Math 466/566 - Midterm Solutions

NOTE: These solutions are for both the 466 and 566 exam. The problems are the same until questions 4 and 5.

1. The moment generating function of a random variable X is

$$M(t) = \left(\frac{1}{1-t}\right)^4$$

(a) Find the mean and variance of X. *Solution:* You may recognize this as the moment generating function of the gamma distribution. If so, you see that $\alpha = 4$ and $\beta = 1$. Then the mean is $\frac{\alpha}{\beta} = 4$ and the variance is $\frac{\alpha}{\beta^2} = 4$.

If you don't recognize the gamma mgf, you can compute:

$$\begin{aligned} M'(t) &= \frac{4}{(1-t)^5}, \qquad M'(0) = 4 \\ M''(t) &= \frac{20}{(1-t)^6}, \qquad M''(0) = 20 \end{aligned}$$

So the mean is 4, and the variance is $20 - 4^2 = 4$.

(b) Let $X_1, X_2, \cdots, X_n$ be i.i.d. with the same distribution as X. Suppose that $n = 100$. Then the probability

$$P(95 \le \sum_{i=1}^{100} X_i \le 110)$$

is approximately equal to $P(a \le Z \le b)$ for some a and b where Z is a standard normal random variable. Find a and b.

Solution: The mean of $\sum_i X_i$ is $n\mu = 400$ and its variance is $n\sigma^2 = 400$. So the standard deviation is 20. If we let

$$Z = \frac{\sum_{i=1}^{100} X_i - 400}{20}$$

then Z is approximately a standard normal. So

$$P(95 \le \sum_{i=1}^{100} X_i \le 110) = P(\frac{95-400}{20} \le Z \le \frac{110-400}{20}) = P(-15.25 \le Z \le -14.5)$$

There was a mistake in the problem. I meant to ask for the probability the sum is between 380 and 430 which becomes $P(-1 \leq Z \leq 1.5)$.

2. We consider the following two populations. Population 1 is all working adults in the US with a college degree. Population 2 is all working adults in the US without a college degree. We consider their annual income in thousands of dollars, and let μ_1 and μ_2 be the means for the two populations. And let σ_1^2 and σ_2^2 be the variances for the two populations. We want to estimate $\mu_1 - \mu_2$, the average increase in salary from a college degree. Samples of size $n_1 = 400$ and $n_2 = 100$ are randomly chosen from the two populations. We find that their sample means and variances are

$$\bar{X}_{1,n_1} = 51.6, \quad s_1^2 = 224, \qquad \bar{X}_{2,n_2} = 27.9, \quad s_2^2 = 63$$

(Remember these are in thousands of dollars, so 51.6 is \$51,600.)

(a) $\bar{X}_{1,n_1} - \bar{X}_{2,n_2}$ is the natural estimator for $\mu_1 - \mu_2$. What is the variance of this estimator in terms of σ_1^2 and σ_2^2 ?

Solution:

$$var(\bar{X}_{1,n_1} - \bar{X}_{2,n_2}) = var(\bar{X}_{1,n_1}) + var(\bar{X}_{2,n_2}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

(b) Find a 95% confidence interval for $\mu_1 - \mu_2$.

Solution: Using the above and using the sample variances to estimate the population variances our estimate of $var(\bar{X}_{1,n_1} - \bar{X}_{2,n_2})$ is

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{224}{400} + \frac{63}{100} = 1.19$$

So the standard deviation is 1.09. The confidence interval is then

$$[(51.6 - 27.9) - 1.96 * 1.09, (51.6 - 27.9) + 1.96 * 1.09] = [21.56, 25.84]$$

Comment: Several students did a pooled estimate of a common population variance. This is appropriate only if you have some reason to think the two populations have the same variance. Here I would expect the college educated population with its larger salaries to have a larger variance.

3. The NRA claims that 40% of the US adult population is opposed to gun control legislation. To test this claim against the hypothesis that the percentage is less than 40%, a random sample of 400 US adults is chosen. It is found that 140 of the 400 are opposed to such legislation.

(a) State the null and alternative hypotheses.

Solution: This is a test on a single population proportion. Let p denote the proportion of the US adult population that is opposed to gun control. The null hypothesis is $p = 0.4$ and the alternative hypothesis is $p < 0.4$.

(b) Specify what the test is if we want a significance level of 0.05, and decide if you accept the null or alternative hypothesis. *Solution:* The statistic is

$$Z = \frac{f - 0.4}{\sqrt{0.4(1 - 0.4)/n}}$$

where f is the sample proportion. It is a one sided alternative with a level of 0.05, so we reject the alternative if $Z < -1.645$. For our data, $f = 140/400 = 0.35$, and we find $Z = -2.04$, so we reject the null hypothesis in favor of the alternative hypothesis.

4. A population has unknown mean μ and known variance $\sigma^2 = 400$. We want to test the null hypothesis $\mu = 100$ against the alternative hypothesis $\mu > 100$. We have a large sample with sample mean $\bar{X}_n$. Let

$$Z = \frac{\bar{X}_n - 100}{\sigma/\sqrt{n}}$$

Our test is that we reject the null hypothesis if $Z > 1.645$.

