Midterm 1 Practice Problems

These are a collection of practice problems for the first midterm exam. If you can do these problems (without looking at solutions), there is a high probability that you will do well on the exam.

---

1. Let $\phi(z)$ and $\Phi(z)$ be the pdf and cdf of the standard normal distribution. Suppose that Z is a standard normal random variable and let $X = 3Z + 1$.

(a) Express $F_X(x) = P(X \leq x)$ in terms of Φ.

(b) Use this to find the pdf of X in terms of ϕ.

(c) Find $P(-1 \leq X \leq 1)$.

(d) Recall that the probability that Z is within one standard deviation of its mean is approximately 68%. What is the probability that X is within one standard deviation of its mean?

2. Suppose $Z \sim N(0, 1)$ and $Y = Z^2$. What is the pdf of Y?

3. Let $X_1, X_2, \ldots X_n$ be iid standard normal random variables. Let

$$Y_n = X_1^2 + X_2^2 + \ldots + X_n^2.$$

(a) Show that $E(X_j^2) = 1$.

(b) Set up an integral for $E(X_j^4)$. For extra credit, use integration by parts to show that

$$E(X_j^4) = 3.$$

(If you can't figure out how to do this, just use this number in the next part)

(c) Deduce from (a) and (b) that $\text{Var}(X_j^2) = 2$.

(d) Use the Central Limit Theorem to approximate $P(Y_{100} > 110)$.

4. Suppose $X_1, X_2, \ldots X_{100}$ are iid with mean $\mu = E(X_j) = 1/5$ and variance $\sigma^2 = \text{Var}(X_j) = 1/9$. Use the central limit theorem to estimate

$$P\left(\sum_{j=1}^{100} X_j < 30\right).$$

5. Let $X \sim \text{Binomial}(100, 1/3)$. Use the Central Limit Theorem to give an approximation of $P(X \leq 30)$.

6. The average IQ in a population is 100 with standard deviation 15 (by definition IQ is actually curved so that this is case). What is the probability that a randomly selected group of 100 people has an average IQ above 115?

7. Let X and Y be independent normal random variables, where $X \sim N(2,5)$ and $Y \sim N(5,9)$, let $W = 3X - 2Y + 1$.

(a) Compute $E(W)$ and $\text{Var}(W)$.

(b) It is known that the sum of independent normal random variables is normal (take this as a given). Compute $P(W \leq 6)$.

8. The SAT math scores across the population of high school seniors follow a normal distribution with mean 500 and standard deviation 100

(a) If five seniors are randomly selected, what is the sampling distribution of the sum of the scores?

(b) If five seniors are randomly selected, what is the sampling distribution of the average of their scores?

9. A random sample of size 16 is taken from a population with mean 10 and variance 4. What is the approximate probability that the sample mean $\overline{X}$ takes a value between 9 and 11?

10. Let $X_1, X_2, \ldots X_5$ be a random sample of size 5 taken from a the uniform distribution on $(0,2)$.

(a) What is the probability density function of $X_{(1)} = \min\{X_1, \ldots, X_5\}$?

(b) What is the probability that the maximum value in the sample is less than 1.5?

11. Suppose that X is a discrete random variable with the following probability mass function:

$$P(X=1) = \theta/2, \quad P(X=2) = \theta/4, \quad P(X=3) = 1 - \frac{3\theta}{4}.$$

We observe a sample of size 5 with values $(1,2,2,3,3)$.

(a) Find the method of moments estimate of θ.

(b) Find the maximum likelihood estimate (MLE) of θ.

12. Let $X_1, X_2, \ldots, X_n$ be a random sample from Binomial(k, p) with pmf

$$p(x) = \binom{k}{x} p^x (1-p)^{k-x}, \quad x = 0, 1, \ldots k.$$

(a) Find a sufficient statistic for p.

(b) Find the maximum likelihood estimate of p.

(c) Show that the MLE is unbiased.

13. There are three boxes containing 0, θ and $\theta + 1$ jellybeans (θ is an unknown parameter here). Each of n people opens one of the boxes uniformly at random and takes all of the jellybeans in that box. The boxes are reset after each person takes their turn. Let $X_1, X_2, \ldots, X_n$ be the number of jellybeans take by each of the n people.

