1 Use the axioms of probability to show the following two statements.

(a) (10 pts) If $E \subset F$, then $P(E) \leq P(F)$.

(b) (10 pts) The probability that exactly one of the events E or F occurs equals $P(E) + P(F) - 2P(E \cap F)$.

Solution. (a) Let F $= E \cup (F \setminus E)$. Since they are disjoint, $P(F) = P(E) + P(F \setminus E)$ **(+5 points)**. This is greater than or equal to $P(E)$ because $P(F \setminus E)$ is nonnegative **(+5 points)**.

- If elements of the sets are enumerated, you will get only +1 point.
- If only one of the two axioms is mentioned, there will be -2 points.
- Without no mentions about axioms, just +5 points.

(b) $P(E \Delta F) = P((E \setminus F) \cup (F \setminus E))$. By mutual exclusiveness of sets **(+5 points)**, we have $P((E \setminus F) \cup (F \setminus E)) = P(E \setminus F) + P(F \setminus E) = P(E) + P(F) - P(E \cap F)$ **(+5 points)**.

- Without no mentions about axioms, just +5 points.
- A solution with a figure is not rigorous. Just +1 point.

2 A genetic disease occurs with probability $0 < p < 1$. A test has been developed to detect the presence of the disease, but it is not perfect. The test has a false positive rate of 1% (meaning that 1% of people without the disease will test positive) and a false negative rate of 5% (meaning that 5% of people with the disease will test negative).

(a) (5 pts) If a person tests positive for the disease, what is the probability that he or she actually has the disease?

(b) (5 pts) If a person tests negative for the disease, what is the probability that he or she actually does not have the disease?

(c) (5 pts) Explain which test result, either positive or negative, is more trustworthy (in terms of probabilities of being accurate), when $p = 1/94$.

Solution. (a)

$$\frac{P(\text{positive}|\text{true})}{P(\text{positive}|\text{true}) + P(\text{positive}|\text{false})} = \frac{0.95p}{0.95p + 0.01(1-p)} \qquad \textbf{(+5 points)}$$

(b)

$$\frac{P(\text{negative}|\text{negative})}{P(\text{negative}|\text{negative}) + P(\text{negative}|\text{true})} = \frac{0.99(1=p)}{0.99(1-p) + 0.05p} \qquad \textbf{(+5 points)}$$

(c) For $p = \frac{1}{94}$, we can conclude that (b) is more reliable. **(+5 points)**

- There are no partial points.
- You get +5 points only when your solutions to (a) and (b) are correct, both.

3 Let N be a Poisson random variable representing the number of event having mean λ, and suppose that each of these events is independently classified as being one of the types 1 and 2 with respective probabilities p and $1-p$. Prove that the number of type 1 and 2 events are independent Poisson random variables with respective means λp and $\lambda(1-p)$.

Solution. Let $X_1 + X_2 = N$ where X_1 (resp. X_2) is the number of type 1 events (resp. type 2). Then the joint density of (X_1, X_2) is

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2) &= P(x_1, x_2, N = x_1 + x_2) \\ &= P(x_1, x_2 | n) P(N = n) \\ &= \frac{e^{-\lambda} \lambda^{x_1 + x_2}}{(x_1 + x_2)!} \binom{n}{x_1} p^{x_1} (1-p)^{x_2} \end{aligned}$$

Marginalizing the last one with respect to x_1 and x_2, respectively, we can point out that the last one is written as the product of two Poisson densities with λp and $\lambda(1-p)$ **(+10 points)**. That is,

$$\begin{aligned} P(X_1 = x_1) = \sum_{x_2 \geq 0} P(X_1 = x_1, X_2 = x_2) &= \frac{e^{-\lambda} \lambda^{x_1} p^{x_1}}{x_1!} \sum_{x_2 \geq 0} \frac{(\lambda(1-p))^{x_2}}{x_2!} \\ &= \frac{e^{-\lambda} \lambda^{x_1} p^{x_1}}{x_1!} e^{\lambda(1-p)} = \frac{e^{-\lambda p} (\lambda p)^{x_1}}{x_1!}, \end{aligned}$$

which is Poisson(λp) and X_2 can be calculated in a similar way. Therefore, they are independent **(+10 points)**.

