18.650. Fundamentals of Statistics
Fall 2023. Problem Set 2

Due Monday, October 16

Problem 1

Let $X \sim N(1, 2.25)$. Compute the following probabilities

1. $\mathbb{P}(X > 1)$. $= \mathbb{P}(N(0, 2.25) > 0) = 1 - \Phi(0) = 0.5$
2. $\mathbb{P}(|X - 2| \leq 1) = \mathbb{P}(1 \leq N(1, 1.5^2) \leq 3) = \mathbb{P}(0 \leq N(0,1) \leq \frac{4}{3}) = \Phi(\frac{4}{3}) - \Phi(0) = 0.9082 - 0.5 = 0.4082$
3. $\mathbb{P}(|X| < 1) = \mathbb{P}(-1 \leq N(1, 1.5^2) \leq 1) = \mathbb{P}(-\frac{4}{3} \leq N(0,1) \leq 0) = \Phi(0) - \Phi(-4/3) = \Phi(0) - (1 - \Phi(\frac{4}{3})) = \Phi(\frac{4}{3}) - .5 = 0.4082$
4. $\mathbb{P}(X^2 - 2X - 1 > 0) = 1 - \mathbb{P}(-\sqrt{2} \leq N(0, 1.5^2) \leq \sqrt{2}) = 2(1 - \Phi(0.9428)) \approx 0.3472.$ The following rearrangement can be useful: $X^2 - 2X - 1 = (X-1)^2 - 2$.

Problem 2 Let

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \right)$$

Compute the following quantities

1. $\mathbb{V}[X] = 1.$
2. $\mathbb{E}[Y^2 + X] = \mathbb{V}(Y) + (\mathbb{E}Y)^2 + \mathbb{E}X = 2 + 0^2 + 1 = 3.$
3. $\mathbb{E}[(X-Y)^2] = \mathbb{E}X^2 + \mathbb{E}Y^2 - 2\mathbb{E}XY = \mathbb{V}(X) + (\mathbb{E}X)^2 + \mathbb{V}(Y) + (\mathbb{E}Y)^2 - 2Cov(X,Y) - 2(\mathbb{E}X)(\mathbb{E}Y) = 1 + 1^2 + 2 + 0^2 - 2 - 0 = 2.$
4. $\mathbb{V}[X + 2Y] = \mathbb{V}(X) + 4\mathbb{V}(Y) + 4Cov(X,Y) = 1 + 8 + 4 = 13.$
5. Find $\alpha > 0$ such that $\alpha X = Y$ with probability 1 or prove that no such α exists. Notice that $\mathbb{E}[\alpha X] = \alpha$ while $\mathbb{E}Y = 0$. Since $A = B$ with probability 1 implies $\mathbb{E}A = \mathbb{E}B$, no $\alpha > 0$ can exist with the desired property.

Problem 3 Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathsf{Ber}(.5)$ and $Y_1, \ldots, Y_n \overset{iid}{\sim} \mathsf{Exp}(1)$. Assume further that all the random variables are mutually independent. Write a central limit theorem (compute the variance/covariance matrix) for each of the following quantities

1. $(\bar{X}_n, \bar{Y}_n)$ $\mathbb{E}\bar{X}_n = 0.5, \mathbb{E}\bar{Y}_n = 1, \mathbb{V}(\bar{X}_n) = 1/(4n), \mathbb{V}(\bar{Y}_n) = 1/n.$
2. $\bar{X}_n - \bar{Y}_n$ $\mathbb{E}[\bar{X}_n - \bar{Y}_n] = -0.5, \mathbb{V}(\bar{X}_n - \bar{Y}_n) = \mathbb{V}(\bar{X}_n) + \mathbb{V}(\bar{X}_n) = 5/(4n)$

