Instructions

- This exam has 7 pages in total, numbered 1 to 7. Make sure your exam has all the pages.
- This exam will be 2 hours and 30 minutes in length.
- This is a closed book, closed-note exam, and no electronic devices (e.g., cellphone, computer) are allowed. The last pages of the exam have some possibly useful reference material.
- For all problems, follow these instructions:
  - Show your work and give reasons.
  - For any problem that asks you to provide an algorithm, be sure to give a step-by-step algorithm. Do not explain how to implement your algorithm in Arena or Excel, but rather, you should provide an algorithm (i.e., pseudo-code) that can be implemented in any language. Also, for any random variates needed by your algorithm, be sure to explicitly provide the steps needed to generate the random variates. You may assume that you have at your disposal a good unif[0,1] random number generator.
  - For all constants or parameter values that can be computed, be sure to explicitly give their values. Be sure to explicitly define any estimators for unknown quantities.
  - In any problem for which you use an approximation based on the central limit theorem (CLT), you can use a quantile point from the normal distribution. Any hypothesis test you perform should be at level $\alpha = 0.05$. Any confidence interval (CI) you provide should be at a 95% confidence level.
1. [10 points] We collected 252 daily closing prices of Microsoft (MSFT) stock over a one-year time period. Let $P_i$ denote the closing price on the $i$th day, $i = 0, 1, \ldots, 251$, and define $Y_i = \ln(P_i/P_{i-1})$, $i = 1, 2, \ldots, 251$. We fit a normal distribution to the sample $Y_1, Y_2, \ldots, Y_{251}$ using Arena’s Input Analyzer, which gave the following output.

![Distribution Summary Image]

- **Distribution Summary**
  - Distribution: Normal
  - Expression: NORM(-0.00159, 0.0341)
  - Square Error: 0.014704

- **Chi Square Test**
  - Number of intervals = 6
  - Degrees of freedom = 3
  - Test Statistic = 17.7
  - Corresponding p-value < 0.005

- **Kolmogorov-Smirnov Test**
  - Test Statistic = 0.0723
  - Corresponding p-value = 0.14

- **Data Summary**
  - Number of Data Points = 251
  - Min Data Value = -0.125
  - Max Data Value = 0.171
  - Sample Mean = -0.00159
  - Sample Std Dev = 0.0342

- **Histogram Summary**
  - Histogram Range = -0.16 to 0.21
  - Number of Intervals = 15

Comment on the appropriateness of fitting a normal distribution to the data $Y_1, \ldots, Y_{251}$. Be sure to discuss the output from the Input Analyzer.
2. [30 points] Let $X$ be a random variable with CDF $F$, and let $\gamma = \mathbb{P}(X \in A)$ for some given set $A$. Assume that $0 < \gamma < 1$. Let $X_1, X_2, \ldots, X_n$ be IID samples from $F$.

(a) [10 points] Give an estimator $\hat{\gamma}_n$ of $\gamma$ that uses all of the $n$ samples $X_1, X_2, \ldots, X_n$. Show that your estimator $\hat{\gamma}_n$ is an unbiased estimator of $\gamma$.

(b) [10 points] Show that your estimator $\hat{\gamma}_n$ satisfies a law of large numbers and satisfies the following central limit theorem (CLT):

$$\frac{\sqrt{n}}{\beta} (\hat{\gamma}_n - \alpha) \overset{D}{\approx} N(0, 1)$$

for large $n$, for some some constants $\alpha$ and $\beta$. Explicitly specify formulas for $\alpha$ and $\beta$.

(c) [10 points] Derive an approximate 95% (two-sided) confidence interval for $\gamma$.

