MATH 3215 Assignment 8

(Note that MLE and MSE are different acronyms.)

1. Suppose that 52 percent of the population are in favor of a policy. Approximate the probability that at least 50 percent of an i.i.d. sample of size n are in favor of the policy, where (a) $n = 10$; (b) $n = 100$; (c) $n = 1000$; (d) $n = 10000$. (It suffices to do normal approximation without the correction for continuity. Compute the numerical values up to 3 digits after the decimal point.)

Consider i.i.d. $X_i \sim \mathsf{Ber}(0.52)$ for $i = 1, \dots, n$. Then

$$\mathbb{P}\Big\{\sum_{i=1}^n X_i \geq 0.5n\Big\} = \mathbb{P}\Big\{\frac{\sum_{i=1}^n X_i - 0.52n}{\sqrt{n \cdot 0.52 \cdot 0.48}} \geq \frac{-0.02n}{\sqrt{n \cdot 0.52 \cdot 0.48}}\Big\} \approx 1 - \Phi\Big(\frac{-0.02n}{\sqrt{n \cdot 0.52 \cdot 0.48}}\Big).$$

Plugging the values for n yields 0.550, 0.656, 0.897, 1.000.

2. The temperature at which a thermostat goes off is normally distributed with standard deviation 0.5 degrees.

(a) If the thermostat is to be tested five times independently, what is $\mathbb{P}\{0.2 \leq S^2 \leq 0.3\}$ where S^2 is the sample variance of the five data values? (Express the result in terms of the CDF of a chi-squared distribution and compute the final numerical value up to 4 digits after the decimal point.)

(b) What is the smallest sample size to ensure that $\mathbb{P}\{0.2 \leq S^2 \leq 0.3\} \geq 0.5$? (It suffices to write down the formula and use a computer to determine the answer.)

(a) We know that $4S^2/0.5^2 = 16S^2$ has distribution χ^2_4. Let F denote the CDF of χ^2_4. Then

$$\mathbb{P}\{0.2 \leq S^2 \leq 0.3\} = \mathbb{P}\{3.2 \leq 16S^2 \leq 4.8\} = F(4.8) - F(3.2) \approx 0.2165.$$

(b) Let the sample size be $m+1$. Then $mS^2/0.5^2 = 4mS^2$ has distribution χ^2_m. Let F_m denote the CDF of χ^2_m. Then

$$\mathbb{P}\{0.2 \leq S^2 \leq 0.3\} = \mathbb{P}\{0.8m \leq 4mS^2 \leq 1.2m\} = F_m(1.2m) - F_m(0.8m).$$

The smallest m to make this larger than 0.5 is 23. The smallest sample size is therefore 24.

3. In an exam, the test scores in a class of 16 students are i.i.d. $\mathcal{N}(517, 120)$ random variables, and the test scores in a class of 25 students are i.i.d. $\mathcal{N}(503, 140)$ random variables. Approximate the probability that the (sample) variance of test scores of the first class exceeds the (sample) variance of test scores of the second class. (Express the result in terms of the CDF of an F-distribution and compute the final numerical value up to 4 digits after the decimal point.)

Let the sample variances be S_1^2 and S_2^2 respectively. Then $15S_1^2/120 \sim \chi^2_{15}$ and $24S_2^2/140 \sim \chi^2_{24}$. As a result,

$$\mathbb{P}\{S_1^2 > S_2^2\} = \mathbb{P}\Big\{\frac{S_1^2/120}{S_2^2/140} > \frac{7}{6}\Big\} = 1 - F(7/6) \approx 0.3577,$$

where F is the CDF of the F-distribution with 15 and 24 degrees of freedom.

Equivalently, we have

$$\mathbb{P}\{S_1^2 > S_2^2\} = \mathbb{P}\left\{\frac{S_2^2/140}{S_1^2/120} < \frac{6}{7}\right\} = \tilde{F}(7/6) \approx 0.3577,$$

where $\tilde{F}$ is the CDF of the F-distribution with 24 and 15 degrees of freedom.

