APMA 1650

Homework 9 - Solutions

Problem 1. (16 pts) Suppose you are trying to measure the temperature θ of a pot of water down to the nearest microKelvin (you don't need to know what microKelvin are). The thermometer you have tends to produce random errors in the measurements down to that scale. To combat this, you make n measurements of the temperature $X_1, X_2, \dots X_n$. Suppose you know that each measurement has a random error E_i given by

$$X_i = \theta + E_i$$

and that $E_1, E_2, \dots E_n$ are iid with $E[E_i] = 0$ and $\text{Var}(E_i) = 9$. You decide to average your measurements to get a better idea of the true temperature,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

a. (8 pts) Use Chebyshev to estimate how large n needs to be at least 90% sure that the error $|\overline{X} - \theta|$ is with in 0.2 units.

b. (8 pts) Improve your answer to (b) instead using the CLT (round your answer to the nearest integer).

Solution:

a. Note that $E\overline{X} = \theta$ and $\text{Var}(\overline{X}) = \frac{9}{n}$. Therefore Chebyshev gives us

$$P(|\overline{X} - \theta| \le 0.2) \ge 1 - \frac{\text{Var}(\overline{X})}{n(0.2)^2} = 1 - \frac{900}{4n}$$

We obtain

$$1 - \frac{900}{4n} \ge 0.9 \Rightarrow n \ge 2250.$$

b. To use the central limit theorem, we note that for n big enough (say more than 100)

$$Z = \frac{\overline{X} - \theta}{3/\sqrt{n}} \approx N(0,1)$$

We get as an approximation

$$P(|\overline{X} - \theta| \le 0.2) = P\left(|Z| \le \frac{0.2\sqrt{n}}{3}\right) \approx \Phi(\sqrt{n}/15) - \Phi(-\sqrt{n}/15) = 2\Phi(\sqrt{n}/15) - 1$$

Therefore

$$2\Phi(\sqrt{n}/15) - 1 \ge 0.9 \Rightarrow n \ge \left(15\Phi^{-1}\left(\frac{1.9}{2}\right)\right)^2 \approx 608.74,$$

and so by rounding up, we need $n \ge 609$.

Problem 2. (24 pts) Let $X_1, X_2, \ldots X_{100}$ be iid standard normal random variables. Let

$$Y = X_1^2 + X_2^2 + \ldots + X_{100}^2.$$

a. (8 pts) Show that X_i^2 is Gamma(α, β) distributed for certain α and β (you may use the fact that $\Gamma(1/2) = \sqrt{\pi}$).

b. (5 pts) Use the known mean and variance of the Gamma distribution and part (a) to determine EX_i^2 and $\mathrm{Var}(X_i^2)$.

c. (8 pts) Use the Central Limit Theorem to find a value y such that $P(Y > y) \approx 0.05$. Approximate to 3 decimal places.

d. (5 pts) It is well known that Y has a $\chi^2(n)$ distribution, compare your answer in part (c) to the one obtained via the $\chi^2(n)$ table provided on the course webpage (or using `chi2inv` in MATLAB)

Solution:

a. To show that X_i^2 is Gamma, we use the CDF method Let $G = X_i^2$

$$F_G(y) = P(X_i^2 \le y) = P(-\sqrt{y} \le X_i \le \sqrt{y}) = 2\Phi(\sqrt{y}) - 1$$

Therefore for $y \in [0, \infty)$

$$f_G(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_G(y) = \frac{1}{\sqrt{y}} \Phi'(\sqrt{y}) = \frac{y^{-1/2}}{\sqrt{2\pi}} e^{-y/2} = \frac{\frac{1}{2}}{\Gamma(1/2)} y^{1/2-1} e^{-(1/2)y}$$

It follows that $G \sim \mathrm{Gamma}(\frac{1}{2}, \frac{1}{2})$

b. Using properties of the Gamma function, we know that the mean and variance are given by

$$EX_i^2 = EG = \frac{\alpha}{\beta} = \frac{\frac{1}{2}}{\frac{1}{2}} = 1$$

and

$$\mathrm{Var}(X_i^2) = \mathrm{Var}(G) = \frac{\alpha}{\beta^2} = \frac{\frac{1}{2}}{\left(\frac{1}{2}\right)^2} = 2$$

c. To use the central limit theorem, we note that $n = 100$ and that Y is a sum of iid random variables with mean 1 and variance 2 and therefore has mean $EY = n$ and variance $\mathrm{Var}(G) = 2n$. It follows by the CLT that

$$Z = \frac{Y - n}{\sqrt{2n}} \approx N(0, 1).$$

Therefore we have

$$P(Y > y) = P\left(Z > \frac{y-100}{10\sqrt{2}}\right) \approx 1 - \Phi\left(\frac{y-100}{10\sqrt{2}}\right)$$

and so

$$1 - \Phi\left(\frac{y-100}{10\sqrt{2}}\right) = 0.05 \quad \Rightarrow \quad y = 100 + 10\sqrt{2}\Phi^{-1}(0.95) \approx 123.262$$

d. Using the table, we see that

$$y = \chi^2_{0.05,100} = 124.342$$

we see that this number is just a little bigger than you would get using the central limit theorem.

