Chapter 5 Solutions

Confidence intervals

1. (a) Let $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$ where both μ and σ^2 are unknown parameters. Using the fact that, the sample variance, S^2, is distributed as $(n-1)S^2/\sigma^2 \sim \mathcal{X}^2_{n-1}$, derive a level $1-\alpha$ confidence interval for σ^2. ($\mathcal{X}^2_k$ denotes the *chi-squared distribution with k degrees of freedom*, which is available in python as `scipy.stats.chi2`. It is a special case of the gamma distribution with shape $k/2$ and rate $1/2$.)

We know that $\dfrac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}^2_{n-1}$ so $Y = \dfrac{(n-1)S^2}{\sigma^2}$ is a pivot function because its distribution does not depend on the unknown parameter σ^2. Let $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ denote the $\alpha/2$ and $1-\alpha/2$ left quantiles of the $\mathcal{X}^2_{n-1}$ distribution respectively. Then we have the following inequality which holds with probability $1-\alpha$

$$\begin{aligned}
&\chi^2_{\alpha/2} < Y < \chi^2_{1-\alpha/2} \\
&\Rightarrow \chi^2_{\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{1-\alpha/2} \\
&\Rightarrow \frac{1}{\chi^2_{1-\alpha/2}} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{\chi^2_{\alpha/2}} \\
&\Rightarrow \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2}},
\end{aligned}$$

so a level $1-\alpha$ confidence interval is $\left[\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}}, \dfrac{(n-1)S^2}{\chi^2_{\alpha/2}}\right]$

(b) Suppose that the following data were observed

$$-1.90,\ -0.89,\ -0.87,\ -0.65,\ -0.32,\ -0.25,\ 0.90,\ 1.00,\ 1.18.$$

For these data $n = 9$, $\bar{x} = -0.2$, and $S^2 = 1.073$. Calculate a 95% confidence interval for σ^2.

From Python, we have

```
>>> import scipy.stats
>>> scipy.stats.chi2.ppf([.025,.975],8)
array([ 2.17973075, 17.53454614])
```

so $\chi^2_{\alpha/2} = 2.18$ and $\chi^2_{1-\alpha/2} = 17.53$. Then,

$$L = \frac{(9-1)(1.073)}{17.53} = 0.49$$
$$U = \frac{(9-1)(1.073)}{2.18} = 3.94,$$

so the 95% confidence interval is $[0.49, 3.94]$.

2. (a) Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{U}(0, \theta)$, $\theta > 0$. By considering an appropriate pivot construct a level $1-\alpha$ confidence interval for θ. *Hint.* Let $W_i = X_i/\theta$.

Let $W_i = X_i/\theta$. Then $W_i \sim U(0,1)$ so if we let $W_{(n)} = \max\{W_1, \ldots, W_n\}$, then the distribution of $W_{(n)}$ does not depend on θ. Note that by letting $X_{(n)} = \max\{X_1, \ldots, X_n\}$, $W_{(n)} = X_{(n)}/\theta$.
The CDF of $W_{(n)}$ is

$$\begin{aligned}
F_W(w) &= \mathbb{P}(W_{(n)} \le w) \\
&= \mathbb{P}(\max\{W_1, \ldots, W_n\} \le w) \\
&= \mathbb{P}(W_1 \le w, \ldots, W_n \le w) \\
&= \mathbb{P}(W_1 \le w) \ldots \mathbb{P}(W_n \le w), \text{ by independence} \\
&= w \times \ldots \times w \\
&= w^n
\end{aligned}$$

So if we let $c_1 = (\alpha/2)^{\frac{1}{n}}$ and $c_2 = (1-\alpha/2)^{\frac{1}{n}}$, with probability $1-\alpha$,

$$\begin{aligned}
&c_1 < W_{(n)} < c_2 \\
&c_1 < X_{(n)}/\theta < c_2 \\
&\frac{1}{c_2} < \theta/X_{(n)} < \frac{1}{c_1} \\
&\frac{X_{(n)}}{c_2} < \theta < \frac{X_{(n)}}{c_1},
\end{aligned}$$

so a level $1-\alpha$ confidence interval is $\left[\frac{X_{(n)}}{c_2}, \frac{X_{(n)}}{c_1}\right]$

(b) Suppose that the following data were observed

$$0.90,\ 1.00,\ 1.18,\ 1.90,\ 2.20.$$

Calculate a 95% confidence interval for θ. Does the confidence interval contain the maximum likelihood estimator for θ?

For the given data, $n = 5$ and $\max x = 2.20$.

For 95% confidence, $\alpha = 0.05$, so $c_1 = (0.025)^(1/5) = 0.48$, and $c_2 = (0.975)^(1/5) = 0.99$. Then,

$$L = \frac{2.20}{0.99} = 2.22$$
$$U = \frac{2.20}{0.48} = 4.58,$$

so the confidence interval is $[2.22, 4.58]$.

We can see that the confidence interval does not contain the maximum likelihood estimator, which equals 2.20. The confidence interval provides a more sensible range of values than the maximum likelihood estimator, which underestimates θ for this model as we have observed in the previous chapter.

Hypothesis testing

1. The author of a weight-loss diet claims that an average adult, weighting 100 Kg, who follows the proposed diet, will lose 20 Kg after 1 month. What are the null and alternative hypotheses?

   Let μ denote the average person's weight after one month of diet. Then H_0: $\mu = 100$ vs H_1: $\mu = 80$.

2. The author of a weight-loss diet claims that an average adult, weighting 100 Kg, who follows the proposed diet, will lose weight after 1 month. What are the null and alternative hypotheses?

   Let μ denote the average person's weight after one month of diet. Then H_0: $\mu = 100$ vs H_1: $\mu < 100$.