4. Suppose that we are given i.i.d. random variables $X_1, \ldots, X_n$ from some distribution $\mathcal{P}_\theta$ where θ is the unknown parameter to be estimated. Let $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ be an estimator of θ computed from the data $X_1, \ldots, X_n$. Let $\mathbb{E}_\theta$ denote the expectation to emphasize that the data is from $\mathcal{P}_\theta$. We define three quantities for the estimator $\hat{\theta}$:

- The *mean squared error (MSE)* of $\hat{\theta}$ is defined as $\mathsf{MSE}(\theta, \hat{\theta}) := \mathbb{E}_\theta[(\hat{\theta} - \theta)^2]$.
- The *bias* of $\hat{\theta}$ is defined as $\mathsf{Bias}_\theta(\hat{\theta}) := \mathbb{E}_\theta[\hat{\theta}] - \theta$.
- The *variance* of the estimator $\hat{\theta}$ is defined as $\mathsf{Var}_\theta(\hat{\theta}) := \mathbb{E}_\theta[(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}])^2]$.

Show that the formula of *bias-variance trade-off* holds:

$$\mathsf{MSE}(\theta, \hat{\theta}) = [\mathsf{Bias}_\theta(\hat{\theta})]^2 + \mathsf{Var}_\theta(\hat{\theta}).$$

In particular, if $\hat{\theta}$ is *unbiased*, that is, the bias is zero, then the MSE is equal to the variance.

We have

$$\begin{aligned}
\mathbb{E}_\theta[(\hat{\theta} - \theta)^2] &= \mathbb{E}_\theta[(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}] + \mathbb{E}_\theta[\hat{\theta}] - \theta)^2] \\
&= \mathbb{E}_\theta[(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}])^2 + (\mathbb{E}_\theta[\hat{\theta}] - \theta)^2 + 2(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}])(\mathbb{E}_\theta[\hat{\theta}] - \theta)] \\
&= \mathbb{E}_\theta[(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}])^2] + (\mathbb{E}_\theta[\hat{\theta}] - \theta)^2 + 2\,\mathbb{E}_\theta[\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}]](\mathbb{E}_\theta[\hat{\theta}] - \theta) \\
&= \mathsf{Var}_\theta(\hat{\theta}) + (\mathsf{Bias}_\theta(\hat{\theta}))^2 + 0.
\end{aligned}$$

5. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two independent estimators of a common parameter θ; for example, $\hat{\theta}_1$ and $\hat{\theta}_2$ may be computed using two independent samples. Suppose that $\mathbb{E}_\theta[\hat{\theta}_i] = \theta$ and $\mathsf{Var}_\theta(\hat{\theta}_i) = \sigma_i^2$ for $i = 1, 2$.

(a) What is $\mathsf{MSE}(\theta, \hat{\theta})$ for the weighted estimator $\hat{\theta} := \lambda\hat{\theta}_1 + (1-\lambda)\hat{\theta}_2$ where $\lambda \in [0, 1]$? Express it in terms of σ_1, σ_2, and λ.

(b) What choice of λ minimizes the MSE?

(a) Since $\mathbb{E}_\theta[\hat{\theta}] = \lambda\theta + (1-\lambda)\theta = \theta$, we have

$$\mathsf{MSE}(\theta, \hat{\theta}) = \mathsf{Var}_\theta(\hat{\theta}) = \lambda^2\,\mathsf{Var}_\theta(\hat{\theta}_1) + (1-\lambda)^2\,\mathsf{Var}_\theta(\hat{\theta}_2) = \lambda^2\sigma_1^2 + (1-\lambda)^2\sigma_2^2.$$

(b) Taking the derivative with respect to λ and setting the result to zero, we obtain

$$2\lambda\sigma_1^2 - 2(1-\lambda)\sigma_2^2 = 0,$$

which yields $\lambda = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$.

6. Suppose that we are given i.i.d. $X_1, \dots, X_n \sim \mathcal{P}_\theta$, where $\mathcal{P}_\theta$ denotes the uniform distribution over $[0, \theta]$.

(a) A natural estimator of θ is $\hat{\theta}_1 := 2\bar{X} = \frac{2}{n}\sum_{i=1}^n X_i$. Compute $\mathsf{Bias}_\theta(\hat{\theta}_1)$, $\mathsf{Var}_\theta(\hat{\theta}_1)$, and $\mathsf{MSE}(\theta, \hat{\theta}_1)$.

(b) It is discussed in a lecture that the maximum likelihood estimator (MLE) of θ is $\hat{\theta}_2 := \max_{1\le i\le n} X_i$. Determine the CDF and the PDF of $\hat{\theta}_2$. (This is similar to a problem in a previous assignment.)

(c) Compute $\mathsf{Bias}_\theta(\hat{\theta}_2)$, $\mathsf{Var}_\theta(\hat{\theta}_2)$, and $\mathsf{MSE}(\theta, \hat{\theta}_2)$.

