Parametric Statistics

Sampling Distributions of Estimators/Confidence Intervals/Hypothesis testing

1. Suppose that against a certain opponent the number of points out local basketball team scores is normally distributed with unknown mean θ and unknown variance, σ^2. Suppose that over the course of the last 10 games between the two teams our team scored the following points: 59, 62, 59, 74, 70, 61, 62, 66, 62, 75

   (a) Compute a 95% confidence interval for θ. Does 95% confidence mean that the probability θ is in the interval you just found is 95%?

   (b) Now suppose that you learn that $\sigma^2 = 25$. Compute a 95% confidence interval for θ. How does this compare to the interval in (a)?

   Answer: We compute the data mean and variance $\bar{x_n} = 65, s^2 = 35.778$. The number of degrees of freedom is 9. We look up the critical value $T_9^{-1}(0.975) = 2.262$ in the table of the t distributions. The 95% confidence interval is $[\bar{x_n} - T_9^{-1}(0.975)s/\sqrt{n}, \bar{x_n} + T_9^{-1}(0.975)s/\sqrt{n}] = [60.721, 69.279]$. 95% confidence means that in 95% of experiments the random interval will contain the true θ. It is not the probability that θ is in the given interval. (b) When the variance is known, then $\sqrt{n}\frac{\bar{x_n}-\mu}{\sigma}$ follows a normal distribution, instead of a t distribution. Hence, the 95% confidence interval is $[\bar{x_n} - \Phi_9^{-1}(0.975)\sigma/\sqrt{n}, \bar{x_n} + \Phi_9^{-1}(0.975)\sigma/\sqrt{n}] = [61.901, 68.099]$. There are two reasons for this, first the true variance 25 is smaller than the sample variance 35.8 and second, the normal distribution has narrower tails than the t distribution.

2. The volume in a set of wine bottles is known to follow a $N(\mu, 25)$ distribution. You take a sample of the bottles and measure their volumes. How many bottles do you have to sample to have a 95% confidence interval for μ with width 1?

   Answer: Let $\bar{x}_n$ be the sample mean for our sample. The 95% confidence interval for the mean is $\bar{x_n} \pm \Phi^{-1}(0.975)\sigma/\sqrt{n}$. The width of this confidence interval is $2 \times \Phi^{-1}(0.975)\sigma/\sqrt{n}$. Setting the width equal to 1, and since $\Phi^{-1}(0.975) = 1.96, \sigma = 5$, we get $n = 19.6^2 = 384$ (since n has to be an integer).

3. We generate a number x from a uniform distribution on the interval $[0, \theta]$. We decide to test $H_0 : \theta = 2$ against $H_1 : \theta = 2$ by rejecting H_0 if $x \leq 0.1$ or $x \geq 1.9$.

   (a) Compute the probability of a type I error.

   (b) Compute the probability of a type II error if the true value of θ is 2.5.

   Answer:

$$P(\text{type I}) = P(\text{reject } H_0 | H_0 \text{ is true}) = P(x \leq 0.1 \text{ or } x \geq 1.9 | \theta = 2) = \frac{0.2}{2} = 0.1$$

$$P(typeII) = P(\text{ do not reject } H_0 | \theta = 2.5) = P(0.1 < x < 1.9 | \theta = 2.5) = 0.72$$

4. Consider a machine that is known to fill soda cans with amounts that follow a normal distribution with (unknown) mean μ and standard deviation $\sigma = 3mL$. We measure the volume of soda in a sample of bottles and obtain the following data (in mL): 352, 351, 361, 353, 352, 358, 360, 358, 359

   (a) Construct a 95% confidence interval for the mean μ.

   (b) Now construct a 98% confidence interval for the mean μ.

   (c) Suppose now that σ is not known. Redo parts (a) and (b), and compare your answers to those above.

Answer: (a) The sample mean is $\bar{x_n} = 356$. $\sqrt{n}\frac{X_n - \mu}{\sigma} \sim N(0,1)$. Since $\frac{1+0.95}{2} = 0.975$, $\Phi^{-1}(0.975) = 1.96, \sigma = 3$ and $n = 9$, the 95% confidence interval is

$$[\bar{x_n} - \Phi^{-1}(0.975)\sigma/\sqrt{n}, \bar{x_n} + \Phi^{-1}(0.975)\sigma/\sqrt{n}] = [354.04, 357.96]$$

(b) Same as above with $\Phi^{-1}(0.99)$ gives us $[353.67, 358.33]$
(c) Using the sample standard deviation $s = \sqrt{15.5} = 3.9$, $\sqrt{n}\frac{X_n - \mu}{s}$ follows the t distribution with $n - 1$=8 degrees of freedom, hence the 95% confidence interval is

$$[\bar{x_n} - T_8^{-1}(0.975)s/\sqrt{n}, \bar{x_n} + T_8^{-1}(0.975)s/\sqrt{n}] = [352.97, 359.03].$$

Similarly the 98% interval is [352.20, 359.80]. These intervals are larger than the corresponding intervals from parts (a) and (b). The new uncertainly regarding the value of σ means we need larger intervals to achieve the same level of confidence. This is reflected in the fact that the t distribution has fatter tails than the normal distribution.