- Independence follows from the joint density. You get +10 points only if you derive the joint density.
- Some students utilized an approximation argument (or Poisson process). It seems to be inappropriate in this case, but such proofs may be valid if you would provide more advanced results of probability theory **(+5 points)**.

4 Suppose that a random variable U has a uniform distribution over $[-1,1]$, and a random variable N has a normal distribution with parameters $(0,1)$. Assume that the random variables U and N are independent.

(a) (10 pts) Are U and UN independent? Justify your answer.

(b) (5 pts) Compute $\mathrm{Cov}(U, UN)$.

Solution. (a) Suppose that U and UN are independent. Then we must have

$$P(UN > 1, U > \frac{1}{2}) = P(UN > 1)P(U > \frac{1}{2})$$

Since we have

$$\begin{aligned} P(UN > 1, U > \frac{1}{2}) &= \int_{\frac{1}{2}}^{1} \frac{1}{2}\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dydx \\ &= 2\sqrt{2\pi} \int_{\frac{1}{2}}^{1} e^{-\frac{y^2}{2}} dydx \end{aligned}$$

and

$$\begin{aligned} P(UN > 1)P(U > \frac{1}{2}) &= \int_{-1}^{1} \frac{1}{2}\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dydx \times \frac{1}{4} \\ &= \frac{1}{4\sqrt{2\pi}} \int_{0}^{1} e^{-\frac{y^2}{2}} dydx \end{aligned}$$

However,

$$\int_{\frac{1}{2}}^{1} e^{-\frac{y^2}{2}} dydx \neq \frac{1}{2}\int_{0}^{1} e^{-\frac{y^2}{2}} dydx$$

(+9 points)

Therefore, U and UN are not independent. **(+1 point)**

Remark 1. (1) If you use wrong statement, I deducted partial points.

(2) For example, if you use $P(UN = x|U = y)$, I deducted 5 points. For continuous random variable X, $P(X = x)$ for constant x is zero. Hence the conditional probability cannot be well-defined.

(3) Also, if you use the statement that $f_{UN}(u,n)f_U(u) \neq f_{UN,U}(u,n)$ from independency, I deducted 2 points. This is slightly wrong statement. The correct equation is $f_{UN}(t)f_U(u) \neq f_{UN,U}(t,n)$ and you need to calculate $P(UN \leq t)$ to find density function of UN.

(4) When only answer is correct, (they are dependent), then you will get 1 point

(b) Since U^2 and N are independent, we have

$$\begin{aligned} Cov(U, UN) &= E[U^2N] - E[U]E[UN] \\ &= E[U^2]E[N] - E[U]E[UN] \\ &= 0 \end{aligned}$$

The last equality holds since $E[U] = 0, E[N] = 0$. **(+5 points)**

Remark 2. (1) If you use statement that U^2N is symmetric without proof, I deducted 2 points.

5 Suppose X and Y are independent discrete random variables with the following probability mass functions:

$$P(X=1)=P(X=-1)=1/2 \quad \text{and} \quad P(Y=1)=P(Y=2)=1/2.$$

(a) (5 pts) Are XY and Y independent? Justify your answer.

(b) (5 pts) Compute $\mathrm{Cov}(XY, Y)$.

(c) (5 pts) Are X^2 and X independent? Justify your answer.

(d) (5 pts) Compute $\mathrm{Cov}(X^2, X)$.

