Problem 1

Define random variable $X = \#$ of arriving automobiles measured per hour and random variable $Y = \#$ of automobiles which don't need equipment violations among X arriving automobiles. We have X has Poisson distribution with parameter $\lambda = 10$. And from independent increments property of Poisson process, the number of arrival automobiles in one hour is independent from next hour. Given $X = x$, for these x automobiles, whether one automobile needs violation is independent from others and the probability of one doesn't need violation is the same, thus Y has Binomial distribution with parameters $n = x$ and $p = 0.5$.

(a) We would like to calculate the probability that $X = 5$. Since $X \sim Pois(10)$,

$$\mathbb{P}(X = 5) = \frac{e^{-10}10^5}{5!}.$$

(b) We would like to calculate the probability that there are $X = 10$ automobiles arriving and all of them don't need violation, i.e. $Y = 10$. We have $X \sim Pois(10)$ and given $X = 10$, $Y \sim Bin(10, 0.5)$. From the conditional probability formula, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A)$, where A, B are two events. We have

$$\begin{aligned}\mathbb{P}(X = 10, Y = 10) &= \mathbb{P}(X = 10) \cdot \mathbb{P}(Y = 10|X = 10) \\ &= e^{-10}\frac{10^{10}}{10!} \cdot (0.5)^{10} = \frac{e^{-10}5^{10}}{10!}.\end{aligned}$$

(c) Similarly as part (b), for fixed $y \geq 10$,

$$\begin{aligned}\mathbb{P}(X = y, Y = 10) &= \mathbb{P}(X = y) \cdot \mathbb{P}(Y = 10|X = y) \\ &= e^{-10}\frac{10^y}{y!} \cdot \binom{y}{10}(0.5)^{10}(0.5)^{y-10} = e^{-10}\frac{10^y}{y!}\frac{y!}{10!(y-10)!}(0.5)^y \\ &= \frac{e^{-10}5^y}{10!(y-10)!}, \quad \text{as desired.}\end{aligned}$$

Problem 2

(a) $L(\theta) = \prod_1^n \frac{1}{(10-2\theta)}\mathbb{1}_{\{2\theta \leq x_i \leq 10\}} = \frac{1}{(10-2\theta)^n}\mathbb{1}_{\{2\theta \leq x_i \leq 10\}}$. We had to put in the indicator function because the range of acceptable values for x_i depends on θ.

(b) Setting the derivative of $L(\theta)$ equal to zero won't lead to a solution - let's maximize $L(\theta)$ directly by looking at it. First of all, for it to be maximized, it must be nonzero, so we need $2\theta \leq x_i \leq 10$. Let's now maximize while that condition holds. We want to minimize the denominator, so we want to minimize $10-2\theta$, so we want 2θ as close to 10 as possible, while still keeping $2\theta \leq x_i$. That means we should make $2\theta = x_{\min}$. Hence $\hat{\theta} = \frac{1}{2}X_{\min}$.

(c) $F_{\frac{1}{2}X_{\min}}(a) = P(\frac{1}{2}X_{\min} \leq a) = P(X_{\min} \leq 2a) = 1 - P(X_{\min} \geq 2a) = 1 - P(X_1 \geq 2a, ..., X_n \geq 2a) = 1 - P(X \geq 2a)^n = 1 - (1 - P(X \leq 2a))^n = 1 - (1 - F_X(2a))^n$. Note we made use of the fact that the X_i are i.i.d. What is $F_X(2a)$? Well, since we have a uniform distribution, it's just the proportion of the interval $[2\theta, 10]$ that $[2\theta, 2a]$ takes up. So it's $\frac{2a-2\theta}{10-2\theta} = \frac{a-\theta}{5-\theta}$. So our answer is $F_{\frac{1}{2}X_{\min}}(a) = 1 - (1 - \frac{a-\theta}{5-\theta})^n$. Note this is only on $[2\theta, 10]$ - it's 0 to the left and 1 to the right of that interval.

