STA2001 Assignment 10

1. (5.4-22). Let X_1 and X_2 be two independent random variables. Let X_1 and $Y = X_1 + X_2$ be $\chi^2(r_1)$ and $\chi^2(r)$, respectively, where $r_1 < r$.

   (a) Find the mgf of X_2.
   (b) What is its distribution?

   Solution:

   (a) We use the mgfs of Y and X_1 to compute it,

$$\begin{aligned}\frac{1}{(1-2t)^{r/2}} &= E\left(e^{tY}\right) \\ &= E\left(e^{tX_1}e^{tX_2}\right) \\ &= E\left(e^{tX_1}\right)E\left(e^{tX_2}\right) \\ &= \frac{1}{(1-2t)^{r_1/2}}M_{X_2}(t)\end{aligned}$$

   Solve this equation yields

$$M_{X_2}(t) = \frac{1}{(1-2t)^{(r-r_1)/2}}, \quad \forall t < 1/2$$

   (b) From the answer to (a), we know that $X_2 \sim \chi^2(r - r_1)$.

2. (5.4-23). Let X be $N(0,1)$. Use the mgf technique to show that $Y = X^2$ is $\chi^2(1)$. Hint: Evaluate the integral representing $E\left(e^{tX^2}\right)$ by writing $w = x\sqrt{1-2t}$.

   Solution:

$$\begin{aligned}E\left(e^{tY}\right) = E\left(e^{tX^2}\right) &= \int_{-\infty}^{+\infty} e^{tx^2}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{tx^2-x^2/2}\,dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\left(x\sqrt{1-2t}\right)^2/2}\,dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-w^2/2}\frac{1}{\sqrt{1-2t}}\,dw \\ &= \frac{1}{\sqrt{1-2t}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-w^2/2}\,dw \\ &= \frac{1}{\sqrt{1-2t}}, \quad t < 1/2\end{aligned}$$

   where we have used change of variable at the 4th equality and the final result is obtained by realizing that the last integral is integrating a standard normal pdf.

   This shows that Y has a $\chi^2(1)$ distribution.

3. (5.5-4). Let X equal the weight of the soap in a 6-pound box. Assume that the distribution of X is $N(6.05, 0.0004)$.

   (a) Find $P(X < 6.0171)$.
   (b) If nine boxes of soap are selected at random from the production line, find the probability that

at most two boxes weigh less than 6.0171 pounds each. Hint: Let Y equal the number of boxes that weigh less than 6.0171 pounds.
(c) Let $\bar{X}$ be the sample mean of the nine boxes. Find $P(\bar{X} \leq 6.035)$.

Solution:

(a)

$$P(X < 6.0.71) = P\left(\frac{X - 6.05}{0.02} < \frac{6.0171 - 6.05}{0.02}\right) = P(Z < -1.645) = 0.05$$

(b) Let W be the number of boxes that weight less than 6.0171 pounds. Then $W \sim b(9, 0.05)$, and

$$\begin{aligned} P(W \leq 2) &= \sum_{k=0}^{2} \binom{9}{k} (0.05)^k (1 - 0.05)^{9-k} \\ &= 0.9916 \end{aligned}$$

(c) Let $\bar{X}$ be the mean of a random sample of the boxes with size $n = 9$. So $\bar{X} \sim N(6.05, 0.02^2/n)$, the probability is given by

$$P(\bar{X} \leq 6.035) = P\left(Z \leq \frac{6.035 - 6.05}{0.02/3}\right) = 0.0122$$

4. (5.5-14). Let T have a t distribution with r degrees of freedom. Show that $E(T) = 0$ provided that $r \geq 2$, and $\text{Var}(T) = r/(r-2)$ provided that $r \geq 3$, by first finding $E(Z)$, $E(1/\sqrt{U}), E\left(Z^2\right)$, and $E(1/U)$.

Solution:

Recall Theorem 5.5-3, a random variable T following t distribution with degree of freedom r is constructed by

$$T = \frac{Z}{\sqrt{U/r}}$$

where $Z \sim N(0, 1)$, $U \sim \chi^2(r)$, and Z and U are independent.

