Math 224 **Exam 3**
Spring 2022 **Thurs Apr 28**

This is a closed book test. You may use a basic scientific calculator. There are 5 pages and 5 problems, each worth 6 points, for a total of 30 points.

Problem 1. Let X be a continuous random variable with the following density function:

$$f_X(x) = \begin{cases} c(x-1) & 1 < x < 2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the value of the constant c.

The total area under a probability density is 1:

$$1 = \int_1^2 c(x-1)\,dx = c\left[\frac{x^2}{2} - x\right]_1^2 = c\left[(4/2-2) - (1/2-1)\right] = c/2.$$

Hence we must have $c = 2$.

(b) Compute the expected value $E[X]$.

The expected value is

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx \\ &= \int_1^2 x \cdot 2(x-1)\,dx \\ &= \int_1^2 2(x^2 - x)\,dx \\ &= 2\left[\frac{x^3}{3} - \frac{x^2}{2}\right]_1^2 \\ &= 2\left[(8/3 - 4/2) - (1/3 - 1/2)\right] = 5/3. \end{aligned}$$

(c) Compute the second moment $E[X^2]$ and the variance $\mathrm{Var}(X)$.

The second moment is

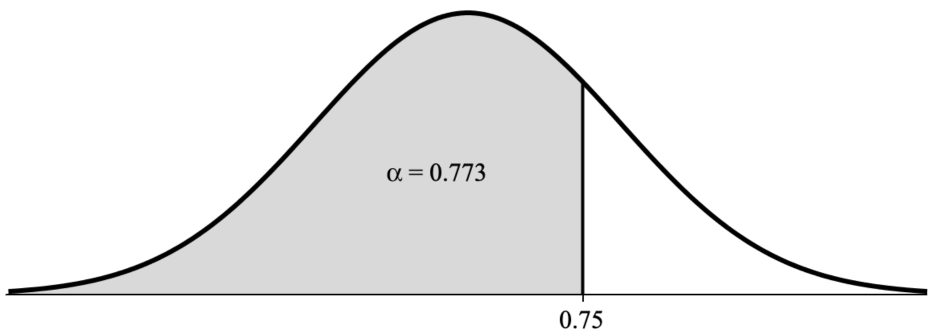
$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 \cdot f_X(x)\,dx \\ &= \int_1^2 x^2 \cdot 2(x-1)\,dx \\ &= \int_1^2 2(x^3 - x^2)\,dx \\ &= 2\left[\frac{x^4}{4} - \frac{x^3}{3}\right]_1^2 \\ &= 2\left[(16/4 - 8/3) - (1/4 - 1/3)\right] = 17/6. \end{aligned}$$

Hence the variance is

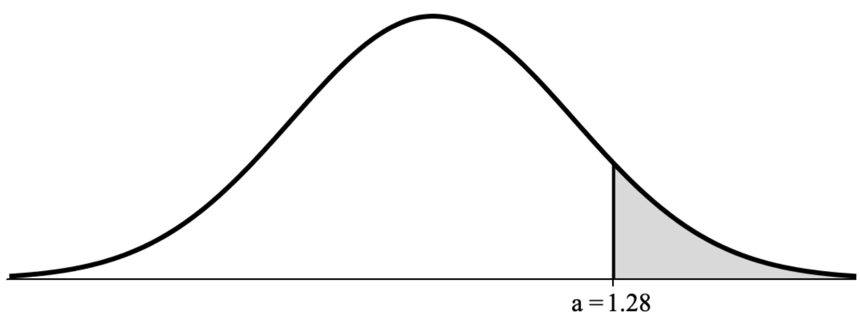
$$\text{Var}(X) = E[X^2] - E[X]^2 = (17/6) - (5/3)^2 = 1/18.$$

Problem 2. Let $Z \sim N(0,1)$ be a standard normal random variable. Use the attached tables to solve the following problems.

