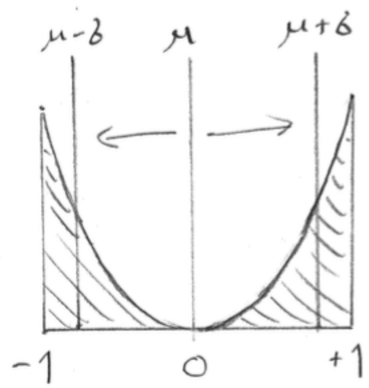
Math 224 **Summer 2019**
Homework 5 **Drew Armstrong**

1. Let $X_1, X_2, \ldots, X_{15}$ be independent and identically distributed (iid) random variables. Suppose that each X_i has pdf defined by the following function:

$$f(x) = \begin{cases} \frac{3}{2} \cdot x^2 & \text{if } -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Compute $E[X_i]$ and $\text{Var}(X_i)$.
(b) Consider the sum $\Sigma X = X_1 + X_2 + \cdots + X_{15}$. Compute $E[\Sigma X]$ and $\text{Var}(\Sigma X)$.
(c) The Central Limit Theorem says that ΣX is approximately normal. Use this fact to estimate the probability $P(-0.18 \leq \Sigma X \leq 0.36)$.

(a) Here is a graph of the pdf of each individual X_i:



Since the distribution is symmetric about zero, we conclude without doing any work that $\mu = E[X_i] = 0$ for each i. To find σ, however, we need to compute an integral. For any i, the variance of X_i is given by

$$\begin{aligned}
\sigma^2 = \text{Var}(X_i) &= E\left[X_i^2\right] - E[X_i]^2 \\
&= E[X_i^2] - 0 \\
&= \int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx - 0 \\
&= \int_{-1}^{1} x^2 \cdot \frac{3}{2} x^2\, dx \\
&= \frac{3}{2} \int_{-1}^{1} x^4\, dx \\
&= \frac{3}{2} \cdot \frac{x^5}{5} \bigg|_{-1}^{1} = \frac{3}{2} \cdot \frac{1}{5} - \frac{3}{2} \cdot \frac{(-1)^5}{5} = \frac{6}{10} = \frac{3}{5}.
\end{aligned}$$

(b) It follows that ΣX has mean and variance given by

$$E[\Sigma X] = E[X_1] + E[X_2] + \cdots + E[X_{15}] = 0 + 0 + \cdots + 0 = 0$$

and

$$\text{Var}(\Sigma X) = \text{Var}(X_1) + \text{Var}(X_2) + \cdots + \text{Var}(X_{15}) = \frac{3}{5} + \frac{3}{5} + \cdots + \frac{3}{5} = 15 \cdot \frac{3}{5} = 9,$$

By the Central Limit Theorem, the sum ΣX is approximately normal and hence

$$\frac{\Sigma X - E[\Sigma X]}{\sqrt{\text{Var}(\Sigma X)}} = \frac{\Sigma X - 0}{\sqrt{9}} = \frac{\Sigma X}{3}$$

is approximately **standard** normal. We conclude that

$$\begin{aligned} P(-0.18 \leq \Sigma X \leq 0.36) &= P\left(\frac{-0.18}{3} \leq \frac{\Sigma X}{3} \leq \frac{0.36}{3}\right) \\ &= P\left(-0.06 \leq \frac{\Sigma X}{3} \leq 0.12\right) \\ &\approx \Phi(0.12) - \Phi(-0.06) \\ &= \Phi(0.12) - [1 - \Phi(0.06)] \\ &= \Phi(0.12) + \Phi(0.06) - 1 \\ &= 0.5478 + 0.5239 - 1 = 7.17\%. \end{aligned}$$

2. Suppose that $n = 48$ seeds are planted and suppose that each seed has a probability $p = 2/3$ of germinating. Let X be the number of seeds that germinate and use the Central Limit Theorem to estimate the probability $P(30 \leq X \leq 33)$ that between 30 and 33 seeds germinate (inclusive). Don't forget to use a continuity correction.

