Stat 252 Winter 2007
Assignment #2 Solutions

1. If the population mean is μ, then $\mathbb{E}(Y_i) = \mu$, $i = 1, \ldots, n$. Hence,

$$\mathbb{E}\left(\overline{Y}\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i) = \frac{n \cdot \mu}{n} = \mu.$$

(That is, $\overline{Y}$ is an unbiased estimator of μ.) In order to show that S^2 is an unbiased estimator of σ^2, we begin by expanding $(Y_i - \overline{Y})^2 = Y_i^2 - 2Y_i\overline{Y} + \overline{Y}^2$. This gives

$$\begin{aligned} S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(Y_i^2 - 2Y_i\overline{Y} + \overline{Y}^2) = \frac{1}{n-1}\left(\sum_{i=1}^{n} Y_i^2 - 2\overline{Y}\sum_{i=1}^{n} Y_i + \overline{Y}^2\sum_{i=1}^{n} 1\right) \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n} Y_i^2 - 2n\overline{Y}^2 + n\overline{Y}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right) \end{aligned}$$

where we used the fact that

$$\sum_{i=1}^{n} Y_i = n\overline{Y}.$$

If the population variance is σ^2, then $\text{Var}(Y_i) = \sigma^2$, $i = 1, \ldots, n$, so that $\mathbb{E}(Y_i^2) = \text{Var}(Y_i) + (\mathbb{E}(Y_i))^2 = \sigma^2 + \mu^2$. Hence,

$$\begin{aligned} \mathbb{E}(S^2) = \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathbb{E}(Y_i^2) - n\mathbb{E}(\overline{Y}^2)\right) &= \frac{1}{n-1}\left(\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n\mathbb{E}(\overline{Y}^2)\right) \\ &= \frac{1}{n-1}\left(n(\sigma^2 + \mu^2) - n\mathbb{E}(\overline{Y}^2)\right) \\ &= \frac{n}{n-1}\left(\sigma^2 + \mu^2 - \mathbb{E}(\overline{Y}^2)\right). \end{aligned}$$

However, we still must compute $\mathbb{E}(\overline{Y}^2)$. As above, $\mathbb{E}(\overline{Y}^2) = \text{Var}(\overline{Y}) + (\mathbb{E}(\overline{Y}))^2 = \text{Var}(\overline{Y}) + \mu^2$ which leaves us with $\text{Var}(\overline{Y})$ to compute. It is common to assume that the data were collected independently of each other; that is, if $i \neq j$, then $\text{Cov}(Y_i, Y_j) = 0$. Therefore, from Theorem 5.12 (that's Stat 251 material)

$$\begin{aligned} \text{Var}(\overline{Y}) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n}\text{Var}(Y_i) + 2\sum\sum_{1 \le i < j \le n}\text{Cov}(Y_i, Y_j)\right) &= \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(Y_i) \\ &= \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}. \end{aligned}$$

Finally, we conclude that $\mathbb{E}(\overline{Y}^2) = \sigma^2/n + \mu^2$ so

$$\mathbb{E}(S^2) = \frac{n}{n-1}\left(\sigma^2 + \mu^2 - \mathbb{E}(\overline{Y}^2)\right) = \frac{n}{n-1}\left(\sigma^2 + \mu^2 - \left(\frac{\sigma^2}{n} + \mu^2\right)\right) = \frac{n}{n-1} \cdot \frac{(n-1)\sigma^2}{n} = \sigma^2$$

meaning that S^2 is an unbiased estimator of σ^2 as required.

2. Let Y denote the life time of a heat lamp in this greenhouse, so that $Y \sim \mathcal{N}(50, 4)$ (in hours). If $Y_1, \ldots, Y_{25}$ represent the 25 heat lamps for this medicinal herb growing operation, then we are interested in

$$P(Y_1 + \cdots + Y_{25} > 1300).$$

Since the Y_i are assumed to be i.i.d. we can conclude that

$$P\left(\frac{Y_1 + \cdots + Y_{25}}{25} > \frac{1300}{25}\right) = P(\overline{Y} > 52)$$

where $\overline{Y} \sim \mathcal{N}(50, 4/25)$. In order to calculate $P(\overline{Y} > 52)$ we normalize and use Table 4 so that

$$P(\overline{Y} > 52) = P\left(\frac{\overline{Y} - 50}{2/5} > \frac{52 - 50}{2/5}\right) = P(Z > 5) \approx 0$$

where $Z \sim\ No(0, 1)$. Hence, we see that

$$P(Y_1 + \cdots + Y_{25} > 1300) \approx 0;$$

that is, it is extremely unlikely that there will be a bulb burning after 1300 hours.

