Problem 1

Assume you are sampling values from a some unknown distribution with finite μ and σ (say a survey asking students to grade their TA on a 0-100 scale).
After getting the first 10 responses (10 samples) you got a mean of $\overline{X}_{10} = 61$.
Give a 95% confidence interval for μ.
After receiving all 50 responses from the students you got a mean of $\overline{X}_{50} = 53$.
Give a 95% confidence interval for μ.
Why would you use 2 different distributions for the two problems?

Solution 1

For the average of 10 samples we use the T distribution with $10 - 1 = 9$ degrees of freedom. We are interested in an interval that we build using the probability $P(-t_{.475} \leq T \leq t_{.475}) = 0.95$. In the T distribution table, for 9 degrees of freedom we get that $t_{.475} \approx 2.26$.

$$0.95 = P(-t_{.475} \leq T \leq t_{.475}) = P(-t_{.475} \leq \frac{\overline{X} - \hat{\mu}}{\hat{\sigma}/\sqrt{n}} \leq t_{.475}) =$$

$$P(\overline{X} - \frac{\hat{\sigma} t_{.475}}{\sqrt{n}} \leq \hat{\mu} \leq \overline{X} + \frac{\hat{\sigma} t_{.475}}{\sqrt{n}}) = P(61 - \frac{2.26\ \hat{\sigma}}{\sqrt{10}} \leq \hat{\mu} \leq 61 + \frac{2.26\ \hat{\sigma}}{\sqrt{10}})$$

Where $\hat{\sigma} = \frac{1}{n-1} \sum_1^n (X_i - \overline{X}_n)$.
So the confidence interval is $[61 - \frac{2.26\ \hat{\sigma}}{\sqrt{10}}, 61 + \frac{2.26\ \hat{\sigma}}{\sqrt{10}}]$.

For the average of 50 2 types of solutions were accepted:
A) Since we don't know σ, do the same as above only with $t_{.475} \approx 2.01$ at 49 degrees of freedom, $n = 50$ and $\overline{X}_n = 53$. This gives $CI = [53 - \frac{2.01\ \hat{\sigma}}{\sqrt{50}}, 53 + \frac{2.01\ \hat{\sigma}}{\sqrt{50}}]$

B) Assume we know σ. We use the fact that the sample is large and hence the CLT can be used. We can now assume that the sample mean is distributed normally with unknown mean μ and known varience σ^2 We build our interval using the probability $P(-z_{.475} \leq Z \leq z_{.475}) = 0.95$. In the Normal distribution table we get that $z_{.475} \approx 1.96$. Now:

$$0.95 = P(-z_{.475} \leq Z \leq z_{.475}) = P(-z_{.475} \leq \frac{\overline{X} - \hat{\mu}}{\sigma/\sqrt{n}} \leq z_{.475}) =$$

$$P(\overline{X} - \frac{\sigma z_{.475}}{\sqrt{n}} \leq \hat{\mu} \leq \overline{X} - \frac{\sigma z_{.475}}{\sqrt{n}}) = P(53 - \frac{1.96\ \sigma}{\sqrt{50}} \leq \hat{\mu} \leq 53 - \frac{1.96\ \sigma}{\sqrt{50}})$$

So the confidence interval is $[53 - \frac{1.96\ \sigma}{\sqrt{50}}, 53 + \frac{1.96\ \sigma}{\sqrt{50}}]$.
Notice that here that the sample is bigger and the approximation closer to normal, the length of the interval is shorter!

For small samples (usually for $n \leq 30$) we can not use the CLT to approximate the mean $(\overline{X}_n)$ as distributed Normally, even when we know the veriance σ of the underlying distribution, but rather follow the T distribution with $n - 1$ degrees of freedom.

Problem 2

You have seen in class the maximum likelihhod estimator $\hat{A} = max(x_i)$ for the parameter A of a Uniform distribution on the interval (0,A).

Find a maximum likelihhod estimator for the parameters A and B of a Uniform distribution on the interval (A,B)?
Find a maximum likelihhod estimator for the parameter A of a Uniform distribution on the interval (-A,A).
Find a maximum likelihhod estimator for the parameter A of a Uniform distribution on the interval (A,2A).

