1. (*18pts*)

(a) (*4pts*) Because

$$E(\overline{X}_{\boldsymbol{c}}) = E\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i E(X_i) = \left(\sum_{i=1}^{n} c_i\right)\mu,$$

The estimator $\overline{X}_{\boldsymbol{c}}$ is unbiased if and only if

$$\sum_{i=1}^{n} c_i = 1.$$

(b) (*6pts*) The variance of $\overline{X}_{\boldsymbol{c}}$ is

$$\begin{aligned} Var(\overline{X}_{\boldsymbol{c}}) &= Var\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i^2 Var(X_i) + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} c_i c_j Cov(X_i, X_j) \\ &= \sum_{i=1}^{n} c_i^2\sigma^2 + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} c_i c_j\left(-\frac{\sigma^2}{N-1}\right) = \left(\sum_{i=1}^{n} c_i^2 - \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{c_i c_j}{N-1}\right)\sigma^2 \end{aligned}$$

(c) (*8pts*) We need to minimizes $\sum_{i=1}^{n} c_i^2 - \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{c_i c_j}{N-1}$ subject to the constraint $\sum_{i=1}^{n} c_i = 1$. We can introduce a Lagrange multiplier λ and define

$$g(c_1,\ldots,c_n,\lambda) = \left(\sum_{i=1}^{n} c_i^2 - \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{c_i c_j}{N-1}\right) + \lambda\left(\sum_{i=1}^{n} c_i - 1\right).$$

By setting

$$\begin{aligned} \frac{\partial g}{\partial c_i} &= 2c_i - \sum_{\substack{j=1\\ j\neq i}}^{n}\frac{c_j}{N-1} + \lambda = 2c_i + \frac{c_i}{N-1} - \frac{1}{N-1}\left(\sum_{j=1}^{n} c_j\right) + \lambda \\ &= \frac{2N-1}{N-1}c_i - \frac{1}{N-1} + \lambda = 0, \quad \text{for } i = 1,\ldots,n \qquad \text{(I)} \\ \frac{\partial g}{\partial \lambda} &= \sum_{i=1}^{n} c_i - 1 = 0 \qquad \text{(II)} \end{aligned}$$

we can obtain $c_i = \frac{N-1}{2N-1}(\frac{1}{N-1} - \lambda)$, $i = 1,\ldots,n$, from (I). That is, we have $c_1 = \cdots = c_n$. Hence, from (II), we know that c_i must be $1/n$, for $i = 1,\ldots,n$.

2. (*30pts*)

(a) (*4pts*) Notice that the population variance σ_f^2 equals $Var(Y_1) = Var[f(X_1)]$. Since $E(Y_1) = E[f(X_1)] = \mu_f$, and the pdf of X_1 is 1/2 when $-1 \le X_1 \le 1$ and 0, otherwise, we have

$$\sigma_f^2 = E(Y_1^2) - [E(Y_1)]^2 = \int_{-1}^{1} f(x)^2 \times \frac{1}{2}\, dx - \mu_f^2. \qquad \text{(III)}$$

(b) (*6pts*) Because $\hat{I}(f) = \overline{Y}$ is the sample mean and $Y_1, \ldots, Y_n$ is a with-replacement simple random sample, the standard error of $\hat{I}(f)$ is

$$\sigma_{\hat{I}(f)} = \sqrt{Var(\overline{Y})} = \sqrt{Var(Y_1)/n} = \sigma_f/\sqrt{n}.$$

Since the population variance σ_f^2 can be estimated by $\hat{\sigma}_f^2 = \frac{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{n-1}$, the standard error $\sigma_{\hat{I}(f)}$ can be estimated by

$$\hat{\sigma}_{\hat{I}(f)} = \frac{\hat{\sigma}_f}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sqrt{\frac{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{n-1}} = \frac{1}{\sqrt{n}}\sqrt{\frac{\sum_{i=1}^{n}[f(X_i)-\hat{I}(f)]^2}{n-1}}.$$

(c) (*4pts*) By the central limit theorem and the law of large number, we have

$$\frac{\hat{I}(f) - I(f)}{\hat{\sigma}_{\hat{I}(f)}} \overset{D}{\approx} N(0,1)$$

when the sample size n is large. The resulting $100(1-\alpha)\%$ confidence interval of $I(f)$ is

$$\hat{I}(f) \pm z(\alpha/2) \times \hat{\sigma}_{\hat{I}(f)},$$

where $z(\alpha/2)$ is the $1-(\alpha/2)$ quantile of $N(0,1)$.