3. [60 points] A software company produces different software titles. Customers having problems with any of the company’s products call the technical-support center, which is open 10 hours each day. Each day, calls arrive to technical support according to a non-homogeneous Poisson process with rate function 

$$\lambda(t) = 12t - t^2 + 15,$$

where $t$ denotes the number of hours since the technical-support center opened for the day. (All times given are in hours.) The technical-support center has enough phone lines and enough staff to answer the phones so that every call received is immediately answered; i.e., no caller ever gets a busy signal or has to wait on hold. A phone call by a customer to technical support consists of two parts.

- First, the customer explains the problem he is having, and the amount of time this takes has the following density function:

$$f(x) = \begin{cases} 
40x & \text{if } 0 \leq x \leq 0.1 \\
5 - 10x & \text{if } 0.1 < x \leq 0.5 \\
0 & \text{otherwise}
\end{cases}$$

- This is followed by the technical-support person providing the solution.

  - In 70% of the cases, the technical-support person knows the answer immediately, and the amount of time it takes to explain it has a uniform[0.1, 0.3] distribution.
  
  - In the other 30% of the cases, the technical-support person needs to research the question and then gives the answer; the total time to research and provide the answer has a Weibull distribution with shape parameter $\alpha = 1/3$ and scale parameter $\beta = 1/12$.

Once the answer is provided, the caller immediately hangs up. No new calls are accepted after $t = 10$, but any call that arrived before $t = 10$ and is still in process at $t = 10$ continues until the solution is provided by the technical-support person. Assume that callers are independent of one another and that the time for a caller to explain his problem is independent of the amount of time it takes for the technical-support person to provide the answer.

For the following questions, you may assume that you have available a good uniform[0,1] random number generator. Also, if one part uses results from previous parts, you do not need to give all of the details from the previous parts.
(a) [10 points] Let $X$ be a random variable having density function $f$. Give an algorithm to generate a sample of $X$.

(b) [10 points] Define $Y$ as the amount of time the technical-support person spends providing the solution. Give an algorithm to generate a sample of $Y$.

(c) [10 points] Define $Z$ as the total amount of time a customer phone call takes. Give an algorithm to generate a sample of $Z$.

(d) [10 points] Give an algorithm to simulate the times of the phone calls that arrive during the time interval $[0, 10]$ and the total number of calls arriving during $[0, 10]$.

(e) [10 points] Define $\theta$ to be the probability that in a day, the number of calls exceeding 1 hour in length is at least 15. More precisely, let $R$ be the number of calls that exceeded 1 hour in length in a day, so $\theta = \mathbb{P}(R \geq 15)$. Using the previous parts of this problem, describe an algorithm to estimate $\theta$ using simulation. Your algorithm should also produce a confidence interval for $\theta$.

(f) [10 points] Let $M$ denote the total number of calls received during the time interval $[0, 10]$. Give an algorithm using $M$ as a control variate to estimate and construct a confidence interval for $\theta$. 


Some possibly useful facts

- The CDF $F$ of random variable $X$ is defined as $F(z) = P(X \leq z)$. If $X$ is a continuous random variable with density function $f$, then $F(z) = \int_{-\infty}^{z} f(x) \, dx$. If $X$ is a discrete random variable with probability mass function $p$ and support $\{x_1, x_2, x_3, \ldots\}$, then $F(z) = \sum_{x_i \leq z} p(x_i)$.

- The expectation or mean $\mu$ of a random variable $X$ is $\mu = \mathbb{E}[X] = \sum_{x_i} x_i p(x_i)$ when $X$ is a discrete random variable with probability mass function $p(\cdot)$ and support $\{x_1, x_2, \ldots\}$, and $\mu = \int_{-\infty}^{\infty} x f(x) \, dx$ when $X$ is a continuous random variable with density function $f$.

- The variance of a random variable $X$ is $\mathbb{V}(X) = \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - \mu^2$.

- The covariance of random variables $X$ and $Y$ with respective means $\mu_X$ and $\mu_Y$ is $\mathbb{C}(X,Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mu_X\mu_Y$.

