率変数 X の期待値を μ とすると、X に関するマルコフの不等式は

$$\Pr[X \geq a] \leq \frac{\mu}{a} \tag{18}$$

となる。ある国の選挙の投票率の期待値が 0.4 であることが事前に分かっているとき、投票率が 0.60 以上になる確率の上限をこの不等式を用いて与えよ。

2. 確率変数 X の期待値を μ、分散を σ^2 とすると、X に関するチェビシェフの不等式は

$$\Pr[|X - \mu| \geq a] \leq \frac{\sigma^2}{a^2} \tag{19}$$

となる。ある国の選挙の投票率の期待値が 0.4、分散が 0.01 であることが事前に分かっているとき、投票率が 0.60 以上になる確率の上限をこの不等式を用いて与えよ。

確率不等式

確率変数 X が、標本空間 $\mathcal{X}$ の任意の部分集合 A に含まれる確率が

$$0 \leq \Pr[X \in A] \leq 1 \tag{20}$$

を必ず満たすのは確率の公理だが、他にも確率論の進展の成果として、確率分布あるいは確率変数が満たすべき条件が得られている。
中でもマルコフの不等式とチェビシェフの不等式は、唐突に確率の評価が必要になったときに持ち出せる、単純だが非常に有用な確率に関する不等式である。

マルコフの不等式とチェビシェフの不等式

マルコフの不等式

$$\Pr\left[X \geq a\right] \leq \frac{\mu}{a} \tag{21}$$

は、非負の確率分布の期待値 μ のみを知っている状況で、確率の範囲を限定できる。
一方、チェビシェフの不等式

$$\Pr\left[|X - \mu| \geq a\right] \leq \frac{\sigma^2}{a^2} \tag{22}$$

は、確率分布の期待値 μ と分散 σ^2 を知っているいる状況で、確率の範囲を限定できる。

確率不等式について

こういう問題は問題文から μ、σ^2、a、a^2 などの数値を正しく読み取る必要がある。
マルコフの不等式については、不等式の方向を逆にする誤りと、割り算の誤りぐらいをたまに見かけるが、概ね驚きのない回答であることが多い。
チェビシェフの不等式については、 a の値の取り間違えが散見される。
　チェビシェフの不等式に記されている事象は

$$|X - \mu| \geq a$$

である。この絶対値記号を理解できていない回答はよく見かける。さらに $-\mu$ という数式を無視した、まるでチェビシェフの不等式を見てマルコフの不等式に見えているような計算もよく見かける。

特にチェビシェフの不等式について

上の不等式は絶対値の中が正の場合と負の場合に分けられ、

$$mu - X \leq a, \; X - \mu \geq a$$

の二つがある。それぞれを X のみの辺が現れるように整理すると、

$$mu - a \leq X, \; X \geq \mu + a$$

である。これらの範囲にある確率が σ^2/a^2 よりも小さい、とチェビシェフの不等式は示している。

宿題３ー１：確率不等式

1. 確率表

X	1	2	$\cdots$	N	総和
確率	p_1	p_2	$\cdots$	p_N	1

と期待値 μ と分散 σ^2 が与えられている。教科書の証明を参考に

$$\begin{aligned}\sigma^2 &= \mathrm{E}\left[(X-\mu)^2\right] \\ &= \sum_{k=1}^{N} (k-\mu)^2 p_k \\ &= \cdots \end{aligned} \tag{23}$$

から始め、不等式

$$\sigma^2 \geq a^2 \mathrm{Pr}\left[|X-\mu| \geq a\right] \tag{24}$$

を示して、チェビシェフの不等式を証明せよ。

宿題３－２：確率不等式

1. マルコフの不等式を用いて、平均年収が 500 万円の世代の中で、年収が 5000 万円以上の人の割合は最大でどれぐらいとなるかを求めよ。
2. マルコフの不等式を用いて、1 企業あたりの平均従業員数が 2000 人の国では、10 万人以上の従業員を抱える企業の割合が最大でどれぐらいとなるかを求めよ。
3. チェビシェフの不等式を用いて、年収の平均が 500 万円、標準偏差が 200 万円の世代の中で、年収が 5000 万円以上の人の割合は最大でどれぐらいとなるかを求めよ。
4. チェビシェフの不等式を用いて、1 企業あたりの従業員数の平均が 500 人、標準偏差が 200 人の国で、10 万人以上の従業員を抱える企業の割合が最大でどれぐらいとなるかを求めよ。

宿題の提出について

提出場所は西５号館３階の「総合情報学専攻事務室」前のポストで、「K演習」というコーナーを見つけて、投函せよ。

提出期限は　　月　　日（　）　　時までとする。

なおこの演習では、提出された宿題は返却しない。その代わりに解答例を、翌日より http://bit.ly/jinlu-teach にて公開する。