Homework 2 Solution

1. Consider i.i.d. observations $X_1, ..., X_n \sim$Bernoulli(p).

   (a) Find the MLE $\hat{p}$.

$$
\begin{aligned}
L(p) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\
l(p) &= \sum_{i=1}^{n} x_i \log p + \sum_{i=1}^{n} 1 - x_i \log(1-p) \\
l'(p) &= \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}(n - \sum_{i=1}^{n} x_i) = 0 \\
\hat{p} &= \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}
\end{aligned}
$$

   (b) Construct a Wald $(1-\alpha)$-confidence interval of p.

$$
\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{\hat{p}(1-\hat{p})}} \rightsquigarrow N(0,1)
$$

   The $(1-\alpha)$ CI is $[\hat{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$

   (c) For the above confidence interval $[L_n, U_n]$, its length is $U_n - L_n$. Find the minimal sample size n_0, such that $U_n - L_n \leq 0.06$ for any n regardless of the value of $\hat{p}$.

$$
\begin{aligned}
U_n - L_n = 2 * z_{1-\frac{\alpha}{2}} * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &\leq 0.06 \\
\frac{\hat{p}(1-\hat{p})}{n} &\leq (\frac{0.06}{2 * z_{1-\frac{\alpha}{2}}})^2 \\
\frac{\hat{p}(1-\hat{p})}{n} \leq \frac{1}{4n} &\leq (\frac{0.06}{2 * z_{1-\frac{\alpha}{2}}})^2 \\
n &\geq \frac{z_{1-\frac{\alpha}{2}}^2}{0.06^2}
\end{aligned}
$$

   Note: $p(1-p)$ has max at $p = \frac{1}{2}$.
   The smallest sample size n_0 is $\lceil \frac{2500}{9} z_{1-\frac{\alpha}{2}}^2 \rceil$.

   (d) Construct a Wilson $(1-\alpha)$-confidence interval of p.

$$|\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{p(1-p)}}| \leq z_{1-\frac{\alpha}{2}}$$

$$(\hat{p}-p)^2 \leq \frac{p(1-p)}{n} z^2_{1-\frac{\alpha}{2}}$$

$$f(p) = (1+\frac{z^2_{1-\frac{\alpha}{2}}}{n})p^2 - (2\hat{p}+\frac{z^2_{1-\frac{\alpha}{2}}}{n})p + \hat{p}^2 \leq 0$$

The two roots are $\frac{-b\pm\sqrt{b^2-4ac}}{2a}$

The CI is $[\frac{\hat{p}+\frac{z^2_{1-\frac{\alpha}{2}}}{2n} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z^2_{1-\frac{\alpha}{2}}}{4n}}}{1+\frac{z^2_{1-\frac{\alpha}{2}}}{n}}, \frac{\hat{p}+\frac{z^2_{1-\frac{\alpha}{2}}}{2n} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z^2_{1-\frac{\alpha}{2}}}{4n}}}{1+\frac{z^2_{1-\frac{\alpha}{2}}}{n}}]$

(e) Find a function g, such that $\sqrt{n}(g(\hat{p})-g(p)) \rightsquigarrow N(0,1)$.

$$\sqrt{n}(\hat{p}-p) \rightsquigarrow N(0, p(1-p))$$

We want $g'(p)^2 p(1-p) = 1$

$$g'(p) = \frac{1}{\sqrt{p(1-p)}}$$
$$g(p) = \arcsin(2p-1) \text{ or } 2\arcsin(\sqrt{p})$$

2. Consider i.i.d. observations $X_1, ..., X_n \sim N(\theta, 1)$.

(a) Find a 0.95-confidence interval of θ.

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp(-\frac{(X_i-\theta)^2}{2})$$
$$l(\theta) = -\frac{\sum(X_i-\theta)^2}{2} + c$$
$$l'(\theta) = \sum_{i=1}^{n} X_i - n\theta = 0$$
$$\hat{\theta} = \bar{X}$$

$$\sqrt{n}(\hat{\theta}-\theta) \sim N(0,1)$$

The 95% CI is $[\hat{\theta} - z_{1-\frac{\alpha}{2}}\frac{1}{\sqrt{n}}, \hat{\theta} + z_{1-\frac{\alpha}{2}}\frac{1}{\sqrt{n}}] = [\hat{\theta} - 1.96\frac{1}{\sqrt{n}}, \hat{\theta} + 1.96\frac{1}{\sqrt{n}}]$

(b) Consider $n = 100$ and $\theta = 3$, generate data and calculate the confidence interval with R. Does the confidence interval cover the true θ?

```
> n = 100
> theta = 3
> set.seed(2018)
> X = rnorm(n, theta, 1)
> mean(X) - qnorm(0.975)/sqrt(n)
[1] 2.825196
> mean(X) + qnorm(0.975)/sqrt(n)
[1] 3.217188
```

The CI cover the true θ.