3. If $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, then the moment generating function of Y_1 is

$$m_{Y_1}(t) = \exp\left\{\mu_1 t + \frac{\sigma_1^2 t}{2}\right\}$$

and, similarly, if $Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, then the moment generating function of Y_2 is

$$m_{Y_2}(t) = \exp\left\{\mu_2 t + \frac{\sigma_2^2 t}{2}\right\}.$$

Therefore, the moment generating function of $Y_1 + Y_2$ is given by

$$m_{Y_1+Y_2}(t) := \mathbb{E}(e^{(Y_1+Y_2)t}) = \mathbb{E}(e^{Y_1 t}) \cdot \mathbb{E}(e^{Y_2 t})$$

since Y_1 and Y_2 are independent. Hence, we find

$$\begin{aligned} m_{Y_1+Y_2}(t) = \mathbb{E}(e^{Y_1 t}) \cdot \mathbb{E}(e^{Y_2 t}) = m_{Y_1}(t) \cdot m_{Y_2}(t) &= \exp\left\{\mu_1 t + \frac{\sigma_1^2 t}{2}\right\} \cdot \exp\left\{\mu_2 t + \frac{\sigma_2^2 t}{2}\right\} \\ &= \exp\left\{(\mu_1 + \mu_2)t + \left(\frac{\sigma_1^2}{2} + \frac{\sigma_2^2}{2}\right) t\right\} \end{aligned}$$

which we recognize as the moment generating function of a normal random variable with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. That is, we have shown $Y_1 + Y_2 \sim\sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ as required.

4. (a) Suppose that $Y_1, \ldots, Y_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$ random variables. As proved on January 12, 2007,

$$\overline{Y} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{\sqrt{n}}\right)$$

so that normalizing we conclude

$$\frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

From the Theorem given on January 15, 2007, we know that the random variable

$$\frac{(n-1)S^2}{\sigma^2}$$

has a $\chi^2(n-1)$ distribution. It also follows from this theorem that $\overline{Y}$ and S^2 are independent. It now follows from the definition of the t distribution that if $Z \sim \mathcal{N}(0,1)$ and $W \sim \chi^2(\nu)$ are independent, then

$$T = \frac{Z}{\sqrt{W/\nu}} \sim t(\nu).$$

Thus, let

$$Z = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \quad \text{and} \quad W = \frac{(n-1)S^2}{\sigma^2}$$

so that

$$\frac{Z}{\sqrt{W/\nu}} = \frac{\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{\overline{Y}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

as required.

4. (b) Similarly,

$$\frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2(n-1) \quad \text{and} \quad \frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2(m-1)$$

and, by assumption, S_1 and S_2 are independent. Hence, from the definition of the F distribution, we conclude that

$$\frac{\frac{(n-1)S_1^2}{\sigma_1^2}/(n-1)}{\frac{(m-1)S_2^2}{\sigma_2^2}/(m-1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n-1, m-1).$$

5. Textbook

(7.38) Suppose that $W_i = X_i - Y_i$. Since $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ are all independent and identically distributed, so too are $W_1, W_2, \ldots, W_n$. Thus we find $E(W_i) = E(X_i - Y_i) = E(X_i) - E(Y_i) = \mu_1 - \mu_2$ and

$$\operatorname{Var}(W_i) = \operatorname{Var}(X_i - Y_i) = \operatorname{Var}(X_i) + \operatorname{Var}(Y_i) - 2\operatorname{Cov}(X_i, Y_i) = \sigma_1^2 + \sigma_2^2$$

using Theorem 5.12 and the fact that $\operatorname{Cov}(X_i, Y_i) = 0$ since X_i and Y_i are independent. If

$$\overline{W} = \frac{1}{n}\sum_{i=1}^{n} W_i,$$

then since the W_i are iid, we conclude

$$E(\overline{W}) = \mu_1 - \mu_2 \quad \text{and} \quad \operatorname{Var}(\overline{W}) = \frac{\sigma_1^2 + \sigma_2^2}{n}.$$

Hence, we can now apply Theorem 7.4 to the normalized random variables

$$U_n = \frac{\overline{W} - E(\overline{W})}{\sqrt{\operatorname{Var}(\overline{W})}} = \frac{\overline{W} - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}} = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}}$$

and conclude that the distribution of U_n converges to $\mathcal{N}(0,1)$.

(7.40) Using the same notation as in **(7.38)**, we find that if the sample sizes differ, then

$$\mathrm{Var}(\overline{W}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Therefore, if

$$U_n = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

then U_n again converges in distribution to $\mathcal{N}(0,1)$. In order to compute the required probability, we simply normalize to obtain a random variable which is (approximately) a standard normal so that we can use Table 4. That is,

$$\begin{aligned}
P(\,|\,(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)| \le 0.05\,) &= P\left(\left|\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right| \le \frac{0.05}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) \\
&= P\left(\left|\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{0.01}{50} + \frac{0.02}{100}}}\right| \le \frac{0.05}{\sqrt{\frac{0.01}{50} + \frac{0.02}{100}}}\right) \\
&\approx P(|Z| \le 2.5) \\
&\approx 1 - 2(0.0062) = 0.9876
\end{aligned}$$

where $Z \sim \mathcal{N}(0,1)$.