3.
$$\bar{XY} := \frac{1}{n}\sum_{i=1}^{n} X_i Y_i \,.$$

$\mathbb{E}\bar{XY} = \mathbb{E}[X_1Y_1] = \mathbb{E}[X_1]\mathbb{E}[Y_1] = 0.5, \mathbb{V}(\bar{XY}) = \mathbb{V}(X_1Y_1)/n$. We further have $\mathbb{V}(X_1Y_1) = \mathbb{E}[X_1^2Y_1^2] - (\mathbb{E}X_1Y_1)^2 = \mathbb{E}[X_1^2]\mathbb{E}[Y_1^2] - (\mathbb{E}X_1)^2(\mathbb{E}Y_1)^2 = (\mathbb{V}(X_1) + (\mathbb{E}X_1)^2)(\mathbb{V}(Y_1) + (\mathbb{E}Y_1)^2) - 1/4 = (1/4 + 1/4)(1 + 1) - 1/4 = 3/4$.

4.
$$\frac{(\bar{X}_n)^2}{\bar{Y}_n}$$

Calculating the expectation directly, we have $\mathbb{E}[(\bar{X}_n)^2/\bar{Y}_n] = \mathbb{E}(\bar{X}_n)^2 \times \mathbb{E}[1/\bar{Y}_n]$. For the first term we have $\mathbb{E}(\bar{X}_n)^2 = \mathbb{V}(\bar{X}_n) + (\mathbb{E}\bar{X}_n)^2 = \frac{1}{4n} + \frac{1}{4}$. For the second note that $\sum_{i=1}^{n} Y_i \sim Gamma(n, 1)$ with density $x^{n-1}e^{-x}/(n-1)!$. We can compute the expectation directly as $\mathbb{E}[1/\bar{Y}_n] = n\int_0^\infty x^{n-2}e^{-x}/(n-1)!dx = \frac{n}{n-1}\int_0^\infty x^{n-2}e^{-x}/(n-2)! = \frac{n}{n-1}$, where we used that the density integrates to 1. Note that if $n = 1$ then $\mathbb{E}[1/\bar{Y}_n] = \infty$ and so $\mathbb{E}[(\bar{X}_n)^2/\bar{Y}_n] = \infty$. For the variance we have $\mathbb{V}((\bar{X}_n)^2/\bar{Y}_n) = \mathbb{E}(\bar{X}_n)^4 \times \mathbb{E}[1/(\bar{Y}_n)^2] - (\mathbb{E}(\bar{X}_n)^2/\bar{Y}_n)^2$. Looking at the terms separately we get

$$\begin{aligned}
\mathbb{E}(\bar{X}_n)^4 &= \frac{1}{n^4}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right)^4 \\
&= \frac{1}{n^4}\left\{\binom{n}{4}4!0.5^4 + \binom{n}{3}\binom{3}{1}\frac{4!}{2!}0.5^3 + \binom{n}{2}(\frac{4!}{2!2!} + 2\frac{4!}{3!})0.5^2 + \binom{n}{1}0.5\right\} \\
&= \frac{1}{n^4}\left(0.0625n^4 + 0.375n^3 + 0.1875n^2 - 0.125n\right).
\end{aligned}$$

For $\bar{Y}_n$, assuming that $n > 2$, we have

$$\begin{aligned}
\mathbb{E}[1/(\bar{Y}_n)^2] &= n^2\int_0^\infty \frac{1}{(n-1)!}x^{n-3}e^{-x}dx \\
&= \frac{n^2}{(n-1)(n-2)}\int_0^\infty \frac{1}{(n-3)!}x^{n-3}e^{-x}dx \\
&= \frac{n^2}{(n-1)(n-2)}.
\end{aligned}$$

If $n = 1, 2$ then $\mathbb{E}[1/(\bar{Y}_n)^2] = \infty$ and so $\mathbb{V}((\bar{X}_n)^2/\bar{Y}_n) = \infty$.

Problem 4 Estimation of a Bernoulli parameter

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli random variables, with unknown parameter $p \in (0, 1)$.

