STAT2001/6039 Final Examination 2012 Solutions

Solution to Problem 1

(a) Let D = "The widget is defective"
and T = "The test indicates that the widget is defective".

Then $P(D) = 3/9$ = 1/3, $P(T \mid D)$ = 0.88 and $P(\bar{T} \mid \bar{D})$ = 0.71.

Therefore

$$P(T) = P(D)P(T \mid D) + P(\bar{D})P(T \mid \bar{D}) = (1/3)(0.88) + (2/3)(0.29) = 0.486667$$

and so $P(D \mid T) = \dfrac{P(D)P(T \mid D)}{P(T)} = \dfrac{(1/3)0.88}{0.486667} = \boxed{\mathbf{0.6027}}$.

(b) Let D_i = "Exactly i of the two widgets are defective" and T_i = "Exactly i of the two widgets are indicated by the tests as being defective" (i = 1,2).

Then we wish to find

$$P(\bar{D}_0 \mid T_0) = 1 - P(D_0 \mid T_0) = 1 - \frac{P(D_0 T_0)}{P(T_0)},$$

where $P(D_0 T_0) = P(D_0)P(T_0 \mid D_0) = \dfrac{6}{9}\left(\dfrac{5}{8}\right)0.71^2$ = 0.210042.

To calculate $P(T_0)$ we partition the sample space as $S = \{AB, A\bar{B}, \bar{A}B, \bar{A}\bar{B}\}$,
where A = "The first widget is defective" and B = "The second widget is defective".

Then, by the law of total probability,

$$P(T_0) = P(AB)P(T_0 \mid AB) + P(A\bar{B})P(T_0 \mid A\bar{B}) + P(\bar{A}B)P(T_0 \mid \bar{A}B) + P(\bar{A}\bar{B})P(T_0 \mid \bar{A}\bar{B})$$

$$= \frac{3}{9}\left(\frac{2}{8}\right)0.12^2 + \frac{3}{9}\left(\frac{6}{8}\right)0.12(0.71) + \frac{6}{9}\left(\frac{3}{8}\right)0.71(0.12) + \frac{6}{9}\left(\frac{5}{8}\right)0.71^2 = 0.253842.$$

It follows that $P(\bar{D}_0 \mid T_0) = 1 - \dfrac{0.210042}{0.253842} = \boxed{\mathbf{0.1725}}$.

R Code for Problem 1 (not required, only for interest)

```
# (a)
PD = (1/3)*0.88+(2/3)*0.29; c(PD,(1/3)*0.88/PD) # 0.4866667 0.6027397

# (b)
PD0T0 = (6/9)*(5/8)*0.71^2
PT0 = (3/9)*(2/8)*0.12^2 + 2*(3/9)*(6/8)*0.12*0.71 + (6/9)*(5/8)*0.71^2
c(PD0T0,PT0,1-PD0T0/PT0) # 0.2100417 0.2538417 0.1725485
```

Solution to Problem 2

(a) Let X be the total number of times that the machine breaks down, and let Y be the total amount paid to Ben, in thousands of dollars.

Then $X \sim Poisson(\lambda)$ with pdf $f(x) = e^{-\lambda}\lambda^x / x!,\ x = 0,1,2,3,\dots$, where λ = 1.5. Also, $Y = \begin{cases} X, & X = 0,1,2 \\ 3, & X = 3,4,5,6,\dots \end{cases}$

So Y has pdf $f(y) = P(Y = y) = \begin{cases} P(X=0) = e^{-\lambda}\lambda^0/0! = 0.223130, & y = 0 \\ P(X=1) = e^{-\lambda}\lambda^1/1! = 0.334659, & y = 1 \\ P(X=2) = e^{-\lambda}\lambda^2/2! = 0.251021, & y = 2 \\ P(X \geq 3) = 1 - 0.223130 - 0.334659 - 0.251021 & \\ \quad = 0.191153, & y = 3 \end{cases}$

So $EY = \sum_{y=0}^{3} yf(y) = 0\times 0.223130 + 1\times 0.334659 + 2\times 0.251021 + 3\times 0.191153$

= 1.410 = $\boxed{\mathbf{1410}}$ dollars.

(b) $P(Y=0) = e^{-\lambda}$. So if y = 0, the likelihood function is $L(\lambda) = e^{-\lambda}$.
This is a strictly decreasing function with a maximum at λ = 0.
So if the insurance company pays Ben nothing, the MLE of λ is $\boxed{\mathbf{0}}$.