- For IID data $X_1, X_2, \ldots, X_n$, the sample mean is $\bar{X}(n) = \frac{1}{n} \sum_{i=1}^{n} X_i$, and the sample variance is $S^2(n) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}(n))^2$.

- The normal distribution with mean $\mu$ and variance $\sigma^2 > 0$ is denoted by $\mathcal{N}(\mu, \sigma^2)$ and has density function $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. Its CDF cannot be evaluated in closed form.

- If $Z_1, Z_2, \ldots, Z_n, n \geq 1$, are IID $\mathcal{N}(0,1)$ random variables, then $X = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ has a chi-squared distribution with $n$ degrees of freedom, which is denoted by $\chi_n^2$.

- If $Z \sim \mathcal{N}(0,1)$ and $U \sim \chi_n^2$ with $Z$ and $U$ independent, then $T = Z/\sqrt{U/n}$ has a Student-$t$ distribution with $n$ degrees of freedom, which is denoted by $t_n$.

- If $X_1, X_2, \ldots, X_n$ are IID $\mathcal{N}(\mu, \sigma^2)$, then $(n-1)S^2(n)/\sigma^2 \sim \chi_{n-1}^2$ and $(\bar{X}(n) - \mu)/\sqrt{S^2(n)/n} \sim t_{n-1}$.

- The uniform distribution over the interval $[a,b]$ has density function $f(x) = 1/(b-a)$ for $a \leq x \leq b$, and $f(x) = 0$ otherwise. Its CDF is $F(x) = (x-a)/(b-a)$ for $a \leq x \leq b$, and $F(x) = 0$ otherwise. Its mean is $\mu = (a+b)/2$ and its variance is $(b-a)/12$.

- The exponential distribution with parameter $\lambda > 0$ has density function $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$, and $f(x) = 0$ for $x < 0$. Its CDF is $F(x) = 1 - e^{-\lambda x}$ for $x \geq 0$, and $F(x) = 0$ for $x < 0$. Its mean is $1/\lambda$ and its variance is $1/\lambda^2$.

- The gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ has density function $f(x) = \beta^{-\alpha} x^{\alpha-1} \exp\{-x/\beta\}/\Gamma(\alpha)$ for $x \geq 0$, and $f(x) = 0$ for $x < 0$, where $\Gamma(\alpha) = \int_{0}^{\infty} t^{\alpha-1} e^{-t} \, dt$. In general, its CDF cannot be evaluated in closed form. Its mean is $\alpha \beta$ and its variance is $\alpha \beta^2$.

- The Weibull distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ has density function $f(x) = \alpha \beta^{-\alpha} x^{\alpha-1} \exp\{-x/\beta\}$ for $x \geq 0$, and $f(x) = 0$ for $x < 0$. Its CDF is $F(x) = 1 - \exp\{-x/\beta\}$ for $x \geq 0$, and $F(x) = 0$ for $x < 0$. Its mean is $\beta/(\alpha)$ and $\alpha\beta$ and its variance is $\beta^2/(\alpha^2)$.

- If $Z \sim \mathcal{N}(\mu, \sigma^2)$, then $X = e^Z$ has a lognormal($\mu, \sigma^2$) distribution. The lognormal($\mu, \sigma^2$) distribution has density function $f(x) = 1/(\sqrt{2\pi}\sigma^2) \exp\{-\ln x - \mu)^2/(2\sigma^2)\}$ for $x \geq 0$, and $f(x) = 0$ for $x < 0$. Its CDF cannot be evaluated in closed form. Its mean is $\exp(\mu + (\sigma^2/2))$ and its variance is $e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$.

- A Poisson distribution has a given parameter $\lambda > 0$, which is also its mean, and has support $\{0,1,2,\ldots\}$. The probability mass function is $p(k) = e^{-\lambda}\lambda^k/k!$ for $k = 0,1,2,\ldots$. 