Therefore, $E(Z) = 0$ and $E\left(Z^2\right) = 1$. Since $U \sim \chi^2(r)$, it follows that

$$\begin{aligned} E\left(1/\sqrt{U}\right) &= \int_0^\infty \frac{1}{\sqrt{u}} \frac{1}{\Gamma(r/2) 2^{r/2}} u^{r/2-1} e^{-u/2} \, du \\ &= \frac{\Gamma[(r-1)/2]}{\sqrt{2}\Gamma(r/2)} \int_0^\infty \frac{1}{\Gamma[(r-1)/2] 2^{(r-1)/2}} u^{(r-1)/2-1} e^{-u/2} \, du \\ &= \frac{\Gamma[(r-1)/2]}{\sqrt{2}\Gamma(r/2)}, \quad r \geq 2 \end{aligned}$$

Note that the last integral is equal to one because the integrand is the pdf of a $\chi^2(r-1)$ random variable.

To find $E(1/U)$, we use a similar approach

$$\begin{aligned} E[1/U] &= \int_0^\infty \frac{1}{u} \frac{1}{\Gamma(r/2) 2^{r/2}} u^{r/2-1} e^{-u/2} \, du \\ &= \frac{\Gamma[(r-2)/2]}{2\Gamma(r/2)} \int_0^\infty \frac{1}{\Gamma[(r-2)/2] 2^{(r-2)/2}} u^{(r-2)/2-1} e^{-u/2} \, du \\ &= \frac{\Gamma(r/2-1)}{2\Gamma(r/2-1)(r/2-1)} \\ &= \frac{1}{r-2}, \quad r \geq 3 \end{aligned}$$

Note that the last integral is equal to one because the integrand is the pdf of a $\chi^2(r-2)$ random variable.

Hence, we could compute the mean and variance as follows

$$\begin{aligned}
E(T) &= E\left[\frac{Z}{\sqrt{U/r}}\right] \\
&= E[Z] \cdot E[1/\sqrt{U/r}] \\
&= 0 \times \left[\frac{\sqrt{r}\Gamma[(r-1)/2]}{\sqrt{2}\Gamma(r/2)}\right] \\
&= 0, \quad r \geq 2
\end{aligned}$$

$$\begin{aligned}
\operatorname{Var}(T) &= E\left(T^2\right) \\
&= E\left[Z^2\right] E[r/U] \\
&= 1 \times \left(r \times \frac{1}{r-2}\right) \\
&= \frac{r}{r-2}, \quad r \geq 3
\end{aligned}$$

5. (5.5-16). Let $n = 9$ in the T statistic defined in Equation 5.5-2. That is,

$$T = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{\overline{X}-\mu}{S/\sqrt{n}}.$$

(a) Find $t_{0.025}$ so that $P\left(-t_{0.025} \leq T \leq t_{0.025}\right) = 0.95$.
(b) Solve the inequality $[-t_{0.025} \leq T \leq t_{0.025}]$ so that μ is in the middle.

Solution:

(a) From Equation 5.5-2 on the textbook, we know that T have a student t distribution with degree of freedom r, which is given by

$$r = n - 1 = 8$$

where n is the sample size.

Let's consider a general case by replacing $t_{0.025}$ with t (as the hint given by the subscript is too obvious),

$$\begin{aligned}
P\left(-t \leq T \leq t\right) &= P(T \leq t) - P(T \leq -t) \\
&= P(T \leq t) - (1 - P(T \leq t)) \\
&= 2P(T \leq t) - 1 \\
&= 0.95
\end{aligned}$$

which means we need to find t such that $P(T \leq t) = 0.975$.

In this equation, since $r = 8$, the value of t is exactly the same as $t_{0.025}(8)$ that can be found in the t distribution table,

$$t = t_{0.025}(8) = 2.306$$

(b) By Equation 5.5-2 on the textbook, $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$, so we rewrite the inequality as

$$-t_{0.025} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{0.025}$$

Rearranging the terms in the inequality,

$$\bar{X} - t_{0.025}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{0.025}\frac{S}{\sqrt{n}}$$

Note that $n = 9$ and $t_{0.025} = t_{0.025}(8) = 2.306$ in this question, so finally we have

$$\bar{X} - 0.7687 \cdot S \le \mu \le \bar{X} + 0.7687 \cdot S$$

6. (5.6-5). Let $X_1, X_2, \ldots, X_{18}$ be a random sample of size 18 from a chi-square distribution with $r = 1$. Recall that $\mu = 1$ and $\sigma^2 = 2$.
(a) How is $Y = \sum_{i=1}^{18} X_i$ distributed?
(b) Using the result of part (a), we see from Table IV in Appendix B that

$$P(Y \le 9.390) = 0.05$$

and

$$P(Y \le 34.80) = 0.99$$

Compare these two probabilities with the approximations found with the use of the central limit theorem

Solution:

(a) Y follows a chi-square distribution with 18 degree of freedom.

