Note. There are 5 problems in total. **To ensure consideration for partial scores, write down necessary intermediate steps.** Correct answers with inadequate or no intermediate steps **may result in zero credit**.

Some useful formula.

- Suppose that X follows a Bernoulli(p) distribution. The probability mass function (pmf) of X is
$$p(x) = p^x(1-p)^{1-x}, \quad \text{for} \quad x = 0, 1.$$
The mean of X is p, and the variance of X is $p(1-p)$.

- Suppose that X follows a binomial(n, p) distribution. The pmf of X is
$$p(x) = \binom{n}{x} p^x(1-p)^{n-x}, \quad \text{for} \quad x = 0, 1, \ldots, n.$$
The mean of X is np, and the variance of X is $np(1-p)$.

- Suppose that X follows a uniform(a, b) distribution. The probability density function (pdf) of X is
$$f(x) = \frac{1}{b-a}, \quad \text{for} \quad a < x < b,$$
and zero otherwise. The mean of X is $\frac{a+b}{2}$, and the variance of X is $\frac{(b-a)^2}{12}$.

- Suppose that X follows a normal(μ, σ^2) distribution. The pdf of X is
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \text{for} \quad -\infty < x < \infty.$$
The mean of X is μ, and the variance of X is σ^2.

---

1. For a without-replacement simple random sample $X_1, \ldots, X_n$ of size n from a population of size N with the population mean μ and the population variance σ^2, consider the following as an estimator of μ:
$$\overline{X}_{\boldsymbol{c}} = \sum_{i=1}^{n} c_i X_i,$$
where the $\boldsymbol{c} = (c_1, \ldots, c_n)$ are fixed numbers.

   (a) (*4pts*) Find a condition on the c_i's such that the estimator $\overline{X}_{\boldsymbol{c}}$ is unbiased.

(b) (*6pts*) What is the variance of $\overline{X}_{\boldsymbol{c}}$? [**Hint**. $Cov(X_i, X_j) = -\frac{\sigma^2}{N-1}$]

(c) (*8pts*) Show that the choice of c_i's that minimizes the variances of the estimator subject to the unbiased condition is $c_i = 1/n$, where $i = 1, \dots, n$. [**Hint**. Introduce a Lagrange multiplier.]

2. Suppose that we wish to calculate the integration

$$I(f) = \frac{1}{2} \int_{-1}^{1} f(x)\, dx\,,$$

where $f(x)$ is a known function. A numerical method, called the *Monte Carlo method*, works in the following way. Generate n independent *uniform* random variables $X_1, X_2, \dots, X_n$ on $[-1, 1]$, and compute the random number:

$$\hat{I}(f) = \frac{1}{n} \sum_{i=1}^{n} f(X_i).$$

Let $Y_i = f(X_i)$, $i = 1, \dots, n$. By the law of large numbers, $\hat{I}(f) = \overline{Y}$ converges in probability to

$$\mu_f = \mathrm{E}(Y) = \mathrm{E}[f(X)] = \int_{-1}^{1} f(x) \times \frac{1}{2}\, dx\; = I(f).$$

This method can be interpreted from the point of view of survey sampling by considering all the numbers in the interval $[-1, 1]$ as an "infinite population," and each $x \in [-1, 1]$ as a population member with a value $f(x)$. The population mean is $\mu_f = I(f)$, and $\hat{I}(f)$ can be interpreted as the sample mean of the with-replacement simple random sample $Y_1, \dots, Y_n$ from this population.

(a) (*4pts*) Show that the population variance σ_f^2 is

$$\int_{-1}^{1} \frac{f(x)^2}{2}\, dx - \mu_f^2.$$

(b) (*6pts*) What is the standard error of $\hat{I}(f)$? How could it be estimated?

(c) (*4pts*) How could a $100(1-\alpha)\%$ confidence interval for $I(f)$ be formed by applying the central limit theorem and the law of large number?

(d) (*12pts*) Suppose that we partition the population into two strata: $[-1, 0)$ and $[0, 1]$, and generate X_i's using a stratified random sampling with proportional allocation, i.e., $n/2$ random observations are generated from $[-1, 0)$ and the other $n/2$ random observations are generated from $[0, 1]$. Consider the following two cases:

- case (i): $f(x) = x^2$,
- case (ii): $f(x) = x(x-1)$.

For both cases, can this stratified random sampling produce a more accurate estimator than the simple random sampling? Report their relative efficiencies. [**Hint**. (i) Denote the stratified sample mean under proportional allocation by $\overline{Y}_{\mathbb{S}}$. Then,

$$Var(\overline{Y}_{\mathbb{S}}) = \frac{1}{n}\sum_{l=1}^{L} W_l \sigma_{f,l}^2,$$

where L is the number of strata, W_l is the lth stratum fraction, and $\sigma_{f,l}^2$ is the subpopulation variance of the lth stratum. (ii) For an X generated from the stratum $[-1, 0)$, $X \sim$ uniform$(-1, 0)$, and for an X generated from the stratum $[0, 1]$, $X \sim$ uniform$(0, 1)$. (iii) The relative efficiency is $Var(\overline{Y})/Var(\overline{Y}_{\mathbb{S}})$.]

