Solutions to June 2006 Examination

1. (a) Since $B \subset A$, then
\[ \mathbb{P}(B) = \mathbb{P}(A \cap B) = 0.6 \]

(b) Testing whether candidate can distinguish between mutual exclusivity and independence.
\[ \mathbb{P}(B)\mathbb{P}(C) = 0.6 \times 0.3 = 0.18 \neq 0 = \mathbb{P}(B \cap C), \]

and, hence, $B$ and $C$ are not independent.

(c) \[
\mathbb{P}((A \setminus B) \cup C) = \mathbb{P}(A \setminus B) + \mathbb{P}(C) - \mathbb{P}((A \setminus B) \cap C) \\
= \mathbb{P}(A \setminus B) + \mathbb{P}(C) - \mathbb{P}(A \cap C) = (0.8 - 0.6) + 0.3 - 0.15 = 0.35
\]

2. See sheet of discrete distributions
(a) Let $X$ be number of horse races in a total of 15 that do not result in at least one serious accident. Then
\[ X \sim Bin(15, 0.93) \]
Hence $E[X] = 15 \times 0.93 = 13.95$. It is expected that about 14 out of every 15 races will pass without incident.

(b) Let $W$ be the number of horse races in a total of 40 that result in at least one serious accident. Then
\[ W \sim Bin(40, 0.07) \]
Setting $Y \sim Po(40 \times 0.07) \sim Po(2.8)$, then
\[ \mathbb{P}(W \geq 2) = 1 - \mathbb{P}(W \leq 1) \approx 1 - \mathbb{P}(Y \leq 1) = 1 - 0.2311 = 0.7689. \]

3. (a) $R_V = \{y : 0 < y < k\}$
(b)
\[ E[Y] = \int_0^k y \times \frac{2y}{k^2} dy = \frac{2}{k^2} \int_0^k y^2 dy \\
= \frac{2}{k^2} \left[ \frac{y^3}{3} \right]_0^k = \frac{2}{k^2} \times \frac{k^3}{3} = \frac{2}{3} k. \]
\[ E[Y^2] = \int_0^k y^2 \times \frac{2y}{k^2} dy = \frac{2}{k^2} \int_0^k y^2 dy = \frac{2}{k^2} \left[ \frac{y^3}{3} \right]_0^k = \frac{2}{k^2} \times \frac{k^4}{4} = \frac{k^2}{2}. \]

\text{Lemma 2.4.2.}

\[ \text{var}(Y) = E[Y^2] - E[Y]^2 = \frac{k^2}{2} - \left( \frac{2}{3} \right) = \frac{k^2}{2} - \frac{4}{9}k^2 = k^2 \left( \frac{9 - 8}{18} \right) = \frac{k^2}{18} = 2 \]
\[ \Rightarrow k^2 = 36 \Rightarrow k = 6, \]

Prop 2.9.6 (iii)

noting that we invoke the assumption that \( k \) is +ve.

4.
(a) \( R_Y = \{ y : 0 < y < 1 \} \).
(b) Assuming that \( y \in R_Y \) (and thus positive), then
\[ F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}((1 - X)^2 \leq y) = \mathbb{P}(1 - X \leq \sqrt{y}) \]
\[ = \mathbb{P}(-X \leq \sqrt{y} - 1) = \mathbb{P}(X \geq 1 - \sqrt{y}) = 1 - F_X(1 - \sqrt{y}). \]

sec. 3.1.

(c) Differentiating \( F_Y(y) \) w.r.t. \( y \) in the above, yields
\[ f_Y(y) = -f_X(1 - \sqrt{y}) \times -\frac{1}{2}y^{-1/2} = \frac{1}{2\sqrt{y}} \times f_X(1 - \sqrt{y}) = \frac{1}{2\sqrt{y}} \times 3y = \frac{3}{2}\sqrt{y}. \]

Sec. 3.1.

5.
(a) Bookwork:
\[ G_X(z) = \sum_{k=0}^{\infty} z^k p_X(k). \]

Defn. 3.2.1.

(b)
\[ G_X(z) = \frac{0.2(z^6 - z)}{z - 1} = \frac{1(z - z^6)}{5(1 - z)} \]
\[ = \frac{1}{5}(z - z^6)(1 + z + z^2 + \ldots) = \frac{1}{5}\{z + z^2 + z^3 + \ldots - z^6 - z^7 - \ldots\} \]
\[ G_X(z) = \frac{1}{5}\{z + z^2 + z^3 + z^4 + z^5\} \]

Rem. 3.2.2 (iii)

Hence, by reading off the coefficient of \( z^3 \), we see that \( p_X(3) = \frac{1}{5} \).
(c) \( R_X = \{1, 2, 3, 4, 5\} \).

6.
(a)
\[ f_X(x) = \int_{y=0}^{1} \frac{2}{5}(2x + 3y) dy = \frac{2}{5} \left[ 2xy + \frac{3y^2}{2} \right]_{y=0}^{y=1} = \frac{2}{5} \left( 2x + \frac{3}{2} \right) = \frac{3}{5} + \frac{4}{5} \]

Ex. 4.3

Q1 (a)
for $0 < x \leq 1$.

