数理統計学試験問題解答と講評

平均点は 59 点でした。以下解説です。

1. $x \geq 0$ とする.

$$P(X^2 \leq x) = P(-\sqrt{x} \leq X \leq \sqrt{x})$$
$$= 2\int_0^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \ (t = \sqrt{u} \text{ と変換する})$$
$$= \int_0^x \sqrt{\frac{2}{\pi}} e^{-\frac{u}{2}} \frac{1}{2\sqrt{u}} du.$$

また $P(X^2 \leq 0) = 0$ である. したがって X^2 の密度関数は $f_{X^2}(x) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} 1_{(0,\infty)}(x)$. 次に $2X+1$ の密度関数を求める. これは平均 1, 分散 $V(2X+1) = 2^2 V(X) = 4$ の正規分布に従うので、$f_{2X+1}(x) = \frac{1}{\sqrt{8\pi}} e^{-\frac{(x-1)^2}{8}}$.
(注) f_{X^2} の密度関数は $x \leq 0$ で 0 ですが、これを注意していない人が多かった。$2X+1$ の方はこれがやはり正規分布に従うことから、平均、分散さえ求めれば、密度関数の形もわかることに注意する.

2. X, Y は独立なパラメータ 1 の指数分布に従うので、2 次元確率変数 (X, Y) の密度関数は $e^{-x-y} 1_{[0,\infty)}(x) 1_{[0,\infty)}(y)$ となる. したがって,

$$E[\sin(X+Y)] = \iint_{x\geq 0, y\geq 0} \sin(x+y) e^{-x-y} dxdy$$
$$= \iint_{x\geq 0, y\geq 0} (\cos x \sin y + \sin x \cos y) e^{-x} e^{-y} dxdy$$
$$= 2\int_0^\infty e^{-x} \sin x dx \int_0^\infty e^{-y} \cos y dy.$$

部分積分を行ない, $\int_0^\infty e^{-x} \sin x dx = \int_0^\infty e^{-x} \cos x dx = 1/2$ がわかるから $E[\sin(X+Y)] = 1/2$.
(注) 2 次元確率変数 (X, Y) の分布が密度関数 $f(x, y)$ をもつとき, ($\varphi(X, Y)$ が可積分ならば)

$$E[\varphi(X, Y)] = \iint_{\mathbb{R}^2} \varphi(x, y) f(x, y) dxdy$$

となるということを使っています.

3.

(1) $E[\bar{X}_n] = \frac{\sum_{i=1}^n E[X_i]}{n} = \frac{10\times n}{n} = 10$. $V[\bar{X}_n] = \frac{1}{n^2}E[\sum_{i=1}(X_i-10)^2] = \frac{1}{n^2}4n = \frac{4}{n}$.
(2)

$$P\left(|\bar{X}_n - 10| \le \varepsilon\right) = P\left(\left|\frac{\bar{X}_n-10}{\sqrt{\frac{4}{n}}}\right| \le \frac{\varepsilon}{\sqrt{\frac{4}{n}}}\right) = P\left(\left|\frac{\bar{X}_n-10}{\sqrt{\frac{4}{n}}}\right| \le \frac{\sqrt{n}}{2}\varepsilon\right)$$
$$= 2\int_0^{\frac{\sqrt{n}}{2}\varepsilon} \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx.$$

ここで $\frac{\bar{X}_n-10}{\sqrt{\frac{4}{n}}}$ は (1) の結果から $\bar{X}_n$ を正規化した確率変数であり，したがって標準正規分布に従うことを用いた．よって $a(n) = \frac{\sqrt{n}}{2}\varepsilon$ である．
(3) 正規分布表より標準正規分布に従う確率変数 T に対して $P(|T| \le 1.96) = 0.95$ なので $\frac{\sqrt{n}}{2} \times 0.04 \ge 1.96$ であればよい．したがって $n \ge 98^2 = 9604$.

(注) (1) の分散の計算で $1/n^2$ が出てくるのを忘れている人がいた。(2) の計算で $\sqrt{n}$ となるべきところが n となったり，$\frac{\sqrt{n}}{2}\varepsilon$ が $\frac{2}{\sqrt{n}}\varepsilon$ と計算間違いして正しい答えに到達していない答案があった．また、98^2 の計算を間違っている人もいた．落ち着いて計算しましょう．

4. サイコロの目が等確率で出る場合を考える．720 回サイコロを投げて，1 の目が出る回数 S は二項分布 $B(720, 1/6)$ に従う．その平均は $720 \times 1/6 = 120$, 分散は $720 \times \frac{1}{6} \times \frac{5}{6} = 100$ である．中心極限定理によりその分布は平均 120, 分散 100 の正規分布で近似できると考えられる．従って $\frac{S-120}{\sqrt{100}} = \frac{S-120}{10}$ の分布は標準正規分布で近似できると考えられる．よって (T を標準正規分布に従う確率変数とすると)

$$P(S \ge 140) = P\left(\frac{S-120}{10} \ge 2\right) = P(T \ge 2) = \frac{1}{2} - P(0 \le T \le 2).$$

正規分布表より $P(0 \le T \le 2) = 0.4773$ なので $P(S \ge 140) = 0.0227$.

5. 標本標準偏差が 10 となっているが、1600 は大きいので大数の法則により，これを標準偏差の近似値と考えてよいことに注意する．母平均を m とする．母分布に従う独立確率変数 X_i およびその標本平均 $\bar{X}_n = \sum_{i=1}^n X_i/n$ を考えると中心極限定理より $\frac{\bar{X}_{1600}-m}{10} \times \sqrt{1600}$ は標準正規分布で近似できる．よって

$$P\left(\left|\frac{\bar{X}_{1600}-m}{10} \times 40\right| \le 1.96\right) = 0.95.$$

従って信頼度 95 %の信頼区間は標本調査の値 $\bar{x}_{1600} = 60$ を用い，不等式 $\left|\frac{\bar{x}_{1600}-m}{10} \times 40\right| \le 1.96$ を解いて $60 - \frac{10}{40} \times 1.96 \le m \le 60 + \frac{10}{40} \times 1.96$. よって $[59.51, 60.49]$.

(注) $\frac{\bar{X}_{1600}-m}{10} \times \sqrt{1600}$ の $\sqrt{1600}$ を忘れている人がいた．