1. Suppose we observe these $n \geq 4$ random variables. Write down a valid statistical model for the resulting data. $\{\text{Ber}(p) : p \in (0, 1)\}$.
2. Define the estimator of p, $\hat{p}_3 = \dfrac{X_1 + X_2 + X_3}{3}$. What is the bias of $\hat{p}_3$? $\mathbb{E}\hat{p}_3 = p$, therefore $\hat{p}_3$ is unbiased.
3. What is the variance of $\hat{p}_3$? $\mathbb{V}(\hat{p}_3) = \mathbb{V}(X_1)/3 = p(1-p)/3$.
4. What is the MSE of the estimator $\hat{p}_3$? $MSE(\hat{p}_3) = \mathbb{E}(\hat{p}_3 - p)^2 = \mathbb{V}(\hat{p}_3) = p(1-p)/3$.
5. What is the MSE of the estimator $\bar{X}_n$. $MSE(\bar{X}_n) = \mathbb{E}(\bar{X}_n - p)^2 = \mathbb{V}(\bar{X}_n) = \mathbb{V}(X_1)/n = p(1-p)/n$.
6. Which estimator is better? Justify your answer. $\bar{X}_n$ is the better estimator because it has lower MSE since $n \geq 4 > 3$.

Problem 5 Let $X_1, \ldots, X_n$ be i.i.d. $\text{Exp}(\lambda)$ random variable, where λ is unknown. Thus, each X_i has density $e^{-x/\lambda}/\lambda, x \geq 0$[1].

1. What is the distribution of $\min_i X_i$ (compute the CDF and take its derivative)? For $x \geq 0$ we get $\mathbb{P}(\min_i X_i \geq x) = \mathbb{P}(X_i \geq x, \forall i) = \mathbb{P}(X_1 \geq x)^n = \exp(-xn/\lambda))$. Differentiating with respect to x gives $d\mathbb{P}(\min_i X_i \geq x)/dx = -(n/\lambda)\exp(-xn/\lambda))$. Therefore $\min_i X_i \sim Exp(\lambda/n)$.
2. Use the previous question to give an unbiased estimator for λ. $n \times \min_i X_i$.
3. What is MSE of the above estimator? $MSE = \mathbb{E}(n \times \min_i X_i - \lambda)^2 = \mathbb{V}(n \times \min_i X_i) = n^2\mathbb{V}(\min_i X_i) = n^2 \times (\lambda/n)^2 = \lambda^2$.
4. What is MSE of the plugin estimator? The MLE for λ is $\hat{\lambda} = \bar{X}_n$. This is unbiased and has $MSE = \mathbb{V}(X_1)/n = \lambda^2/n$.

Problem 6 Let $X_n \sim \text{Unif}\big(-\frac{1}{n}, \frac{1}{n}\big)$ and let X be a random variable such that $\mathbb{P}(X = 0) = 1$.

1. Compute and draw the CDF $F_n(x)$ and $F(x)$ of X_n and X respectively. $F_n(x) = \min(1, \max(0, (nx+1)/2))$. The graph of F_n is a straight line from $(-1/n, 0)$ to $(1/n, 1)$, 0 before $-1/n$ and 1 after $1/n$. The graph of F is 1 on $[0, \infty)$ and 0 elsewhere.
2. Does $X_n \xrightarrow{\mathsf{P}} X$? (prove or disprove) For any $\epsilon > 0$ we have $\mathsf{P}(|X_n - X| > \epsilon) = \mathsf{P}(|X_n| > \epsilon) \leq \delta_{n \leq 1/\epsilon}$, where $\delta_A = 1$ if A holds and 0 otherwise. In particular, X_n converges to X in probability.

[1] Some people use the alternative definition with density $\lambda e^{-\lambda x}$, but we stick with Wasserman's convention (see Section 2.4 in the textbook).

3. Does $X_n \rightsquigarrow X$? (prove or disprove) X_n converges to X in distribution because it converges in probability.