Solution. (a) They are not independent since

$$P(XY=1, Y=1) = P(X=1, Y=1) = \frac{1}{4} \neq \frac{1}{8} = P(XY=1)P(Y=1) \quad \textbf{(+5 points)}$$

(b) By the definition of expectation and covariance, we have

$$\begin{aligned} Cov(XY,Y) &= E[XY^2] - E[XY]E[Y] \\ &= (-4\times\frac{1}{4}+4\times\frac{1}{4}-1\times\frac{1}{4}+1\times\frac{1}{4}) - (-2\times\frac{1}{4}+2\times\frac{1}{4}-1\times\frac{1}{4}+1\times\frac{1}{4})(1\times\frac{1}{2}+2\times\frac{1}{2}) \\ &= 0 \quad \textbf{(+5 points)} \end{aligned}$$

(c) We have

$$P(X^2=1, X=1) = P(X=1) = \frac{1}{2} = 1\times\frac{1}{2} = P(X^2=1)P(X=1)$$

and

$$P(X^2=1, X=-1) = P(X=-1) = \frac{1}{2} = 1\times\frac{1}{2} = P(X^2=1)P(X=-1)$$

Hence, X^2 and X are independent. **(+5 points)**

(d) Since X^2 and X are independent, the covariance is zero. i.e., $Cov(X^2, X) = 0$ **(+5 points)**

6 Consider a set $\{1,\dots,n\}$. Suppose that its permutation π is chosen at random among all possible choices. Let N be the number of integer values i in $\{1,\dots,n\}$ such that the ith number of the permutation is i itself. For example, if $n=5$ and the permutation π is given by $\pi(1)=1$, $\pi(2)=3$, $\pi(3)=5$, $\pi(4)=4$, and $\pi(5)=2$, then $N=2$. Compute $\mathbb{E}[N]$ and $\mathrm{Var}(N)$.

Solution. Let X_i $(1\le i\in\mathbb{N}\le n)$ be a random variable, taking value 1 if $\pi(i)=i$, or taking value 0 otherwise($\pi(i)\neq i$). In other words, X_i is a indicator random variable of the event $\{\pi=i\}$. Then

$$N=\sum_{i=1}^{n}X_i\textbf{(+5 points)}.$$

$$\mathbb{E}[X_i]=P(\{X_i=1\})=\frac{(n-1)!}{n!}=\frac{1}{n}\textbf{(+1 point)},$$

$$\mathbb{E}[N]=\sum_{i=1}^{n}\mathbb{E}[X_i]=n\cdot\frac{1}{n}=1\textbf{(+1 point)}.$$

$$\mathbb{E}[X_i^2]=\mathbb{E}[X_i]=1\textbf{(+1 point)},$$

$$i\neq j\Rightarrow\mathbb{E}[X_iX_j]=P(\{X_iX_j=1\})=\frac{(n-2)!}{n!}=\frac{1}{n(n-1)}\textbf{(+1 point)},$$

$$\begin{aligned}\mathbf{Var}(N)=\mathbb{E}[N^2]-(\mathbb{E}[N])^2&=(\sum_{i=1}^{n}\mathbb{E}[X_i^2]+\sum_{1\le i,j\le n,i\neq j}\mathbb{E}[X_iX_j])-1^2\\&=(n\cdot\frac{1}{n}+n(n-1)\cdot\frac{1}{n(n-1)})-1=(1+1)-1=1\textbf{(+1 point)}.\end{aligned}$$

[Another solution]: Let d_n be the number of derangement of $\{1,\dots,n\}$. In other words, d_n is the number of permutations of $\{1,\dots,n\}$ with no fixed point. Then for each $0\le m\in\mathbb{N}\le n$,

$$P(\{N=m\})=\frac{\binom{n}{m}\cdot d_{n-m}}{n!}\textbf{(+5 points)}.$$

We have

$$n!=\sum_{m=0}^{n}\binom{n}{m}\cdot d_{n-m}\textbf{(+2 points)}.$$

$$\mathbb{E}[N]=\sum_{m=0}^{n}\frac{m\binom{n}{m}d_{n-m}}{n!}=\sum_{m=1}^{n}\frac{\binom{n-1}{m-1}d_{(n-1)-(m-1)}}{(n-1)!}=\frac{(n-1)!}{(n-1)!}=1\textbf{(+1 point)}.$$

$$\mathbb{E}[N(N-1)]=\sum_{m=0}^{n}\frac{m(m-1)\binom{n}{m}d_{n-m}}{n!}=\sum_{m=2}^{n}\frac{\binom{n-2}{m-2}d_{(n-2)-(m-2)}}{(n-2)!}=\frac{(n-2)!}{(n-2)!}=1\textbf{(+1 point)},$$

$$\mathbf{Var}(N)=\mathbb{E}[N(N-1)]+\mathbb{E}[N]-(\mathbb{E}[N])^2=1+1-1^2=1\textbf{(+1 point)}.$$

