BUEC 333, Answers to problem set exercises
Part 1: Probability and Statistics

1

a. $P(X=1)=½$, $P(X=2)=P(X=3)=\ \ldots\ =P(X=6)\ =\frac{1}{10}$
b. $P(A)=P(X=4)+P(X=5)+P(X=6)=\frac{3}{10}$
c. $P(B)=P(X=1)+P(X=3)+P(X=5)=\frac{1}{2}+\frac{1}{10}+\frac{1}{10}=\frac{7}{10}$
d. $\mu_x=\sum_i X_i * P(X_i)=\ 2.5$

$$Unfair\ die:\ \ Var(X)=\sum_i (X_i-\mu_x)^2 * P(X_i)$$
$$=\tfrac{1}{2}*(1-2.5)^2+\tfrac{1}{10}*\{(2-2.5)^2+(3-2.5)^2+(4-2.5)^2+(5-2.5)^2+(6-2.5)^2\}=\tfrac{1}{2}*2.25+\tfrac{1}{10}*21.25=2.25$$
$$\mu_{x'}=\sum_i X'_i * P(X'_i)=\ 3.5$$
$$Fair\ die:\ Var(X')=\sum_i (X'_i-\mu_{x'})^2 * P(X'_i)$$
$$=\tfrac{1}{6}*\{(1-3.5)^2+(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2+(6-3.5)^2\}=2.91$$

2

	X1	X2
HHH	3	3
HHT	2	1
HTH	2	1
THH	2	1
HTT	1	-1
TTH	1	-1
THT	1	-1
TTT	0	-3

a.

X1	3	2	1	0
Prob	⅛	⅜	⅜	⅛

b.

X2	3	1	-1	-3
Prob	⅛	⅜	⅜	⅛

3

a. Given $P(X=0)=p$, we have $P(X=1)=1-p$
b. $E(X)=p(0)+(1-p)(1)=1-p$
c. $var(X)=E(X^2)-(E(X))^2=(p(0^2)-(1-p)(1^2))-(1-p)^2=p(1-p)$

4. The variance question.

- Table 3. E(X)=½*(-1)+½*2=0.5. So the variance is (-1-0.5)^2*0.5+(2-0.5)^2*0.5=2.25. Alternatively, call the RV in Table 1 “X”, and the one in Table 3 “Y”. Note that Y=1.5*X+0.5. Therefore, Var(Y)=1.5^2 Var(X)

- Table 4. E(X)=8/10. Then Var(X)=1/10*(-1-8/10)^2 + 9/10*(1-8/10)^2 = 1/10*(-18/10)*(-18/10)+9/10*2/10*2/10 = 360/1000=0.36

7
Let X1 be the first shot, X2 be the second shot.
X1 = 1 if has a bullet, 0 otherwise
X2 = 1 if has a bullet, 0 otherwise

Spin again:
Pr(X2=0) = ½
Pull the trigger without spinning:
Pr(X2=0|X1=0) = ⅔

Better not to spin.

8

a. Recall $P(X,Y) = P(X|Y)P(Y)$ and $P(X|Y) = P(X)$ if X and Y are independent. Thus $P(X = 10, Y = 20) = 0.3 \times 0.6 = 0.18$
b. $P(X = 10|Y = 20) = P(X = 10, Y = 20)/P(Y = 20) = 0.18/0.6 = 0.3$, which is just the same as $P(X = 10)$ unconditionally.
c. No.

9

a.

$$E(V) = E(20 - 7Y) = 20 - 7E(Y) = 20 - 7 * 0.78 = 14.54$$
$$E(W) = E(3 + 6X) = 3 + 6E(X) = 3 + 6 * 0.7 = 7.2$$

b.

$$\sigma_V^2 = Var(20 - 7Y) = (-7)^2\sigma_Y^2 = 8.4084 \ \sigma_W^2 = Var(3 + 6X) = 6^2\sigma_X^2 = 7.56$$

c.

$$\sigma_{WV} = cov(3 + 6X,\ 20 - 7Y) = 6(-7)con(X,Y) = -42 * 0.084 = -3.52$$
$$corr(W,V) = \frac{\sigma_{WV}}{\sigma_V * \sigma_W} = -\frac{3.528}{\sqrt{7.56 * 8.4084}} = -0.4425$$