Problem 7

In a simple model of inheritance, a given person has genotypes AA, aa or Aa. Each person has two alleles (each of which is either a or A), and the first and the second allele are inherited independently and are identically distributed. Note that Aa is the same as aA: the order of alleles does not matter. Therefore $\mathbb{P}(\mathsf{AA}) = \theta^2$, $\mathbb{P}(\mathsf{aa}) = (1-\theta)^2$ where $\theta \in (0,1)$ is an unknown parameter. Our goal here is to estimate this parameter.

1. Compute $\mathbb{P}(\mathsf{Aa})$ in terms of θ.
   The probability is $\mathbb{P}(\mathsf{Aa}) = \boxed{2\theta(1-\theta)}$ as we have $\theta, 1-\theta$ chance of getting A, a respectively and there are two possible combinations Aa, aA that we can get.
2. Define the random vector $X = \{0,1\}^3$ associated to a random person with genotype g by
   $$X = \begin{cases} (1,0,0) & \text{if } g = \mathsf{AA} \\ (0,1,0) & \text{if } g = \mathsf{aa} \\ (0,0,1) & \text{if } g = \mathsf{Aa} \end{cases}$$
   What is the pmf $p(x)$ of X for $x = (x_1, x_2, x_3) \in \{(1,0,0), (0,1,0), (0,0,1)\}$?
   - $p(x) = \theta^x(1-\theta)^{1-x}$
   - $p(x) = (\theta^2)^{x_1}(1-\theta^2)^{x_2}(1-2\theta)^{x_3}$
   - $p(x) = (1-\theta)^{2x_1}(\theta)^{2x_2}([2\theta(1-\theta)]^{x_3}$
   - $\boxed{p(x) = (\theta)^{2x_1}(1-\theta)^{2x_2}[2\theta(1-\theta)]^{x_3}}$
   - $p(x) = (\theta)^{2x_3}(1-\theta)^{2x_1}[2\theta(1-\theta)]^{x_2}$
3. A sample of 942 males from Zimbabwe was collected, and the following genotypes were observed: 501 of type AA, 83 of type aa and 358 of type Aa. Compute the maximum likelihood estimator for θ.
   By the previous part, the likelihood function for this data is:
   $$\mathcal{L}_n(\theta) = (\theta)^{2\cdot 501}(1-\theta)^{2\cdot 83}[2\theta(1-\theta)]^{358} = 2^{358}\theta^{1360}(1-\theta)^{524}$$
   Setting the derivative of the log-likelihood to 0 we get:
   $$0 = \frac{\partial l_n(\theta)}{\partial \theta} = \frac{1360}{\theta} - \frac{524}{1-\theta}$$
   Lastly, solving for theta gives:
   $$\hat{\theta}_{MLE} = \boxed{\frac{340}{471}}$$

Problem 8 Let $X_1, \ldots, X_n$ be independent copies of the random variable X where X is a mixture of two uniform random variables and has pdf:

$$f(x) = \frac{1}{4\theta}\mathbf{1}(x \in [0, 2\theta]) + \frac{1}{4\theta}\mathbf{1}(x \in [\theta, 3\theta])$$

for some unknown $\theta > 0$. For this problem, we call $\mathsf{Unif}[0, 2\theta)$, the first component.

1. Compute the proportion π of the first component?

   The proportion is $\frac{2\theta - 0}{4\theta} = \boxed{\frac{1}{2}}$.

2. Compute $\mathsf{E}[X]$ and $\mathsf{V}[X]$.
   Note that the density has value 2 on $[\theta, 2\theta]$ and takes value 1 on the rest of $[0, 3\theta]$. As both intervals have mean $\frac{3\theta}{2}$, $\mathsf{E}[X] = \boxed{\frac{3\theta}{2}}$. We also have:

$$\begin{aligned}\mathsf{E}[X^2] &= \int_0^{3\theta} \frac{x^2}{4\theta}dx + \int_\theta^{2\theta} \frac{x^2}{4\theta}dx \\ &= \frac{17}{6}\theta^2\end{aligned}$$

