CE2100/CZ2100
Probability and Statistics for Computing

Tutorial 1 - Sampling Distributions

Problem 1

To determine whether a bottling machine is working satisfactorily, a production line manager randomly samples ten 12-ounce bottles every hour and measures the amount of beverage in each bottle. The mean $\bar{x}$ of the 10 fill measurements is used to decide whether to readjust the amount of beverage delivered per bottle by the filling machine.

If records show that the amount of fill per bottle is normally distributed, with a standard deviation of .2 ounce, and if the bottling machine is set to produce a mean fill per bottle of 12.1 ounces, what is the approximate probability that the sample mean $\bar{x}$ of the 10 test bottles is less than 12 ounces?

Problem 2

A college C would like to have 1050 freshmen, and cannot accommodate more than 1060. Assume that each applicant accepts with probability $p = 0.6$ and that the acceptance can be modeled with Binomial distribution. If the college accepts 1700 freshmen, what is the probability that it will have too many acceptances?

Problem 3

The proportion of individuals with an Rh-positive blood type is 85%. You have a random sample of $n = 500$ individuals. What is the probability that the sample proportion $\hat{p}$ lies between 83% and 88%?

Problem 4

A housing survey was conducted to determine the price of a typical home in Topanga, CA. The mean price of a house was roughly \$1.3 million with a standard deviation of \$300,000. There were no houses listed below \$600,000 but a few houses above \$3 million.
What is the probability that the mean of 60 randomly chosen houses in Topanga is more than \$1.4 million?

Problem 5

Suppose that an insurance company has 10,000 policy holders. The expected yearly claim per policyholder is \$240 with a standard deviation of \$800. What is the approximate probability that the total yearly claims $S_{10,000} > \$2.6$ million?

Additional Drill Questions (Do not discuss in the tutorial)

Problem 6

Suppose you roll a 6-sided die 10 times. Let X be the total value of all 10 dice, *i.e.*, $X = X_1 + X_2 + \cdots + X_{10}$. You win the game if $X \leq 25$ or $X \geq 45$. Use the central limit theorem to calculate the probability that you win.

Problem 7

Suppose you have a new algorithm and want to test its running time. You have an idea of the variance of the algorithm's run time: $\sigma^2 = 4\,\text{second}^2$ but you want to estimate the mean: $\mu = t\,\text{second}$. You can run the algorithm repeatedly. How many trials do you have to run so that your estimated runtime is $t \pm 0.5$ with 95% certainty?

Problem 8

Suppose $X_1, X_2, \ldots, X_{30}$ are independent Poisson random variables with mean $\mathbb{E}(X_i) = 2$ and $Var(X_i) = 2$. Use the central limit theorem to approximate

$$P\left(\sum_{i=1}^{30} X_i > 50\right).$$

Problem 9 (Not included in quizzes)

Let X_i =weight of car i and Y_i =fuel in gallons to go 100 miles. We use the model $Y_i = \theta X_i + \epsilon_i$ where ϵ_i are independent errors with

$$\mathbb{E}[\epsilon_i] = 0, Var(\epsilon_i) = \sigma^2$$

How do we estimate θ from data? We minimize the least squares criterion

$$SS(\theta) = \sum_{i=1}^{n} (Y_i - \theta X_i)^2$$

which is minimized by

$$\hat{\theta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

What is the distribution of $\hat{\theta} - \theta$? (Note X_i is not a random variable in this question.)