BIOSTAT 704 - Homework 3

Austin Allen

February 15th, 2024

Problem 1: BE Exercise 8.15

For BE 8.15, you must justify each result using a theorem(s) and/or definition(s); that is, you will NOT get full credit for just listing the distribution.

Supose that $X_i \sim N(\mu, \sigma^2), i = 1, ..., n$ and $Z_i \sim N(0, 1), i = 1, ..., k$, and all variables are independent. State the distribution of each of the following variables if it is a "names" distribution or otherwise state "unknown."

Part (a): $X_1 - X_2$

Theorem 8.3.1 describes the linear combination of independent normal random variables.

Let $Y = X_1 - X_2$.

$$
\begin{aligned}
&Y \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right) \\
&\implies Y \sim N\left((1)\mu + (-1)\mu, (1)^2\sigma^2 + (-1)^2\sigma^2\right) \\
&\implies Y \sim N(0, 2\sigma^2)
\end{aligned}
$$

Part (b): $X_2 + 2X_3$

By applying the same results from theorem 8.3.1, we can see that for $Y = X_2 + 2X_3$:

$$
\begin{aligned}
&Y \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right) \\
&\implies Y \sim N\left((1)\mu + (2)\mu, (1)^2\sigma^2 + (2)^2\sigma^2\right) \\
&\implies Y \sim N(3\mu, 5\sigma^2)
\end{aligned}
$$

Part (c): $\frac{X_1 - X_2}{\sigma S_Z \sqrt{2}}$

$(X_1 - X_2)/\sqrt{2}\sigma \sim N(0, 1)$. Thus, by theorem 8.4.2, $\frac{X_1 - X_2}{\sigma S_Z \sqrt{2}} \sim t(k-1)$ (note: $\mu = 0$).

Part (d): Z_1^2

We know from theorem 8.3.5 that if $Z \sim N(0, 1)$, then $Z^2 \sim \chi^2(1)$. Thus, $Z_1^2 \sim \chi^2(1)$.

Part (e): $\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma S_Z}$

Similarly to part c), $(\bar{X} - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$. Thus, $\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma S_Z} \sim t(k-1)$

Part (f): $Z_1^2 + Z_2^2$

We know that $Z^2 \sim \chi^2(1)$ (see theorem 8.3.5). We can combine this with the fact that the sum of independent χ^2 random variables follows a χ^2 distribution with $\sum_{i=1}^n \nu_i$ degrees of freedom. Thus, for $Y = Z_1^2 + Z_2^2, Y \sim \chi^2(2)$.

Part (g): $Z_1^2 - Z_2^2$

Unfortunately, the exact distribution of $Z_1^2 - Z_2^2$ is not possible to derive with the theorems we've learned to this point.

Part (h): $\frac{Z_1}{\sqrt{Z_2^2}}$

From theorem 8.4.1, we know that for independent random variables V and Z, $T = \frac{Z}{\sqrt{V/\nu}} \sim t(\nu)$. We also know and have used theorem 8.3.5 which shows that $Z^2 \sim \chi^2(1)$. Thus, for $T = \frac{Z_1}{\sqrt{Z_2^2}}, T \sim t(1)$.

Part (i): $\frac{Z_1^2}{Z_2^2}$

Theorem 8.4.4 describes a random variable $X = \frac{V_1/\nu_1}{V_2/\nu_2}, X \sim F(\nu_1, \nu_2)$, where $V_1 \sim \chi^2(\nu_1)$ and $V_2 \sim \chi^2(\nu_2)$. Combining this with the fact that $Z^2 \sim \chi^2(1)$, we can deduce that for $X = \frac{Z_1^2}{Z_2^2}, X \sim F(1, 1)$.

Part (k): $\frac{\bar{X}}{\bar{Z}}$

Unfortunately, we do not have enough information to derive the exact distribution of $\frac{\bar{X}}{\bar{Z}}$.

Part (l): $\frac{\sqrt{nk}(\bar{X}-\mu)}{\sigma\sqrt{\sum_{i=1}^k Z_i^2}}$

We can rewrite this to be in this form:

$$\frac{\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}}{\frac{\sqrt{\sum_{i=1}^k Z_i^2}}{\sqrt{k}}}$$

The denominator follows a standard normal distribution and $\sum_{i=1}^k Z_i^2$ follows a χ^2 distribution with k degrees of freedom. Thus, applying theorem 8.4.1, for $Y = \frac{\sqrt{nk}(\bar{X}-\mu)}{\sigma\sqrt{\sum_{i=1}^k Z_i^2}}, Y \sim t(k)$.