   Yielding:

$$V(X) = \mathsf{E}[X^2] - \mathsf{E}[X]^2 = \boxed{\frac{7}{12}\theta^2}$$

3. Assume that we start the k-th E-step of the EM algorithm with a candidate θ_k from the previous M step. Let $w_1, \ldots, w_n$ be the weights obtained in the E-step of the EM algorithm. Compute these weights and show that they can take only three values depending on X_i and θ_k.
   Define the density:

$$f_\theta(x, z) = \begin{cases} \frac{1}{4\theta}\mathbf{1}(x \in [0, 2\theta]) & z = 1 \\ \frac{1}{4\theta}\mathbf{1}(x \in [\theta, 3\theta]) & z = 0 \end{cases}$$

   The expectation minimization algorithm then asks us to compute the value of θ maximizing:

$$\begin{aligned}\mathsf{E}\left[\log \prod_{i=1}^n f_\theta(X_i, Z_i)|X, \theta_k\right] &= \sum_{i=1}^n \mathsf{E}[\log f_\theta(X_i, Z_i)|X, \theta_k] \\ &= \sum_{i=1}^n (P(Z_i = 1|X_i, \theta_k)\log f_\theta(X_i, 1) + P(Z_i = 0|X_i, \theta_k)\log f_\theta(X_i, 0))\end{aligned}$$

As such our weights are $P(Z_i = 1|X_i, \theta_k)$. Now, if $Z_i \in [0, \theta_k]$, then we must have $Z_i = 1$. Similarly if $Z_i \in [2\theta_k, 3\theta_k]$ we must have $Z_i = 0$. Lastly, if $Z_i \in [\theta_k, 2\theta_k]$, then $Z_i \sim Ber(1/2)$ by part 1. In these cases the weights are respectively $\boxed{1, 0, .5}$

4. Assume that $n = 8$ and the observations are (in order)

$$1.01 \quad 1.02 \quad 1.19 \quad 1.19 \quad 1.28 \quad 2.39 \quad 2.56 \quad 2.58$$

and that the the EM algorithm is initialized at $\theta_0 = 3$, what are the values of the iterates: θ_1, θ_2 and θ_3?

Consider the expression we want to maximize from the previous part:

$$\sum_{i=1}^{n} \left(P(Z_i = 1|X_i, \theta_k) \log f_\theta(X_i, 1) + P(Z_i = 0|X_i, \theta_k) \log f_\theta(X_i, 0)\right)$$

By our analysis of the weights the terms $\log f_\theta(X_i, 0)$ will have nonzero weight whenever $X_i \in [\theta_k, 3\theta_k]$ and $\log f_\theta(X_i, 1)$ will have nonzero weight whenever $X_i \in [0, 2\theta_k]$. As the logarithm of 0 is negative infinity, whenever the corresponding weight is nonzero we must have the corresponding pdf $f_\theta(X_i, 0)$ or $f_\theta(X_i, 1)$ be nonzero. This implies that we need the following implications:

$$\begin{aligned} X_i \in [0, 2\theta_k] &\implies X_i \in [0, 2\theta] \\ X_i \in [\theta_k, 3\theta_k] &\implies X_i \in [\theta, 3\theta] \end{aligned}$$

So long as these hold the value of $\log f_\theta(X_i, 0/1)$ will always be $-\log(2\theta)$. Thus we need to find the minimum value of θ at each step such that the above implications hold.

For the first step every value lies in $[0, \theta_0]$ so we need every value to remain in $[0, 2\theta]$. Thus $\boxed{\theta_1 = 2.58/2 = 1.29}$. For the next step, $2\theta_1 = 2.58$ so the minimum possible value θ_2 can take is 1.29. This satisfies the implications and thus $\boxed{\theta_2 = 1.29}$. Finally, the same logic shows $\boxed{\theta_3 = 1.29}$.

Problem 9

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, I)$, where $\mu \in \mathbb{R}^p$ and I is the $p \times p$ identity matrix (the X_i are random vectors).

1. Let A be a $p \times p$ matrix such that $A\mu = 0$ and $\mathsf{Tr}(A) = 1$ (here $\mathsf{Tr}(A)$ denotes the trace of A). Compute $\mathsf{E}[X^\top AX]$.