Note that $P(\{N=m\})=\frac{\binom{n}{m}\cdot d_{n-m}}{n!}\neq\frac{\binom{n}{m}(n-m)!}{n!}-\frac{\binom{n}{m+1}(n-(m+1))!}{n!}$ for $1\le m\le n-1$ since $P(\{N\ge m\})\neq\frac{\binom{n}{m}(n-m)!}{n!}$ since $|\{\pi:N\ge m\}|\neq\binom{n}{m}(n-m)!$ since $\binom{n}{m}(n-m)!$ counts each permutation making $N\ge m$ several times, not exactly one time.

7 Suppose that buses arrive at a given bus stop in accordance with a Poisson process with rate three per hour.

(a) (5 pts) What is the probability there will be at least two buses from 1:00 PM to 1:30 PM?

(b) (5 pts) What are the mean and variance of the waiting time of a person at a bus stop if he or she just missed a bus?

(c) (5 pts) Assuming that the event in part (a) occurs, what is the probability that there will be at least three buses from 1:00 PM to 2:00 PM?

Solution. The counting process $\{N(t)\}_{0\le t\in\mathbb{R}}$ counting the number of arrivals of buses with hour as the unit of the time index is a Poisson process with rate $\lambda = 3$.

(a) The desired probability is $P(\{N(\frac{1}{2}) \ge 2\})$**(+2 points)**.

$$P(\{N(\frac{1}{2}) \ge 2\}) = 1 - P(\{N(\frac{1}{2}) = 0\}) - P(\{N(\frac{1}{2}) = 1\})\textbf{(+1 point)}.$$

$N(t) \sim \text{Poisson}(t\lambda)$, $N(\frac{1}{2}) \sim \text{Poisson}(\frac{3}{2})$, so

$$P(\{N(\frac{1}{2}) = 0\}) = e^{-\frac{3}{2}}\textbf{(+1 point)},$$

$$P(\{N(\frac{1}{2}) = 1\}) = \frac{3}{2}e^{-\frac{3}{2}}\textbf{(+1 point)}.$$

(b) Let X be the waiting time of a person at a bus stop if he or she just missed a bus. Then X is the interarrival time of Poisson process with rate $\lambda = 3$, so X has exponential distribution with rate parameter 3 with its probability density function $f_X(x) = 1_{\{x\ge 0\}}3e^{-3x}$**(+3 points)**. Then

$$\mathbb{E}[X] = \frac{1}{3}\textbf{(+1 point)},$$

$$\text{Var}(X) = \frac{1}{3^2} = \frac{1}{9}\textbf{(+1 point)}.$$

(c) The desired probability is $P(N(1) \ge 3|N(\frac{1}{2}) \ge 2)$**(+1 point)**.

$$P(N(1) \ge 3|N(\frac{1}{2}) \ge 2) = \frac{P(\{N(\frac{1}{2}) \ge 2, N(1) \ge 3\})}{P(\{N(\frac{1}{2}) \ge 2\})}\textbf{(+1 point)}.$$

The value of $P(\{N(\frac{1}{2}) \ge 2\})$ was found in part (a).