(e) (*4pts*) What is the optimal allocation of n for the case (ii) in the problem (d)?

3. Two independent samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are to be compared to see if there is a difference in the population means. Assume that the observations in the two samples are normally distributed with *known* variances, i.e., X_i's $\sim$ i.i.d. $N(\mu_X, \sigma_X^2)$ and Y_j's $\sim$ i.i.d. $N(\mu_Y, \sigma_Y^2)$, where σ_X^2 and σ_Y^2 are *known*. We have shown in the lectures that when $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, the confidence interval of $\mu_X - \mu_Y$ is

$$(\overline{X} - \overline{Y}) \pm z(\alpha/2) \times \sigma\sqrt{\frac{1}{n} + \frac{1}{m}},$$

where $z(\alpha/2)$ is the $1 - (\alpha/2)$ quantile of $N(0, 1)$, and the power function of the z-test for $H_0 : \mu_X = \mu_Y$ is

$$\beta_\Delta = 1 - \Phi\left(z(\alpha/2) - \frac{\mu_X - \mu_Y}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}\right) + \Phi\left(-z(\alpha/2) - \frac{\mu_X - \mu_Y}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}\right),$$

where Φ is the cdf of $N(0, 1)$. We also have shown in the lectures that when $\sigma_X^2 \neq \sigma_Y^2$, the confidence interval of $\mu_X - \mu_Y$ is

$$(\overline{X} - \overline{Y}) \pm z(\alpha/2) \times \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}},$$

Suppose that a total of N (i.e., $n+m = N$) subjects are available for the experiment.

(a) (*5pts*) When $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, how should this total N be allocated between the two samples in order to provide the shortest confidence interval for $\mu_X - \mu_Y$? Explain how you get your answer.

(b) (*3pts*) When $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, how should this total N be allocated between the two samples in order to make the z-test of $H_0 : \mu_X = \mu_Y$ as powerful as possible? Explain how you get your answer.

(c) (*6pts*) When $\sigma_X^2 \neq \sigma_Y^2$, how should this total N be allocated between the two samples in order to provide the shortest confidence interval for $\mu_X - \mu_Y$? Explain how you get your answer.

4. Let $X_1, \ldots, X_n$ be a random sample from an $N(0,1)$ distribution and let $Y_1, \ldots, Y_n$ be an independent random sample from an $N(1,1)$ distribution.

   (a) (*8pts*) Determine the expection of the rank sum statistic W_X of the X_i's. Use Φ, the cdf of $N(0,1)$, to express your answer.

   (b) (*12pts*) Determine the variance of the rank sum statistic W_X of the X_i's. Define

   $$\tau = P(X_i > Y_j \text{ and } X_i > Y_l) = P(X_i > Y_j \text{ and } X_k > Y_j),$$

   where $j \neq l$ and $i \neq k$. Use τ to express your answer.

   [**Hint**: (i) For the rank sum statistic W_X and the Mann-Whitney test statistic U_X, we have $W_X = U_X + \frac{n(n+1)}{2}$; (ii) $U_X = n^2 \times \hat{\pi}$, where $\hat{\pi}$ is an unbiased estimator of $\pi = P(X > Y)$; (iii) U_X can be expressed as $\sum_{i=1}^{n} \sum_{j=1}^{n} Z_{ij}$, where $Z_{ij} = 1$, if $X_i > Y_j$, and 0, otherwise; (iv) $Z_{ij} \sim \text{Bernoulli}(\pi)$]

5. Suppose that $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, where both μ and σ^2 are parameters. To test the null and alternative hypotheses

   $$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_A : \mu \neq \mu_0$$

   where μ_0 is a known constant, the t-test is often used:

   $$t = \frac{\overline{X} - \mu_0}{s_{\overline{X}}},$$

   where $s_{\overline{X}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2} \Big/ \sqrt{n}$. The test rejects H_0 when $|t|$ is large, and under H_0, t follows a t distribution with degrees of freedom $n-1$. Let

   $$\Omega = H_0 \cup H_A = \{(\mu, \sigma^2) : -\infty < \mu < \infty, 0 < \sigma^2 < \infty\},$$

   and

   $$\omega = H_0 = \{(\mu_0, \sigma^2) : 0 < \sigma^2 < \infty\}.$$

   Under Ω, the MLEs of μ and σ^2 are

   $$\hat{\mu}_\Omega = \overline{X} \quad \text{and} \quad \hat{\sigma}^2_\Omega = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

   and under ω, the MLE of σ^2 is

   $$\hat{\sigma}^2_\omega = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_0)^2.$$

   Use the following steps to show that the likelihood ratio test of this H_0 is equivalent to the t test.

(a) (*2pts*) Show that the log-likelihood function under Ω is

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

(b) (*10pts*) Let $\Lambda = \frac{\sup_\omega \mathcal{L}}{\sup_\Omega \mathcal{L}}$ be the likelihood ratio test statistic, where $\mathcal{L}$ is the likelihood function. Show that

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

(c) (*6pts*) Show that the likelihood ratio test rejects H_0 if and only if $|t|$ is large.