We start by computing $\mathsf{E}[X^\top AX]$. Since $X^\top AX$ is a scalar, it follows that $\mathsf{E}[X^\top AX] = \mathsf{E}[Tr(X^\top AX)]$. Since the trace is invariant under circular shits, we have that

$\mathsf{E}[Tr(X^\top AX)] = \mathsf{E}[Tr(XX^\top A)]$. The trace is a linear operation so we can use linearity of expectation to write: $\mathsf{E}[Tr(XX^\top A)] = Tr(\mathsf{E}[XX^\top A]) = Tr(\mathsf{E}[XX^\top]A)$. Notice that $\mathsf{E}[XX^\top] = Cov(X) + \mathsf{E}[X]\mathsf{E}[X^\top] = I + \mu\mu^\top$, which can be derived from the definition of the covariance matrix of X. In summary, we have that $\mathsf{E}[X^\top AX] = Tr((I + \mu\mu^\top)A) = Tr(IA) + Tr(\mu\mu^\top A) = Tr(A) + Tr(\mu^\top A\mu) = 1 + 0 = 1$.

2. What is the likelihood function for μ? The likelihood function is:

$$\mathcal{L}_n(X_1, \ldots, X_n; \mu) = \boxed{\frac{1}{(2\pi)^{np/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} ||X_i - \mu||^2\right)}$$

3. Compute the maximum likelihood estimator $\hat{\mu}_{MLE}$ for μ. Prove that it actually maximizes the likelihood function.
Setting the derivative of the log-likelihood to 0 we get:

$$0 = \frac{\partial l_n(\mu)}{\partial \mu} = -\sum_{i=1}^{n}(X_i - \mu)$$

Solving yields:

$$\hat{\mu}_{MLE} = \boxed{\overline{X}_n}$$

To see that this actually maximizes the likelihood function we note that $||X_i - \mu||^2$ is a strictly convex function and thus has a unique minimizer which must be where the derivative equals zero.

4. What is the distribution of $\hat{\mu}_{MLE}$? As it's an average of n i.i.d. $\mathcal{N}(\mu, I)$ random variables, $\hat{\mu}_{MLE} \sim \boxed{\mathcal{N}(\mu, I/n)}$.

5. Let B be a fixed $m \times p$ matrix. What is the variance of $B\hat{\mu}_{MLE}$? Note that the variance matrix satisfies the following nice property:

$$\text{Var}(BX) = \mathsf{E}[(BX - \mathsf{E}[BX])(BX - \mathsf{E}[BX])^T] = \mathsf{E}[B(X - \mathsf{E}[X])(X - \mathsf{E}[X])^T B^T] = B\text{Var}(X)B^T$$

Using this:

$$\text{Var}(B\hat{\mu}_{MLE}) = B\text{Var}(\hat{\mu}_{MLE})B^T = \boxed{\frac{1}{n}BB^T}$$

6. Define the function $g(X) = \|X\|^2$. What is the asymptotic variance of $g(\hat{\mu}_{MLE})$? Since, $\hat{\mu}_{MLE} = \overline{X}_n$, the central limit theorem tells us:

$$\sqrt{n}(\hat{\mu}_{MLE} - \mu) \to \mathcal{N}(0, I)$$

Applying the delta method with g from the problem statement yields:

$$\sqrt{n}(g(\hat{\mu}_{MLE}) - g(\mu)) \to \mathcal{N}(0, \nabla g(\mu) I \nabla g(\mu)^T)$$

We have $\nabla g(\mu) = 2\mu$, so the asymptotic variance is:

$$\nabla g(\mu) I \nabla g(\mu)^T) = 4\mu I \mu^T = \boxed{4\mu\mu^T}$$

Problem 10

The lifetime (in months) of a cell phone battery is modeled by a random variable X that has pdf

$$f_\theta(x) = K\theta^x \mathbf{1}(x > 0)$$

for an unknown $\theta \in (0, 1)$. Assume that we have n independent observations $X_1, \ldots, X_n$ of the lifetime of n cell phones. We want to use them to estimate $\theta \in (0, 1)$.