Part (m): $\frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma^2} + \sum_{i=1}^k (Z_i - \bar{Z})^2$

We can recognize $\frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma^2}$ as the sample variance, which we know follows a χ^2 distribution with $n-1$ degrees of freedom. We also know that $\sum_{i=1}^k (Z_i - \bar{Z})^2$ follows a χ^2 distribution with k degrees of freedom. Lastly, we know that the sum of two χ^2 random variables with degrees of freedom ν_1 and ν_2 results in a χ^2 random variable with $\nu_1 + \nu_2$ degrees of freedom. Thus, for $Y = \frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma^2} + \sum_{i=1}^k (Z_i - \bar{Z})^2, Y \sim \chi^2(k+n-1)$.

Part (n): $\frac{\bar{X}}{\sigma^2} + \frac{\sum_{i=1}^k Z_i}{k}$

We know that $\frac{\bar{X}}{\sigma^2} \sim N(\mu/\sigma^2, 1/\sigma^2 n)$ (sum of independent Normal random variables). We also know that $\frac{\sum_{i=1}^k Z_i}{k} \sim N(k\mu/k, k/k^2) \implies \frac{\sum_{i=1}^k Z_i}{k} \sim N(0, 1/k)$ (also by the sum of normal independent random variables). Thus, by adding these two expression together, we get $Y = \frac{\bar{X}}{\sigma^2} + \frac{\sum_{i=1}^k Z_i}{k}, Y \sim N(\mu/\sigma^2, 1/n\sigma^2 + 1/k)$.

Part (o): $k\bar{Z}^2$

$k\bar{Z}^2 = (k)(1/k)\sum_{i=1}^{k} Z_i^2 \sim \chi^2(k)$ (sum of independent χ^2 random variables).

Part (p): $\frac{(k-1)\sum_{i=1}^{n}(X_i-\bar{X})^2}{(n-1)\sigma^2\sum_{i=1}^{k}(Z_i-\bar{Z})^2}$

Note that the numerator is the sample variance for X_i which is distributed as a χ^2 random variable with $n-1$ degrees of freedom. Similarly, the denominator is the sample variance of Z_i and is distributed as a χ^2 random variable with $k-1$ degrees of freedom. Thus, by theorem 8.4.4, $\frac{(k-1)\sum_{i=1}^{n}(X_i-\bar{X})^2}{(n-1)\sigma^2\sum_{i=1}^{k}(Z_i-\bar{Z})^2} \sim F(n-1, k-1)$.

Problem 2: Let X_i for $i = 1, 2, 3$ be independent random variables with $N(i, i^2)$ distributions.

For each of the following situations, use the X_i's to construct a statistic with the indicated distribution. That is, your solution to each part below should be a function of **ALL THREE** X_i's.

For reference, I will explicitly state the distribution of all three X_i's:

- $X_1 \sim N(1, 1)$
- $X_2 \sim N(2, 4)$
- $X_3 \sim N(3, 9)$

Part (a): Chi-square distribution with 3 degrees of freedom.

$(X_1 - 1)^2 + \frac{(X_2-2)^2}{4} + \frac{(X_3-3)^2}{9} \sim \chi^2(3)$

Part (b): Student t distribution with 2 degrees of freedom.

$\frac{\frac{X_3-3}{3}}{\sqrt{\left[(X_1-1)^2+\frac{(X_2-2)^2}{4}\right]/2}} \sim t(2)$

Part (c): F distribution with 1 and 2 degrees of freedom.

$\frac{\frac{(X_3-3)^2}{9}}{\left[(X_1-1)^2+\frac{(X_2-2)^2}{4}\right]/2} \sim F(1, 2)$

Problem 3: BE Exercises: 8.13, 8.18.

For the BE exercises listed above, you may use R or statistical tables to find the probabilities. However, you are encouraged to do it both ways for practice.

8.13: consider independent random variables $Z_i \sim N(0, 1), i = 1, ..., 16$, and let $\bar{Z}$ be the sample mean.

Find the following:

Part (a): $P[\bar{Z} < 1/2]$

For $Y = \bar{Z}, Y \sim N(0, 1/16)$

```r
# Set variables
mu <- 0
standard_deviation <- sqrt(1/16)
q <- 1/2
# Calculate probability
prob <- pnorm(q, mu, standard_deviation)
# Print probability
print(paste("The probability that the sample mean is less than 0.5 is", round(prob, 3)))
```

```
## [1] "The probability that the sample mean is less than 0.5 is 0.977"
```

Part (b): $P[Z_1 - Z_2 < 2]$

For $Y = Z_1 - Z_2, Y \sim N(0, 2)$.

```
# Set variables
mu <- 0
standard_deviation <- sqrt(2)
q <- 2
# Calculate probability
prob <- pnorm(q, mu, standard_deviation)
# Print probability
print(paste("The probability that the sample mean is less than 0.5 is", round(prob, 3)))
```

```
## [1] "The probability that the sample mean is less than 0.5 is 0.921"
```

Part (c): $P[Z_1 + Z_2 < 2]$

For $Y = Z_1 + Z_2, Y \sim N(0, 2)$. Note that we should get to the same probability that we got in part (b). However, this time I'm going to do it with a table (just to shake things up).

