Problem 1. [20 points: 5+10+5]
(a) First note that:
\[ \theta = \mathbb{P}(X_i = 0) = e^{-\lambda} \]
In HW5, problem 3, we saw that the MLE for \( \lambda \) is given by the sample mean \( \bar{X} \). So by equivariance we have that the MLE \( \hat{\theta} \) for \( \theta \) is \( e^{-\bar{X}} \).

(b) By the Central Limit Theorem:
\[ \sqrt{n}(\bar{X} - \lambda) \sqrt{\lambda} \xrightarrow{d} \mathcal{N}(0,1) \]
So by the Delta Method with \( g(t) = e^{-t} \) we have that the limiting distribution of \( \hat{\theta} = e^{-\bar{X}} \) is defined by:
\[ \sqrt{n}(e^{-\bar{X}} - e^{-\lambda}) e^{-\lambda} \sqrt{\lambda} \xrightarrow{d} \mathcal{N}(0,1) \]

(c) By the Weak Law of Large numbers \( \bar{X} \xrightarrow{P} \lambda \). Since function \( e^{-t} \) is continuous, by the Continuous Mapping Theorem \( \hat{\theta} = e^{-\bar{X}} \xrightarrow{P} e^{-\lambda} \), i.e. \( \hat{\theta} \) is consistent.

Note: notice that we also have \( \sqrt{n}(\bar{X} - \lambda) \sqrt{\bar{X}} \xrightarrow{d} \mathcal{N}(0,1) \). How would you formally show it?

Problem 2. [30 points: 5+5+10+10]
(a)
\[ \mathbb{E}[X_1] = \frac{\theta}{2} \implies \hat{\theta}_1 = 2\bar{X} \]
The likelihood function is \( L(\theta) = \frac{1}{\theta^n} I_{(X(n),\infty)}(\theta) \). This function is equal to zero on \(( -\infty, X(n) )\), then at \( X(n) \) jumps to \( 1/(X(n)) > 0 \) and decreases asymptotically to zero as \( \theta \to \infty \). This means that the supremum is at \( \theta = X(n) \), i.e. the MLE for \( \theta \) is
\[ \hat{\theta}_2 = X(n) \]

(b) By the Weak Law of Large Numbers we know that \( \bar{X} \xrightarrow{P} \theta/2 \) so by Continuous Mapping Theorem \( \hat{\theta}_1 \) is consistent. To show consistency of \( \hat{\theta}_2 \), let \( \epsilon > 0 \) and then, using results from previous HWs and Test1:
\[ \mathbb{P}(|\theta - X(n)| > \epsilon) = \mathbb{P}(\theta - \epsilon > X(n)) = \left(1 - \frac{\epsilon}{\theta} \right)^n \to 0 \]
as \( n \to \infty \), which proves the claim.
(c) MOME. We have that \( \text{Var}(\hat{\theta}_1) = \frac{\theta^2}{3n} \). Thus, by the Central Limit Theorem:

\[
\frac{\hat{\theta}_1 - \theta}{\text{sd}(\hat{\theta}_1)} = \frac{\sqrt{n}(2\bar{X} - \theta)}{\theta/\sqrt{3}} \rightsquigarrow N(0, 1)
\]

MLE. To find the limiting distribution \( \hat{\theta}_2 = X(n) \):

\[
P\left( \frac{n(\theta - X(n))}{\theta} \leq c \right) = P(\theta - X(n) \leq \theta c/n) = 1 - \left( 1 - \frac{c}{n} \right)^n \rightarrow 1 - e^{-c}
\]

which implies that:

\[
\frac{n(\theta - X(n))}{\theta} \rightsquigarrow \text{Exp}(1)
\]

(d) MOME. For every \( \epsilon > 0 \) we can find \( K \) such that

\[
P\left( n|\hat{\theta}_1 - \theta| > K \right) = P\left( |\hat{\theta}_1 - \theta| > K/\sqrt{n} \right) \leq \frac{n\text{Var}(\hat{\theta}_1)}{K^2} = \frac{\theta^2}{3K^2} < \epsilon
\]

in particular for every \( K > \frac{\theta}{\sqrt{3}\epsilon} \). Thus, \( \hat{\theta}_1 - \theta = O_P(1/\sqrt{n}) \).