$P(Y=1) = \lambda e^{-\lambda}$. So if y = 1, $L(\lambda) = \lambda e^{-\lambda}$ $\Rightarrow l(\lambda) = \log L(\lambda) = \log\lambda - \lambda$

$\Rightarrow l'(\lambda) = (1/\lambda) - 1 \overset{\text{set}}{=} 0 \Rightarrow \lambda = 1$.

$P(Y=2) = \lambda^2 e^{-\lambda}/2$. So if y = 2, $L(\lambda) = \lambda^2 e^{-\lambda}$ $\Rightarrow l(\lambda) = \log L(\lambda) = 2\log\lambda - \lambda$

$\Rightarrow l'(\lambda) = (2/\lambda) - 1 \overset{\text{set}}{=} 0 \Rightarrow \lambda = 2$.

$P(Y=3)=1-e^{-\lambda}-\lambda e^{-\lambda}-\lambda^2 e^{-\lambda}/2$. So if y = 3, $L(\lambda)=1-e^{-\lambda}-\lambda e^{-\lambda}-\lambda^2 e^{-\lambda}/2$

$$\Rightarrow \quad L'(\lambda)=0-e^{-\lambda}(-1)-\{\lambda e^{-\lambda}(-1)+1e^{-\lambda}\}-(1/2)\{\lambda^2 e^{-\lambda}(-1)+2\lambda e^{-\lambda}\}$$
$$=e^{-\lambda}+\lambda e^{-\lambda}-e^{-\lambda}+(1/2)\lambda^2 e^{-\lambda}-\lambda e^{-\lambda}$$
$$=\lambda^2 e^{-\lambda}/2.$$

Thus $L'(\lambda)$ is positive for $\lambda>0$, and therefore $L(\lambda)$ is strictly increasing for $\lambda>0$. Thus, if y = 3, then $L(\lambda)$ is maximised at the largest possible value of λ, namely 2.

So the MLE of λ is $\boxed{\hat{\lambda}=\begin{cases}0 & \text{if } y=0 \\ 1 & \text{if } y=1 \\ 2 & \text{if } y=2 \text{ or } 3\end{cases}}$ (which occurs if $x=0$) (which occurs if $x=1$) (which occurs if $x\geq 2$)

Hence the pdf of this MLE is $\boxed{f(\hat{\lambda})=\begin{cases}e^{-\lambda}, & \hat{\lambda}=0 \\ \lambda e^{-\lambda}, & \hat{\lambda}=1 \\ 1-e^{-\lambda}-\lambda e^{-\lambda}, & \hat{\lambda}=2\end{cases}}$

Therefore the expected value of the MLE is

$$E\hat{\lambda}=\sum_{\hat{\lambda}}\hat{\lambda}f(\hat{\lambda})=0\left(e^{-\lambda}\right)+1\left(\lambda e^{-\lambda}\right)+2\left(1-e^{-\lambda}-\lambda e^{-\lambda}\right)$$
$$=\lambda e^{-\lambda}+2-2e^{-\lambda}-2\lambda e^{-\lambda}$$
$$=\boxed{2-(2+\lambda)e^{-\lambda},\ 0\leq\lambda\leq 2}$$ (general expression as a function of λ)
$$=2-(2+1)e^{-1}=2-3/e = \boxed{\mathbf{0.8964}}$$ for the case λ = 1.

R Code for Problem 2 (not required, only for interest)

```
# (a)
lam=1.5; xv=0:2; fxv=exp(-lam)*(lam^xv)/factorial(xv)
fxv # 0.2231302 0.3346952 0.2510214
sum(fxv) # 0.8088468

yv=0:3; fyv=c(fxv,1-sum(fxv))
fyv # 0.2231302 0.3346952 0.2510214 0.1911532
sum(yv*fyv) # 1.410198

# (b)
2-3/exp(1) # 0.8963617
```

Solution to Problem 3

(a) Here: $w = 1/(1-y)$ is a strictly increasing function for $0 < y < 1$, which means we may apply the *transformation method* (as follows)

$$w = \frac{1}{1-y} \Rightarrow y = 1 - \frac{1}{w} = 1 - w^{-1} \Rightarrow \frac{dy}{dw} = 0 - (-1)w^{-2} = \frac{1}{w^2}$$

$$f_Y(y) = \frac{y^{2-1}(1-y)^{1-1}}{\Gamma(2)\Gamma(1)/\Gamma(2+1)} = 2y,\ 0 \le y \le 1$$

$$f_W(w) = f_Y(y)\left|\frac{dy}{dw}\right| = 2\left(1-\frac{1}{w}\right)\left|\frac{1}{w^2}\right| = 2\left(\frac{1}{w^2}-\frac{1}{w^3}\right) = 2(w^{-2}-w^{-3})$$

$$y = 0 \Rightarrow w = 1/(1-0) = 1, \qquad y \to 1 \Rightarrow w \to \infty .$$

It follows that W has density $\boxed{f_W(w) = 2\left(\frac{1}{w^2}-\frac{1}{w^3}\right), 1 \le w < \infty}$.