1. Show that $K = \log(1/\theta)$.
   Note that $X_1 \sim \mathsf{Exp}(1/\log(1/\theta))$. Thus $K = \log(1/\theta)$. Another way to get this result is by remembering that the pdf must integrate to 1. K is used as a normalization constant to make sure that this is true.
2. Compute the expected value and the variance of X.
   By the observation that X_1 is an exponential random variable, $\mathsf{E}[X_1] = \boxed{1/\log(1/\theta)}$ and $V(X_1) = \boxed{1/\log(1/\theta)^2}$
3. Compute the maximum likelihood estimator $\hat{\theta}$ of θ.
   The likelihood function is

$$\mathcal{L}_n(\theta) = \log(1/\theta)^n \theta^{\sum_i X_i}$$

   Setting the derivative of the log-likelihood to 0:

$$0 = \frac{\partial l_n}{\partial \theta} = -\frac{n}{\theta \log(1/\theta)} + \frac{1}{\theta}\sum_{i=1}^{n} X_i$$

   Solving gives us the maximum likelihood estimator is:

$$\hat{\theta}_{MLE} = \boxed{\exp\left(-\frac{1}{\overline{X}_n}\right)}$$

4. Using the Delta method, show that $\sqrt{n}(\hat{\theta}-\theta)$ converges in distribution to $\mathcal{N}(0,\sigma^2)$, where σ^2 is to be made explicit.
By the central limit theorem,

$$\sqrt{n}(\overline{X}_n - \log(1/\theta)^{-1}) \to \mathcal{N}(0, \log(1/\theta)^{-2})$$

Let $g(x) = \exp(-1/x)$. Note that $g'(1/\log(1/\theta)) = \log(\theta)^2\theta$. Applying the delta method with g yields that the asymptotic variance is

$$(\log(\theta)^2\theta)^2 \log(1/\theta)^{-2} = \boxed{\theta^2 \log(\theta)^2}$$

5. Compute the Fisher information $I(\theta)$.
As the Fisher information is the inverse of the asymptotic variance we have:

$$I(\theta) = \boxed{\frac{1}{\theta^2 \log(\theta)^2}}$$

6. We observe $\hat{\theta} = 0.62$ for $n = 100$. Show that $(0, 0.6647]$ is a valid confidence interval at asymptotic level 95% for θ.
By part 4 combined with Slutsky's theorem,

$$\sqrt{n\hat{\theta}_n^2 \log(\hat{\theta}_n)^2}\left(\hat{\theta}_n - \theta\right) \to \mathcal{N}(0,1)$$

asymptotically. Since, $\Phi(1.64) \approx .95$, this means that:

$$P\left(\sqrt{n\hat{\theta}_n^2 \log(\hat{\theta}_n)^2}\left(\hat{\theta}_n - \theta\right)\right) \leq 1.64) \approx .95$$

This yields the following confidence interval at asymptotic level 95%:

$$\left(0, .62 + 1.64\sqrt{\frac{.62^2 \log(.62)^2}{100}}\right] = \boxed{(0, .6647]}$$

where we can set the left endpoint to 0 as we know $\theta > 0$.

Problem 11 Predicting heat waves

The Boston Health Department is monitoring the average daily temperatures of a city over the summer months to understand patterns and prepare for potential heatwaves. They've collected daily temperature data for the past two months. The department believes that the daily temperatures follow a normal distribution. As an 18.650 student, the city council decided to hire you to help them with some very important tasks.

1. Model the daily temperatures as a random variable and specify its potential distribution. That is, formalize the conjecture made by the health department. Justify why this might be a suitable distribution.
   We can model the daily temperatures as a random variable T that follows a normal distribution, $T \sim \mathcal{N}(\mu, \sigma^2)$, where μ is the mean daily temperature and σ^2 is the variance. This choice is based on the Central Limit Theorem which suggests that the sum (or average) of a large number of independent and identically distributed random variables will be approximately normally distributed. Daily temperatures, resulting from various independent climatic factors, can thus be modeled as a normal distribution, especially when looking at averages.

