ECE 315 - INTRODUCTION TO STATISTICS - INVEST 22
CONFIDENCE INTERVALS

WINTER 2004 **A.P. FELZER**

To do "well" on this investigation you must not only get the right answers but must also do neat, complete and concise writeups that make obvious what each problem is, how you're solving the problem and what your answer is. You also need to include drawings of all circuits as well as appropriate graphs and tables.

From the last Investigation we know how to calculate sample means $\overline{X}$ and sample variances s^2. We also know that the smaller the variance of the sample means $\overline{X}$ the more likely it is that $\overline{X}$ is close to the population mean μ. The main objective of this Investigation is to make use of the Central Limit Theorem to quantify how close $\overline{X}$ is to μ.

1. What does the Central Limit Theorem tell us about the probability distribution of the sample mean $\overline{X}$ as follows

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

2. From Problem (1) we know that $\overline{X}$ has a Gaussian distribution. And from the last Investigation we know that

$$E\left[\overline{X}\right] = \mu$$

   Now show that

$$Var\left[\overline{X}\right] = \sigma_{\overline{X}}^2 = \frac{1}{n} Var[X]$$

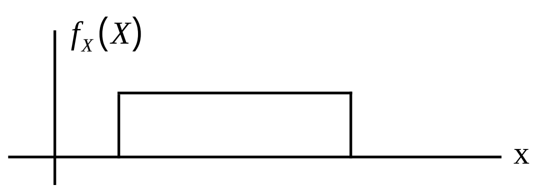
   Hint - make use of the fact that

$$Var[aX] = a^2 Var[X]$$

   and when $X_1, X_2, \cdots, X_n$ are independent then

$$Var\left[X_1 + X_2 + \cdots + X_n\right] = Var\left[X_1\right] + Var\left[X_2\right] + \cdots + Var\left[X_n\right]$$

3. Given the results of Problem (2)
   a. Sketch the probability distribution $f_{\overline{X}}\left(\overline{X}\right)$ of a random variable X with probability density as follows



   b. Describe how increasing the size of n affects $f_{\overline{X}}\left(\overline{X}\right)$. Draw graphs to illustrate

4. The objective of this problem is to apply our result that the sample means

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

have a Gaussian distribution with variance

$$Var\left[\bar{X}\right] = \sigma_{\bar{X}}^2 = \frac{1}{n} Var\left[X\right]$$

to estimate the population mean μ. Suppose the sample mean is $\bar{X} = 7.8$

a. What's the probability that the population mean μ is in the interval 7.8 ± 0.2 if n = 100 samples
b. What is the value of a in the following expression $P\left[7.8 - a \le \mu \le 7.8 + a\right] = 0.99$ if n = 100 samples
c. How large does n have to be for the probability to be 0.99 that the population mean μ is in the interval 7.8 ± 0.1

5. The objective of this and the next problem is to make use of hypothesis testing to estimate the probability p of a coin coming up heads. The trick is to simply define each flip of the coin to be a random variable X_i as follows

$$X_i = \begin{cases} 1 & \text{i'th toss is heads} \\ 0 & \text{i'th toss is tails} \end{cases}$$

If we now flip the coin n times our estimate $\hat{p}$ for the probability p that our coin comes up heads is simply

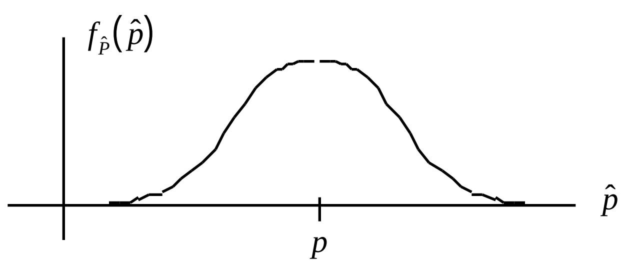
$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$$

a. Make use of the fact that

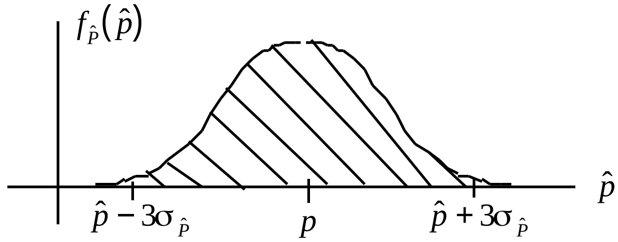
$$Var\left[\hat{p}\right] = \frac{pq}{n} = \frac{p(1-p)}{n} \qquad \text{to show that} \qquad \sigma_{\hat{p}} = \frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

b. Describe how $\sigma_{\hat{p}}$ depends on n equal to the number of tosses in the sample

6. From Problem (5) we know that $\hat{p}$ has a Gaussian distribution (when n ≥ 30) as follows



and we have an equation for $\sigma_{\hat{p}}$. Therefore we can say that 99.9% of the time our estimate $\hat{p}$ is approximately within $3\sigma_{\hat{p}}$ of p as follows



We call this the confidence interval. **Memorize** this result

a. Why is $3\sigma_{\hat{p}}$ only an approximation of the 3σ-point
b. Find the confidence interval when p = 0.4 and n = 100
c. Find n for a confidence interval of ±0.05

7. Up to now we've been assuming that the sample size n is large enough (at least around 30) for us to apply the Central Limit Theorem in our calculations of confidence intervals for sample means. But when n is less than 30 the Gaussian approximation is no longer valid. But as long as the underlying population is Gaussian - as it is for sample means $\bar{X}$ - we can use **Student's t-distribution** $f_T(t)$ with t as follows

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

Student's t-distribution is similar to the Gaussian but "flatter" and more spread out. Note in particular that the Student t-Distribution has a different curve for each value of n. And that $n - 1$ is referred to as the **degrees of freedom (df)** of the distribution.

Use a table of Student's t-distribution to find the 90% confidence interval for a sample of size $n = 20$ with $\bar{X} = 7.8$ and $s = 3$. Go to the table and find the value of $t_{0.05}$ for 19 degrees of freedom. And then substitute into

$$P(-t_{0.05} < t < t_{0.05}) = 0.90$$

and find the corresponding confidence interval for μ