Note: This result can also be obtained via the *cdf method*, i.e.

$$F_W(w) = P(W \le w) = P\left(\frac{1}{1-Y} \le w\right) = P\left(\frac{1}{w} \le 1-Y\right) = P\left(Y \le 1-\frac{1}{w}\right) = \int_0^{1-1/w} 2y\,dy$$

$$= \left(1-\frac{1}{w}\right)^2 \Rightarrow f_W(w) = F_W'(w) = 2\left(1-w^{-1}\right)^1(-(-w^{-2})) = 2\left(\frac{1}{w^2}-\frac{1}{w^3}\right), w \ge 1.$$

To find the mode of W we note that $f'(w) = 2(-2w^{-3}+3w^{-4})$.

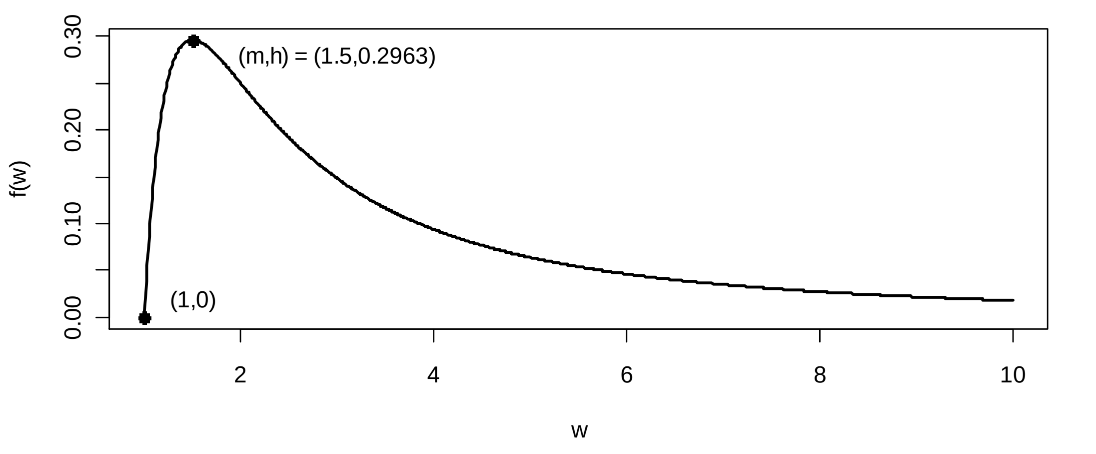
Setting this to zero leads to $w = m = Mode(W) = 3/2 = \boxed{\mathbf{1.5}}$.

Then also $h = f_W(m) = 2\left(\frac{1}{(3/2)^2}-\frac{1}{(3/2)^3}\right) = \frac{8}{27} = \boxed{\mathbf{0.2963}}$.

Some other features of W's density are that $f_W(1) = 0$ and $f_W(w) \to 0$ as $w \to \infty$.

These facts lead to the sketch of W's density shown in Figure 1.

Figure 1: Density of W



(b) R has cdf

$$F(r) = P\{(X-Y)^2 \le r\}$$
$$= P(-\sqrt{r} \le X - Y \le \sqrt{r})$$
$$= P(X - \sqrt{r} \le Y \le X + \sqrt{r})$$
$$= \iint_{x-\sqrt{r} \le y \le x+\sqrt{r}} f(x,y)\,dxdy\,.$$

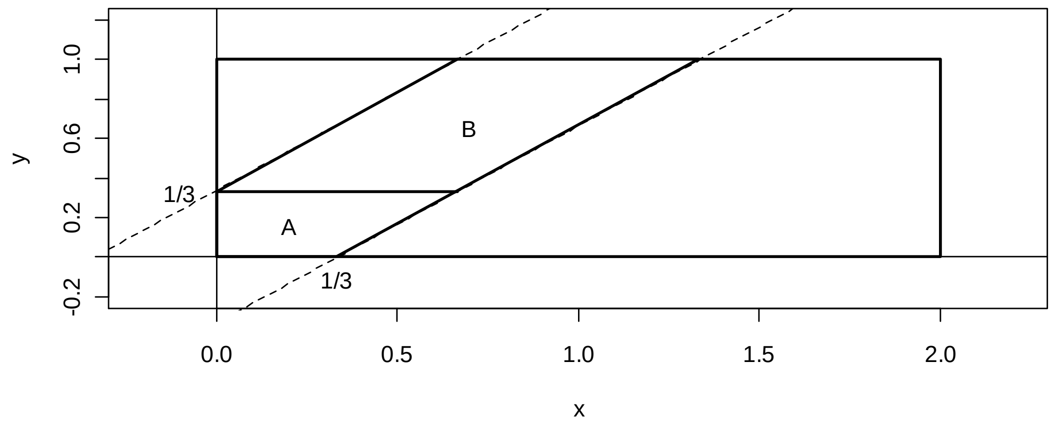
We now need to consider two separate cases, namely $0 < r < 1$ and $1 < r < 4$.