(d) (*12pts*) Let $l = 1$ and $l = 2$ represent the strata $[-1, 0)$ and $[0, 1]$, respectively. It is clear that for the two strata, their stratum fractions W_l's are 1/2. Denote the subpopulation mean of the lth stratum by $\mu_{f,l}$.

- Case (i). When $f(x) = x^2$, from (III), we can obtain that
  - the population mean and variance are

$$\mu_f = \int_{-1}^{1} x^2 \times \frac{1}{2}\, dx = \frac{1}{3} \quad \text{and} \quad \sigma_f^2 = \int_{-1}^{1} \frac{x^4}{2} dx - \left(\frac{1}{3}\right)^2 = \frac{4}{45},$$

  - for the first stratum $[-1, 0)$,

$$\mu_{f,1} = \int_{-1}^{0} x^2 \times 1\, dx = \frac{1}{3} \quad \text{and} \quad \sigma_{f,1}^2 = \int_{-1}^{0} \frac{x^4}{1} dx - \left(\frac{1}{3}\right)^2 = \frac{4}{45},$$

  - for the first stratum $[0, 1]$,

$$\mu_{f,2} = \int_{0}^{1} x^2 \times 1\, dx = \frac{1}{3} \quad \text{and} \quad \sigma_{f,2}^2 = \int_{0}^{1} \frac{x^4}{1} dx - \left(\frac{1}{3}\right)^2 = \frac{4}{45}.$$

Therefore, $Var(\overline{Y}) = \frac{4}{45n}$ and $Var(\overline{Y}_{\mathbb{S}}) = \frac{1}{n}(\frac{1}{2} \times \frac{4}{45} + \frac{1}{2} \times \frac{4}{45}) = \frac{4}{45n}$. Because the relative efficiency is 1, this stratified random sampling cannot produce a more accurate estimator in this case.

- Case (ii). When $f(x) = x(x-1)$, from problem (III), we can obtain that
  - the population mean and variance are

$$\mu_f = \int_{-1}^{1} x(x-1) \times \frac{1}{2}\, dx = \frac{1}{3} \quad \text{and} \quad \sigma_f^2 = \int_{-1}^{1} \frac{x^2(x-1)^2}{2} dx - \left(\frac{1}{3}\right)^2 = \frac{19}{45},$$

  - for the first stratum $[-1, 0)$,

$$\mu_{f,1} = \int_{-1}^{0} x(x-1) \times 1\, dx = \frac{5}{6} \quad \text{and} \quad \sigma_{f,1}^2 = \int_{-1}^{0} \frac{x^2(x-1)^2}{1} dx - \left(\frac{5}{6}\right)^2 = \frac{61}{180},$$

  - for the first stratum $[0, 1]$,

$$\mu_{f,2} = \int_{0}^{1} x(x-1) \times 1\, dx = -\frac{1}{6} \quad \text{and} \quad \sigma_{f,2}^2 = \int_{0}^{1} \frac{x^2(x-1)^2}{1} dx - \left(-\frac{1}{6}\right)^2 = \frac{1}{180}.$$

  Therefore, $Var(\overline{Y}) = \frac{19}{45n}$ and $Var(\overline{Y}_{\mathbb{S}}) = \frac{1}{n}(\frac{1}{2} \times \frac{61}{180} + \frac{1}{2} \times \frac{1}{180}) = \frac{31}{180n}$. Because the relative efficiency is $\frac{19 \times 180}{45 \times 31} \approx 2.4516 > 1$, this stratified random sampling can produce a much more accurate estimator than the simple random sample in this case.

(e) (*4pts*) In the optimal allocation, the subsample sizes n_1 and n_2 for the two strata are proportional to $W_1\sigma_{f,1}$ and $W_2\sigma_{f,2}$. Therefore,

$$n_1 : n_2 = \left(\frac{1}{2} \times \sqrt{\frac{61}{180}}\right) : \left(\frac{1}{2} \times \sqrt{\frac{1}{180}}\right) = \sqrt{61} : 1 \approx 7.81 : 1,$$

and $n_1 = \frac{7.81}{8.81} \times n$, $n_2 = \frac{1}{8.81} \times n$.