2. Given your model, describe how you would estimate the parameters of the distribution using the provided data. Discuss any assumptions you're making. For the mean μ we can use the sample mean, which is the average of all the observed daily temperatures from the data collected over the two months. For the variance σ^2 we can use the sample variance, calculated as the average of the squared differences from the mean, taking into consideration Bessel's correction (dividing by $n-1$ instead of n for an unbiased estimator). Both the biased and unbiased estimators will be accepted as possible estimators. The assumptions we are making include:
   - Independence: Each day's temperature is an independent observation.
   - Identical distribution: The underlying factors affecting temperature remain consistent over the two months.
   - Normality: As mentioned earlier, we are assuming that the daily temperatures are normally distributed.

3. The department is particularly concerned about extremely high temperatures. Using your model, explain how you would compute the probability that the temperature on a given day exceeds 100F. Also, discuss the implications of your model's assumptions on this probability estimate.
   Given the normal distribution model, the probability that the temperature on a given day exceeds a critical threshold of 100F is given by: $\mathsf{P}(T > 100) = 1 - \mathsf{P}(T \leq 100)$. This can be found using the cdf of the standard normal distribution (denoted as Φ): $\mathsf{P}(T > 100) = 1 - \Phi\left(\frac{100-\mu}{\sigma}\right)$. The implications of our model's assumptions on this probability estimate include:
   - If the assumption of normality doesn't hold, then our probability estimate might be inaccurate.
   - The independence assumption implies that previous days' temperatures don't impact future temperatures, which might not be entirely accurate in real-world scenarios.
   - If there are external factors (e.g., sudden climatic changes, volcanic eruptions) that significantly influence the temperature and were not present in the past

two months, our model might not capture the true probability of extreme temperatures.

Hedge fund interview question

In every PSet, we have an additional question taken from a hedge fund interview. This question is not mandatory and does not hold any point but you are welcome to give it a shot.

Problem 12

Given i.i.d. random variables $X_1, \ldots, X_n$, what is $\mathsf{E}[\max_i X_i]$ if:

1. $X_1, \ldots, X_n \sim \mathsf{Unif}\big(0,1\big)$?
   Consider the cdf of the distribution of $max_i X_i$:
   $\mathsf{P}(max_i X_i \leq x) = \mathsf{P}(X_1 \leq x, X_2 \leq x, ..., X_n \leq x) = \mathsf{P}(X_1 \leq x)\mathsf{P}(X_2 \leq x) \cdots P(X_n \leq x)$
   , because the X_i are i.i.d. Given that X_i follows a uniform distribution on [0, 1] the cdf is: $\mathsf{P}(X_i \leq x) = x$ for $0 \leq x \leq 1$.
   Therefore, $\mathsf{P}(max_i X_i \leq x) = x^n$. We get the pdf by differentiating the pdf with respect to x: $f(x) = nx^{n-1}$. The expected value is given by:
   $\mathsf{E}[max_i X_i] = \int_0^1 xf(x)dx = \int_0^1 xnx^{n-1}dx = \frac{n}{n+1}$.
2. $X_1, \ldots, X_n \sim \mathsf{Exp}\big(1\big)$?
   We proceed with a similar approach to part 1 of the problem. The cdf for a single X_i is $\mathsf{P}(X_i \leq x) = 1 - e^{-x}$. Therefore, $\mathsf{P}(max_i X_i \leq x) = (1-e^{-x})^n$ and $f(x) = ne^{-x}(1-e^{-x})^{n-1}$. Finally ,$\mathsf{E}[max_i X_i] = \int_0^\infty xf(x)dx = \int_0^\infty xne^{-x}(1-e^{-x})^{n-1}dx$

	Second decimal place of Z									
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

The table lists $P(Z \leq z)$ where $Z \sim N(0,1)$ for positive values of z.