(c) Repeat the above experiments for 10000 times, for how many times the confidence interval cover the true θ? Explain your discovery.

```
> n = 100
> theta = 3
> set.seed(2018)
> result = numeric(10000)
> for(i in 1:10000){
+   X = rnorm(n, theta, 1)
+   L = mean(X) - qnorm(0.975)/sqrt(n)
+   U = mean(X) + qnorm(0.975)/sqrt(n)
+   if(L <= 3 && U >= 3){
+     result[i] = 1
+   }
+ }
> sum(result)
[1] 9489
```

In my experiment, 9489 out of 10000 times the confidence interval cover the true θ. There are about 95% confidence levels that contain the true value.

(d) Run the experiments above for $n = 50$ and $\theta = 20$. Does the conclusion change? Why?

```
> n = 50
> theta = 20
> set.seed(2018)
> result = numeric(10000)
> for(i in 1:10000){
+   X = rnorm(n, theta, 1)
+   L = mean(X) - qnorm(0.975)/sqrt(n)
+   U = mean(X) + qnorm(0.975)/sqrt(n)
+   if(L <= theta && U >= theta){
+     result[i] = 1
+   }
+ }
> sum(result)
[1] 9489
```

The conclusion does not change. The X has normal distribution, so the normality of $\hat{\theta}$ will not be influenced by the smaller sample size. Using the correct distribution, the confidence interval contains the true value 95% of the time.

3. Consider i.i.d. observations $X_1, ..., X_n \sim N(0, \sigma^2)$.

(a) Find the MLE $\hat{\sigma}^2$.

$$L(\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp(-\frac{\sum_{i=1}^n X_i^2}{2\sigma^2})$$
$$l(\sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{\sum_{i=1}^n X_i^2}{2\sigma^2} + c$$
$$l'(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n X_i^2}{2(\sigma^2)^2} = 0$$
$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n X_i^2}{n}$$

(b) Use CLT to construct a Wald $(1-\alpha)$-confidence interval of σ^2.

$$E(\frac{\sum_{i=1}^n X_i^2}{n}) = \sigma^2$$
$$Var(\frac{\sum_{i=1}^n X_i^2}{n}) = \frac{2\sigma^4}{n}$$
$$\frac{\sqrt{n}(\hat{\sigma}^2 - \sigma^2)}{\sqrt{2}\sigma^2} \rightsquigarrow N(0,1)$$

The $(1-\alpha)$ CI is $[\hat{\sigma}^2 - z_{1-\frac{\alpha}{2}}\frac{\sqrt{2}\hat{\sigma}^2}{\sqrt{n}}, \hat{\sigma}^2 + z_{1-\frac{\alpha}{2}}\frac{\sqrt{2}\hat{\sigma}^2}{\sqrt{n}}]$

Note: Ways to compute var(X^2):

- var(X^2) = E(X^4) – E(X^2)^2
  Compute $E(X^4)$ by definition:
  $$E(X^4) = \int_{-\infty}^{\infty} x^4 f(x)dx$$
  Or using χ^2
- Easier way: $X \sim N(0,\sigma^2) \to \frac{X}{\sigma} \sim N(0,1) \to (\frac{X}{\sigma})^2 \sim \chi_1^2$
  $\Rightarrow$
  $$\mathrm{var}((\frac{\mathrm{X}}{\sigma})^2) = 2$$
  $$\mathrm{var}(\mathrm{X}^2) = 2\sigma^4$$
  $$Var(\frac{\sum_{i=1}^n X_i^2}{n}) = \frac{2\sigma^4}{n}$$

(c) What is the exact distribution of $\frac{1}{\sigma^2}\sum_{i=1}^n X_i^2$?

$$\frac{X_i}{\sigma} \sim N(0,1)$$
$$\frac{X_i^2}{\sigma^2} \sim \chi_1^2$$
$$\frac{\sum_{i=1}^n X_i^2}{\sigma^2} \sim \chi_n^2$$

(d) Use the above fact to construct a confidence interval for σ^2.

$$\chi^2_{n,\frac{\alpha}{2}} \leq \frac{\sum_{i=1}^n X_i^2}{\sigma^2} \leq \chi^2_{n,1-\frac{\alpha}{2}}$$
$$\frac{\sum_{i=1}^n X_i^2}{\chi^2_{n,1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{\sum_{i=1}^n X_i^2}{\chi^2_{n,\frac{\alpha}{2}}}$$