(b) 

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{\frac{2}{3}(2x + 3y)}{\frac{1}{3}(3 + 4x)} = \frac{2(2x + 3y)}{3 + 4x}$$

for $0 < y \leq 1$.

(c) 

$$f_{Y|X}(y|1) = \frac{2(2 + 3y)}{7} = \frac{4}{7} + \frac{6}{7}y.$$ 

Hence

$$P \left( Y \leq \frac{3}{4} \mid X = 1 \right) = \frac{1}{7} \int_0^{3/4} (4 + 6y)dy = \frac{1}{7}[4y + 3y^{3/4}]_0^{3/4} = \frac{1}{7} \left( 3 + 3 \times \frac{9}{16} \right)$$

$$= \frac{1}{7} \left( 3 + \frac{27}{16} \right) = \frac{1}{7} \left( \frac{48 + 27}{16} \right) = \frac{75}{7 \times 16} = \frac{75}{112} = 0.670$$

(to 3 s.f.)

7. $\mu = 3320$, and $\sigma = 660$.

$$\mathbb{P}(3223 \leq \bar{Y} \leq 3407) = \mathbb{P} \left( \frac{3223 - 3320}{660/15} \leq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq \frac{3407 - 3320}{660/15} \right)$$

$$\approx \mathbb{P} \left( \frac{-97}{44} \leq N(0, 1) \leq \frac{87}{44} \right) = \mathbb{P}(-2.20 \leq N(0, 1) \leq 1.977) = \Phi(1.98) - \Phi(-2.20)$$

$$= \Phi(1.98) + \Phi(2.20) - 1 = 0.9761 + 0.98610 - 1 = 0.9622$$

(to 3 s.f.)

8. $H_0 : \ p_A = p_B = p_C = p_D = 0.25$. Therefore, $e_j = 224 \times 0.25 = 56$ for each $j$.

$$\sum_{j=1}^{4} \frac{(o_j - e_j)^2}{e_j} = \sum_{j=1}^{4} \frac{o_j^2}{e_j} - n = \frac{1}{56} \left( 42^2 + 64^2 + 53^2 + 65^2 \right) - 224$$

$$= 230.25 - 224 = 6.25$$

Under $H_0$, the above statistic comes from the $\chi^2$ distribution. Since $\chi^2_{obs} = 6.25 < 7.815 = \chi^2_{3,0.05}$, then there is no evidence to reject $H_0$ at the 5% level. Software is working well!
9.
(a) 
\[ m = 9, \quad \sum_{i=1}^{m} x_i = 189.3, \quad \sum_{i=1}^{m} x_i^2 = 3984.55, \quad \bar{x} = 189.3/9 = 21.03333 \]
\[ s_X^2 = \frac{1}{m-1} \left( \sum_{i=1}^{m} x_i^2 - m\bar{x}^2 \right) = \frac{1}{8} \left( 3984.55 - 9 \times (21.03333)^2 \right) = 0.367658 \]
\[ n = 13, \quad \sum_{i=1}^{n} y_i = 271.6, \quad \sum_{i=1}^{n} y_i^2 = 5686.52, \quad \bar{y} = 271.6/13 = 20.8923 \]
\[ s_Y^2 = \frac{1}{n-1} \left( \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 \right) = \frac{1}{12} \left( 5686.52 - 13 \times (20.8923)^2 \right) = 1.01445 \]

Hence
\[ \frac{s_Y^2}{s_X^2} = \frac{1.01445}{0.367658} = 2.759 \]
(to 4 s.f.)

Testing
\[ H_0 : \sigma_X^2 = \sigma_Y^2 \text{ vs. } H_1 : \sigma_X^2 \neq \sigma_Y^2 \]
at the 5% level. Thus \( \alpha = 0.05 \), and so \( \alpha/2 = 0.025 \).
\[ F_{12,8,0.025} = 4.200 \]

Since \( s_Y^2/s_X^2 \) is greater than 1 (thus no need to consider the lower critical value), and does not exceed the upper 2.5% point of the \( F_{12,8} \) distribution, then we conclude that there is no evidence to reject \( H_0 \) at the 5% level.

(b) Testing
\[ H_0 : \mu_X = \mu_Y \text{ vs. } H_1 : \mu_X \neq \mu_Y \]
The pooled sample variance is:
\[ s^2 = \frac{(m - 1)s_X^2 + (n - 1)s_Y^2}{m + n - 2} = \frac{8 \times 0.367658 + 12 \times 1.01445}{20} = \frac{15.114664}{20} = 0.755733 \]
and thus
\[ \hat{s} = \sqrt{s^2} = 0.869329 \]

Hence
\[ T_{obs} = \frac{\bar{x} - \bar{y}}{\hat{s} \sqrt{\frac{1}{m} + \frac{1}{n}}} = \frac{21.03333 - 20.8923}{0.869329 \sqrt{\frac{1}{9} + \frac{1}{13}}} = \frac{0.14103}{0.869329 \times \sqrt{0.188034}} \]
\[ = \frac{0.14103}{0.869329 \times 0.433629} = 0.374 \]
to 3 s.f.