Problem 3. (30 pts) Let $X_1, X_2, \ldots X_n$ be a random sample from a distribution associated to random variable X with parameter θ.

a. (10 pts) If the distribution of X has PDF

$$f_X(x;\theta) = \begin{cases} (\theta+1)x^\theta & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

For $\theta > -1$. Find the maximum likelihood estimator $\hat{\Theta}_{MLE}$ of θ.

b. (10 pts) If the distribution of X is Poisson, find the maximum likelihood estimator $\hat{\Theta}_{MLE}$ for $\theta = E[2^X]$.

c. (8 pts) Use the WLLN and properties of convergence in probability to show that $\hat{\Theta}_{MLE}$ is a consistent estimator of θ for both (a) and (b) above.

Solution:

a. To find the MLE for θ we write the log-likelihood function

$$\ln(L(\theta)) = \sum_{i=1}^{n} \ln(f(x_i;\theta)) = \theta\left(\sum_{i=1}^{n} \ln(x_i)\right) + n\ln(1+\theta).$$

Taking the derivative in θ and setting equal to zero gives

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln(L(\theta)) = \sum_{i=1}^{n} \ln(x_i) + \frac{n}{1+\theta} = 0.$$

Solving for θ gives

$$\theta = \frac{-1}{\frac{1}{n}\sum_{i=1}^{n}\ln(x_i)} - 1.$$

Therefore the MLE is given by

$$\hat{\Theta}_{MLE} = \frac{-1}{\frac{1}{n}\sum_{i=1}^{n}\ln(X_i)} - 1.$$

b. Let λ be the Poisson rate so that

$$P(X = x) = \frac{\lambda^x}{x!}e^{-\lambda}.$$

We want to find how θ relates to λ. To see this, we note that

$$\theta = E(2^X) = \sum_{x=1}^{\infty} \frac{2^x \lambda^x}{x!} e^{-\lambda} = \left(\sum_{x=1}^{\infty} \frac{(2\lambda)^x}{x!} \right) e^{-\lambda} = e^{2\lambda} e^{-\lambda} = e^{\lambda}$$

We can can find the MLE directly by writing $\lambda = \ln(\theta)$ so that

$$f(x;\theta) = \frac{(\ln(\theta))^x}{x!} \frac{1}{\theta}.$$

The log likelihood function is given by

$$\ln(L(\theta)) = \ln \left(\prod_{i=1}^{n} \left(\frac{(\ln(\theta))^{x_i}}{x_i!} \frac{1}{\theta} \right) \right) = \left(\sum_{i=1}^{n} x_i \right) \ln(\ln(\theta)) - n \ln \theta - \sum_{i=1}^{n} \ln(x_i!)$$

Taking the derivative in θ and setting equal to zero gives

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \ln(L(\theta)) = \left(\sum_{i=1}^{n} x_i \right) \frac{1}{\theta \ln \theta} - \frac{n}{\theta} = 0$$

This implies since $\theta > 0$ that

$$\left(\sum_{i=1}^{n} x_i \right) \frac{1}{\ln \theta} = n$$

solving for θ gives

$$\theta = e^{\frac{1}{n} \sum_{i=1}^{n} x_i}$$

so that the MLE is given by

$$\hat{\Theta}_{MLE} = e^{\overline{X}}$$

c. For (a), to show that

$$\hat{\Theta}_{MLE} = \frac{1}{-\frac{1}{n} \sum_{i=1}^{n} \ln(X_i)} - 1$$

is consistent, we note that $Y_i = \ln(X_i)$ for $i = 1, \ldots, n$ are iid RVs with mean

$$\begin{aligned} EY_i = E \ln(X_i) &= \int_0^1 x^{\theta} \ln(x)(1+\theta)\mathrm{d}x \\ &= x^{1+\theta} \ln(x) \Big|_0^1 - \int_0^1 x^{\theta} \mathrm{d}x \\ &= \frac{-1}{1+\theta} \end{aligned}$$

A similar calculation shows that the variance of Y_i is finite (indeed Y_i is just an exponential random variable). Therefore by the WLLN

$$\frac{1}{n}\sum_{i=1}^{n} \ln(X_i) \xrightarrow{p} \frac{-1}{1+\theta}.$$

Using the fact that convergence in probability can be passed through continuous functions, we obtain

$$\hat{\Theta}_{MLE} = \frac{1}{-\frac{1}{n}\sum_{i=1}^{n} \ln(X_i)} - 1 \xrightarrow{p} \frac{1}{\frac{1}{1+\theta}} - 1 = \theta$$

Therefore it is consistent.

