Confidence intervals

Exercise 1 (Gaussian model – unknown mean, known variance)**:**
Let $\bar{X}$ be the empirical mean. We know that $Z = \frac{\bar{X}-\theta}{\sqrt{\frac{1}{n}}} \sim \mathcal{N}(0,1)$. Let Q be the quantile function of the standard normal distribution. We look for confidence intervals at level $1-\alpha$, with $\alpha = 10\%$.

1. Let $c = Q(1-\frac{\alpha}{2}) = Q(0.95) \approx 1.64$.
   Then $P(-c < Z < c) = 1 - 2P(Z > c) = 1 - \alpha$.
   We get :
   $$P_\theta(-c < \frac{\bar{X}-\theta}{\sqrt{\frac{1}{n}}} < c) = 1 - \alpha.$$
   We obtain the confidence interval $(\bar{X} - \frac{c}{\sqrt{n}}, \bar{X} + \frac{c}{\sqrt{n}})$.

2. Let $c' = Q(1-\alpha) = Q(0.9) \approx 1.28$.
   Then $P(Z < -c') = 1 - \alpha$ and
   $$P_\theta(-c' < \frac{\bar{X}-\theta}{\sqrt{\frac{1}{n}}}) = 1 - \alpha.$$
   We obtain the upper confidence bound $\bar{X} + \frac{c'}{\sqrt{n}}$.

3. Confidence interval $(1.67, 2.32)$, Upper confidence bound 2.26.

Exercise 2 (Gaussian model – unknown mean and variance)**:**

1. The average daily consumption is $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ with $\theta = (\mu, \sigma^2)$. We have:
   $$\frac{\bar{X}-\mu}{\sqrt{V/(n-1)}} \sim \mathrm{St}(n-1),$$
   where V is the empirical variance. Let Q be the quantile function of $T \sim \mathrm{St}(n-1)$. Then
   $$P(-c < T < c) = 1-\alpha \quad \Longleftrightarrow \quad c = Q(1-\frac{\alpha}{2})$$
   so that
   $$P(\mu \in [\bar{X} - \sqrt{\frac{V}{n-1}}Q(1-\frac{\alpha}{2}), \bar{X} + \sqrt{\frac{V}{n-1}}Q(1-\frac{\alpha}{2})]) = 1-\alpha.$$
   For $n = 10$, $Q(1-\frac{\alpha}{2}) \approx 2.26$ for $\alpha = 5\%$. The interval is $[85, 115]$.

2. Since 90 is in the confidence interval, you accept the null hypothesis that the mean is 90.

Exercise 3 (Gaussian model – known mean, unknown variance)**:**
The empirical variance V satisfies $\frac{nV}{\theta} \sim \chi^2(n)$. Let Q be the quantile function of $Z \sim \chi^2(n)$.

1. Let c_1, c_2 such that $P(Z < c_1) = P(Z > c_2) = \frac{\alpha}{2}$. Then:

$$P(\theta < \frac{nV}{c_2}) = \frac{\alpha}{2}, \quad P(\theta > \frac{nV}{c_1}) = \frac{\alpha}{2}.$$

   Confidence interval $[\frac{nV}{c_2}, \frac{nV}{c_1}]$. For $n = 50$, $c_1 \approx 28$, $c_2 \approx 79$ and confidence interval $(0.0063, 0.0178)$ (that is, $(0.079, 0.134)$ for the standard deviation).

2. Let c be such that $P(Z < c) = \alpha$. Then:

$$P(\theta > \frac{nV}{c}) = \alpha.$$

   Upper confidence bound $\frac{nV}{c}$. For $n = 50$, $c \approx 29$ and upper confidence bound 0.017 (that is, 0.130 for the standard deviation).

Exercise 4 (Poisson model)**:**
The total number of emails satisfies $S \sim \mathcal{P}(n\theta) \approx \mathcal{N}(n\theta, n\theta)$, that is $S \approx n\theta + \sqrt{n\theta}Z$ with $Z \sim \mathcal{N}(0,1)$. To get an upper bound on θ, we need a lower bound on Z. Let Q be the quantile function of Z:

$$P(Z > c) = 1 - \alpha \quad \iff \quad c = Q(\alpha).$$

We get:

$$P(\frac{S - n\theta}{\sqrt{n\theta}} > c) = 1 - \alpha.$$

For $\alpha < 0.5$, we have $c < 0$ so the inequality is satisfied if and only if $S > n\theta$ or:

$$(S - n\theta)^2 < n\theta c^2,$$

that is, with $\bar{x} = S/n$, either $\theta < \bar{x}$ or

$$\theta < \bar{x} + \frac{c^2}{2n} + \sqrt{\frac{c^2}{n}(\bar{x} + \frac{c^2}{4n})} \approx \bar{x} + \sqrt{\frac{c^2}{n}\bar{x}}$$

for large n. Upper confidence bound $\bar{x} + \sqrt{\frac{c^2}{n}\bar{x}}$. For $\alpha = 0.05$, $c = -1.64$ so 22.