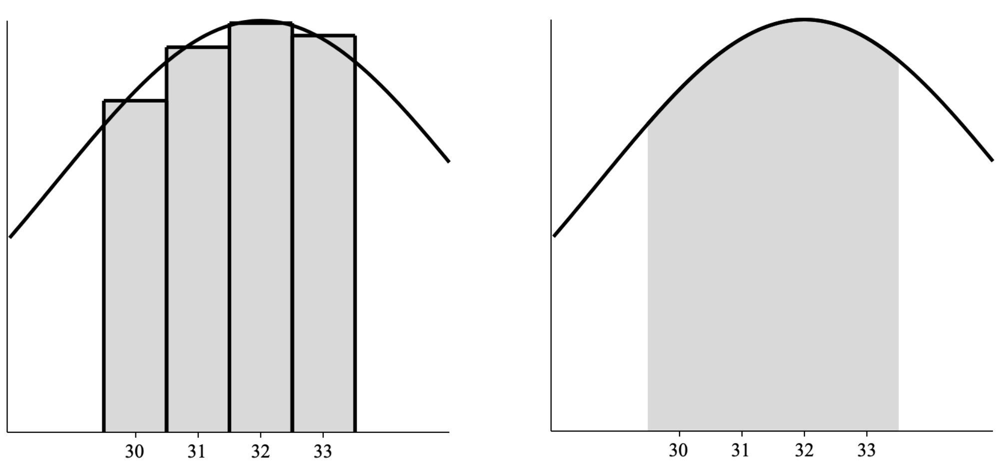
We observe that X is a binomial random variable with the following pmf:

$$P(X = k) = \binom{48}{k}(2/3)^k(1/3)^{48-k}.$$

My laptop tells me that the exact probability is

$$P(30 \leq X \leq 33) = \sum_{k=30}^{33} P(X = k) = \sum_{k=30}^{33} \binom{48}{k}(2/3)^k(1/3)^{48-k} = 45.18\%.$$

To compute an approximation by hand we will use the de Moivre-Laplace Theorem, which says that X is approximately normal with mean $np = 32$ and variance $\sigma^2 = np(1-p) = 32/3$, i.e., standard deviation $\sigma = \sqrt{32/3} = 3.266$. Let X' be a **continuous** random variable with $X' \sim N(\mu = 32, \sigma^2 = 32/3)$. Here is a picture comparing the probability **mass** function of the discrete variable X to the probability **density** function of the continuous variable X':



The picture suggests that we should use the following continuity correction:[1]

$$P(30 \leq X \leq 33) \approx P(29.5 \leq X' \leq 33.5).$$

And then because $(X' - 32)/3.266$ is **standard** normal we obtain

$$\begin{aligned}
P(29.5 \leq X' \leq 33.5) &= P(-2.5 \leq X' - 32 \leq 1.5) \\
&= P\left(-0.77 \leq \frac{X' - 32}{3.266} \leq 0.46\right) \\
&= \Phi(0.46) - \Phi(-0.77) \\
&= \Phi(0.46) - [1 - \Phi(0.77)] \\
&= \Phi(0.46) + \Phi(0.77) - 1 = 0.6772 + 0.7794 - 1 = 45.66\%.
\end{aligned}$$

Not too bad.

3. Suppose that a coin is flipped 100 times and let X be the number of heads. We will use $\hat{p} = X/100$ as an (unbiased) estimator for the unknown probability of heads: $p = P(H)$.

(a) Assuming that the coin is **fair** ($p = 1/2$), compute $E[\hat{p}]$ and $\text{Var}(\hat{p})$.
(b) In order to test the hypothesis $H_0 =$ "$p = 1/2$" against $H_1 =$ "$p > 1/2$" we flip the coin 100 times and get heads 60 times. Should you reject H_0 in favor of H_1 at the 5% level of significance? At the 2.5% level of significance? At the 1% level of significance? [Hint: $\hat{p}$ is approximately normal.]

(a) In general we know that $E[X] = np$ and $\text{Var}(X) = np(1-p)$ hence $E[\hat{p}] = p$ and $\text{Var}(\hat{p}) = p(1-p)/n$. If $n = 100$ and $p = 1/2$ then we obtain

$$E[\hat{p}] = \frac{1}{2} \qquad \text{and} \qquad \text{Var}(\hat{p}) = \frac{(1/2)(1-1/2)}{100} = \frac{1}{400}.$$

(b) Now consider the null hypothesis:

$$H_0 = \text{“}p = \frac{1}{2}.\text{”}$$

To test this against the alternative hypothesis

$$H_1 = \text{“}p > \frac{1}{2}\text{”}$$

we will assume that H_0 is true and compute the value of $\hat{p}$ that would be so far above $1/2$ to give us an α surprise. In other words: We want to find the critical value k_α such that

$$P(\hat{p} > k_\alpha | H_0) = \alpha.$$

[1] If you don't do this then you will still get a reasonable answer, it just won't be as accurate.

So let us assume that H_0 is true. Then since $n = 100$ is large enough (I guess) we can assume that $(\hat{p} - 1/2)/\sqrt{1/400}$ is approximately standard normal. Thus we obtain

$$P\left(\frac{\hat{p} - 1/2}{\sqrt{1/400}} > z_\alpha \;|H_0\right) = \alpha,$$

$$P\left(\hat{p} - \frac{1}{2} > z_\alpha \cdot \sqrt{\frac{1}{400}} \;|H_0\right) = \alpha,$$

$$P\left(\hat{p} > \frac{1}{2} + z_\alpha \cdot \sqrt{\frac{1}{400}} \;|H_0\right) = \alpha.$$

We conclude that the critical value is

$$k_\alpha = \frac{1}{2} + z_\alpha \cdot \sqrt{\frac{1}{400}}.$$

and the critical region is

$$\hat{p} > k_\alpha,$$

$$\hat{p} > 1/2 + z_\alpha \cdot \sqrt{1/400}.$$

If we perform the experiment and get a value of $\hat{p}$ in this region then we will reject H_0 in favor of H_1. [Reason: If H_0 is true then this result would only have an α chance of happening. If α is small then this means that either (1) something extremely rare happened, or (2) H_0 is wrong.]