(d) Now, to find the density, we just take the derivative of the CDF. We get $f_{\frac{1}{2}X_{\min}}(x) = 2n(1 - F_X(2x))^{n-1} f_X(2x) = 2n(1 - \frac{x-\theta}{5-\theta})^{n-1} \frac{1}{10-2\theta}$. Note this is only on $[2\theta, 10]$ - it's 0 to the left and to the right of that interval.

(e) Now, $E(\frac{1}{2}X_{\min}) = \int_\theta^5 x 2n(\frac{5-x}{5-\theta})^{n-1} \frac{1}{10-2\theta} dx = n \int_\theta^5 x (\frac{5-x}{5-\theta})^{n-1} \frac{1}{5-\theta} dx$. That's it, you just need to set up the integral.

In case you're curious, here's the computation worked out. $n \int_\theta^5 x \frac{(5-x)^{n-1}}{(5-\theta)^n} dx =$
$\frac{n}{(5-\theta)^n} \int_\theta^5 x(5-x)^{n-1} dx = -\frac{n}{(5-\theta)^n} \int_{5-\theta}^0 (5-u) u^{n-1} du =$
$\frac{n}{(5-\theta)^n} \int_0^{5-\theta} (5-u) u^{n-1} du = \frac{n}{(5-\theta)^n} \int_0^{5-\theta} (5u^{n-1} - u^n) du =$
$\frac{n}{(5-\theta)^n} \left(\frac{5}{n}(5-\theta)^n - \frac{1}{n+1}(5-\theta)^{n+1} \right) = \frac{5-n\theta}{n+1}$. That is not equal to θ, so our estimator is biased.

`Problem 3`

(a) The number of people take more than 3 minutes actually follows Binomial Distribution. If we denote it by X, then $X \sim Bin(n, p)$, where n=50, and p=0.3.

34% people is actually 17 people. Thus we have:

$$P(X > 17) = \sum_{k=18}^{50} \binom{50}{k} 0.3^k 0.7^{50-k}$$

We can also use Central Limit Theorem:

$$\begin{aligned}
P(X > 17) &= P\Big(\frac{X - E(X)}{\sqrt{Var(X)}} > \frac{17 - E(X)}{\sqrt{Var(X)}}\Big) \\
&\approx P\Big(z > \frac{17 - 15}{\sqrt{10.5}}\Big) \\
&= 1 - \Phi(0.617) \\
&= 1 - 0.7291 \\
&= 0.2709
\end{aligned}$$

(b) Denote Y the time customs take. Although we don't know what distribution does Y follow, we have already known the mean is 2 and standard deviation is 0.6, then by Central Limit Theorem:

$$\begin{aligned}
P(1.8 \le \bar{Y} \le 2.25) &= P\Big(\sqrt{n}\frac{1.8 - E(\bar{Y})}{\sqrt{\bar{Y}}} \le \sqrt{n}\frac{\bar{Y} - E(\bar{Y})}{\sqrt{\bar{Y}}} \le \sqrt{n}\frac{2.25 - E(\bar{Y})}{\sqrt{\bar{Y}}}\Big) \\
&\approx P\Big(\sqrt{n}\frac{1.8 - 2}{0.6} \le z \le \sqrt{n}\frac{2.25 - 2}{0.6}\Big) \\
&= \Phi\Big(\frac{5}{12}\sqrt{50}\Big) - \Phi\Big(-\frac{1}{3}\sqrt{50}\Big) \\
&= \Phi(2.94) - \Phi(-2.53) \\
&= 0.9984 - 0.0094 \\
&= 0.989
\end{aligned}$$