3. (*14pts*)

(a) (*5pts*) When α and σ are fixed, the length of the confidence interval is proportional to $\sqrt{\frac{1}{n} + \frac{1}{m}}$. The solution of minimizing $\frac{1}{n} + \frac{1}{m}$ subject to the constraint $n + m = N$ gives the confidence interval with the shortest length. By substituting $m = N - n$ into $\frac{1}{n} + \frac{1}{m}$, we have

$$\frac{1}{n} + \frac{1}{N-n} = \frac{N}{n(N-n)} = -\frac{N}{[n-(N/2)]^2 - (N^2/4)},$$

which is minimized when $n = N/2$. The answer is $n = m = N/2$.

(b) (*3pts*) When α is fixed, the power β_Δ increases along with the increase of $\left|\frac{\mu_x - \mu_Y}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}\right|$. Because μ_X, μ_Y, σ are fixed parameters, the solution of minimizing $\frac{1}{n} + \frac{1}{m}$ subject to the constraint $n + m = N$ also gives the most powerful test. The answer is also $n = m = N/2$.

(c) (*6pts*) The solution of minimizing $\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$ subject to the constraint $n+m=N$ gives the confidence interval with the shortest length. By substituting $m = N - n$ into $\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$, differentiating it with respect to n, and setting it to be 0, we have

$$\frac{d}{dn}\left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{N-n}\right) = -\frac{\sigma_X^2}{n^2} + \frac{\sigma_Y^2}{(N-n)^2} = \frac{n^2\sigma_Y^2 - (N-n)^2\sigma_X^2}{n^2(N-n)^2} = 0.$$

Because

$$n^2\sigma_Y^2 - (N-n)^2\sigma_X^2 = 0 \quad \Leftrightarrow \quad \left(\frac{N}{n} - 1\right)^2 = \frac{\sigma_Y^2}{\sigma_X^2} \quad \Leftrightarrow \quad \frac{n}{N} = \frac{\sigma_X}{\sigma_X + \sigma_Y}$$

The answer is $n = \frac{\sigma_X}{\sigma_X+\sigma_Y}N$ and $m = \frac{\sigma_Y}{\sigma_X+\sigma_Y}N$. The population with large population variance should be allocated more subjects, and the optimal sample sizes should be proportional to the population standard deviations.

4. (*20pts*)

(a) (*8pts*) Using the hints, we can get

$$E(W_X) = E(U_X) + \frac{n(n+1)}{2} = n^2E(\hat{\pi}) + \frac{n(n+1)}{2} = n^2\pi + \frac{n(n+1)}{2}.$$

Because $X \sim N(0,1)$, $Y \sim N(1,1)$, and (X,Y) are independent, we know that $Y - X \sim N(1,2)$ and $\frac{(Y-X)-1}{\sqrt{2}} \sim N(0,1)$. Since $\pi = P(X > Y)$, we have

$$\pi = P(Y - X < 0) = P\left(\frac{(Y-X)-1}{\sqrt{2}} < -\frac{1}{\sqrt{2}}\right) = \Phi\left(-\frac{1}{\sqrt{2}}\right),$$

and

$$E(W_X) = n^2 \times \Phi\left(-\frac{1}{\sqrt{2}}\right) + \frac{n(n+1)}{2}.$$

(b) (*12pts*) Using the hints, we can get

$$Var(W_X) = Var(U_X) = Var\left(\sum_{i=1}^{n}\sum_{j=1}^{n} Z_{ij}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} Cov(Z_{ij}, Z_{kl}).$$

For the value of $Cov(Z_{ij}, Z_{kl})$, we need to consider the following four cases.

- If $i = k$ and $j = l$, then $Cov(Z_{ij}, Z_{kl}) = Var(Z_{ij}) = \pi(1-\pi)$. There are n^2 covariances of this case.
- If $i \neq k$ and $j \neq l$, then Z_{ij} and Z_{kl} are independent so that $Cov(Z_{ij}, Z_{kl}) = 0$.
- If $i = k$ but $j \neq l$, then

$$Cov(Z_{ij}, Z_{kl}) = E(Z_{ij}Z_{kl}) - E(Z_{ij})E(Z_{kl}) = E(Z_{ij}Z_{il}) - \pi^2.$$

Because

$$E(Z_{ij}Z_{il}) = 1 \times P(X_i > Y_j \text{ and } X_i > Y_l) + 0 \times [1 - P(X_i > Y_j \text{ and } X_i > Y_l)] = \tau,$$

we have $Cov(Z_{ij}, Z_{kl}) = \tau - \pi^2$ when $i = k$ but $j \neq l$. There are $n^2(n-1)$ covariances of this case.