$$P(\{N(\frac{1}{2}) \ge 2, N(1) \ge 3\}) = P(\{N(\frac{1}{2}) \ge 3\}) + P(\{N(\frac{1}{2}) = N[0,\frac{1}{2}] = 2, N\left(\frac{1}{2},1\right] \ge 1\})\textbf{(+1 point)}.$$

$$P(\{N(\frac{1}{2}) \ge 3\}) = 1 - \sum_{i=0}^{2} P(\{N(\frac{1}{2}) = i\}) = 1 - (1 + \frac{3}{2} + \frac{(\frac{3}{2})^2}{2!})e^{-\frac{3}{2}} = 1 - \frac{29}{8}e^{-\frac{3}{2}}\textbf{(+1 point)}.$$

$$\begin{aligned} P(\{N(\frac{1}{2}) = N[0,\frac{1}{2}] = 2, N\left(\frac{1}{2},1\right] \ge 1\}) &= P(\{N(\frac{1}{2}) = N[0,\frac{1}{2}] = 2\})P(\{N\left(\frac{1}{2},1\right] \ge 1\}) \\ &= (\frac{(\frac{3}{2})^2}{2!}e^{-\frac{3}{2}})(1 - e^{-\frac{3}{2}}) = \frac{9}{8}e^{-\frac{3}{2}}(1 - e^{-\frac{3}{2}})\textbf{(+1 point)}. \end{aligned}$$

8 Let X, Y, Z be a sample from a normal distribution with mean μ and variance σ^2. Construct a random variable by using all X, Y, Z (*e.g.*, $(X+Y+Z)/\sigma^2$ or $XY(Z-\mu)$) that has the following distributions.

(a) (3 pts) Chi-square distribution with 1 degrees of freedom

(b) (3 pts) Chi-square distribution with 2 degrees of freedom

(c) (3 pts) Chi-square distribution with 3 degrees of freedom

(d) (3 pts) t distribution with 2 degrees of freedom

(e) (3 pts) F distribution with 2 and 1 degrees of freedom

Solution. (a) For any choice of a, b, c with $abc \neq 0$,

$$\left(\frac{aX + bY + cZ - (a+b+c)\mu}{\sigma\sqrt{a^2+b^2+c^2}}\right)^2$$

will be an answer. Especially,

$$\left(\frac{X+Y+Z-3\mu}{\sigma\sqrt{3}}\right)^2 = \frac{3(\overline{X}-\mu)^2}{\sigma^2}$$

is an answer.

(b) For any choice of a, b with $ab \neq 0$,

$$\left(\frac{aX + bY - (a+b)\mu}{\sigma\sqrt{a^2+b^2}}\right)^2 + \left(\frac{Z-\mu}{\sigma}\right)^2,$$
$$\left(\frac{aY + bZ - (a+b)\mu}{\sigma\sqrt{a^2+b^2}}\right)^2 + \left(\frac{X-\mu}{\sigma}\right)^2,$$
$$\left(\frac{aZ + bX - (a+b)\mu}{\sigma\sqrt{a^2+b^2}}\right)^2 + \left(\frac{Y-\mu}{\sigma}\right)^2$$

will be an answer. Moreover,

$$\frac{2S^2}{\sigma^2} = \frac{(X-\overline{X})^2 + (Y-\overline{X})^2 + (Z-\overline{X})^2}{\sigma^2}$$

is also be an answer.

(c)

$$\left(\frac{X-\mu}{\sigma}\right)^2 + \left(\frac{Y-\mu}{\sigma}\right)^2 + \left(\frac{Z-\mu}{\sigma}\right)^2$$

is an answer. Also, since $\overline{X}$ and S are independent, you may choose

$$\frac{3(\overline{X}-\mu)^2 + 2S^2}{\sigma^2}.$$

(d)

$$\frac{\frac{X-\mu}{\sigma}}{\sqrt{\frac{1}{2}\left(\left(\frac{Y-\mu}{\sigma}\right)^2 + \left(\frac{Z-\mu}{\sigma}\right)^2\right)}} = \frac{\sqrt{2}(X-\mu)}{\sqrt{(Y-\mu)^2 + (Z-\mu)^2}},$$

$$\frac{\frac{Y-\mu}{\sigma}}{\sqrt{\frac{1}{2}\left(\left(\frac{Z-\mu}{\sigma}\right)^2+\left(\frac{X-\mu}{\sigma}\right)^2\right)}}=\frac{\sqrt{2}(Y-\mu)}{\sqrt{(Z-\mu)^2+(X-\mu)^2}},$$