Problem 4

(a) First note that $P(-2.3 < Z < 2.3) \approx .98$ where Z has the standard normal distribution. If $\mathcal{B}(n,p) = k$, the likelihood function for p is given by $L(p) = \binom{n}{k}p^k(1-p)^{n-k}$. Taking derivative and setting to zero (done in section, lecture), gives us that the maximum likelihood estimate is given by $p = \frac{k}{n}$, so the maximum likelihood estimator is given by $\hat{p} = \frac{X}{n}$. Now, we know that for n large, $\mathcal{B}(n,p)$ approximates the normal distribution with parameters np and $np(1-p)$. We normalize to have

$$\frac{\frac{X}{n} - np}{\sqrt{p(1-p)/n}} \sim Z$$

We also want to replace the p's in the denominator by $\frac{X}{n}$. We can do this by the Law of Large numbers, so

$$\frac{\frac{X}{n} - np}{\sqrt{\frac{X}{n}(1 - \frac{X}{n})/n}} \sim Z$$

Thus

$$.98 = P(-2.3 < Z < 2.3) = P\left(-2.3 < \frac{\frac{X}{n} - np}{\sqrt{\frac{X}{n}(1-\frac{X}{n})/n}} < 2.3\right) =$$

$$P\left(-2.3\sqrt{\frac{X}{n}(1-\frac{X}{n})/n} < \frac{X}{n} - np < 2.3\sqrt{\frac{X}{n}(1-\frac{X}{n})/n}\right) =$$

$$P\left(-2.3\sqrt{\frac{X}{n}(1-\frac{X}{n})/n} + \frac{X}{n} < p < 2.3\sqrt{\frac{X}{n}(1-\frac{X}{n})/n} + \frac{X}{n}\right)$$

(b) By the above, we can see that a one sided interval will be given by

$$1 - \alpha = P\left(p < z_\alpha\sqrt{\frac{X}{n}(1-\frac{X}{n})/n} + \frac{X}{n}\right)$$

(c) Recall that z_α is the upper quantile of standard normal distribution such that $\mathbb{P}(Z > z_\alpha) = \alpha$. (1) will give a confidence interval with only $\alpha(100)\%$ confidence level so it is incorrect. (2) is incorrect. It gives $100(\alpha)\%$ confidence interval. (3) will give $(\frac{\alpha}{2})(100)\%$ confidence, so it is incorrect. (4) is incorrect because it is a two sided interval.

(d) Note that $P(-2.4 > Z) \approx .99)$. So we do everything as in part (a) but with only one side (notice that the Z score is different)

$$.98 \approx P\left(-2.4 < \frac{\frac{X}{n} - p}{\sqrt{\frac{X}{n}(1-\frac{X}{n}/n}}\right) = P\left(p < \frac{X}{n} + 2.4\sqrt{\frac{X}{n}(1-\frac{X}{n})/n}\right).$$

Problem 5

First, by looking up the table, we will find:

$$\Phi(-1.64) = 0.05,$$
$$\Phi(2.33) = 0.99,$$
$$\Phi(2.58) = 0.995$$

(a) This is a two sided 94% confidence interval, the left tail is 5%, right tail is 1%;

(b) This is a right sided (one sided) 99.5% confidence interval, the left tail is 0%, right tail is 0.5%;

(c) This is a two sided 45% confidence interval, the left tail is 5%, right tail is 50%;

Problem 6

(a) A person either will get vaccinated or not. We have a binomial/Bernoulli variable. The propotion of the population that will is our parameter p, ie, the probability that a random person from the population will get vaccinated. We have 350 samples from the Bernoulli, so the MLE for p is $\frac{126}{350}$. We know how to construct confidence intervals for binomials using the normal approximation (done in lecture and in Larsen and Marx). Just plug in $\frac{126}{350}$ for $\frac{k}{n}$. We get $[\frac{126}{350} - z_{.05}\sqrt{\frac{\frac{126}{350}(1-\frac{126}{350})}{350}}, \frac{126}{350} + z_{.05}\sqrt{\frac{\frac{126}{350}(1-\frac{126}{350})}{350}}]$.

