1. The random vector $(X, Y, Z)^T$ follows a multivariate Normal distribution with mean vector $\mathbf{0} = (0, 0, 0)^T$ and covariance matrix

$$
\begin{pmatrix}
1 & \rho & 0 \\
\rho & 1 & \gamma \\
0 & \gamma & 1
\end{pmatrix}.
$$

In particular, $X$ and $Z$ are independent.

(a) Define $U = Y - Z$ and $W = Y + Z$. What are respectively the marginal distributions of $U$ and $W$?

(b) Compute $\text{Cov}(U, W)$. Are $U$ and $W$ independent? Explain your answer.

(c) Obtain the conditional distribution of $X$, given $W = Y + Z$.

2. There are $n$ balls labeled by $1, 2, \ldots, n$ and $n$ bags labeled by $1, 2, \ldots, n$.

(a) Randomly put $n$ balls in $n$ bags (allowing more than one ball in one bag), what is the probability that there is at least one ball in the bag of the same label?

(b) Randomly put $n$ balls in $n$ bags such that there is exactly one ball in each bag, what is the probability that there is no ball in the bag of the same label?

3. Let $X_1, X_2, \cdots$ be independent and identically distributed (IID) variables from uniform distribution on $(1, 2)$, and let $H_n$ denote the harmonic average of the first $n$ variables

$$H_n = \frac{n}{\sum_{i=1}^{n} X_i^{-1}}.$$

(a) Show that $H_n \xrightarrow{p} c$ as $n \to \infty$, and identify the constant $c$.

(b) Show that $\sqrt{n}(H_n - c)$ converges in distribution, and identify the limit distribution.

4. Let $X_1, \cdots, X_n$ be an IID sample of from Beta distribution $\text{Beta}(\theta, 1)$, where $\theta > 0$ is an unknown parameter.

(a) Find the maximum likelihood estimator (MLE) of $1/\theta$. 

(b) Calculate the information inequality lower bound for $1/\theta$. Does the MLE obtained in (a) achieve the inequality lower bound? Show your answer.

(c) Find an unbiased estimator of $\theta/(\theta + 1)$. Check whether the unbiased estimator achieves the information inequality variance bound.

5. Let $X_1, \ldots, X_n$ be an IID sample from $N(\mu, \sigma^2)$, where $\mu$ and $\sigma^2 > 0$ are unknown parameters.

(a) Obtain a complete and sufficient statistic for $\theta = (\mu, \sigma^2)$.

(b) Find the UMVUE of $\sigma^r$ for $r > 0$.

(c) Find the UMVUE of $\frac{\mu}{\sigma}$.

6. Let $X_1, \ldots, X_n$ be an IID sample from $N(\mu, 1)$ with an unknown $\mu$. Suppose that one forgets to record the values of $X_1, \ldots, X_n$ in a study and instead only records $Y_i = I(X_i > 0)$ for $i = 1, \ldots, n$.

(a) Find the MLE of $\mu$ based on the observed data, $Y = (Y_1, \ldots, Y_n)$.

(b) Is $\sum_{i=1}^n Y_i$ a sufficient statistic for $\mu$? Justify your answer.

(c) Is $\sum_{i=1}^n Y_i$ a complete statistic? Explain.

(d) Use the observed data $Y$ to construct a level-$\alpha$ uniformly most powerful (UMP) test for testing $H_0 : \mu \leq \mu_0$, vs $H_1 : \mu > \mu_0$.

Please describe the form of the rejection region in terms of $Y$, and simplify the expression as much as you can. You can use the normal approximation to compute the cut-off value for the rejection region.
Solutions:

1. (a) \( P(\text{desired event}) = 1 - P(\text{there is no ball in the bag of the same label}) = 1 - (n-1)^n/n^n. \)

   (b) This is the same as envelop matching problem.

   \[
P(\text{desired event}) = 1 - P(\text{at least one match}) = 1 - \left( \sum_{k=0}^{n} \left( \begin{array}{c} n \\ k \end{array} \right) \frac{(n-k)!}{n!} \right)
\]

   \[
   = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n-1} \frac{n!}{n!}
\]

2. (a) \( U \) is distributed as \( N(0, 2 - 2\gamma) \). \( W \) is distributed as \( N(0, 2 + 2\gamma) \).

   (b) \( \text{Cov}(U, W) = \text{Var}(Y) - \text{Var}(Z) = 0. \) Since they are jointly bivariate normal, they are independent.

   (c) The joint distribution of \( X \) and \( W \) is bivariate Normal distribution with mean zero and variance-covariance matrix

   \[
   \begin{pmatrix}
   1 \\
   \rho \\
   2 + 2\gamma
   \end{pmatrix}
   \]

   Hence, \( X \) conditional on \( W \) follows a Normal distribution with mean \( \frac{\rho W}{\sqrt{2+2\gamma}} \) and variance \( 1 - \rho \).