For (b), to show that

$$\hat{\Theta}_{MLE} = e^{\overline{X}}$$

is consistent, we note that by the WLLN

$$\overline{X} \xrightarrow{p} \lambda$$

and therefore using properties of convergence in probability

$$\hat{\Theta}_{MLE} \xrightarrow{p} e^{\lambda} = \theta,$$

and is therefore consistent.

Problem 4. (40 pts) Let $X_1, X_2, \ldots, X_n$ be a random sample from Uniform$(0, \theta)$.

a. (8 pts) Show that the order statistic

$$\hat{\Theta} = \max\{X_1, X_2, \ldots, X_n\}$$

is a maximum likelihood estimator for θ. (Hint: taking derivatives isn't always the best strategy to finding a maximum)

b. Show that $\hat{\Theta}$ in part (a) has a CDF given by

$$F_{\hat{\Theta}}(x) = \begin{cases} 0 & x \leq 0 \\ \left(\frac{x}{\theta}\right)^n & x \in [0, \theta] \\ 1 & \text{otherwise} \end{cases}$$

(Hint: $\{\hat{\Theta} \leq x\} \Leftrightarrow \{X_1 \leq x\} \cap \{X_2 \leq x\} \cap \ldots \cap \{X_n \leq x\}$)

c. (6 pts) What is the bias $B(\hat{\Theta})$? Explain intuitively why this makes sense.

d. (8 pts) Find a value c so that $\tilde{\Theta} = c\hat{\Theta}$ is unbiased.

e. (8 pts) Find the mean square errors $MSE(\hat{\Theta})$, $MSE(\tilde{\Theta})$. Which one is better?

Solution.

a. The likelihood function is given by

$$L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f_{X_i}(x_i) = \begin{cases} \frac{1}{\theta^n} & x_1, x_2, \ldots, x_n \in [0, \theta] \\ 0 & \text{otherwise.} \end{cases}$$

To maximize this, we note that if θ is such that $\max\{x_1, x_2, \ldots x_n\} \leq \theta$ then $x_1, x_2, \ldots, x_n \in [0, \theta]$ and the likelihood function is $L(\theta) = \theta^{-n}$, which gets bigger the smaller you take θ. However if we take θ too small $\theta \leq \max\{x_1, x_2, \ldots x_n\}$ we violate $x_1, x_2, \ldots, x_n \in [0, \theta]$ and therefore $L(\theta) = 0$. This means that $L(\theta)$ has a jump in θ at the value $\theta = \max x_1, x_2, \ldots x_n$ and can be expressed as

$$L(\theta) = \begin{cases} \frac{1}{\theta^n} & 0 \leq \theta \leq \max\{x_1, x_2, \ldots, x_n\} \\ 0 & \max\{x_1, x_2, \ldots, x_n\} < \theta. \end{cases}$$

From this, it is easy to see that $L(\theta)$ is maximized at the jump

$$\hat{\theta} = \max\{x_1, x_2, \ldots, x_n\},$$

and therefore

$$\hat{\Theta} = \max\{X_1, X_2, \ldots, X_n\}$$

b. Using the hint and the fact that X_i are independent we see that

$$F_{\hat{\Theta}}(x) = P(\hat{\Theta} \leq x) = P(X_1 \leq x, X_2 \leq x, \ldots X_n \leq x) = \prod_{i=1}^{n} P(X_i \leq x)$$

which gives us the answer upon substituting

$$P(X_i \leq x) = \begin{cases} 0 & x \leq 0 \\ \left(\frac{x}{\theta}\right) & x \in [0, \theta] \\ 1 & \text{otherwise.} \end{cases}$$

c. Note that the PDF of $\hat{\Theta}$ is

$$f_{\hat{\Theta}}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{\hat{\Theta}}(x) = \begin{cases} n\frac{x^{n-1}}{\theta^n} & x \in [0, \theta] \\ 0 & \text{otherwise} \end{cases}$$