$$\frac{\frac{Z-\mu}{\sigma}}{\sqrt{\frac{1}{2}\left(\left(\frac{X-\mu}{\sigma}\right)^2+\left(\frac{Y-\mu}{\sigma}\right)^2\right)}}=\frac{\sqrt{2}(Z-\mu)}{\sqrt{(X-\mu)^2+(Y-\mu)^2}}$$

are answers. Also, you may choose

$$\frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{3}}}{\sqrt{\frac{1}{2}\cdot 2S^2/\sigma^2}}=\frac{\overline{X}-\mu}{S/\sqrt{3}}.$$

(e)

$$\frac{\left(\left(\frac{Y-\mu}{\sigma}\right)^2+\left(\frac{Z-\mu}{\sigma}\right)^2\right)/2}{\left(\frac{X-\mu}{\sigma}\right)^2}=\frac{(Y-\mu)^2+(Z-\mu)^2}{2(X-\mu)^2},$$

$$\frac{\left(\left(\frac{Z-\mu}{\sigma}\right)^2+\left(\frac{X-\mu}{\sigma}\right)^2\right)/2}{\left(\frac{Y-\mu}{\sigma}\right)^2}=\frac{(Z-\mu)^2+(X-\mu)^2}{2(Y-\mu)^2},$$

$$\frac{\left(\left(\frac{X-\mu}{\sigma}\right)^2+\left(\frac{Y-\mu}{\sigma}\right)^2\right)/2}{\left(\frac{Z-\mu}{\sigma}\right)^2}=\frac{(X-\mu)^2+(Y-\mu)^2}{2(Z-\mu)^2},$$

are answers. Also, you may choose

$$\frac{S^2/\sigma^2}{\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{3}}\right)^2}=\frac{S^2}{3(\overline{X}-\mu)^2}.$$

- Without shifting, in other words, without $-\mu$: **(-1 point)**.
- Wrong scaling : **(-1 point)**.
- Not using all of X, Y, Z : **(-1 point)**.
- With independence issue : No points at all.

9 Find $P\{|X| \leq 1\}$, when X has the following moment generating function.

(a) (5 pts) $\phi(t) = \frac{1}{6}e^{-2t} + \frac{1}{3}e^{-t} + \frac{1}{4}e^{t} + \frac{1}{4}e^{2t}$

(b) (10 pts) $\phi(t) = (1 - p + pe^t)^n$ for some $p \in (0,1)$ and a positive integer n.

Solution. Note that if $\phi(t) = \sum_i p_i e^{x_i t}$, then it is the momentum generating function of distribution where $P\{X = x_i\} = p_i$.

(a) $P\{|X| \leq 1\} = P\{X = -1\} + P\{X = 1\} = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$.

(b) $\phi(t) = \sum_{k=0}^{n} \binom{n}{k}(1-p)^{n-k}p^k e^{kt}$. Thus,

$$P\{|X| \leq 1\} = P\{X = 0\} + P\{X = 1\} = \binom{n}{0}(1-p)^n + \binom{n}{1}(1-p)^{n-1}p = (1-p)^n + np(1-p)^{n-1}.$$

- **If you try to use the central limit theorem or to integrate the momentum generating function, you will never get a point even you wrote something correct.**
- For (b), if you detect that X follows the binomial distribution, you will get **(+7 points)**.

10 (a) (10 pts) Let $X_1, X_2, \ldots, X_n$ be a sample of values from a population having mean μ and variance σ^2. By using Chebyshev's inequality, derive a lower bound of the probability that the difference between the sample mean and μ is less than a positive constant ϵ.

(b) (10 pts) We plan a survey to estimate the proportion p, $0 < p < 1$, of the population who favors a certain candidate in an upcoming election. Based on (a), how many people should we survey, regardless of the value p, so that our guess (sample mean) has no less than a 0.95 probability of being within 0.02 of the true population proportion p?