• Law of large numbers (LLN): if $X_1, X_2, \ldots, X_n$ is a sequence of IID random variables with finite mean $\mu$, then $\bar{X}(n) \approx \mu$ for large $n$.

• Central limit theorem (CLT): if $X_1, X_2, \ldots, X_n$ is a sequence of IID random variables with finite mean $\mu$ and positive, finite variance $\sigma^2$, then $\frac{\sqrt{n}(\bar{X}(n) - \mu)}{\sigma} \overset{D}{\rightarrow} \mathcal{N}(0, 1)$ for large $n$, where $\overset{D}{\rightarrow}$ means approximately equal to in distribution. Also, $\frac{\sqrt{n}}{\sqrt{n(n-1)}}(\bar{X}(n) - \mu) \overset{D}{\rightarrow} \mathcal{N}(0, 1)$ for large $n$.

• Goodness-of-fit tests: Suppose $X_1, X_2, \ldots, X_n$ is a sequence of IID random variables with finite mean $\mu$ and positive, finite variance.

• Law of large numbers (LLN): if $X_1, X_2, \ldots, X_n$ is a sequence of IID random variables with finite mean $\mu$ and positive, finite variance $\sigma^2$, then $\frac{\sqrt{n}(\bar{X}(n) - \mu)}{\sigma} \overset{D}{\rightarrow} \mathcal{N}(0, 1)$ for large $n$, where $\overset{D}{\rightarrow}$ means approximately equal to in distribution. Also, $\frac{\sqrt{n}}{\sqrt{n(n-1)}}(\bar{X}(n) - \mu) \overset{D}{\rightarrow} \mathcal{N}(0, 1)$ for large $n$.

• Confidence intervals (CIs): if $X_1, X_2, \ldots, X_n$ is a sequence of IID random variables with finite mean $\mu$ and positive, finite variance $\sigma^2$, then an approximate $100(1 - \alpha)$% confidence interval for $\mu$ is $[\bar{X}(n) \pm z_{1-\alpha/2}S(n)/\sqrt{n}]$, where $z_{1-\alpha/2}$ is the constant such that $\mathcal{P}(\mathcal{N}(0, 1) > z_{1-\alpha/2}) = \alpha/2$. For example, $\mathcal{P}(\mathcal{N}(0, 1) > 1.65) = 0.05$, $\mathcal{P}(\mathcal{N}(0, 1) > 1.96) = 0.025$, and $\mathcal{P}(\mathcal{N}(0, 1) > 2.60) = 0.005$.

• Logarithm: $\ln e = 1$, $\ln(ab) = \ln(a) + \ln(b)$, $\ln(a^b) = b \ln(a)$, and $e^{a \ln(b)} = b^a$.

• Quadratic formula: The roots of $ax^2 + bx + c = 0$ are $x = [-b \pm \sqrt{b^2 - 4ac}]/(2a)$.

• Parameter estimation: Suppose $X_1, \ldots, X_n$ are IID with distribution $F_{\theta_1, \ldots, \theta_d}$, where $\theta_1, \ldots, \theta_d$ are unknown and to be estimated.

  - MLE: If $F_{\theta_1, \ldots, \theta_d}$ is a continuous distribution with density function $f_{\theta_1, \ldots, \theta_d}$, then compute the likelihood $L(X_1, \ldots, X_n) = \prod_{i=1}^{n} f_{\theta_1, \ldots, \theta_d}(X_i)$. If $F_{\theta_1, \ldots, \theta_d}$ is a discrete distribution with probability mass function $p_{\theta_1, \ldots, \theta_d}$, then compute the likelihood $L(X_1, \ldots, X_n) = \prod_{i=1}^{n} p_{\theta_1, \ldots, \theta_d}(X_i)$. Then choose the values of $\theta_1, \ldots, \theta_d$ that maximize $L(X_1, \ldots, X_n)$, or equivalently that maximize $\ln L(X_1, \ldots, X_n)$.