(b) Now we are told that there are 3500 people in the town and to use the finite correction factor. Read about the finite correction factor at the very end of the section (the last comment before the exercises for the section). Essentially, because people are selected without replacement, we have to tweak our formula for the variance to be $\frac{p(1-p)}{n}(\frac{N-n}{N-1})$ instead of just $\frac{p(1-p)}{n}$, where N is the size of the total population, and n is the number drawn. Well, this change leads to a change in the formula for our confidence interval. We now have
$[\frac{126}{350} - z_{.05}\sqrt{\frac{\frac{126}{350}(1-\frac{126}{350})}{350}(\frac{3500-350}{3499})}, \frac{126}{350} + z_{.05}\sqrt{\frac{\frac{126}{350}(1-\frac{126}{350})}{350}(\frac{3500-350}{3499})}]$.

Problem 7

We have X $\sim$ Pois(λ). We want to know if $\hat{\lambda} = \frac{1}{n}\sum X_i$ is efficient. Well, var($\hat{\lambda}$) $= \frac{1}{n}var(X) = \frac{\lambda}{n}$. We need to check if that's the Cramer-Rao lower bound or not. The lower bound is given by $(-nE[\frac{\partial^2 ln f_X(X;\lambda)}{\partial\lambda^2}])^{-1}$. Well, $lnf_X(X) = -\lambda + Xln(\lambda) - ln(X!)$. Taking two derivatives with respect to λ results in $-\frac{1}{\lambda^2}X$. The expected value of that is simply $-\frac{1}{\lambda^2}\lambda = -\frac{1}{\lambda}$. Now, we have to multiply by $-n$ and take the reciprocal. The result is $\frac{\lambda}{n}$, which is what we wanted!

Challenge Problem

(a) First we need to know what is $Var(X)$:

$$
\begin{aligned}
Var(X) &= EX^2 - (EX)^2 \\
&= \int_{2\theta}^{10} \frac{x^2}{10-2\theta}dx - \Big(\int_{2\theta}^{10} \frac{x}{10-2\theta}dx\Big)^2 \\
&= \frac{1}{3(10-2\theta)} \int_{2\theta}^{10} dx^3 - \Big(\frac{1}{2(10-2\theta)} \int_{2\theta}^{10} dx^2\Big)^2 \\
&= \frac{10^3 - (2\theta)^3}{3(10-2\theta)} - \Big(\frac{10^2 - (2\theta)^2}{2(10-2\theta)}\Big)^2 \\
&= \frac{10^2 + 10 \cdot 2\theta + (2\theta)^2}{3} - \frac{(10+2\theta)^2}{4} \\
&= \frac{1}{12}(10-2\theta)^2 \\
&= \frac{(5-\theta)^2}{3}
\end{aligned}
$$

Thus, by method of Plug-in, the MLE of $Var(X)$ is $\frac{1}{3}(5 - \frac{X_{min}}{2})^2$.

(b) We claim it's biased. To see why, recall the PDF of $\frac{1}{2}X_{min}$, we have:

$$
\begin{aligned}
E\Big(\frac{1}{3}(5 - \frac{X_{min}}{2})^2\Big) &= \int_{\theta}^{5} \frac{1}{3}(5-x)^2 f_{\frac{1}{2}X_{min}}(x)dx \\
&= \int_{\theta}^{5} \frac{1}{3}(5-x)^2 \cdot 2n\Big(\frac{5-x}{5-\theta}\Big)^{n-1} \frac{1}{10-2\theta}dx \\
&= \frac{n}{3}\int_{\theta}^{5} \frac{(5-x)^{n+1}}{(5-\theta)^n}dx \\
&= -\frac{n}{3}\int_{5-\theta}^{0} \frac{y^{n+1}}{(5-\theta)^n}dy \\
&= \frac{n}{3(n+2)(5-\theta)^n}\int_{0}^{5-\theta} dy^{n+2} \\
&= \frac{n}{n+2} \cdot \frac{1}{3}(5-\theta)^2
\end{aligned}
$$

By the result, we can see, it's biased.