I first have to standardize Y by subtracting the mean (0) and dividing by the standard deviation ($\sqrt{2}$). This will help me search for the probability in the standard normal CDF table.

The quantile I'm interested in is $Z = 2/\sqrt{2} \approx 1.41$. Sure enough, at this spot in the table I see the value of 0.9207 which rounds nicely to 0.921, which was the answer in the previous section.

Part (d): $P[\sum_{i=1}^{16} Z_i^2 < 32]$

For $Y = \sum_{i=1}^{16} Z_i^2, Y \sim \chi^2(16)$.

```
# Set variables
degrees_of_freedom <- 16
q <- 32
# Calculate probability
prob <- pchisq(q, degrees_of_freedom)
# Print probability
print(paste("The probability that the sample mean is less than 0.5 is", round(prob, 3)))
```

```
## [1] "The probability that the sample mean is less than 0.5 is 0.99"
```

Part (e): $P[\sum_{i=1}^{16} (Z_i - \bar{Z})^2 < 25]$

For $Y = \sum_{i=1}^{16} (Z_i - \bar{Z})^2, Y \sim \chi^2(15)$.

```
# Set variables
degrees_of_freedom <- 15
q <- 25
# Calculate probability
prob <- pchisq(q, degrees_of_freedom)
# Print probability
print(paste("The probability that the sample mean is less than 0.5 is", round(prob, 3)))
```

```
## [1] "The probability that the sample mean is less than 0.5 is 0.95"
```

8.18: Assume that Z, V_1, and V_2 are independent random variables with $Z \sim N(0, 1), V_1 \sim \chi^2(5)$, and $V_2 \sim \chi^2(9)$.

Find the following:

Part (a): $P[V_1 + V_2 < 8.6]$

For $Y = V_1 + V_2, Y \sim \chi^2(14)$.

```r
# Set variables
degrees_of_freedom <- 14
q <- 8.6
# Calculate probability
prob <- pchisq(q, degrees_of_freedom)
# Print probability
print(paste("The probability that the sample mean is less than 0.5 is", round(prob, 3)))
```

```
## [1] "The probability that the sample mean is less than 0.5 is 0.144"
```

Part (b): $P[Z/\sqrt{V_1/5} < 2.015]$

For $Y = Z/\sqrt{V_1/5}, Y \sim t(5)$.

```r
# Set variables
degrees_of_freedom <- 5
q <- 2.015
# Calculate probability
prob <- pt(q, degrees_of_freedom)
# Print probability
print(paste("The probability that the sample mean is less than", q, "is", round(prob, 3)))
```

```
## [1] "The probability that the sample mean is less than 2.015 is 0.95"
```

Part (c): $P[Z > 0.611\sqrt{V_2}]$

We can rewrite this as $1 - P[Z/\sqrt{V_2/9} < 1.833]$. For $Y = Z/\sqrt{V_2/9}, Y \sim t(9)$.

```r
# Set variables
degrees_of_freedom <- 9
q <- 1.833
# Calculate probability
prob <- 1 - pt(q, degrees_of_freedom)
# Print probability
print(paste("The probability that the sample mean is less than", q, "is", round(prob, 3)))
```

```
## [1] "The probability that the sample mean is less than 1.833 is 0.05"
```

Part (d): $P[V_1/V_2 < 1.450]$

Let's also rewrite this one as $P[(V_1/V_2)((1/5)/(1/9)) < 1.450((1/5)/(1/9))] = P[(V_1/5)/(V_2/9) < 2.61]$

For $Y = (V_1/5)/(V_2/9), Y \sim F(5, 9)$.

```r
# Set variables
degrees_of_freedom_1 <- 5
degrees_of_freedom_2 <- 9
q <- 2.61
# Calculate probability
prob <- pf(q, degrees_of_freedom_1, degrees_of_freedom_2)
# Print probability
print(paste("The probability that the sample mean is less than", q, "is", round(prob, 3)))
```

```
## [1] "The probability that the sample mean is less than 2.61 is 0.9"
```

Part (e): The value b such that $P[\frac{V_1}{V_1+V_2} < b] = 0.9$

Let $X = V_1 + V_2, X \sim \chi^2(14)$.