- If $i \neq k$ but $j = l$, then similarly, we can get $Cov(Z_{ij}, Z_{kl}) = \tau - \pi^2$. There are $n^2(n-1)$ covariances of this case.

Therefore,

$$Var(W_X) = n^2\pi^2 + 2n^2(n-1)(\tau - \pi^2) = 2n^2(n-1)\tau - n^2(2n-3)\pi^2,$$

where $\pi = \Phi\left(-1/\sqrt{2}\right)$.

5. (*18pts*)

(a) (*2pts*) Because the joint pdf of the data $X_1, \ldots, X_n$ is

$$f(x_1, \ldots, x_n; \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = (2\pi)^{-n/2} \times (\sigma^2)^{-n/2} \times e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (x_i-\mu)^2\right]},$$

the log-likelihood function is

$$l(\mu, \sigma^2) = \log(f(x_1, \ldots, x_n; \mu, \sigma^2)) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2.$$

(b) (*10pts*) Because $\log(\Lambda) = \sup_\omega \log(\mathcal{L}) - \sup_\Omega \log(\mathcal{L}) = \sup_\omega l(\mu, \sigma^2) - \sup_\Omega l(\mu, \sigma^2)$, and

$$\begin{aligned}
\sup_\Omega l(\mu, \sigma^2) &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}_\Omega^2) - \frac{1}{2\hat{\sigma}_\Omega^2}\sum_{i=1}^n (x_i - \hat{\mu}_\Omega)^2 \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}_\Omega^2) - \frac{1}{2} \times \frac{n}{\sum_{i=1}^n (x_i - \hat{\mu}_\Omega)^2} \times \sum_{i=1}^n (x_i - \hat{\mu}_\Omega)^2 \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}_\Omega^2) - \frac{n}{2},
\end{aligned}$$

and

$$\begin{aligned}
\sup_\omega l(\mu, \sigma^2) &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}_\omega^2) - \frac{1}{2\hat{\sigma}_\omega^2}\sum_{i=1}^n (x_i - \mu_0)^2 \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}_\omega^2) - \frac{1}{2} \times \frac{n}{\sum_{i=1}^n (x_i - \mu_0)^2} \times \sum_{i=1}^n (x_i - \mu_0)^2 \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}_\omega^2) - \frac{n}{2},
\end{aligned}$$

we have

$$\log(\Lambda) = \sup_\omega l(\mu, \sigma^2) - \sup_\Omega l(\mu, \sigma^2) = -\frac{n}{2}\log(\hat{\sigma}_\omega^2) + \frac{n}{2}\log(\hat{\sigma}_\Omega^2) = -\frac{n}{2}\log\left(\frac{\hat{\sigma}_\omega^2}{\hat{\sigma}_\Omega^2}\right).$$

(c) (*6pts*) The likelihood ratio test rejects H_0 when Λ (or $\log(\Lambda)$) is small. Because

$$\begin{aligned}
\frac{\hat{\sigma}_\omega^2}{\hat{\sigma}_\Omega^2} &= \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sum_{i=1}^n (X_i - \overline{X})^2} = \frac{\sum_{i=1}^n (X_i - \overline{X} + \overline{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \overline{X})^2} \\
&= \frac{[\sum_{i=1}^n (X_i - \overline{X})^2] + n(\overline{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \overline{X})^2} = 1 + \frac{1}{n-1} \times \frac{(\overline{X} - \mu_0)^2}{[\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2]/n} \\
&= 1 + \frac{1}{n-1} \times \frac{(\overline{X} - \mu_0)^2}{s_{\overline{X}}^2} = 1 + \frac{t^2}{n-1},
\end{aligned}$$

the likelihood ratio test rejects $H_0 \Leftrightarrow \frac{\hat{\sigma}_\omega^2}{\hat{\sigma}_\Omega^2}$ is large $\Leftrightarrow t^2$ is large $\Leftrightarrow |t|$ is large.