3. The author of a weight-loss diet claims that an average adult, weighting 100 Kg, who follows the proposed diet, will notice a change in their weight after 1 month. What are the null and alternative hypotheses?

Let μ denote the average person's weight after one month of diet. Then H_0: $\mu = 100$ vs H_1: $\mu \neq 100$.

4. The author of a weight-loss diet claims that an average adult, weighting 100 Kg, who follows the proposed diet, will lose weight after 1 month. An experiment was conducted to verify this claim. Three adults, who weighted 100 Kg, followed the diet for one month and their weights at the end of the month were recorded. The experimenters would accept the author's claim if the sample mean $\bar{X}$ of the three measured weights is less than 90. Suppose that the population standard deviation is $\sigma = 15$.

   (a) The three people's weights after the end of the month were: 82, 86, and 93. What is the experimenters' conclusion?

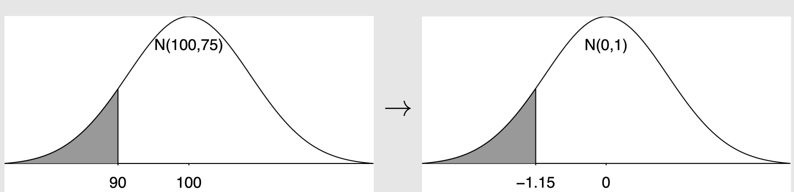
   In this case $\bar{x} = (82 + 86 + 93)/3 = 87$. Because $\bar{x} < 90$ the experimenters will conclude that there is evidence that the diet helps people loose weight.

   (b) According to the central limit theorem, what is the asymptotic distribution of the sample mean of $n = 3$ measurements from a population with mean $\mu = 100$ and standard deviation $\sigma = 15$?

   The central limit theorem says that the distribution of the sample mean $\bar{X}$ approximates the normal distribution with mean μ and variance $\sigma^2/n = 15^2/3 = 75$. In this case it will be the $\mathrm{N}(100, 75)$ distribution.

   (c) Use the central limit theorem to calculate the probability of Type I error of the experimenters' decision rule.

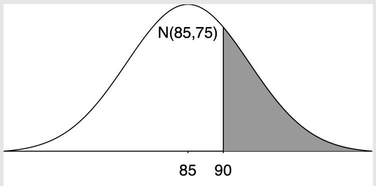
   The experimenters' rule is "reject H_0 if $\bar{X} < 90$".
   $\mathbb{P}(\text{Type I error}) = \mathbb{P}(\bar{X} < 90 | \mu = 100)$.
   Using the central limit theorem: $\bar{X} \sim \mathrm{N}(100, 75)$. Then to find the probability:



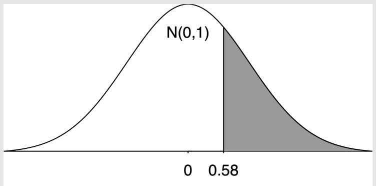
$z = \dfrac{90 - 100}{\sqrt{75}} = -1.15 \Rightarrow \Phi(-1.15) = \mathbf{0.1251}$ is the probability of Type I error.

(d) Use the central limit theorem to calculate the probability of Type II error of the experimenters' decision rule assuming that the average weight after one month is 85.

The experimenters' rule is "reject H_0 if $\bar{X} < 90$".
$\mathbb{P}(\text{Type II error}) = \mathbb{P}(\bar{X} \geq 90 | \mu = 85)$.
Using the central limit theorem: $\bar{X} \sim \text{N}(85, 75)$ because now $\mu = 85$. Then to find the probability:



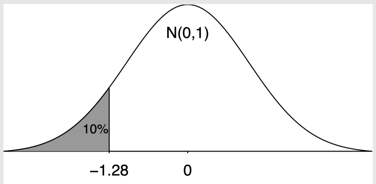
$\rightarrow$



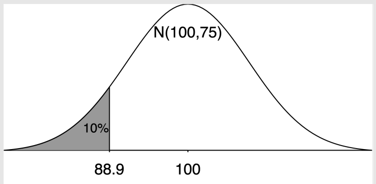
$z = \dfrac{90 - 85}{\sqrt{75}} = 0.58 \Rightarrow \Phi(0.58) = 0.7190 \Rightarrow 1 - 0.7190 = \mathbf{0.2810}$ is the probability of Type II error.

(e) Propose a rule of the form "accept the author's claim if $\bar{X} < c$" (in other words find c) such that the probability of Type I error is 10%.

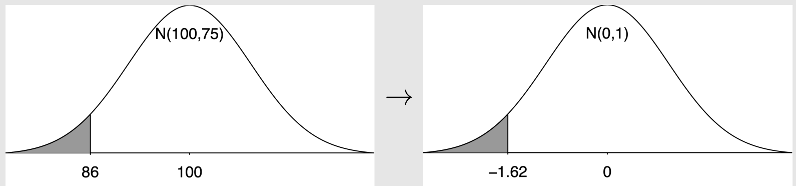
In this case we have to do the reverse calculation from 4c. If we want $\Phi(z) = 0.10$, then we must choose $z = -1.28$. In this case $\dfrac{c - 100}{\sqrt{75}} = -1.28 \Rightarrow c = 100 - (1.28)\sqrt{75} = \mathbf{88.9}$.



$\rightarrow$



(f) Suppose that in the sample we find that $\bar{x} = 86$. Find the p-value. What is your conclusion at significance level $\alpha = 5\%$?



$z = \dfrac{86 - 100}{\sqrt{75}} = -1.62 \Rightarrow \Phi(-1.62) = \mathbf{0.0526}$ is the p-value.

Since p-value $\nless \alpha$, we do not reject the null hypothesis at the 5% level.