  - Method-of-moments: Let $\mu_j = \mathbb{E}[X^j]$ be the $j$th (theoretical) moment of $F_{\theta_1, \ldots, \theta_d}$, and let $M_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j$ be the $j$th sample moment of the data. Then set $\mu_j = M_j$ for $j = 1, 2, \ldots, d$, and solve for $\theta_1, \ldots, \theta_d$.

• Goodness-of-fit tests: Suppose $X_1, \ldots, X_n$ are IID from some hypothesized distribution $F$.

  - Chi-squared test: Divide samples into $k$ bins. Let $O_i$ be the number of observations in the $i$th bin, and let $E_i$ be the expected number of observations in the $i$th bin. Compute test statistic $\chi^2 = \sum_{i=1}^{k} (O_i - E_i)^2/E_i$, which has a chi-squared distribution with $\nu$ degrees of freedom, where $\nu = \text{(number of bins)} - (1 + \text{number of estimated parameters})$.

  - Kolmogorov-Smirnov (K-S) test: Compute empirical CDF $\hat{F}_n(x) = (\# \text{ observations } \leq x)/n$. Then compute $D_n = \max_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)|$. Obtain the test statistic by centering and scaling $D_n$, where the centering and scaling depend on the hypothesized $F$.

• LCG: $Z_i = (aZ_{i-1} + c) \mod m$ and $U_i = Z_i/m$, for constants $a$, $c$ and $m$.

• Inverse-transform method to generate sample $X$ having CDF $F$. Generate $U \sim \text{unif}[0, 1]$, and return $X = F^{-1}(U)$.

• Convolution method: Suppose $X = Y_1 + Y_2 + \cdots + Y_n$, where $Y_1, Y_2, \ldots, Y_n$ are independent, and $Y_i \sim F_i$. We generate a sample of $X$ by generating $Y_1 \sim F_1, Y_2 \sim F_2, \ldots, Y_n \sim F_n$, where $Y_1, \ldots, Y_n$ are generated independently, and returning $X = Y_1 + \cdots + Y_n$.

• Composition method: Suppose we can write CDF $F$ as $F(x) = p_1 F_1(x) + p_2 F_2(x) + \cdots + p_d F_d(x)$ for all $x$, where $p_i > 0$ for each $i$ and $p_1 + \cdots + p_d = 1$. We then generate a sample $X \sim F$ as follows. First generate $I$ having pmf $(p_1, p_2, \ldots, p_d)$. If $I = j$, then independently generate $Y_j \sim F_j$, and return $X = Y_j$. 

6
• Acceptance-rejection (AR) method: To generate $X$ having density $f$, suppose $Y$ is a RV with density $g$ such that there exists a constant $c$ for which $f(x) \leq cg(x)$ for all $x$. Then we can generate a sample of $X$ using the AR method as follows. (i) Generate $Y$ with density $g$. (ii) Generate $U \sim \text{unif}[0, 1]$, independent of $Y$. (iii) If $cg(Y)U \leq f(Y)$, then return $X = Y$ and stop; else, reject $(Y, U)$ and return to step (i).

• Box-Muller method for generating standard normals: Generate independent $U_1, U_2 \sim \text{unif}[0, 1]$, and return $X_1 = \sqrt{-2\ln(U_1)} \cos(2\pi U_2)$ and $X_2 = \sqrt{-2\ln(U_1)} \sin(2\pi U_2)$ as IID $N(0, 1)$.

• If $Z = (Z_1, \ldots, Z_d)'$ is a vector of IID standard normals, then $AZ + \mu \sim N_d(\mu, \Sigma)$ for a constant vector $\mu$ and matrix $A$, where $\Sigma = AA'$ and prime denotes transpose.

• A Poisson process $[N(t) : t \geq 0]$ with parameter $\lambda > 0$ is a counting process, where $N(t)$ is the number of events in the time interval $[0, t]$, and $N(t)$ has a Poisson distribution with parameter $\lambda t$. The interevent times are IID exponential with mean $1/\lambda$.