This means that

$$E\hat{\Theta} = \int_0^{\theta} xn \left(\frac{x^{n-1}}{\theta^n}\right) \mathrm{d}x = \frac{n}{n+1}\theta.$$

and consequently the Bias is

$$B(\hat{\Theta}) = \frac{n}{n+1}\theta - \theta = \frac{-\theta}{n+1} \neq 0$$

Therefore $\hat{\Theta}$ is biased. However this isn't very surprising, since is is clear that because $X_i \leq \theta$ then

$$\hat{\Theta} = \max\{X_1, X_2, \ldots, X_n\} \leq \theta$$

and so $\hat{\Theta}$ always undershoots the value it is estimating unless one of the X_i hits θ exactly(which never happens).

d. To make the estimator unbiased, we choose

$$c = \frac{\theta}{B(\hat{\Theta}) + \theta} = \frac{n+1}{n}$$

so that

$$\tilde{\Theta} = \left(\frac{n+1}{n}\right) \max\{X_1, X_2, \ldots, X_n\}$$

Therefore

$$E\tilde{\Theta} = \left(\frac{n+1}{n}\right) E\hat{\Theta} = \left(\frac{n+1}{n}\right)\left(\frac{n}{n+1}\right)\theta = \theta,$$

and so $\tilde{\Theta}$ is unbiased.

e. To find the mean square error, we use the formula

$$\text{MSE}(\hat{\Theta}) = \text{Var}(\hat{\Theta}) + B(\hat{\Theta})^2$$

and note that

$$E\hat{\Theta}^2 = \int_0^\theta x^2 n \left(\frac{x^{n-1}}{\theta^n}\right) \mathrm{d}x = \frac{n}{n+2}\theta^2$$

so

$$\text{Var}(\hat{\Theta}) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\right)^2 \theta^2 = \frac{n}{(n+2)(n+1)^2}\theta^2.$$

This gives

$$\text{MSE}(\hat{\Theta}) = \frac{n}{(n+2)(n+1)^2}\theta^2 + \left(\frac{\theta}{n+1}\right)^2 = \frac{\theta}{(n+2)(n+1)}$$

Similarly for the unbiased estimator $\tilde{\Theta}$ we have

$$\begin{aligned}\text{MSE}(\tilde{\Theta}) &= \text{Var}(\tilde{\Theta}) = c^2 \text{Var}(\hat{\Theta}) \\ &= \left(\frac{n+1}{n}\right)^2 \left(\frac{n}{(n+2)(n+1)^2}\right)\theta^2 = \frac{1}{n(n+2)}\theta^2.\end{aligned}$$

To see which one is better, lets consider the ratio

$$\frac{\text{MSE}(\tilde{\Theta})}{\text{MSE}(\hat{\Theta})} = \frac{(n+2)(n+1)}{2n(n+2)} = \frac{n+1}{2n} = \frac{1}{2} + \frac{1}{2n} < 1$$

if $n > 1$ and equality if $n = 1$. Therefore

$$\text{MSE}(\tilde{\Theta}) < \text{MSE}(\hat{\Theta}) \quad \text{for} \quad n > 1.$$

So the unbiased estimator is better for $n > 1$.

Problem 5. (20 pts) Suppose you have two coins, one is fair, the other produces heads with probability 3/4. You are handed one of the coins and decide to flip the coin n times to try to figure out if it's fair or not.

a. (4 pts) Assuming you know which coin was chosen, explain what the WLLN predicts the proportion of heads should look like as you take n large.

b. (8 pts) Use Chebyshev to estimate the number of coin flips you need to take to be 95% sure you know which coin was chosen.

c. (8 pts) Use the CLT to estimate how many coin flips you need to take to be 95% sure you know which coin was chosen.

Solution:

a. Let p be the probability of heads, then by the WLLN the proportion of heads out of n coin tosses should converge to p in probability as $n \to \infty$.

b. Please note that there are many approaches one could take to this problem. Below is just one approach.

   Note that $p_1 = 1/2$ and $p_2 = 3/4$ are $p_2 - p_1 = 1/4$ apart. Let X be the proportion of heads in n tosses. We split the distance between p_1 and p_2 in half and consider two cases:

   i Either $X - 1/2 \le \frac{p_2 - p_1}{2} = \frac{1}{8}$ in which case was are assured that $3/4 - X > 1/8$

   ii Or $3/4 - X \le \frac{1}{8}$ in which case $X - 1/2 > 1/8$.