(d) Does $\hat{\theta}_1$ or $\hat{\theta}_2$ achieve the smaller MSE for $n \ge 10$? (This example shows the power of the MLE and that sometimes a biased estimator can be better.)

(a) We have

$$\mathsf{Bias}_\theta(\hat{\theta}_1) = \Big(\frac{2}{n}\sum_{i=1}^n \mathbb{E}_\theta[X_i]\Big) - \theta = \Big(\frac{2}{n}\sum_{i=1}^n \frac{\theta}{2}\Big) - \theta = 0,$$

so

$$\mathsf{MSE}(\theta, \hat{\theta}_1) = \mathsf{Var}_\theta(\hat{\theta}_1) = \frac{4}{n^2}\sum_{i=1}^n \mathsf{Var}_\theta(X_i) = \frac{4}{n^2}\sum_{i=1}^n \frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

(b) Let $F(t)$ denote the CDF of $\hat{\theta}_2$. We have $F(t) = 0$ for $t \le 0$, $F(t) = 1$ for $t \ge \theta$, and

$$F(t) = \mathbb{P}\{\hat{\theta}_2 \le t\} = (t/\theta)^n, \qquad \text{for } t \in (0, \theta).$$

The PDF of $\hat{\theta}_2$ is $f(t) = \frac{n}{\theta}\left(\frac{t}{\theta}\right)^{n-1}$ for $t \in (0, 1)$ and $f(t) = 0$ otherwise.

(c) We have

$$\mathsf{Bias}_\theta(\hat{\theta}_2) = \mathbb{E}_\theta[\hat{\theta}_2] - \theta = \int_0^\theta t\,\frac{n}{\theta}\Big(\frac{t}{\theta}\Big)^{n-1} dt - \theta = \frac{n\,\theta}{n+1} - \theta = \frac{-\theta}{n+1},$$

$$\mathsf{Var}_\theta(\hat{\theta}_2) = \mathbb{E}_\theta\Big[\Big(\hat{\theta}_2 - \frac{n\,\theta}{n+1}\Big)^2\Big] = \int_0^\theta \Big(t - \frac{n\,\theta}{n+1}\Big)^2 \frac{n}{\theta}\Big(\frac{t}{\theta}\Big)^{n-1} dt = \frac{n\,\theta^2}{(n+1)^2(n+2)},$$

and

$$\mathsf{MSE}(\theta, \hat{\theta}_2) = \Big(\frac{-\theta}{n+1}\Big)^2 + \frac{n\,\theta^2}{(n+1)^2(n+2)} = \frac{2\,\theta^2}{2 + 3n + n^2}.$$

(d) The MLE has smaller MSE for $n \ge 3$.

7. Suppose that we are given i.i.d. $X_1, \dots, X_n$ from the exponential distribution with density $f(x) = e^{-(x-\theta)}$ for $x \ge \theta$ and $f(x) = 0$ otherwise. What is the MLE of θ?

The likelihood is $e^{n\theta}e^{-\sum_{i=1}^n x_i}$ for $\theta \le x_i$ for all $i = 1, \dots, n$ and is zero otherwise. To maximize this function over θ, we should take $\theta = \min_{1\le i\le n} x_i$. Therefore, the MLE is $\min_{1\le i\le n} X_i$.

8. The gamma distribution has PDF

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta} \qquad \text{for } x > 0,$$

and $f(x) = 0$ for $x \leq 0$, where $k > 0$ is a fixed number assumed to be known, $\Gamma(k)$ is the gamma function (its specific form is not important here), and $\theta > 0$ is the parameter to be estimated. Given i.i.d. observations $X_1, \ldots, X_n$ from the gamma distribution, what is the MLE of θ?

The joint PDF of $(X_1, \ldots, X_n)$ is

$$\frac{1}{\Gamma(k)^n \theta^{kn}} \prod_{i=1}^{n} x_i^{k-1} e^{-x_i/\theta}.$$

The log-likelihood is

$$-n \log \Gamma(k) - kn \log \theta + \sum_{i=1}^{n} \left[(k-1) \log x_i - \frac{x_i}{\theta} \right].$$

Take the derivative with respect to θ and set the result to zero:

$$-\frac{kn}{\theta} + \sum_{i=1}^{n} \frac{x_i}{\theta^2} = 0.$$

It follows that the MLE of θ is

$$\hat{\theta} = \frac{1}{kn} \sum_{i=1}^{n} X_i.$$