MLE. Recall that if \( W_i \sim \text{Uniform}(0, 1) \), then \( W(n) \sim \text{Beta}(n, 1) \) such that \( E[W(n)] = \frac{n}{n+1} \) and \( \text{Var}(W(n)) = \frac{n}{(n+1)^2(n+2)} \). Thus, \( \bar{X}(n) = \frac{\theta n}{n+1} \) and \( \text{Var}(\hat{\theta}_2) = \frac{\theta^2 n}{(n+1)^2(n+2)} \) and \( \text{Bias}(\hat{\theta}_2) = \frac{\theta}{n+1} \).

Therefore, for every \( \epsilon > 0 \) we can find \( K \) such that

\[
P(n|\hat{\theta}_1 - \theta| > K) \leq \frac{n^2}{K^2} \text{E}[|\hat{\theta}_1 - \theta|^2] = \frac{n^2}{K^2} \left( \frac{\theta^2}{(n+1)^2} + \frac{\theta^2 n}{(n+1)^2(n+2)} \right)
\]

\[
\leq \frac{n^2}{K^2} \left( \frac{\theta^2}{n^2} + \frac{\theta^2 n}{n^3} \right) \leq \frac{2\theta^2}{K^2} < \epsilon
\]

in particular for every \( K > \frac{\sqrt{2}\theta}{\epsilon} \). This proves that \( \hat{\theta}_2 - \theta = O_P(1/n) \).

In terms of speed of convergence the \( \hat{\theta}_2 \) is preferable.

**Problem 3.** [10 points]

For the Normal distribution we know the MLE for its mean is given by \( \bar{X} \). We also have that in this problem \( \text{sd}(\bar{X}) = \frac{1}{\sqrt{n}} \). Thus, the Wald statistic for this problem is given by:

\[
W = \sqrt{n}(|\bar{X} - \mu_0|)
\]

and under the null hypothesis \( H_0 : \mu = \mu_0 \) we have \( W \sim N(0, 1) \) (exact distribution).

The alternative hypothesis is \( H_1 : \mu \neq \mu_0 \), which means that we are doing a two-sided test. Thus, in the case of equal tails, the rejection rule of a size \( \alpha \) Wald test has the form:

\[
W < -z_{\alpha/2} \text{ or } W > z_{\alpha/2} \iff |W| > z_{\alpha/2} \iff |\sqrt{n}(\bar{X} - \mu_0)| > z_{\alpha/2}
\]

where \( z_{\beta} = \Phi^{-1}(1 - \beta) \), and \( \Phi(*) \) is the cdf of \( Z \sim N(0,1) \). Notice that the absolute value comes from symmetry about zero.

The p-value for the observed \( w = \sqrt{n}(\bar{x} - \mu) \) is:

\[
P_{\mu_0}(|W| > w) = 2\Phi(-|w|)
\]
where $\Phi$ is the CDF of the standard normal distribution. If $X_i \sim N(\mu, 1)$ then we have that $W \sim N(\sqrt{n}(\mu - \mu_0), 1)$, where $\mu$ is the true value. Thus, the CDF of the p-value is given by:

$$
P_{\mu}(2\Phi(-|W|) \leq \alpha) = P_{\mu}(\Phi(|W|) \leq \alpha/2) = P_{\mu}(|W| \leq -z_{\alpha/2}) = P_{\mu}(W \leq -z_{\alpha/2}) + P_{\mu}(W \geq z_{\alpha/2})$$

Writing $W = Z + \sqrt{n}(\mu - \mu_0)$, where $Z \sim N(0, 1)$, the above equals to

$$= P_{\mu}(Z \leq -z_{\alpha/2} - \sqrt{n}(\mu - \mu_0)) + P_{\mu}(Z \geq z_{\alpha/2} - \sqrt{n}(\mu - \mu_0))$$

$$= \Phi(-z_{\alpha/2} - \sqrt{n}(\mu - \mu_0)) + \Phi(-z_{\alpha/2} + \sqrt{n}(\mu - \mu_0))$$

In the particular case of $\mu = \mu_0$ the previous equation equals $\alpha$ which proves that the p-value has Uniform distribution under the null hypothesis.