• A nonhomogeneous Poisson process $[N'(t) : t \geq 0]$ with rate function $\lambda(t), t \geq 0$, can be generated by generating a (homogeneous) Poisson process $[N(t) : t \geq 0]$ with constant rate $\lambda_0 \geq \lambda(t)$ for all $t \geq 0$, and then accepting event time $V_i$ of the (homogeneous) Poisson process with probability $\lambda(V_i)/\lambda_0$, independent of all other events. For each $t$, $\mathbb{E}[N'(t)] = \int_0^t \lambda(s) \, ds$.

• Common random numbers (CRN): Suppose that $X$ and $Y$ are outputs from different simulation models, and let $\mu_x = \mathbb{E}[X]$ and $\mu_y = \mathbb{E}[Y]$. Assume that we can express $X = h_x(U_1, U_2, \ldots, U_d)$ and $Y = h_y(U_1, U_2, \ldots, U_d)$, where $h_x$ and $h_y$ are given functions, and $U_1, U_2, \ldots, U_d$ are IID $\text{unif}[0, 1]$. Then to estimate and construct a CI for $\alpha = \mu_x - \mu_y$ using CRN, generate $n$ IID replications of $D = h_x(U_1, U_2, \ldots, U_d) - h_y(U_1, U_2, \ldots, U_d)$, and form a confidence interval for $\alpha$ by computing the sample mean and sample variance of the IID replicates of $D$. A variance reduction is guaranteed when $h_x$ and $h_y$ are monotone in the same direction for each argument.

• Antithetic variates (AV): Suppose that $X$ is an output from a simulation model, and let $\alpha = \mathbb{E}[X]$. Assume that we can express $X = h_x(U_1, U_2, \ldots, U_d)$, where $h_x$ is a given function, and $U_1, U_2, \ldots, U_d$ are IID $\text{unif}[0, 1]$. Then to estimate and construct a CI for $\alpha$ using AV, we generate $n$ IID replications of $D = [h_x(U_1, U_2, \ldots, U_d) + h_x(1 - U_1, 1 - U_2, \ldots, 1 - U_d)]/2$, and form a confidence interval for $\alpha$ by computing the sample mean and sample variance of the IID replicates of $D$. A variance reduction is guaranteed (compared to $2n$ IID replications of just $X$) when $h_x$ is monotone in each argument.

• Control variates (CV): Let $(X, Y)$ be outputs from a simulation, where $\mathbb{E}[Y]$ is known. To estimate $\alpha = \mathbb{E}[X]$ using CV, generate $n$ IID replications $(X_i, Y_i)$, $i = 1, 2, \ldots, n$, of $(X, Y)$. Let $X'_i = X_i - \hat{\lambda}_n(Y_i - \mathbb{E}[Y])$, where $\hat{\lambda}_n$ is an estimate of $\text{Cov}[X, Y]/\mathbb{V}[Y]$ from the $n$ replications. Construct a CI for $\alpha$ by taking the sample mean and sample variance of the $X'_i$.

• Importance sampling (IS): Let $X$ be a random variable with density $f$, and let $\alpha = \mathbb{E}_f[h(X)]$, where $h$ is a function and $\mathbb{E}_f$ denotes expectation when $X$ has density $f$. Let $g$ be another density function such that $h(x)f(x) > 0$ implies $g(x) > 0$. Then $\alpha = \mathbb{E}_g[h(X)L(X)]$, where $L(x) = f(x)/g(x)$ is the likelihood ratio and $\mathbb{E}_g$ denotes expectation when $X$ has density $g$. To estimate and construct a CI for $\alpha$ using IS, use $g$ to generate IID samples of $h(X)L(X)$, and compute the sample mean and sample variance of the $h(X)L(X)$. 

7