10 (a) P(X=6|Y=1) = P(X=6,Y=1) / P(Y=1) = ¼ / ½ =½.
(b) P(Y=0|X=4) = P(X=4,Y=0) / P(X=4) = 1/12 / (1/12+1/20) = ⅝
(c) P(X=3) = P(X=3,Y=0) + P(X=3,Y=1) = 1/12 + 1/20 = 2/15.
(d) E(X|Y=0) = 3.5, because the conditional prob P(X=x|Y=0)=⅙
P(X=6|Y=1) = ½, P(X=x|Y=1)=1/10 for x=1,...,5, so

E(X|Y=0) = ½*6 + 1/10*(1+...+5) = 3 + 15/10 = 4.5

(e) 3.5*½ + 4.5*½ = 4

(f) 1/10*((1-4.5)^2 + (2-4.5)^2 + (3-4.5)^2 + (4-4.5)^2 + (5-4.5)^2)) + 0.5*(6-4.5)^2 = 3.25.

11

a. $E(M|A=0) = \sum_i M_i Pr(M = M_i|A=0)$

$= 0 * 0.35/0.5 + 1 * 0.065/0.5 + 2 * 0.05/0.5 + 3 * 0.025/0.5 + 4 * 0.01/0.5 = 0.56$

$E(M|A=1) = \sum_i M_i Pr(M = M_i|A=1)$

$= 0 * 0.45/0.5 + 1 * 0.035/0.5 + 2 * 0.01/0.5 + 3 * 0.005/0.5 + 4 * 0.00/0.5 = 0.14$

b. Old computers are more likely to crash

c.

$$Var(M|A=0) = \sum_i (M_i - E(M|A=0))^2 - Pr(M = M_i|A=0)$$

$$= (0-0.56)^2 * 0.7 + \ ... \ + (4-0.56)^2 * 0.02 = 0.99$$

12

Cov(X,Y) = E(XY)-E(X)E(Y) = 171.7 - 5.33 x 30.15 = 11

Corr(X,Y) = Cov(X, Y)/[SD(X)SD(Y)] = 11/(2.6 x 14.77) = 0.286

13

a. Recall that if $Y \sim N(\mu, \sigma^2)$, then $\frac{Y-\mu}{\sigma} \sim N(0,1)$. Thus, given $Y \sim N(1,4)$, $\Pr(Y \le 3) = \Pr(\frac{Y-1}{2} \le \frac{3-1}{2}) = \Phi(1) = 0.8413$.

b. Given $Y \sim (3,9)$, $\Pr(Y > 0) = 1 - \Pr(Y \le 0) = 1 - \Pr(\frac{Y-3}{3} \le \frac{0-3}{3}) = 1 - \Phi(-1) = \Phi(1) = 0.8413$

c. Given $Y \sim (50,25)$, $\Pr(40 \le Y \le 52) = \Pr(\frac{40-50}{5} \le \frac{Y-50}{5} \le \frac{52-50}{5}) = \Phi(0.4) - \Phi(-2) = \Phi(0.4) - [1 - \Phi(2)] = 0.6326$

d. Given $Y \sim (5,2)$, $\Pr(6 \le Y \le 8) = \Pr(\frac{6-5}{\sqrt{2}} \le \frac{Y-5}{\sqrt{2}} \le \frac{8-5}{\sqrt{2}}) = \Phi(2.1213) - \Phi(0.7071) = 0.2229$.

14 P(X^2 > 4) = P(X< -2) + P(X>2)

= P((X-3) / 2 < -5/2) + P((X-3) / 2 > -½)

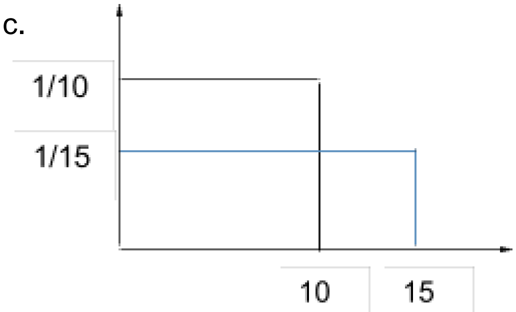
= P(Z< -5/2) + P(Z> -½) where Z is a standard normal.