(a) What is the significance level of this test?

Solution: The level is $P(Z > 1.645)$ if the null hypothesis is true. In this case Z is a standard normal and so the level is 5%.

(466 b) Recall that the power is the probability we reject the null hypothesis. It depends on μ. Suppose that $n = 100$. What is the power when $\mu = 105$?

Solution: We want to compute $P(Z > 1.645)$ assuming that $\mu = 105$. Note that in this case Z is not a standard random normal. Instead,

$$Z' = \frac{\bar{X}_n - 105}{\sigma/\sqrt{n}}$$

is a standard random normal. So

$$\begin{aligned} P(Z > 1.645) &= P(\frac{\bar{X}_n - 100}{\sigma/\sqrt{n}} > 1.645) \\ &= P(\frac{\bar{X}_n - 105}{\sigma/\sqrt{n}} > \frac{100 - 105}{\sigma/\sqrt{n}} + 1.645) \\ &= P(Z' > -2.5 + 1.645) = P(Z > -0.855) \approx 0.804 \end{aligned}$$

(566 b) Recall that the power is the probability we reject the null hypothesis. It depends on μ. Suppose we want the power to be at least 90% when $\mu \geq 102$. How large must the sample size n be ?

Solution: In the range $\mu \geq 102$, the power is smallest at $\mu = 102$. So we want $P(Z > 1.645) = 0.9$ when $\mu = 102$. The standard normal RV now is not Z but rather

$$Z' = \frac{\bar{X}_n - 102}{\sigma/\sqrt{n}}$$

So

$$\begin{aligned} 0.9 &= P(Z > 1.645) = P(Z' > 1.645 + \frac{100 - 102}{\sigma/\sqrt{n}}) \\ &= P(Z' > 1.645 - \sqrt{n}/10) \end{aligned}$$

From the table, $P(Z' > -1.28) = 0.9$, so $1.645 - \sqrt{n}/10 = -1.28$, which give $n = 856$.

5. (466) A random sample $\{X_1, X_2, X_3\}$ of size 3 is drawn from a population with mean μ and variance σ^2. (So X_1, X_2, X_3 are i.i.d.) Let

$$\begin{aligned} T_1 &= \frac{1}{3}(X_1 + X_2 + X_3) \\ T_2 &= \frac{1}{4}(X_1 + 2X_2 + X_3) \end{aligned}$$

(a) Show that T_1 and T_2 are both unbiased estimators of the population mean μ.

Solution: Since $E[X_1] = E[X_2] = E[X_3] = \mu$,

$$\begin{aligned} E[T_1] &= \frac{1}{3}(E[X_1] + E[X_2] + E[X_3]) = \frac{1}{3}3\mu = \mu \\ E[T_2] &= \frac{1}{4}(E[X_1] + 2E[X_2] + E[X_3]) = \frac{1}{4}4\mu = \mu \end{aligned}$$

(b) Compute the variances of T_1 and T_2.

Solution: Since the X_i are independent,

$$\begin{aligned} var(T_1) &= \frac{1}{9}(var(X_1) + var(X_2) + var(X_3)) = \frac{1}{3}\sigma^2 \\ var(T_2) &= \frac{1}{16}(var(X_1) + 4var(X_2) + var(X_3)) = \frac{3}{8}\sigma^2 \end{aligned}$$

(c) Which estimator would you use? Explain your answer.

Solution: Both are unbiased, so their mean square error (risk) is equal to their variance. So I would use T_1 since it has smaller variance and hence smaller risk.

5. (566) A random sample $\{X_1, X_2, X_3\}$ of size 3 is drawn from a population with mean μ and variance σ^2. (So X_1, X_2, X_3 are i.i.d.) Let

$$T_a = \frac{1}{4}(X_1 + aX_2 + X_3)$$

(a) For what value of a is T_a an unbiased estimator of the population mean μ.
Solution:

$$E[T_a] = \frac{1}{4}(E[X_1] + aE[X_2] + E[X_3]) = \frac{2+a}{4}\mu$$

For an unbiased estimator this must equal μ, so a must be 2.
(b) Compute the variance of T_a.

Solution:

$$var[T_a] \quad = \quad \frac{1}{16}(var(X_1) + a^2 var(X_2) + var(X_3)) = \frac{2+a^2}{16}\sigma^2$$

(c) Recall that the mean square error is $E[(T_a - \mu)^2]$. Find the value of a which gives the best estimator in the sense of minimizing this mean square error. *Solution:* The bias is $\frac{2+a}{4}\mu - \mu = \frac{a-2}{4}\mu$. So the risk is

$$\left(\frac{a-2}{4}\right)^2 \mu^2 + \frac{2+a^2}{16}\sigma^2$$

With $\mu = 1$ and $\sigma^2 = 1$, this becomes

$$\left(\frac{a-2}{4}\right)^2 + \frac{2+a^2}{16}$$

This attains its mininum at $a = 1$.