Solution. (a) The Chebyshev inequality is

$$\mathbf{P}(|X - \mathbf{E}X| > \epsilon) \leq \mathbf{E}|X - \mathbf{E}X|^2/\epsilon^2.\textbf{(+3 points)}$$

Since $\mathbf{E}\overline{X} = \mu$ and $\mathrm{Var}(\overline{X}) = \sigma^2/n$, we have

$$\mathbf{P}(|\overline{X} - \mu| \leq \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon^2}.\textbf{(+7 points)}$$

(b) $\epsilon = 0.02$ and

$$\frac{\sigma^2}{n\epsilon^2} \leq 0.05,$$

so we need to specify σ^2. Let X be the number of response that prefer a certain candidate. Then we can regard X as a binomial random variable. Let p be the true ratio, then X/n is our estimated ratio. Note that $\mathrm{Var}(X/n) = np(1-p)/n^2 = p(1-p)/n = \sigma^2/n$. **(+3 points)** Also, $1 = p + (1-p) \geq 2\sqrt{p(1-p)}$. **(+4 points)** Therefore

$$n \geq \frac{1}{4 \cdot 0.05 \cdot 0.02^2} \geq \frac{p(1-p)}{0.05 \cdot 0.02^2} \qquad \forall p \in [0, 1].$$

Thus, $n \geq 12500$ guarantees the requirement. **(+3 points)**

11 A real-valued function $\varphi(x)$ is called convex if and only if for any $-\infty < x_1 < \infty$, $-\infty < x_2 < \infty$, and $0 \le t \le 1$,

$$\varphi(tx_1 + (1-t)x_2) \le t\varphi(x_1) + (1-t)\varphi(x_2).$$

(a) (5 pts) Let $X \sim \text{Bernouill}(p)$, where $0 \le p \le 1$. Prove that

$$\varphi(\mathbb{E}[X]) \le \mathbb{E}[\varphi(X)].$$

(b) (10 pts) Let X be a discrete random variable such that

$$P(X = x_i) = p_i, \quad i = 1, 2, \dots, 10, \quad \text{and} \quad \sum_{i=1}^{n} p_i = 1,$$

where $n \ge 1$ is an integer. Prove that

$$\varphi(\mathbb{E}[X]) \le \mathbb{E}[\varphi(X)].$$

(Hint: You can apply the mathematical induction for n to prove the statement.)

Solution. (a) Consider $X = x_1$ with probability p and $X = x_2$ with probability $1-p$. Then

$$\phi(\mathbf{E}X) = \phi(px_1 + (1-p)x_2) \le p\phi(x_1) + (1-p)\phi(x_2) = \mathbf{E}\phi(X).\text{(+5 points)}$$

Note that the first equality follows from the definition, the first inequality follows from the convexity, and the second equality again follows from the definition. Now put $x_1 = 1$ and $x_2 = 0$.
(b) We'll use the induction. The base case ($n = 2$) is already shown in (a). Assume the statement holds for $n = k-1$. Now suppose $p_1 + \cdots + p_k = 1$, $p_i \ge 0$, and $\mathbf{P}(X = x_i) = p_i$. Define X' such that

$$\mathbf{P}\left(X' = \frac{\sum_{i=1}^{k-1} p_i x_i}{1-p_k}\right) = 1 - p_k, \qquad \mathbf{P}(X' = x_k) = p_k.$$

By the base case (applied to X'),

$$\phi(\mathbf{E}X') = \phi\left(\sum_{i=1}^{k} p_i x_i\right) \le (1-p_k)\phi\left(\frac{\sum_{i=1}^{k-1} p_i x_i}{1-p_k}\right) + p_k\phi(x_k).\text{(+5 points)}$$

Note that

$$\sum_{i=1}^{k-1} \frac{p_i}{1-p_k} = 1.$$

By the induction hypothesis,

$$(1-p_k)\phi\left(\frac{\sum_{i=1}^{k-1} p_i x_i}{1-p_k}\right) \le \sum_{i=1}^{k-1} p_i \phi(x_i),$$

which completes the proof. **(+5 points)**