(b) Each X_i has $\mu = 1$ and $\sigma^2 = 2$ and they are i.i.d (since they forms a random sample), so for $Y = \sum_{i=1}^{18} X_i$ we have

$$E(Y) = 18 \times 1 = 18, \qquad \mathrm{Var}(Y) = 18 \times 2 = 36$$

We use $N(18, 36)$ to approximate the following

$$\mathrm{P}(Y \le 9.390) \approx P\left(Z \le \frac{9.390 - 18}{\sqrt{36}}\right) = 0.0756$$

$$\mathrm{P}(Y \le 34.8) \approx P\left(Z \le \frac{34.8 - 18}{\sqrt{36}}\right) = 0.9974$$

7. (5.6-8). Let X equal the weight in grams of a miniature candy bar. Assume that $\mu = E(X) = 24.43$ and $\sigma^2 = \mathrm{Var}(X) = 2.20$. Let $\bar{X}$ be the sample mean of a random sample of $n = 30$ candy bars. Find
(a) $E(\bar{X})$.
(b) $\mathrm{Var}(\bar{X})$.
(c) $P(24.17 \le \bar{X} \le 24.82)$, approximately.

Solution:

Note that $X_1, X_2, \ldots, X_{30}$ is a random sample of size 30. So they are i.i.d.

(a)

$$E(\bar{X}) = E\left(\frac{1}{30}\sum_{i=1}^{30} X_i\right) = \frac{1}{30}\sum_{i=1}^{30} E(X_i) = \mu = 24.43$$

(b)

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{1}{30}\sum_{i=1}^{30} X_i\right) = \frac{1}{30^2}\sum_{i=1}^{30} \mathrm{Var}(X_i) = \frac{1}{30^2} \times 30 \times \sigma^2 = \frac{2.20}{30} = 0.0733$$

(c) Let $Z \sim N(0,1)$. Using central limit theorem, we can approximately compute

$$
\begin{aligned}
P(24.17 \le \bar{X} \le 24.82) &\approx P\left(\frac{24.17 - 24.43}{\sqrt{0.0733}} \le Z \le \frac{24.82 - 24.43}{\sqrt{0.0733}}\right) \\
&= P(-0.96 \le Z < 1.44) \\
&= 0.7566
\end{aligned}
$$

8. (5.6-14). Suppose that the sick leave taken by the typical worker per year has $\mu = 10$, $\sigma = 2$, measured in days. A firm has $n = 20$ employees. Assuming independence, how many sick days should the firm budget if the financial officer wants the probability of exceeding the number of days budgeted to be less than 20%?

   Solution:

   Let X_i be the length of sick leave (measure in days) taken by a typical worker per year. So we have $E(X_i) = 10$ and $\mathrm{Var}(X_i) = 4$.

   Let $Y = \sum_{i=1}^{20} X_i$, and note that $X_1, X_2, \dots X_{20}$ are i.i.d, we have

$$
E(Y) = 20 \times 10 = 200
$$

$$
\mathrm{Var}(Y) = 20 \times 4 = 80
$$

   Since the distribution of Y is unknown but it's a sum of i.i.d random variables, we could use the central limit theorem to approximate

$$
\begin{aligned}
P(Y > y) &= P\left(\frac{Y - 200}{\sqrt{80}} > \frac{y - 200}{\sqrt{80}}\right) \\
&= 1 - P\left(\frac{Y - 200}{\sqrt{80}} \le \frac{y - 200}{\sqrt{80}}\right) \\
&\approx 1 - P\left(Z \le \frac{y - 200}{\sqrt{80}}\right)
\end{aligned}
$$

   Setting $1 - P\left(Z \le \frac{y-200}{\sqrt{80}}\right) = 0.2$ and checking the normal distribution table we have

$$
\frac{y - 200}{\sqrt{80}} = 0.842 \;\Rightarrow\; y = 207.5
$$

   As we want the probability less than 0.2 and note that Y is measured in days, we take $y = 208$.

9. Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. Bernoulli random variables with parameter 1/2. Let

$$
Y_n := \frac{\sum_{i=1}^{n^2} X_i - (n^2/2)}{(n/2)}.
$$

   Let $\mathbb{Q}$ be the set of rational numbers.

   (i) Show that Y_n converge in distribution to $Y \sim \mathcal{N}(0,1)$. Hint: consider $Y_{\sqrt{n}}$.

   (ii) Find $\Pr(Y_n \in \mathbb{Q})$ and $\Pr(Y \in \mathbb{Q})$. Does $\Pr(Y_n \in \mathbb{Q})$ converge to $\Pr(Y \in \mathbb{Q})$?