**Problem 4.** [40 points: 10+10+10+10]

(a) We compute the MLE when both $\mu, \sigma^2$ are unknown for a given sample $X_1, \ldots, X_n$:

$$L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2\right]$$

Taking log and finding the partial derivatives:

$$\frac{\partial}{\partial \mu} \log L(\mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)$$

$$\frac{\partial}{\partial \sigma^2} \log L(\mu, \sigma^2) = -\frac{n}{\sigma^2} + \frac{1}{\sigma^4} \sum_{i=1}^{n} (X_i - \mu)^2$$

Setting both partial derivatives equal to 0 and solving the system of equations we obtain the estimators:

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

This is exactly the same as example 4 in lecture 7. See also Lecture 9, Example 9 to see that the second order conditions are satisfied. Since the MLE $\hat{\sigma}$ is a consistent estimator of $\sigma$, by Slutsky Theorem

$$\frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}} \sim N(0, 1)$$

Thus, the test statistic of the Wald test for $H_0: \mu = \mu_0$ versus $H_a: \mu \neq \mu_0$ is

$$W = \frac{\bar{X} - \mu_0}{\hat{\sigma}/\sqrt{n}}$$

Hence, under $H_0$ we have $W \approx N(0, 1)$. The alternative hypothesis is $H_1 : \mu \neq \mu_0$, which means that we are doing a two-sided test. Thus, in the case of equal tails, the rejection rule of a size $\alpha$ Wald test has the form:

$$W < -z_{\alpha/2} \text{ or } W > z_{\alpha/2} \iff |W| > z_{\alpha/2}$$
(b) Likelihood Ratio Test for $H_0 : \mu = \mu_0$ Vs $H_1 : \mu \neq \mu_0$.

We reject $H_0$ if:

$$\lambda(X^n) = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} = \frac{L(\hat{\theta}_0)}{L(\theta)} < k$$

(1)

where $\Theta_0 = \mu_0 \times (0, +\infty)$, and $k$ is some constant that will make $\mathbb{P}$(TYPE I error) $\leq \alpha$.

We have:

$$L(\mu_0; \sigma^2) = (2\pi \sigma^2)^{-n/2} \exp \left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu_0)^2 \right\}$$

(2)

$$l(\mu_0; \sigma^2) = -\frac{n}{2} \log(2\pi \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu_0)^2$$

(3)

Moreover, let $\xi = \sigma^2$:

$$\frac{\partial l}{\partial \xi} = -\frac{n}{2\xi} + \frac{1}{2} \frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{\xi^2} = 0$$

(4)

solved for $\hat{\sigma}^2 = \xi^* = \frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{n}$. The second order condition is satisfied:

$$\frac{\partial^2 l}{\partial \xi^2} = \frac{n}{2\xi^2} - \frac{n}{\xi^2} < 0$$

(5)

Thus

$$L(\hat{\theta}_0) = L(\mu_0, \hat{\sigma}^2) = \left(2\pi \frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{n} \right)^{-n/2} \exp \left\{-\frac{n}{2} \right\}$$

(6)

The MLE estimators for $\mu$ and $\sigma^2$ are $\bar{X}$ and $\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$, respectively. Thus:

$$L(\hat{\theta}) = \left(2\pi \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n} \right)^{-n/2} \exp \left\{-\frac{n}{2} \right\}$$

(7)

Therefore, we will reject $H_0$ if

$$\lambda(X^n) = \frac{L(\hat{\theta}_0)}{L(\theta)} = \frac{\left(2\pi \frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{n} \right)^{-n/2} \exp \left\{-\frac{n}{2} \right\}}{\left(2\pi \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n} \right)^{-n/2} \exp \left\{-\frac{n}{2} \right\}} = \left(\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sum_{i=1}^{n} (X_i - \mu_0)^2} \right)^{n/2}$$

(8)

$$= \left(\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2 + n(X - \mu_0)^2} \right)^{n/2} = \left(1 + \frac{n(X - \mu_0)^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \right)^{-n/2} < k$$

(9)

i.e. we reject $H_0$ if

$$\frac{n(X - \mu_0)^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

is sufficiently large. Equivalently we reject $H_0$ when $T = \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}}$ is very large or very small,

where $S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1}$. Since $T \sim T_{n-1}$, for a fixed $\alpha$ we reject $H_0$ if $|T| > t_{n-1, \alpha/2}$. 
However, under the null hypothesis, for large \( n \), \( T \approx N(0,1) \), such that we can also use \( z_{\alpha/2} \) as the critical value. Thus, compare results in (a).