At the 5% level, \( \alpha = 0.05 \), and so \( t_{m+n-2, \alpha/2} = t_{20.0.025} = 2.086 \)

Since \[ |T_{\text{obs}}| < t_{20.0.025} \]

then we have \textbf{no} evidence to reject \( H_0 \) at the 5% level.

10.

(a)

i) \( X \) has a symmetric distribution about 0, and so obviously \( E[X] = 0 \). More formally,

\[
E[X] = \int_{-1}^{1} x \times \frac{3}{2} x^2 dx = \frac{3}{2} \times \left[ \frac{x^4}{4} \right]_{-1}^{1} = \frac{3}{2} \times \frac{1}{4} (1 - 1) = 0.
\]

Also

\[
E[X^2] = \int_{-1}^{1} x^2 \times \frac{3}{2} x^2 dx = \frac{3}{10} \left[ x^5 \right]_{-1}^{1} = \frac{3}{10} (1 - (-1)) = \frac{6}{10} = \frac{3}{5}.
\]

Hence

\[
\text{var}(X) = E[X^2] - E[X]^2 = \frac{3}{5}.
\]

ii) \( n = 30, \quad n \times \mu = 0, \quad n \times \sigma^2 = 30 \times \frac{3}{5} = 18. \)

\[
P(-1.2 \leq Y \leq 1.5) \approx P \left( \frac{-1.2 - 0}{\sqrt{18}} \leq N(0, 1) \leq \frac{1.5 - 0}{\sqrt{18}} \right)
\]

\[
= P(-0.28 \leq N(0, 1) \leq 0.35) = \Phi(0.35) - \Phi(-0.28)
\]

\[
= \Phi(0.35) + \Phi(0.28) - 1 = 0.6368 + 0.6103 - 1 = 0.2471.
\]

(b)

\[
\frac{\sigma^2}{n} = \frac{(5.8)^2}{40} = 0.841
\]

\[
P(52.5 \leq \bar{X} \leq 54.6) = P \left( \frac{52.5 - 54.03}{\sqrt{0.841}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{54.6 - 54.03}{\sqrt{0.841}} \right)
\]

\[
\approx P(-1.67 \leq N(0, 1) \leq 0.62) = \Phi(0.62) - \Phi(-1.67) = \Phi(0.62) + \Phi(1.67) - 1
\]

\[
= 0.7324 + 0.9525 - 1 = 0.6849.
\]
11.
(a) The end points of the C.I. are given by

\[
\bar{x} \pm \frac{sX}{\sqrt{m}} t_{m-1,a/2} = \bar{x} \pm \frac{sX}{\sqrt{m}} t_{9.0.025} = 74.4 \pm \frac{14.33876}{\sqrt{10}} \times 2.262
\]

\[
= 74.4 \pm 10.2566 = (64.1, 84.7).
\]

(b) The end points of the C.I. are given by

\[
\bar{y} \pm \frac{sY}{\sqrt{n}} t_{n-1,a/2} = \bar{y} \pm \frac{sY}{\sqrt{n}} t_{6.0.05} = 77.14286 \pm \frac{9.118271}{\sqrt{7}} \times 1.943
\]

\[
= 77.14286 \pm 6.69632 = (70.4, 83.8).
\]

(c) Testing \(H_0: \mu_X = \mu_Y\) vs. \(H_1: \mu_X \neq \mu_Y\).

\[
s_X^2 = (14.33876)^2 = 205.600,
\]

\[
s_Y^2 = (9.118271)^2 = 83.1429
\]

Hence the pooled sample variance is given by

\[
s^2 = \frac{(m-1)s_X^2 + (n-1)s_Y^2}{m+n-2} = \frac{9 \times 205.600 + 6 \times 83.1429}{15} = 156.61716
\]

We derive a 95% C.I. for \(\mu_X - \mu_Y\), the difference between the population means:

\[
\bar{x} - \bar{y} \pm s \sqrt{\frac{1}{m} + \frac{1}{n}} \times t_{m+n-2,a/2}
\]

\[
= 74.4 - 77.14286 \pm 12.5147 \sqrt{\frac{1}{10} + \frac{1}{7}} \times 2.131
\]

\[
= -2.74286 \pm 12.5147 \times 0.492805 \times 2.131 = -2.74286 \pm 13.1425 = (-15.9, 10.4)
\]

(to 3 s.f.)

Since 0 lies inside the interval, then there is no evidence to reject \(H_0\). This follows from the duality between 100(1 - \(\alpha\)% equal tails C.I.'s and 100\(\alpha\)% two-sided hypothesis tests. So there is no difference between the two methods.

see Remarks 9.3.1(iii) for a slightly different scenario.