Solution 2

For Uniform (A,B) the likelihhod function is $L(x_1, ..., x_n|A, B) = (\frac{1}{B-A})^n$ for any sample. To maximize this we must minimize the value of $(B - A)$ (interval length), yet we must keep all samples with in the range, i.e. $\forall x_i, x_i \in (A, B)$.
An MLE for A and B would then be $\hat{A} = min(X_i), \quad \hat{B} = max(X_i)$.
These values yield the minimal length since it's the smallest interval to include all sampled points.

For Uniform (-A,A) the likelihhod function is $L(x_1, ..., x_n|A) = (\frac{1}{2A})^n$ for any sample. To maximize this we must minimize the value of A , yet we must keep all samples with in the range, i.e. $\forall x_i, -A \leq x_i \leq A$.
An MLE for A would be $\hat{A} = max(|x_i|)$.
This is the smallest value that promises that all sampled points are in the requiered range.

For Uniform (A,2A) the likelihhod function is $L(x_1, ..., x_n|A) = (\frac{1}{A})^n$ for any sample. To maximize this we must minimize the value of A , yet we must keep all samples with in the range, i.e. $\forall x_i, A \leq x_i \leq 2A$.
Two interesting estimators would be $\hat{A}_1 = min(x_i)$ and $\hat{A}_2 = \frac{max(x_i)}{2}$.
Since $\hat{A}_2 = \frac{max(x_i)}{2} \leq \frac{2A}{2} = A \leq min(x_i)$, all samples are between $\hat{A}_2$ and 2 $\hat{A}_2$.
Since 2 $\hat{A}_2 \geq 2$ $A \geq max(x_i)$ all samples are also between $\hat{A}_1$ and 2 $\hat{A}_1$. Clearly both of these estimators try to minimize $\hat{A}$ and keep all points within $(\hat{A}, 2\hat{A})$, but is one of them better?
Since $min(x_i) > A$ and $max(x_i) < 2A$, $\frac{max(x_i)}{2} < A < min(x_i)$ thus $\hat{A}_2 < \hat{A}_1$ so $\hat{A}_2 = \frac{max(x_i)}{2}$ gives higher likelihood for the sample.
In general, any other estimator $\hat{A}_3$ for A would be either larger than $\hat{A}_2$ (hence with lower likelihhod) or smaller (hence with zero likelihhod since 2 $\hat{A}_3 < max(x_i)$). So $\hat{A}_2 = \frac{max(x_i)}{2}$ is indeed the MLE.
Note that the fact that the likelihhod of $\hat{A}_2$ is greater than that of $\hat{A}_1$ for any sample and any value of θ, it does not mean that $\hat{A}_2$ is also a tighter estimator than $\hat{A}_1$.

Problem 3

Estimator $\hat{\Theta}$ is called an unbiased estimator for Θ if $E(\hat{\Theta}) = \Theta$ (notice that $\hat{\Theta}$ is indeed a random variable!).

Consider a Uniform distribution on the interval (0,A).
Is the maximum likelihood estimator for A unbiased?
Is $\hat{A}_1 = 2\overline{X}_n$ an estimator for A? Is it a reasonable estimator for A?
Is the above defined $\hat{A}_1$ an unbiased estimator for A?
Is $\hat{A}_2 = 2$ an estimator for A? Is it a reasonable estimator for A?
Is the above defined $\hat{A}_2$ an unbiased estimator for A?