(c) We have that
\[
\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}
\]
so that for the MLE
\[
\hat{\sigma}^2 = \frac{\sigma^2 \sum_{i=1}^{n} (X_i - \bar{X})^2}{n} \sim \frac{\sigma^2}{n} \chi^2_{n-1}
\]
Thus, \( sd(\hat{\sigma}^2) = \frac{\sigma^2}{n} \sqrt{\frac{2}{n-1}} \) which we can estimate by
\[
\hat{sd}(\hat{\sigma}^2) = \frac{\hat{\sigma}^2}{n} \sqrt{\frac{2}{n-1}}
\]
So by asymptotic normality of MLE and Slutsky Theorem, we have
\[
\frac{\hat{\sigma}^2 - \sigma^2}{\sigma^2 \sqrt{2/(n-1)}} \approx \frac{\hat{\sigma}^2 - \sigma^2}{\hat{\sigma}^2 \sqrt{2/n}} \sim N(0,1)
\]
So the test statistic is
\[
W = \frac{\hat{\sigma}^2 - \sigma_0^2}{\sigma^2 \sqrt{2/n}}
\]
and the rejection region of the \( \alpha \) level Wald Test is defined by
\[
|W| > z_{\alpha/2}
\]
for similar reasons explained in (a).

(d) Likelihood Ratio Test for \( H_0 : \sigma = \sigma_0 \) Vs. \( \sigma \neq \sigma_0 \).
We reject \( H_0 \) if:
\[
\lambda(X^n) = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} < k \tag{11}
\]
where \( \Theta_0 = \sigma_0 \times \mathbb{R} \) and \( k \) is a constant. For every fixed value of \( \sigma^2 \), we have that the MLE for \( \mu \) is \( \bar{X} \). Thus:
\[
L(\hat{\theta}_0) = (2\pi \sigma_0^2)^{-n/2} \exp \left\{ -\frac{1}{2} \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma_0^2} \right\} \tag{12}
\]
The MLE estimators for \( \mu \) and \( \sigma^2 \) are \( \bar{X} \) and \( \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \), respectively. Thus:
\[
L(\hat{\theta}) = \left( \frac{2\pi \sum_{i=1}^{n} (X_i - \bar{X})^2}{n} \right)^{-n/2} \exp \left\{ -\frac{n}{2} \right\} \tag{13}
\]
Thus:
\[
\lambda(X^n) = \frac{(2\pi \sigma_0^2)^{-n/2} \exp \left\{ -\frac{1}{2} \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma_0^2} \right\}}{(2\pi \sum_{i=1}^{n} (X_i - \bar{X})^2)^{-n/2} \exp \left\{ -\frac{n}{2} \right\}} = \left( \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n \sigma_0^2} \right)^{n/2} \exp \left\{ -\frac{n}{2} \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n \sigma_0^2} + \frac{n}{2} \right\} \tag{14}
\]
We can see that $\lambda(X^n)$ is a function of the kind $(g(y))^{n/2}$, where $g(y) = ye^{1-y}$. Function $g$ is positive with unique maximum in $y = 1$ (see figure below). We will reject $H_0$ if $g(y) < k_1$ or $g(y) > k_2$, where $y = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n\sigma_0^2}$. We can then use statistic $T = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma_0^2} \sim \chi^2_{n-1}$, under the null hypothesis $H_0$. In particular, we will reject $H_0$ if $T < a$ or $T > b$, for some particular $a, b$ such that $\mathbb{P}_{\sigma_0}(\{T < a\} \cup \{T > b\}) = \alpha$. The exact choice would be $\{a, b\} = g^{-1}(k_\alpha^{2/n})$, where $k_\alpha$ is such that $\mathbb{P}_{\sigma_0}(\lambda(X^n) < k_\alpha) = \alpha$ and $g^{-1}$ is the generalized inverse of $g$. However, we might simply consider the equal tailed quantiles of the $\chi^2_{n-1}$ distribution (i.e. $a = \chi^2_{n-1;1-\alpha/2}, b = \chi^2_{n-1;\alpha/2}$), or choosing $a, b$ delimiting the shortest interval with coverage $1 - \alpha$ of the $\chi^2_{n-1}$ distribution (see the figure below). Those $a, b$ satisfy the following system of equations:

$$\begin{cases} 
0 < a < b \\
f(a) = f(b) \\
\int_a^b f(t)dt = 1 - \alpha 
\end{cases}$$

(16)

where $f$ is the p.d.f. of $\chi^2_{n-1}$. These values can be found numerically (e.g. by using R). Alternatively, we could use the asymptotic result $\kappa = -2\log \lambda(X^n) \approx \chi^2_1$ (1 df because under $H_1$ there are 2 free parameters, while under $H_0$ we have 1 free parameter) and rejecting $H_0$ if $\kappa > \chi^2_{1,\alpha}$. Yet, in this exercise we prefer the previous non-asymptotic solutions.