   In either of these cases, we can use the mid-point value $1/2 + 1/8 = 3/4 - 1/8 = 5/8$ to decide which coin was flipped. If $X \le 5/8$ then we say it was the fair coin, if $X \ge 5/8$ then we say it was the biased coin.

   It follows that we will be 95% sure that the fair coin was chosen if

$$P(X - 1/2 \le 1/8) \ge .95$$

and similarly we will be 95% sure that the biased coin is chosen if

$$P(3/4 - X \le 1/8) \ge .95.$$

In general X has mean $p \in \{1/2, 3/4\}$ and variance $p(1-p)/n$. Therefore Chebyshev gives

$$P(X - p \le 1/8) \text{ or } P(p - X \le 1/8) \ge P(|X - p| \le 1/8) \ge 1 - \frac{64p(1-p)}{n}$$

Since we don't know whether $p = 1/2$ or $3/4$ we have no choice but to bound $p(1-p)$ by the worst case $p = 1/2$, $p(1-p) \le 1/4$ which gives

$$P(|X - p| \le 1/8) \ge 1 - \frac{16}{n}$$

We see that we need

$$1 - \frac{16}{n} \ge .95 \Rightarrow n \ge 16(20) = 320$$

in order to have 95% confidence that we know which coin was chosen.

c. If instead we want to use the CLT, we note (recall HW5) that we need to take n bigger than at least

$$n > 9\frac{\max\{p, 1-p\}}{\min\{p, 1-p\}} = 9\frac{3/4}{1/4} = 27$$

whereby we can approximate

$$Z = \frac{X - p}{\sqrt{p(1-p)/n}} \approx N(0, 1).$$

This gives

$$P(|X - p| \le 1/8) = P\left(|Z| \le \frac{\sqrt{n}}{8\sqrt{p(1-p)}}\right) \ge P\left(|Z| \le \frac{\sqrt{n}}{4}\right) \approx 2\Phi(\sqrt{n}/4) - 1$$

where we again used the worst case variance $p(1-p) \le 1/4$. Therefore we see that a good estimate is

$$2\Phi(\sqrt{n}/4) - 1 \ge 0.95 \Rightarrow n \ge \left(4\Phi^{-1}(1.95/2)\right)^2 \approx 61.463.$$

Therefore we need to take $n \ge 62$ to be able to tell which coin is which with 95% certainty. This is certainly a lot better than what we got in part b.

Note: In the CLT case you can improve your answer even more by realizing that $X - p$ and $p - X$ have the same distribution and therefore

$$P(X - p \le 1/8) = P(p - X \le 1/8) \ge P\left(Z \le \frac{\sqrt{n}}{4}\right) = \Phi(\sqrt{n}/4) = .95$$

Which means that you can actually take

$$n \geq (4 * \Phi^{-1}(.95))^2 \approx 43.2887$$

Rounding up gives $n = 44$.

Problem 6. (20 pts) You are trying to guess the mean score on the latest midterm in a large class. You randomly sampled 10 students and ask them what their score was. The responses you get are

$$73, 82, 91, 50, 68, 77, 92, 81, 75, 69.$$

Assume the grades are normally distributed (just roll with it...) with unknown mean μ and variance σ^2.

a. (10 pts) Find a 90% confidence interval for the variance σ^2

b. (10 pts) Find a 95% confidence interval for the mean μ.

You may use one of the tables provided on the course webpage or `chi2inv` and `tinv` in MATLAB.

Solution. We first compute the sample mean and sample variance

$$\overline{X} = 75.8, \quad S^2 = 149.0667$$

a. The $(1-\alpha)100\%$ confidence interval for σ^2 is

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right]$$

For $n = 10$ and $\alpha = 0.1$, we find

$$\chi^2_{0.05,9} = 16.919, \quad \chi^2_{0.95,9} = 3.3251$$

and therefore the 90% confidence interval is

$$\left[\frac{9(149.0667)}{16.919}, \frac{9(149.0667)}{3.3251}\right] = [79.2955, 403.4767].$$

b. The $(1-\alpha)100\%$ confidence interval for μ is

$$\left[\overline{X} - \frac{t_{\alpha/2,n-1}S}{\sqrt{n}}, \overline{X} + \frac{t_{\alpha/2,n-1}S}{\sqrt{n}}\right]$$

For $n = 10$ and $\alpha = 0.05$, we find

$$t_{0.025,9} = 2.2622$$

and therefore the 95% confidence interval is

$$\left[75.8 - \frac{2.2622(\sqrt{149.0667})}{\sqrt{10}}, 75.8 + \frac{2.2622(\sqrt{149.0667})}{\sqrt{10}}\right] = [67.0658, 84.5342].$$