(a) Calculate the Bias of $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i$ as an estimator for θ.

(b) For what value of θ is $\hat{\theta}$ and unbiased estimator?

(c) Based on part (a), give an estimator for θ in terms of $\hat{\theta}$ that is always unbaised.

14. Let $Y_1, Y_2, \ldots Y_n$ be a random sample from a Poisson(θ) distribution.

a. Find the maximum likelihood estimator $\hat{\Theta}_{MLE}$ for θ.

b. Show that $\hat{\Theta}_{MLE}$ is an unbiased and consistent estimator for θ.

c. Show that the sample variance S^2 is also an unbiased and consistent estimator for θ.

15. A random sample of 300 CitiBank VISA cardholder accounts indicated a sample mean debt of 1220 with a sample standard deviation of 840. Construct an approximate 95% confidence interval for the average debt of all cardholders, using the fact that $P(Z > 1.96) = 0.025$ for a standard normal random variable Z.

16. Five smokers between the ages of 25 and 30 who were participating in a heart study carried out in Framingham, Massachusetts, were randomly selected. The following observations are readings of their systolic blood pressure: $124, 134, 136, 125, 133$. The sample mean was 130.4 with standard deviation 5.50. a) Find a 95% confidence interval for the mean μ. b) Based on your result in part (a), would you conclude at the $\alpha = 0.05$ significance level that the mean systolic blood pressure for smokers is greater than 125?

17. A political candidate in a close race asks her staff to take a poll to estimate the proportion of voters supporting her. If she assumes that the true proportion p is close to 0.5, what sample size is required so that the standard error of $\hat{p}$ is less than 0.02?

18. The fracture strength of tempered glass averages 14 (measured in thousands of pounds per square inch) and has a standard deviation of 2. Use the central limit theorem to:

a. Estimate the probability that the average fracture strength of 100 randomly selected pieces of the glass exceeds 14.5.

b. Find an interval that includes, with probability 0.95, the average fracture strength of 100 randomly selected pieces of this glass.

19. The fracture strength of tempered glass averages 14 (measured in thousands of pounds per square inch) and has a standard deviation of 2. Use the central limit theorem to:

a. Estimate the probability that the average fracture strength of 100 randomly selected pieces of the glass exceeds 14.5.

b. Find an interval that includes, with probability 0.95, the average fracture strength of 100 randomly selected pieces of this glass.

20. In a study of the effect of glucose on memory in elderly people, scientists tested 16 volunteers for long-term memory, recording the number of words recalled from a list read to each person. Each person was assigned a score based on how well they performed. The sample mean and standard deviation of these score were given by $\overline{x} = 79.47$ and $s = 25.25$. Assume the scores are normally distributed.

a. Give a 99% confidence interval for the mean of the scores.

b. Give a 90% confidence interval for the standard deviation of the score.

21. Let $Y_1, Y_2, \ldots Y_n$ denote a random sample from the from a distribution with PDF given by

$$f(y;\theta) = \begin{cases} \frac{1}{\Gamma(\alpha)\theta^\alpha} y^{\alpha-1} e^{-y/\theta} & y > 0 \\ 0 & \text{otherwise} \end{cases},$$

where $\alpha > 0$ is a known parameter (note that this is just a $\Gamma(\alpha, 1/\theta)$ distribution).

a. Find the maximum likelihood estimator $\hat{\Theta}$ for θ.

b. Show that $\hat{\Theta}$ is a consistent estimator for θ.

c. Show that $Q = \alpha n \hat{\Theta}/\theta$ is a pivotal quantity. What is its distribution?

d. Use the pivotal quantity from part (c) to construct a $100(1-\alpha)\%$ confidence interval for θ. You should write things in terms of the inverse CDF values $\gamma_{\alpha,\beta,p} = F_Y^{-1}(1-p)$ for particular choices of α, β, and p, where Y is a Gamma(α, β) random variable.

22. It is known that the probability p of tossing heads on an unbalanced coin is either $1/4$ or $3/4$. The coin is tossed twice and a value for Y, the number of heads, is observed. For each possible value of Y, which of the two values for p ($1/4$ or $3/4$) maximizes the probability that $Y = y$? Depending on the value of y actually observed, what is the MLE of p?