16

a. Because X_1 has a uniform probability distribution. X_1 is time and as a result is continuous.

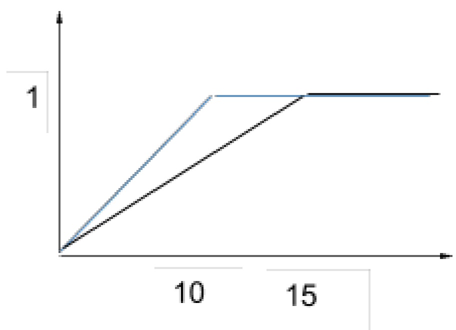
b. 0, 15

c.



e. $P(1 \leq X_1 \leq 3) = \frac{2}{15}$, $P(1 \leq X_2 \leq 3) = 2/10$

f.



Statistics

1 SW 2.17

$\mu_{Y=0.4}$ and $\sigma_Y{}^2 = 0.4 * 0.6 = 0.24$

a. i. $P(\bar{Y} \geq 0.43) = Pr(\frac{\bar{Y} - 0.4}{\sqrt{0.24/n}} \geq \frac{0.43 - 0.4}{\sqrt{0.24/n}}) = Pr(\frac{\bar{Y} - 0.4}{\sqrt{0.24/n}} \geq 0.6124) = 0.27$

ii. $P(\bar{Y} \leq 0.37) = Pr(\frac{\bar{Y} - 0.4}{\sqrt{0.24/n}} \leq \frac{0.37 - 0.4}{\sqrt{0.24/n}}) = Pr(\frac{\bar{Y} - 0.4}{\sqrt{0.24/n}} \leq -1.22) = 0.27$

b. Pr(-1.96≤Z≤1.96)=0.95

therefore we need to find *n* such that:

$0.41 = \frac{0.41 - 0.4}{\sqrt{0.24/n}} > -1.96\ and\ \frac{0.39 - 0.4}{\sqrt{0.24/n}} < -1.96.$ solving this gives us: $n \geq 9220$

SW 3.1

The CLT suggests that when the sample size (n) is large, the distribution of the sample average ($\bar{Y}$) is approximately $N(\mu_Y, \sigma_{\bar{Y}}{}^2)$ with $\sigma_{\bar{Y}}{}^2 = \sigma_Y{}^2/n$.

Given population $\mu_Y = 100$, $\sigma_Y{}^2$ =43 , we have

a. n=100, $\sigma_{\bar{Y}}{}^2 = \sigma_Y{}^2/n = 43/100 = 0.43$ and

$P(\bar{Y} < 101) = Pr(\frac{\bar{Y} - 100}{\sqrt{0.43}} < \frac{101 - 100}{\sqrt{0.43}}) \approx \Phi(1.525) = 0.9364$

b. n=64, $\sigma_{\bar{Y}}{}^2 = \sigma_Y{}^2/n = 43/64 = 0.6719,$ and

$P(101 < \bar{Y} < 103) = Pr(\frac{101 - 100}{\sqrt{0.6719}} < \frac{\bar{Y} - 100}{\sqrt{0.6719}} < \frac{103 - 100}{\sqrt{0.6719}}) \approx \Phi(3.6599) - \Phi(1.2200) = 0.9999 - 0.8888 = 0.1111$

c. n=165, $\sigma_{\bar{Y}}{}^2 = \sigma_Y{}^2/n = 43/165 = 0.2606$ and

$P(\bar{Y} > 98) = 1 - P(\bar{Y} \leq 98) = 1 - Pr(\frac{\bar{Y} - 100}{\sqrt{0.2606}} < \frac{98 - 100}{\sqrt{0.2606}}) \approx 1 - \Phi(-3.9178) = \Phi(3.9178) \approx 1$

2

3

<u>S&W 3.8</u>

Given that n = 1000, $\bar{X}$ = 1110, s_X= 123, a 95% confidence interval for the population mean is $\left\{ \bar{X} \pm 1.96\left(\frac{s_X}{\sqrt{n}}\right)\right\} = \left\{1110 \pm 1.96\left(\frac{123}{\sqrt{1000}}\right)\right\} = (1102.38,\ 1117.62)$