   Solution:

   Observe that $Y_n = Z_{n^2}$, where

$$Z_n := \frac{\sum_{i=1}^{n} X_i - n/2}{\sqrt{n}/2} = \frac{\sum_{i=1}^{n} (X_i - \mathbb{E}[X_i])}{\sqrt{n \operatorname{Var}[X_i]}}$$

By Central Limit Theorem, $Z_n \xrightarrow{D} Y$. Hence, as a subsequence of Z_n, $Y_n \xrightarrow{D} Y$.

(ii) Bernoulli random variables are integer valued, hence $\frac{\sum_{i=1}^{n^2} X_i - (n^2/2)}{(n/2)}$ is always a rational number. We have $\Pr(Y_n \in \mathbb{Q}) = 1$ for all n.

$Y \sim \mathcal{N}(0,1)$ is a continuous random variable. For any rational number q, we have $\Pr(Y = q) = 0$.

The set of rational numbers is countable. Hence by countable additivity of probability measures we have

$$\Pr(Y \in \mathbb{Q}) = \sum_{q \in \mathbb{Q}} \Pr(Y = q) = 0$$

We observe that $\Pr(Y_n \in \mathbb{Q})$ does not converge to $\Pr(Y \in \mathbb{Q})$.

10. Suppose X is a continuous random variable with probability density function:

$$f(x) = \frac{x^2}{3},$$

where $-1 \leq x \leq 2$. Define a random variable $Y = X^2$.

(a) What is the pdf of Y? Hint: The function $Y = X^2$ is not one-to-one mapping for all X.

(b) Use the Central Limit Theorem to find the probability $P(\sum_{i=1}^{175} Y_i \geq 385)$, where $Y_1, \cdots, Y_{175}$ is a random sample of size 175 from the common distribution of Y.

(c) Suppose that two random numbers 0.8 and 0.2 are randomly generated from the uniform distribution $U(0,1)$. Transform the two random numbers 0.8 and 0.2 into two random numbers y_1 and y_2 drawn from the distribution of Y.

Solution:

(a) The sample space of Y, $\bar{S}_Y$ is $[0,4]$. First consider $0 \leq y \leq 1$, $P(Y < y) = P(-\sqrt{y} < X < \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$. Then

$$\begin{aligned} f_Y(y) &= f_X(\sqrt{y})\frac{1}{2\sqrt{y}} - f_X(-\sqrt{y})\frac{-1}{2\sqrt{y}} \\ &= \frac{y}{3} \times \frac{1}{2\sqrt{y}} - \frac{y}{3} \times \frac{-1}{2\sqrt{y}} \\ &= \frac{\sqrt{y}}{3} \end{aligned}$$

When $1 < y \leq 4$, the function is one-to-one mapping, and can easily check that $f_Y(y) = \frac{\sqrt{y}}{6}$.

(b)

$$\begin{aligned} E(Y) &= \int_0^1 \frac{1}{3} y^{3/2} dy + \int_1^4 \frac{1}{6} y^{3/2} dy = 11/5 \\ E(Y^2) &= \int_0^1 \frac{1}{3} y^{5/2} dy + \int_1^4 \frac{1}{6} y^{5/2} dy = 43/7 \\ Var(Y) &= \frac{1922}{175} \end{aligned}$$

$$
\begin{aligned}
P(\sum_{i=1}^{175} Y_i \geq 385) &= P(\frac{\sum_{i=1}^{175} Y_i - 175 \times \frac{11}{5}}{\sqrt{175 \times \frac{1922}{175}}} \geq \frac{385 - 175 \times \frac{11}{5}}{\sqrt{175 \times \frac{1922}{175}}} = 0) \\
&= 0.5
\end{aligned}
$$

(c) From part(a) we know that the cdf of Y is

$$
F_Y(y) = \begin{cases} (\frac{2y}{9})^{3/2}, & 0 \leq y \leq 1 \\ \frac{1}{9} + \frac{1}{9} y^{3/2}, & 1 < y \leq 4 \end{cases} \tag{1}
$$

and the inverse of it would be

$$
F_Y^{-1}(x) = \begin{cases} (\frac{9x}{2})^{2/3}, & 0 \leq x \leq \frac{2}{9} \\ (9x-1)^{2/3}, & \frac{2}{9} < x \leq 1 \end{cases} \tag{2}
$$

Hence, the two random numbers $y_1 = (9 \times 0.8 - 1)^{2/3} = (31/5)^{2/3} = 3.375$ and $y_2 = (9 \times 0.2/2)^{2/3} = (0.9)^{2/3} = 0.932$.