(7.41) If $n_1 = n_2 = n$, then we are trying to find the value of n such that

$$P(\,|\,(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)| \le 0.04\,) = 0.90.$$

Now, if we normalize (and write $Z \sim \mathcal{N}(0,1)$), then we obtain

$$P\left(|Z| \le \frac{0.04}{\sqrt{\frac{0.01}{n} + \frac{0.02}{n}}}\right) = 0.90.$$

But from Table 4 we find that $P(|Z| \le 1.645) \approx 0.90$, which implies that

$$\frac{0.04}{\sqrt{\frac{0.01}{n} + \frac{0.02}{n}}} \approx 1.645.$$

Solving for n gives $n \approx 50.74$. Thus, we need each sample to contain at least 51 data points.

6. Textbook

(6.64) If $Y_1, Y_2, \ldots, Y_n$ are all independent and identically distributed beta$(2,2)$ random variables, then each has density function

$$f_Y(y) = \frac{\Gamma(4)}{\Gamma(2)\Gamma(2)} y(1-y) = 6y(1-y), \quad 0 < y < 1.$$

(a) If $Y_{(n)} = \max\{Y_1, \ldots, Y_n\}$, then

$$P(Y_{(n)} \le t) = P(Y_1 \le t, \ldots, Y_n \le t) = P(Y_1 \le t) \cdots P(Y_n \le t) = [P(Y_1 \le t)]^n$$

since the Y_i are independent (the second equality) and identically distributed (the third equality). Now, for any $0 < t < 1$,

$$P(Y_1 \le t) = \int_0^t f_Y(y)\,dy = \int_0^t 6y(1-y)\,dt = \int_0^t 6y\,dy - \int_0^t 6y^2\,dy = 3t^2 - 2t^3 = t^2(3-2t)$$

so that the distribution function of $Y_{(n)}$ is

$$F(t) = [P(Y_{(n)} \le t)]^n = t^{2n}(3-2t)^n, \;\; 0 < t < 1.$$

(Of course, $F(t) = 0$ for $t \le 0$, and $F(t) = 1$ for $t \ge 1$.

(b) The density function of $Y_{(n)}$ is therefore

$$f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}t^{2n}(3-2t)^n = 2nt^{2n-1}(3-2t)^n - 2nt^{2n}(3-2t)^{n-1} = 6nt^{2n-1}(3-2t)^{n-1}(1-t)$$

for $0 < t < 1$ and 0 otherwise.

(c) For $n = 2$, the expected value $E(Y_{(2)})$ is

$$E(Y_{(2)}) = \int_0^1 t\,f(t)\,dt = \int_0^1 12t^4(3-2t)(1-t)\,dt = \int_0^1 (36t^4 - 60t^5 + 24t^6)\,dt = \frac{36}{5} - \frac{60}{6} + \frac{24}{7} = \frac{22}{35}.$$

(6.65) If $Y_1, Y_2, \ldots, Y_n$ are all independent and identically distributed exponential(β) random variables, then each has density function

$$f_Y(y) = \frac{1}{\beta}e^{-y/\beta}, \;\; 0 < y < \infty.$$

(a) If $Y_{(1)} = \min\{Y_1, \ldots, Y_n\}$, then

$$P(Y_{(1)} > t) = P(Y_1 > t, \ldots, Y_n > t) = P(Y_1 > t) \cdots P(Y_n > t) = [P(Y_1 > t)]^n$$

since the Y_i are independent (the second equality) and identically distributed (the third equality). Now, for any $0 < t < \infty$,

$$P(Y_1 > t) = \int_t^\infty f_Y(y)\,dy = \frac{1}{\beta}\int_t^\infty e^{-y/\beta}\,dy = e^{-t/\beta}$$

so that

$$P(Y_{(1)} > t) = e^{-tn/\beta}.$$

Thus, the distribution function of $Y_{(n)}$ is

$$F(t) = P(Y_{(1)} \le t) = 1 - P(Y_{(1)} > t) = 1 - e^{-tn/\beta} = 1 - e^{-t/(\beta/n)}$$

which is the distribution function of an exponential random variable with mean β/n.

(b) If $n = 5$, $\beta = 2$, then the distribution function of $Y_{(1)}$ is

$$F(t) = 1 - e^{-5t/2}, \quad 0 < t < \infty$$

so that the corresponding density function is

$$f(t) = \frac{5}{2}e^{-5t/2}, \quad 0 < t < \infty.$$

Hence,

$$P(Y_{(1)} \leq 3.6) = \frac{5}{2}\int_0^{3.6} e^{-5t/2}\,dt = -e^{-5t/2}\Big|_0^{3.6} = 1 - e^{-9}.$$