<u>S&W 3.13 (a)</u>

Given that n=420, $\bar{Y} = 646.2$, $s_Y = 19.5$, a 95% confidence interval for the population mean is $\left\{ \bar{Y} \pm 1.96\left(\frac{s_Y}{\sqrt{n}}\right)\right\} = \left\{646.2 \pm 1.96\left(\frac{19.5}{\sqrt{420}}\right)\right\} = (644.34,\ 648.06)$

4 SW 3.11

Assume that n is an even number. Then $\widehat{Y}$ is constructed by applying a weight of ½ to the n/2 “odd” observations and a weight of 3/2 to the remaining n/2 observations.

$E(\widehat{Y}) = 1/n\ [1/2\ E(Y_1) + 3/2E(Y_2) + ... + 1/2\ E(Y_{n-1}) + 3/2E(Y_n)]$

$= 1/n\ [1/2\ * n/2 * \mu_Y + 3/2\ * n/2 * \mu_Y] = \mu_Y$

$Var(\widehat{Y}) = 1/n^2\ [1/4\ Var(Y_1) + 9/4\ Var(Y_2) + ... + 1/4\ VarE(Y_{n-1}) + 9/4\ Var(Y_n)]$

$= 1/n^2[1/4 \ * n/2 * \sigma_Y^2 \ + 9/4 \ * n/2 * \sigma_Y^2] = \sigma_Y^2 \ = 1.25 \ \sigma_Y^2/n$

5.

1. E(1/2(Y1+Y2))=1/2 E(Y1+Y2) = 1/2 * (E(Y1) + E(Y2)) = 1/2 * 2 mu = mu
2. 1/8 var(Y), which is smaller than var(Y-bar) , so you prefer Y-tilde
3. It is biased
4. The estimator is based on the unknown quantity that you are trying to estimate!

6 Stock and Watson 3.2.

a. Let s be the number of successes in the trial. Then the fraction of success in n trials is

   $$\hat{p} = \tfrac{s}{n} = \tfrac{1}{n}\sum_i Y_i = \overline{Y}$$

b. $$E(\hat{p}) = E(\tfrac{1}{n}\sum_i Y_i) = \tfrac{1}{n}\sum_i E(Y_i) = \tfrac{1}{n}\sum_i p = p$$

c. $$var(\hat{p}) = var(\tfrac{1}{n}\sum_i Y_i) = \tfrac{1}{n^2}\sum_i var(Y_i) = \tfrac{1}{n^2}\sum_i p(1-p) = \tfrac{1}{n}p(1-p)$$

7

8

9 Stock and Watson 2.7. Let M and F denote the randomly selected male and female earnings, respectively. Then $\mu_C = \mu_M + \mu_F$ and $\sigma_C^2 = \sigma_M^2 + \sigma_F^2 + 2cov(M,F)$.

a. $\mu_C = 40000 + 45000 = 85000$
b. From the correlation formula we get $cov(M,F) = \sigma_M\sigma_F corr(M,F) = 12 \times 18 \times 0.8 = 172.8$ where the units are squared thousands of dollars per year.
c. $\sigma_C = \sqrt{\sigma_M^2 + \sigma_F^2 + 2cov(M,F)} = \sqrt{12^2 + 18^2 + 2 \times 172.8} = \sqrt{813.6} = 28.524$
d. Let the exchange rate be E Euros per dollar. Then the mean combined income would be $E \times \mu_C$ Euros per year, with standard deviation $E \times \sigma_C \times 1000$ Euros per year. The correlation is unit free, and thus unchanged after conversion.

10

1. 3
2. (2-3)^2+(3-3)^2+(4-3)^2 / (n-1) = 1.
3. [3-1.64*SE(Xbar), 3+1.64*SE(Xbar)] = [3-1.64/sqrt(3), 3+1.64/sqrt(3)]
4. Incorrect answer: “contains μ with 90% probability”. Correct: “If we would repeat this procedure many, many times, it would contain μ ~90% of the time.”.

11 No. Note X-tilde = (n-2) /n X-bar, so that E[X-tilde] = (n-2)/n * mu

12