Exercise 5 (Uniform model)**:**
We have:

$$M(X) = \max(X_1, \dots, X_n) \sim \theta Z$$

with:

$$Z = \max(U_1, \dots, U_n)$$

and $U_1, \dots, U_n$ i.i.d. uniform random variables on $[0,1]$. Then:

$$P(Z < u) = u^n.$$

We get:
$$P(Z < c) = 1 - \alpha$$
for $c = (1-\alpha)^{1/n}$. Now:
$$P(\theta > M(X)/c) = 1 - \alpha.$$
Application: $M(x) = 3$ years, $c \approx 0.98$ so $\theta > 3.06$ years.

Exercise 6 (Bernoulli model)**:**
The total number of electric cars satisfies $S \sim \mathcal{B}(n,\theta) \approx \mathcal{N}(n\theta, n\theta(1-\theta))$, that is $S \approx n\theta + \sqrt{n\theta(1-\theta)}Z$ with $Z \sim \mathcal{N}(0,1)$. Let Q be the quantile function of Z.

1. Upper confidence bound:
$$P(Z > c) = 1 - \alpha \quad \iff \quad c = Q(\alpha).$$
We get:
$$P(\frac{S - n\theta}{\sqrt{n\theta(1-\theta)}} > c) = 1 - \alpha.$$
For $\alpha < 0.5$, we have $c < 0$ so the inequality is satisfied if $S > n\theta$ or:
$$(S - n\theta)^2 < n\theta(1-\theta)c^2,$$
that is, with $\bar{x} = S/n$, if $\theta < \bar{x}$ or
$$\theta < \left(\bar{x} + \frac{c^2}{2n} + \sqrt{\frac{c^2}{n}(\bar{x}(1-\bar{x}) + \frac{c^2}{4n})}\right) / (1 + \frac{c^2}{n}) \approx \bar{x} + \sqrt{\frac{c^2}{n}\bar{x}(1-\bar{x})}.$$
For $\alpha = 0.01$, $c \approx -2.3$ and $\theta < 0.29$.

2. Lower confidence bound:
$$P(Z < c) = 1 - \alpha \quad \iff \quad c = Q(1-\alpha).$$
We get:
$$P(\frac{S - n\theta}{\sqrt{n\theta(1-\theta)}} < c) = 1 - \alpha.$$
For $\alpha < 0.5$, we have $c > 0$ so the inequality is satisfied if $S < n\theta$ or:
$$(S - n\theta)^2 < n\theta(1-\theta)c^2,$$
that is,
$$\theta > \left(\bar{x} + \frac{c^2}{2n} - \sqrt{\frac{c^2}{n}(\bar{x}(1-\bar{x}) + \frac{c^2}{4n})}\right) / (1 + \frac{c^2}{n}) \approx \bar{x} - \sqrt{\frac{c^2}{n}\bar{x}(1-\bar{x})}.$$
For $\alpha = 0.01$, $c \approx 2.3$ and $\theta > 0.11$.

3. Confidence interval:

$$P(|Z| < c) = 1 - \alpha \quad \Longleftrightarrow \quad c = Q(1 - \frac{\alpha}{2}).$$

We get:

$$P(\left|\frac{S - n\theta}{\sqrt{n\theta(1-\theta)}}\right| < c) = 1 - \alpha.$$

The inequality is satisfied if:

$$(S - n\theta)^2 < n\theta(1-\theta)c^2,$$

that is,

$$\theta \in \left[\left(\bar{x} + \frac{c^2}{2n} \pm \sqrt{\frac{c^2}{n}(\bar{x}(1-\bar{x}) + \frac{c^2}{4n})}\right)/(1 + \frac{c^2}{n})\right] \approx \left[\bar{x} \pm \sqrt{\frac{c^2}{n}\bar{x}(1-\bar{x})}\right].$$

For $\alpha = 0.01$, $c \approx 2.6$ and $\theta \in [0.1, 0.3]$.