Solution 3

To check if the MLE is biased or not we must first find its Expected value. For this we must first find the p.d.f. of $max(x_i)$. Denote $Y = max(x_i)$:

$$F_Y(y) = P(Y \leq y) = P(max(x_i) \leq y) = P(x_1 \leq y, ..., x_n \leq y) =$$

$$P(x_1 \leq y) \; ... \; P(x_n \leq y) = P^n(x \leq y) = \frac{y}{A}^n$$

So $f_Y(y) = F_Y'(y) = n\frac{1}{A}^n y^{n-1}$. Now we can find $E(max(x_i)$:

$$E(max(x_i)) = E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = \int_0^A y\, n(\frac{1}{A})^n y^{n-1}\, dy = \int_0^A n(\frac{y}{A})^n\, dy =$$

$$n(\frac{1}{A})^n(\frac{y^{n+1}}{n+1})]_0^n = (\frac{n}{n+1})(\frac{1}{A})^n A^{n+1} = (\frac{n}{n+1})A < A$$

So the MLE $max(x_i)$ is biased since $E(max(x_i)) \neq A$.

$\hat{A}_1 = 2\overline{X}_n$ is a good estimator for A since we expect the mean of the sample to fall close to the midpoint of zero and A, hence $2\overline{X}_n$ should be close to A. To find out if this is an unbiased estimator we must calculate its expected value:

$$E(\hat{A}_1) = E(2\overline{X}_n) = 2E(\overline{X}_n) = 2E(\frac{1}{n}\sum_1^n x_i) = \frac{2}{n}\sum_1^n E(x_i) = \frac{2}{\cancel{n}}\cancel{n}E(x) = 2\frac{A}{2} = A$$

So $\hat{A}_1 = 2\overline{X}_n$ is an unbiased estimator.

$\hat{A}_2 = 2$ is indeed an estimator if we call it so, but it doesn't make a lot of sense since it ignores the samples!
$E(\hat{A}_2) = E(2) = 2$, so for $A \neq 2$ it is a biased estimator.

Problem 4

Estimator $\hat{\Theta}_1$ is considered a tighter estimator for Θ than the estimator $\hat{\Theta}_2$ if for any value of Θ and any sequence of samples $x_1, x_2, ..., x_n$ $|\hat{\Theta}_1 - \Theta| \leq |\hat{\Theta}_2 - \Theta|$.

Once again consider a Uniform distribution on the interval (0,A).
Is the maximum likelihood estimator for A tighter than the above defined $\hat{A}_1$?
Is the maximum likelihood estimator for A tighter than the above defined $\hat{A}_2$?
Can you find an estimator for A that is tighter than $\hat{A}_1$?

Solution 4

By examples we can show that $max(x_i)$ is not tighter than $2\overline{X}_n$ and vice versa.
Take an example of a Uniform (0,1) distribution ($A = 1$).
For the single sample $\{1\}$, the MLE $max(x_i) = 1$ has the exact value of A while $\hat{A}_1 = 2\overline{X}_n = 2 * 1 = 2$ is at a distance from A.
For the single sample $\{\frac{1}{2}\}$, $\hat{A}_1 = 2\overline{X}_n = 2 * \frac{1}{2} = 1$ has the exact value of A while the MLE $max(x_i) = \frac{1}{2}$ is at a distance from it.

By examples we can show that $max(x_i)$ isn't even tighter than the senseless estimator $\hat{A}_2 = 2$ (and vice versa).
Take an example of a Uniform (0,1) distribution ($A = 1$).
For the single sample $\{1\}$, the MLE $max(x_i) = 1$ has the exact value of A while $\hat{A}_2 = 2$ is at a distance from A.
But for a Uniform (0,2) distribution, almost any sample (e.g. $\{1\}$), $\hat{A}_2 = 2$ gives the exact value of A while the MLE does not (in the example $max(x_i) = 1 \neq 2$).

$\hat{A}_3 = max(2\bar(X)_n, max(x_i))$ is a tighter estimator for A than $\hat{A}_1 = 2\bar(X)_n$.
Proof:
If $2\bar(X)_n \geq max(x_i)$ then both estimators are equal ($|A - \hat{A}_1| = |A - \hat{A}_3|$).
If $max(x_i) > 2\bar(X)_n$, then since $A \geq max(x_i)$ we get $A \geq max(x_i) > 2\bar(X)_n$ thus

$$|A - \hat{A}_3| = |A - max(2\bar(X)_n, max(x_i))| = |A - max(x_i)| < |A - 2\bar(X)_n| = |A - \hat{A}_1|$$