おまけの解説

（ア）

■そもそも確率変数とは何ぞや？

　統計学の確率論において，起こりうることがらに割り当てている値を取る変数……とこれで分かる人は少ないと思うので，サイコロやコインで考えてみる。

<例　一般的なサイコロの出る目の確率分布>

さいころの出る目（確率変数 X）	1	2	3	4	5	6
確率	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

　この場合，確率変数の値（＝さいころの出る目）をと X おくと次のように表すことができる。右側のカッコの中は X がとる値の範囲であり，この例では「確率変数が 1 から 6 までの整数の値を取る」ことを表す。上の表のように，確率変数のとる値にその値をとる確率を対応させたものを，この確率変数の確率分布または単に分布という。

$$P(X)=\frac{1}{6} \quad (X=1,2,3,4,5,6)$$

サイコロを投げて 4 の目が出る事象の確率は $\frac{1}{6}$ であることは，

$$P(X=3)=\frac{1}{6}，\text{または} P(3)=\frac{1}{6}$$

のように書くことができる。

　世の中にはなぜか 1～3 の出目が出ないサイコロ（大槻ハンチョウの四五六賽）も存在するから，その場合は，

$$P(X)=\frac{1}{3} \quad (X=4,5,6)$$

と書くことができる。

＜例　一般的なコインの出る目の確率分布＞

コインの裏表（確率変数 X）	裏(=0)	表(=1)
確率	$\frac{1}{2}$	$\frac{1}{2}$

$$P(X)=\frac{1}{2}\quad (X=0,1)$$

サイコロは出る目の値をそのまま確率変数とすることができますが，コインのように事象に数字が無い場合でも，それぞれ事象に数値を設定することで確率変数がとる値とすることができる。上の表では，表が出たら 1，裏が出たら 0 としている。別に 1 じゃなくても，200 とか 500 億とかでも良いが，基本は計算しやすい，理解しやすい値にする。

今回の問題では，このコインと同じように「**晴れ**の場合は 1，**晴れ**以外の場合は 0 の値をとる」確率変数を X と定義しているだけである。

■期待値とは何ぞや？

一般に，確率変数 X が次の表の確率分布に従うとき

$$x_1p_1+x_2p_2+\cdots+x_np_n$$

を確率変数の平均または期待値といい，$E(X)$で表す。

X(確率変数)	x_1	x_2	$\cdots$	x_n	計
P(確率)	p_1	p_2	$\cdots$	p_n	1

こんな抽象的な説明で分かるはずもないので，またサイコロで考える。

さいころの出る目（確率変数 X）	1	2	3	4	5	6	計
確率 P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

サイコロを振ったら平均でどのくらい大きい出目がでるかを知りたい，それを知ることができるのが期待値である。先ほどの公式に当てはめて計算すると，

$$E(X)=1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6}=3.5$$

となる。よく桃鉄のプレイ動画を観ていると「サイコロを 1 個振ると期待値 3.5，サイコロを 2 個振ると期待値 7，……，サイコロを 8 個振ると 28」なんてコメントをよく見かけると思うが，その正体はこれである。ちなみにサイコロ 2 個（急行カード，プロペラカード）の場合の期待値は，

さいころの出る目（確率変数 X）	2	3	4	5	6	7	8	9	10	11	12	計
確率 P	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$	1

$$E(X)=2\times\frac{1}{36}+3\times\frac{1}{18}+4\times\frac{1}{9}+5\times\frac{1}{9}+6\times\frac{5}{36}+7\times\frac{1}{6}+8\times\frac{5}{36}+9\times\frac{1}{9}$$

$$+10\times\frac{1}{12}\times 11\times\frac{1}{18}+12\times\frac{1}{36}=7$$

となる。

　ということで，今回の問題の場合も，公式を覚えていても良いし，サイコロで期待値の計算方法を身体に染み付けているはずなので，

X	0	1	計
確率	$1-p$	p	1

この確率変数 X の平均（期待値）を m とすると

$$0\times(1-p)+1\times p=p$$

となり，$m=$ $\boxed{\text{ア}⓪\,p}$ となる。

（イ）

■正規分布に従うとは？

　中心極限定理（母集団がどんな確率分布であっても繰り返し実験を行うなら，標準化された正規分布に収束するという定理）の話で，**高校生はこれを丸暗記**している。そうは言っても，統計や数学なんてただの道具だから，大学生以上でも丸暗記している人間の方が多そうではある。

　「標本平均は母平均 m の周りで、母分散 σ^2 を標本の大きさ n で割った分散の大きさ σ^2/n でブレる」とでも覚えておけば良いんじゃないだろうか。みんながみんなが数学やるわけじゃないんだから……て，誰かが言ってた。

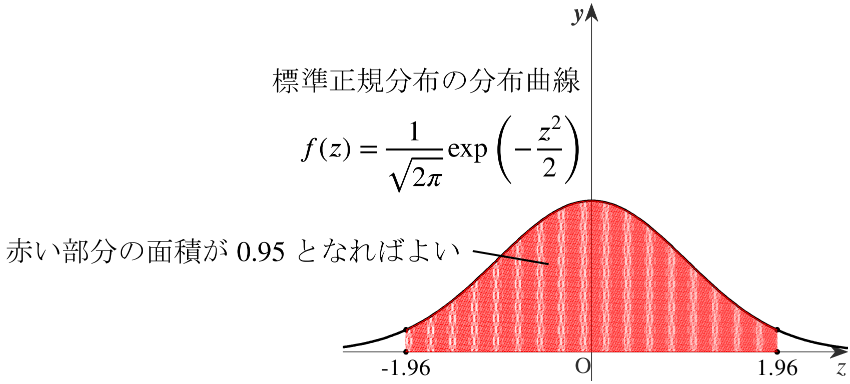
（オ）

信頼度 95%で「$\overline{X} - a \leq m \leq \overline{X} + a$」となるような $a(>0)$の値を求める。すなわち，$P\left(\overline{X} - a \leq m \leq \overline{X} + a\right) = 0.95$ となるaを求める。先ほど言った通り，標本の大きさnが大きい時，$\overline{X}$は近似的に正規分布$N\left(m, \dfrac{\sigma^2}{n}\right)$に従う。よって$\overline{X}$を標準化した$Z = \dfrac{\overline{X} - m}{\dfrac{\sigma}{\sqrt{n}}}$は標準正規分布$N(0, 1)$に従う。

$$P\left(\bar{X}-a \leq m \leq \bar{X}+a\right)=P\left(m-a \leq \bar{X} \leq m+a\right)$$

$$=P\left(-a\times\frac{\sqrt{n}}{\sigma} \leq Z \leq a\times\frac{\sqrt{n}}{\sigma}\right)$$

$$=P\left(0 \leq Z \leq a\times\frac{\sqrt{n}}{\sigma}\right)\times 2=0.95$$



正規分布表より，$a\times\frac{\sqrt{n}}{\sigma}=1.96$ となるので，$a=1.96\times\frac{\sigma}{\sqrt{n}}$

母平均 m に対する信頼度 95%の信頼区間は,

$\left[\bar{X}-1.96\times\frac{\sigma}{\sqrt{n}}, \bar{X}+1.96\times\frac{\sigma}{\sqrt{n}}\right]$ となる。これを覚えている高校生や大学生が多いはず，というか今の私もそう。