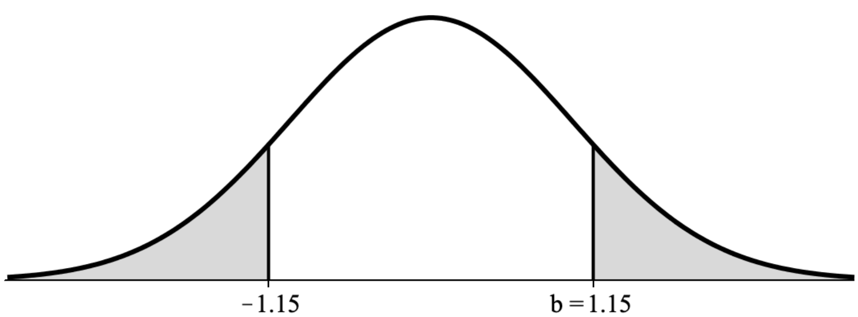
(a) Find α such that $P(Z < 0.75) = \alpha$ and draw a picture to illustrate your answer.



(b) Find a such that $P(Z > a) = 10\%$ and draw a picture to illustrate your answer.



(c) Find b such that $P(|Z| > b) = 25\%$ and draw a picture to illustrate your answer.



Problem 3. Let $X_1, \ldots, X_{16}$ be an iid sample from a distribution with unknown mean μ and variance σ^2. Consider the sample mean:

$$\overline{X} = \frac{1}{16}\left(X_1 + X_2 + \cdots + X_{16}\right).$$

(a) Compute the expected value of the sample mean, $E[\overline{X}]$. [Hint: Use linearity.]

$$\begin{aligned} E[\overline{X}] &= \frac{1}{16}\left(E[X_1] + E[X_2] + \cdots + E[X_{16}]\right) \\ &= \frac{1}{16}\underbrace{(\mu + \mu + \cdots + \mu)}_{\text{16 times}} = \mu \end{aligned}$$

(b) Compute the variance of the sample mean, $\mathrm{Var}(\overline{X})$. [Hint: The X_i are independent.]

$$\begin{aligned} \mathrm{Var}(\overline{X}) &= \frac{1}{16^2}\left(\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_{16})\right) \\ &= \frac{1}{16^2}\underbrace{(\sigma^2 + \sigma^2 + \cdots + \sigma^2)}_{\text{16 times}} = \frac{\sigma^2}{16} \end{aligned}$$

(c) Use the Central Limit Theorem to estimate $P(\mu - 0.1\sigma < \overline{X} < \mu + 0.1\sigma)$. [Hint: Yes, this is possible. The unknown values of μ and σ will cancel out.]

The CLT says that $\overline{X}$ is approximately normal. Since $E[\overline{X}] = \mu$ and $\mathrm{Var}(\overline{X}) = \sigma^2/16$, this tells us that $\left(\overline{X} - \mu\right)/\sqrt{\sigma^2/16} = \left(\overline{X} - \mu\right)/(\sigma/4) = 4(\overline{X} - \mu)/\sigma$ is approximately standard normal. Hence

$$\begin{aligned} P(\mu - 0.1\sigma < \overline{X} < \mu + 0.1\sigma) &= P(-0.1\sigma < \overline{X} - \mu < 0.1\sigma) \\ &= P\left(-0.1 < \frac{\overline{X} - \mu}{\sigma} < 0.1\right) \\ &= P\left(-0.4 < \frac{4(\overline{X} - \mu)}{\sigma} < 0.4\right) \\ &\approx \Phi(0.4) - \Phi(-0.4) \\ &= \Phi(0.4) - (1 - \Phi(0.4)) \\ &= 2 \cdot \Phi(0.4) - 1 \\ &= 2(0.6554) - 1 \\ &\approx 31.08\%. \end{aligned}$$

Problem 4. There are 5 different colors of Skittles. Let p be the unknown proportion of red Skittles in a very large bag of Skittles. In order to estimate p you took a sample of $n = 100$ Skittles from the bag and found that $Y = 25$ were red. Use the attached table to solve the following problems.

(a) Compute a symmetric two-sided 95% confidence interval for the unknown p.