Suppose we performed the experiment and got $\hat{p} = 0.6$. Here are three tests:

- At $\alpha = 5\%$ and $z_\alpha = 1.645$ the rejection region is

  $$\hat{p} > 1/2 + z_\alpha \cdot \sqrt{1/400} = 1/2 + 1.645 \cdot \sqrt{1/400} = 0.582.$$

  Thus we reject H_0 in favor of H_1.
- At $\alpha = 2.5\%$ and $z_\alpha = 1.96$ the rejection region is

  $$\hat{p} > 1/2 + z_\alpha \cdot \sqrt{1/400} = 1/2 + 1.96 \cdot \sqrt{1/400} = 0.598.$$

  Thus we reject H_0 in favor of H_1.
- At $\alpha = 1\%$ and $z_\alpha = 2.33$ the rejection region is

  $$\hat{p} > 1/2 + z_\alpha \cdot \sqrt{1/400} = 1/2 + 2.33 \cdot \sqrt{1/400} = 0.617.$$

  Thus we **do not reject** H_0 in favor of H_1.

4. Let p be the true proportion of voters in a population who intend to vote for candidate A. Suppose that we polled 1000 voters and 522 told us that they intend to vote for A. Use this to compute confidence intervals for p at the 90%, 95% and 99% confidence levels. [Hint: You may assume that the voters' responses are iid coin flips.]

Let X be the number of sampled voters who say "yes." We will assume that this is a binomial random variable with $E[X] = np = 1000p$ and $\text{Var}(X) = np(1-p) = 1000p(1-p)$. We will use the sample proportion $\hat{p} = X/1000$ as an estimator for p. Recall that $E[\hat{p}] = p$ and $\text{Var}(\hat{p}) = p(1-p)/1000$. Then since $n = 1000$ is rather large we will assume that

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/1000}} \text{ is approximately standard normal.}$$

Recall that this leads to the following approximate $(1-\alpha)100\%$ confidence interval:

$$\begin{aligned}
\hat{p} - z_{\alpha/2} \cdot \sqrt{\tfrac{\hat{p}(1-\hat{p})}{1000}} &< p < \hat{p} + z_{\alpha/2} \cdot \sqrt{\tfrac{\hat{p}(1-\hat{p})}{1000}}, \\
0.522 - z_{\alpha/2} \cdot \sqrt{\tfrac{(0.522)(0.478)}{1000}} &< p < 0.522 + z_{\alpha/2} \cdot \sqrt{\tfrac{(0.522)(0.478)}{1000}}, \\
0.522 - z_{\alpha/2} \cdot 0.0158 &< p < 0.522 + z_{\alpha/2} \cdot 0.0158.
\end{aligned}$$

Here are three specific levels of confidence:

- If $(1-\alpha)100\% = 90\%$ then $\alpha = 5\%$ and $z_{\alpha/2} = 1.645$, hence
$$\begin{aligned}
0.522 - 1.645 \cdot 0.0158 &< p < 0.522 + 1.645 \cdot 0.0158, \\
49.6\% &< p < 54.8\%.
\end{aligned}$$
- If $(1-\alpha)100\% = 95\%$ then $\alpha = 2.5\%$ and $z_{\alpha/2} = 1.96$, hence
$$\begin{aligned}
0.522 - 1.96 \cdot 0.0158 &< p < 0.522 + 1.96 \cdot 0.0158, \\
49.1\% &< p < 55.3\%.
\end{aligned}$$
- If $(1-\alpha)100\% = 99\%$ then $\alpha = 1\%$ and $z_{\alpha/2} = 2.575$, hence
$$\begin{aligned}
0.522 - 2.575 \cdot 0.0158 &< p < 0.522 + 2.575 \cdot 0.0158, \\
48.1\% &< p < 56.3\%.
\end{aligned}$$

5. Let $X_1, \ldots, X_{10}$ be an iid sample from a **normal distribution** with variance $\sigma^2 = 36$ and unknown mean μ. Suppose that the sample mean is measured to be

$$\overline{X} = \frac{1}{10}(X_1 + \cdots + X_{10}) = 50.$$

Use this to compute confidence intervals for μ at the 90%, 95% and 99% confidence levels. [Hint: $(\overline{X} - \mu)/\sqrt{\sigma^2/n}$ has an exactly standard normal distribution.]