The $1-\alpha$ CI is $[\frac{\sum_{i=1}^n X_i^2}{\chi^2_{n,1-\frac{\alpha}{2}}}, \frac{\sum_{i=1}^n X_i^2}{\chi^2_{n,\frac{\alpha}{2}}}]$

(e) For the two confidence intervals above, compare them with numerical experiments in R. For example, take $n = 100$ and $\sigma^2 = 1$, conduct repeated experiments and compare the coverage and the length of the two confidence intervals.

n	10	100	1000
Wald Coverage	0.8731	0.9452	0.9505
Chi Coverage	0.9505	0.9521	0.95
Wald Length	1.7629	0.5560	0.1754
Chi Length	2.6062	0.5772	0.1760

As n increases, the coverage of Wald converge to 95%. This is caused by the CLT, the distribution converges to normal as n increases. The coverage for chi square is always 95%, since this is the exact distribution of $\hat{\sigma}^2$. We don't need a large n.

Wald CI has shorter length in the experiments.

R code:

```
Coverage.Length = function(n){
sigma2 = 1
CI1 = matrix(0, nrow = 10000, ncol=3)
CI2 = matrix(0, nrow = 10000, ncol=3)
colnames(CI1) = c('L','U','cover')
colnames(CI2) = c('L','U','cover')
set.seed(2018)
for(i in 1:10000){
X = rnorm(n, 0, sigma2)
sigma2hat = sum(X^2)/n
CI1[i,1:2] = c(sigma2hat - qnorm(0.975)*sqrt(2)*sigma2hat/sqrt(n),
    sigma2hat + qnorm(0.975)*sqrt(2)*sigma2hat/sqrt(n))
CI2[i,1:2] = c(sum(X^2)/qchisq(0.975,df=n), sum(X^2)/qchisq(0.025,
    df=n))
if(CI1[i,1] <= 1 && CI1[i,2] >= 1){
CI1[i,3] = 1
}
if(CI2[i,1] <= 1 && CI2[i,2] >= 1){
CI2[i,3] = 1
}
}
return(list('wald.converage' = mean(CI1[,'cover']), 'chi.converage'
    = mean(CI2[,'cover']),
```

```
'wald.length' = mean(CI1[,'U'] - CI1[,'L']),
'chi.length' = mean(CI2[,'U'] - CI2[,'L'])))
}
n10 = Coverage.Length(10)
n100 = Coverage.Length(100)
n1000 = Coverage.Length(1000)
```

4. Given an estimator $\hat{\theta}$ of θ. The performance is measured by the mean-squared error, defined as $\mathbb{E}(\hat{\theta} - \theta)^2$.

   (a) Show $\mathbb{E}(\hat{\theta} - \theta)^2 = \text{var}(\hat{\theta}) + (\mathbb{E}\hat{\theta} - \theta)^2$. In other words, the error is the sum of variance and bias squared.

$$\begin{aligned} E(\hat{\theta} - \theta)^2 &= E(\hat{\theta} - E\hat{\theta} + E\hat{\theta} - \theta)^2 \\ &= E(\hat{\theta} - E\hat{\theta})^2 + (E\hat{\theta} - \theta)^2 + 2E\left[(\hat{\theta} - E\hat{\theta})(E\hat{\theta} - \theta)\right] \\ &= E(\hat{\theta} - E\hat{\theta})^2 + (E\hat{\theta} - \theta)^2 + 2E\left[(\hat{\theta} - E\hat{\theta})\right](E\hat{\theta} - \theta) \\ &= var(\hat{\theta}) + (E\hat{\theta} - \theta)^2 \end{aligned}$$

   Note: $E(\hat{\theta} - E\hat{\theta}) = 0$

   (b) Consider i.i.d. observations $X_1, ..., X_n \sim N(\mu, \sigma^2)$. Show that $\hat{\sigma}_1^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ is an unbiased estimator of σ^2.

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n X_i^2 - 2\bar{X}\sum_{i=1}^n X_i + \bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - n\bar{X}^2 \end{aligned}$$

$$\begin{aligned} E\hat{\sigma}^2 &= \frac{1}{n-1}\left[\sum_{i=1}^n EX_i^2 - nE\bar{X}^2\right] \\ &= \frac{n}{n-1}(\sigma^2 + \mu^2) - \frac{n}{n-1}(\frac{\sigma^2}{n} + \mu^2) \\ &= \sigma^2 \end{aligned}$$

   (c) Calculate $\mathbb{E}(\hat{\sigma}_1^2 - \sigma^2)^2$.