The general form[1] of the interval is

$$|p - \hat{p}| < z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}$$

$$\text{or} \qquad \text{“}p = \hat{p} \pm z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}\text{”}.$$

In our case we have $\alpha = 5\%$, so that $z_{\alpha/2} = z_{2.5\%} = 1.96$. We also have $\hat{p} = Y/n = 25/100 = 1/4$, so our confidence interval is

$$\begin{aligned} p &= 1/4 \pm 1.96 \cdot \sqrt{(1/4)(3/4)/100} \\ &= 0.25 \pm 0.085 \\ &= 25\% \pm 8.5\%. \end{aligned}$$

(b) Test the hypothesis $H_0 =$ “$p = 1/5$” against the alternative $H_1 =$ “$p > 1/5$” at the 5% level of significance.

The rejection region for a general one-sided test is

$$\hat{p} > p_0 + z_\alpha \cdot \sqrt{p_0(1-p_0)/n}.$$

In our case we have $\alpha = 5\%$, so that $z_\alpha = z_{5\%} = 1.645$, and $p_0 = 1/5$. So our rejection region is

$$\begin{aligned} \hat{p} &> 1/5 + 1.645 \cdot \sqrt{(1/5)(2/5)/100} \\ &= 0.2 + 0.0658 \\ &= 0.2658. \end{aligned}$$

In other words, any value of $\hat{p}$ greater than 0.2658 should cause us to reject the hypothesis “$p = 1/5$” in favor of “$p > 1/5$” at the 5% level of significance. In our case we had $\hat{p} = 0.25$, so we **fail to reject** the hypothesis “$p = 1/5$”.

Problem 5. Suppose that the following iid sample comes from a **normal distribution** with unknown mean μ and variance σ^2:

1.0	2.0	3.0	4.0

Use the attached table to solve the following problems.

(a) Compute a symmetric two-sided 95% confidence interval for the unknown μ.

Let X be the unknown distribution from which the sample was drawn. The general form[2] of the confidence interval is

$$|\mu - \overline{X}| < t_{\alpha/2}(n-1) \cdot \sqrt{S^2/n}$$

$$\text{or} \qquad \text{“}\mu = \overline{X} \pm t_{\alpha/2}(n-1) \cdot \sqrt{S^2/n}\text{”}.$$

[1] This is just the most basic form. It can be made more accurate in various ways.

[2] This will be accurate only to the extent that X has a normal distribution.

In our case we have $\alpha = 5\%$ and $n = 4$, so that $t_{\alpha/2}(n-1) = t_{2.5\%}(3) = 3.182$. The sample mean and sample variance are computed as follows:

$$\overline{X} = \frac{1}{4}(1+2+3+4) = 5/2,$$
$$S^2 = \frac{1}{3}\left((1-5/2)^2 + (2-5/2)^2 + (3-5/2)^2 + (4-5/2)^2\right) = 5/3.$$

Hence our confidence interval is

$$\begin{aligned}\mu &= 5/2 \pm 3.182 \cdot \sqrt{(5/3)/4} \\ &= 2.5 \pm 2.054.\end{aligned}$$

That's not great, but what do you expect with only four data points?

(b) Test the hypothesis $H_0 =$ "$\mu = 4$" against the alternative $H_1 =$ "$\mu < 4$" at the 5% level of significance.

The rejection region for a general one-sided test is

$$\overline{X} < \mu_0 - t_\alpha(n-1) \cdot \sqrt{S^2/n}.$$

In our case we have $\alpha = 5\%$ and $n = 4$ so that $t_\alpha(n-1) = t_{5\%}(3) = 2.353$, and we have $\mu_0 = 4$. So our rejection region is

$$\begin{aligned}\overline{X} &< 4 - 2.353 \cdot \sqrt{(5/3)/4} \\ &= 4 - 1.52 \\ &= 2.48.\end{aligned}$$

In other words, any value of $\overline{X}$ less than 2.48 should cause us to reject the hypothesis "$\mu = 4$" in favor of "$\mu < 4$" at the 5% level of significance. In our case we have $\overline{X} = 2.5$, which is not less than 2.48, so we **fail to reject** the hypothesis "$\mu = 4$". You might object and say that the true mean is surely less than 4, but we can't be 95% confident about this because of the small sample size.