3. (a) It is easy to find that \( \frac{1}{X_1}, \frac{1}{X_2}, \ldots, \frac{1}{X_n} \) is a random sample from density \( 1/y^2, \frac{1}{2} < y < 1. \) By LLN,

   \[
   \frac{1}{n} \sum_{i=1}^{n} \frac{1}{X_i} \xrightarrow{P} \frac{\rho}{E \frac{1}{X_1}} = \ln 2,
   \]

   hence \( H_n \xrightarrow{P} \frac{1}{\ln 2} = c. \)

   (b) By CLT, \( \sqrt{n}(H_n^{-1} - \ln 2) \xrightarrow{D} N(0, \frac{1}{2} - (\ln 2)^2)). \) Then by Delta method,

   \[
   \sqrt{n}(H_n - c) \xrightarrow{D} N \left( 0, \frac{\frac{1}{2} - (\ln 2)^2}{\ln 2^4} \right).
   \]
4. Let $X_1, \ldots, X_n$ be a sample of from Beta distribution $\text{Beta}(\theta, 1)$, where $\theta > 0$.

(a) The MLE of $\theta$ is $\hat{\theta} = -\sum_{i=1}^{n} \log X_i$. By invariance principle, the MLE of $1/\theta$ is $1/\hat{\theta} = -\sum_{i=1}^{n} \log X_i/n$.

(b) Since $E(-\log X_1) = \frac{1}{\theta}$, the MLE is unbiased. The Cramer-Rao variance lower bound is $1/(n\theta^2)$. Since $\text{Var}(\log X) = \frac{1}{\theta^2}$, the MLE achieves the CR bound.

(c) $E X = \frac{\theta}{\theta + 1}$, so $\bar{X}$ is unbiased for $\frac{\theta}{\theta + 1}$. And $\text{Var}(\bar{X}) = \frac{\theta(\theta + 1)^2}{n(\theta + 2)(\theta + 1)^2}$. The CR bound for estimating $\frac{\theta}{\theta + 1}$ is $\frac{\theta^2}{n(\theta + 1)^2}$. The estimator $\bar{X}$ does not achieve the bound.

5. Let $X_1, \ldots, X_n$ be a random sample from $\text{N}(\mu, \sigma^2)$.

(a) $(\bar{X}, S^2)$ is complete and sufficient. (need to show details)

(b) Note $T = S^2 \sim \text{Gamma}(\frac{n-1}{2}, \frac{2\sigma^2}{n})$. Denote $\alpha = \frac{n-1}{2}$ and $\beta = \frac{2\sigma^2}{n}$. We have

$$E(S^2) = E[(S^2)^{\alpha/2}] = \int_0^{\infty} t^{\alpha/2} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} e^{-t/\beta} t^{\alpha-1} dt = \sigma^2 \frac{\Gamma(\frac{r+n-1}{2})}{\Gamma(\frac{r}{2})} \left[ \frac{2}{n-1} \right]^{\frac{r}{2}}.$$

Using the Rao-Blackwell, the estimator $S^2 \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n-2}{2})} \left[ \frac{n-1}{2} \right]^{\frac{r}{2}}$ is the UMVUE for $\sigma^2$.

(c) $\bar{X}$ is independent of $S^2$. Therefore

$$E \left[ \frac{\bar{X}}{S^2} \right] = E \left[ \frac{1}{S} \right] = \frac{\mu \sigma}{\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n-2}{2})} \left[ \frac{n-1}{2} \right]^{\frac{r}{2}}}.$$

The UMVUE of $\mu/\sigma$ is $\frac{\bar{X}}{S^2} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n-2}{2})} \left[ \frac{n-1}{2} \right]^{\frac{r}{2}}$.

6. (a) $Y_i$ follows $\text{Bin}(1, p)$ with $p = P(X_i > 0) = \Phi(\mu)$, where $\Phi(\cdot)$ is the CDF of $\text{N}(0, 1)$. Then $\hat{\mu}_{\text{MLE}} = \bar{Y}$ and $\hat{\mu}_{\text{MLE}} = \Phi^{-1}(\bar{Y})$.

(b) $Y_i \sim \text{Bin}(1, p)$. By the factorization theorem, $T = \sum_{i=1}^{n} Y_i$ sufficient for $p$. The conditional distribution $P(Y|T)$ is free of $p$, and free of $\mu$ as well. So $T = \sum_{i=1}^{n} Y_i$ is sufficient for $\mu$.

(c) Since $T = \sum_{i=1}^{n} Y_i \sim \text{Bin}(n, p)$, the distribution family $\{\text{Binomial}(n, p)\}$ is complete. So $T$ is a complete statistic.

(d) For any $\mu_2 > \mu_1$, we have $p_2 = \Phi(\mu_2) > p_1 = \Phi(\mu_1)$ and

$$f(y|\mu_2) = \left\{ \begin{array}{ll} p_2 - p_1 & \text{if } 1 \leq y < n \left( 1 - \frac{p_2}{p_1} \right) \\ p_1 & \text{otherwise} \end{array} \right\} \left( \frac{1 - p_2}{1 - p_1} \right)^n,$$

which is non-decreasing in $T$. Define $p_0 = \Phi(\mu_0)$. By Karlin-Rubin theorem, the size $\alpha$ test is: reject $H_0$ if $T > t_0$, where $t_0$ is chosen such that

$$P_{p_0}(T > t_0) = \alpha.$$

Using the normal approximate $T \sim \text{N}(np_0, np_0(1-p_0))$. We can get $t_0 = np_0 + z_{\alpha} \sqrt{np_0(1-p_0)}$. 

1