$$\begin{aligned} E(\hat{\sigma}_1^2 - \sigma^2)^2 &= \text{var}(\hat{\sigma}_1^2) + (\mathbb{E}\hat{\sigma}_1^2 - \sigma^2)^2 \\ &= \text{var}(\hat{\sigma}_1^2) \\ &= \frac{2\sigma^4}{n-1} \quad \text{since} \quad \left(\frac{(n-1)\hat{\sigma}_1^2}{\sigma^2} \sim \chi^2_{n-1}\right) \end{aligned}$$

Optional:
The last equality comes from the fact that $\frac{\sum_{i=1}^n (X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$. Therefore,

$$Var(\sum_{i=1}^n (X_i - \bar{X})^2) = 2(n-1)\sigma^4$$

To prove $\frac{\sum_{i=1}^n (X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$,

$$\begin{aligned}
\sum_{i=1}^n (\frac{X_i - \mu}{\sigma})^2 &= \sum_{i=1}^n (\frac{X_i - \bar{X}}{\sigma} + \frac{\bar{X} - \mu}{\sigma})^2 \\
&= \sum_{i=1}^n (\frac{X_i - \bar{X}}{\sigma})^2 + 2\frac{\bar{X} - \mu}{\sigma} \sum_{i=1}^n \frac{X_i - \bar{X}}{\sigma} + n(\frac{\bar{X} - \mu}{\sigma})^2 \\
&= \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 + (\frac{\bar{X} - \mu}{\sigma/\sqrt{n}})^2
\end{aligned}$$

The left hand side is χ^2_n, the last term on the right hand side is χ^2_1. Based on the moment generating function,

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2_{n-1}$$

(d) An alternative estimator is $\hat{\sigma}_2^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$. Calculate $\mathbb{E}(\hat{\sigma}_2^2 - \sigma^2)^2$.

$$\begin{aligned}
E\hat{\theta}_2^2 &= \frac{1}{n} \sum_{i=1}^n EX_i^2 - E\bar{X}^2 \\
&= \sigma^2 + \mu^2 - (\frac{\sigma^2}{n} + \mu^2) \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(\hat{\sigma}_2^2 - \sigma^2)^2 &= \mathrm{var}(\hat{\theta}_2^2) + \frac{\sigma^2}{\mathrm{n}^2} \\
&= \frac{2(n-1)\sigma^4}{n^2} + \frac{\sigma^2}{n^2} = \frac{2n-1}{n^2}\sigma^4
\end{aligned}$$

(e) Consider a general estimator $\hat{\sigma}_c^2 = c \sum_{i=1}^n (X_i - \bar{X})^2$. Calculate $\mathbb{E}(\hat{\sigma}_c^2 - \sigma^2)^2$. Which c is the best?

$$E(\hat{\sigma}_c^2) = cn(\sigma^2 + \mu^2) - cn(\frac{\sigma^2}{n} + \mu^2) = c(n-1)\sigma^2$$

$$\mathrm{var}(\hat{\sigma}_{\mathrm{c}}^2) = \mathrm{c}^2 2(\mathrm{n}-1)\sigma^4$$

$$\mathbb{E}(\hat{\sigma}_c^2 - \sigma^2)^2 = c^2 2(n-1)\sigma^4 + (c(n-1)-1)^2\sigma^4$$

Take derivative with respect to c

$$4c(n-1)\sigma^4 + 2(cn-c-1)\sigma^4(n-1) = 0$$
$$c = \frac{1}{n+1}$$

With $c = \frac{1}{n+1}$

$$\begin{aligned}\mathbb{E}(\hat{\sigma}_c^2 - \sigma^2)^2 &= c^2 2(n-1)\sigma^4 + (c(n-1)-1)^2\sigma^4 \\ &= \frac{2\sigma^4}{n+1}\end{aligned}$$

(f) Discuss your results.
$\hat{\sigma}_2^2$ is biased, $\hat{\sigma}_1^2$ is unbiased.
$\hat{\sigma}_2^2$ has smaller MSE than $\hat{\sigma}_1^2$ by trading off variance for bias.
We have

$$MSE_1 \geq MSE_2 \geq MSE_c$$

5. Consider i.i.d. observations $X_1, ..., X_n \sim N(\mu, \mu^2)$. For $\hat{\mu} = \bar{X}$, find a function g, such that $\sqrt{n}(g(\hat{\mu}) - g(\mu)) \rightsquigarrow N(0,1)$.

$$\sqrt{n}(\hat{\mu} - \mu) \rightsquigarrow N(0, \mu^2)$$

We want $g'(\mu)^2\mu^2 = 1$

$$g'(\mu) = \sqrt{\frac{1}{\mu^2}}$$
$$g(\mu) = \log \mu$$