First note that $E[\overline{X}] = \mu$ and $\mathrm{Var}(\overline{X}) = \sigma^2/n$. Since the sample comes from a normal distribution this implies that $(\overline{X} - \mu)/\sqrt{\sigma^2/n}$ is standard normal. It follows that for all $0 < \alpha < 1$ we have

$$\begin{aligned}
P\left(-z_{\alpha/2} < \frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} < z_{\alpha/2}\right) &= 1 - \alpha, \\
P\left(-z_{\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}} < \overline{X} - \mu < z_{\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}}\right) &= 1 - \alpha, \\
P\left(-z_{\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}} < \mu - \overline{X} < z_{\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}}\right) &= 1 - \alpha, \\
P\left(\overline{X} - z_{\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}} < \mu < \overline{X} + z_{\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}}\right) &= 1 - \alpha.
\end{aligned}$$

Thus we obtain the following $(1-\alpha)100\%$ confidence interval:

$$\begin{aligned}
50 - z_{\alpha/2} \cdot \sqrt{\tfrac{36}{10}} &< \mu < 50 + z_{\alpha/2} \cdot \sqrt{\tfrac{36}{10}}, \\
50 - z_{\alpha/2} \cdot 1.897 &< \mu < 50 - z_{\alpha/2} \cdot 1.897.
\end{aligned}$$

Here are three specific levels of confidence:

- If $(1-\alpha)100\% = 90\%$ then $\alpha = 10\%$ and $z_{\alpha/2} = 1.645$, hence

$$\begin{aligned} 50 - 1.645 \cdot 1.895 &< \mu < 50 + 1.645 \cdot 1.895, \\ 46.88 &< \mu < 53.12. \end{aligned}$$

- If $(1-\alpha)100\% = 95\%$ then $\alpha = 5\%$ and $z_{\alpha/2} = 1.96$, hence

$$\begin{aligned} 50 - 1.645 \cdot 1.895 &< \mu < 50 + 1.645 \cdot 1.895, \\ 46.28 &< \mu < 53.72. \end{aligned}$$

- If $(1-\alpha)100\% = 99\%$ then $\alpha = 1\%$ and $z_{\alpha/2} = 2.575$, hence

$$\begin{aligned} 50 - 1.645 \cdot 1.895 &< \mu < 50 + 1.645 \cdot 1.895, \\ 45.11 &< \mu < 54.86. \end{aligned}$$

6. In order to estimate the average weight μ of a chocolate bar, a random sample of $n = 9$ bars from a production line are weighed, yielding the following results in grams:

21.40	18.85	18.55	19.40	19.15	22.45	22.80	22.20	23.15

Use this data to compute a 98% confidence interval for the average weight μ. [Hint: Assume that the weight of each chocolate bar is normally distributed and let $\overline{X}, S^2$ be the sample mean and sample variance. Since $n = 9$ is a small number you should use the fact that $(\overline{X} - \mu)/\sqrt{S^2/n}$ has a t-distribution with 8 degrees of freedom.]

First we compute the sample mean:

$$\overline{X} = \frac{X_1 + \cdots + X_9}{9} = \frac{21.40 + \cdots + 23.15}{9} = 20.883.$$

Then we compute the sample variance:

$$S^2 = \frac{1}{8}\left[(21.40 - 20.883)^2 + \cdots + (23.15 - 20.883)^2\right] = 3.506.$$

Since the 9 data points come from a normal distribution with mean μ we may assume that $(\overline{X} - \mu)/\sqrt{S^2/9}$ has a t-distribution with 8 degrees of freedom. Thus we obtain the following $(1-\alpha)100\%$ confidence interval for μ:

$$\begin{aligned} \overline{X} - t_{\alpha/2}(8) \cdot \sqrt{\tfrac{S^2}{9}} &< \mu < \overline{X} + t_{\alpha/2}(8) \cdot \sqrt{\tfrac{S^2}{9}}, \\ 20.883 - t_{\alpha/2}(8) \cdot \sqrt{\tfrac{3.506}{9}} &< \mu < \overline{X} + t_{\alpha/2}(8) \cdot \sqrt{\tfrac{3.506}{9}}, \\ 20.883 - t_{\alpha/2}(8) \cdot 0.6242 &< \mu < \overline{X} + t_{\alpha/2}(8) \cdot 0.6242. \end{aligned}$$

For a $(1-\alpha)100\% = 98\%$ confidence interval we have $\alpha = 2\%$ and $t_{\alpha/2}(8) = 2.896$, hence

$$\begin{aligned} 20.883 - 2.896 \cdot 0.6242 &< \mu < 20.883 + 2.896 \cdot 0.6242. \\ 19.08 &< \mu < 22.69. \end{aligned}$$