For $0 < r < 1$, $F(r) = \int_A f(x,y)dxdy + \int_B f(x,y)dxdy$

for regions A and B shown in Figure 2

$$= \int_{y=0}^{\sqrt{r}} y\left(\int_{x=0}^{y+\sqrt{r}} dx\right)dy + \int_{y=\sqrt{r}}^{1} y\left(\int_{x=y-\sqrt{r}}^{y+\sqrt{r}} dx\right)dy$$
$$= \int_{0}^{\sqrt{r}} y(y+\sqrt{r})dy + \int_{\sqrt{r}}^{1} y\left\{(y+\sqrt{r}) - (y-\sqrt{r})\right\}dy$$
$$= \left[\frac{y^3}{3} + \frac{y^2\sqrt{r}}{2}\Bigg|_{y=0}^{\sqrt{r}}\right] + \cancel{2}\sqrt{r}\left[\frac{y^2}{\cancel{2}}\Bigg|_{y=\sqrt{r}}^{1}\right]$$
$$= \left\{\left(\frac{r^{3/2}}{3} + \frac{rr^{1/2}}{2}\right) - (0+0)\right\} + \sqrt{r}(1-r) = \sqrt{r}\left(1 - \frac{r}{6}\right).$$

Figure 2: The case 0 < r < 1 for deriving f(r) with the example r = 1/9

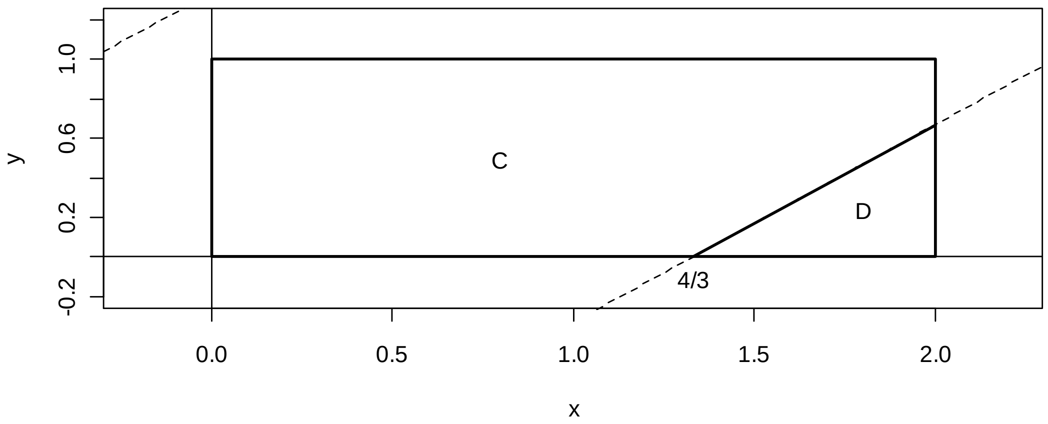


For $1 < r < 4$, $\quad F(r) = \int_C f(x,y)dxdy = 1 - \int_D f(x,y)dxdy$

for regions C and D shown in Figure 3

$$
\begin{aligned}
&= 1 - \int_{y=0}^{2-\sqrt{r}} y \left(\int_{x=y+\sqrt{r}}^{2} dx \right) dy \\
&= 1 - \int_{0}^{2-\sqrt{r}} y(2 - y - \sqrt{r}) dy \\
&= 1 - \left[(2-\sqrt{r})\frac{y^2}{2} - \frac{y^3}{3} \Big|_{y=0}^{2-\sqrt{r}} \right] \\
&= 1 - \left\{ (2-\sqrt{r})\frac{(2-\sqrt{r})^2}{2} - \frac{(2-\sqrt{r})^3}{3} \right\} = 1 - \frac{1}{6}(2-\sqrt{r})^3.
\end{aligned}
$$

Figure 3: The case 1 < r < 4 for deriving f(r) with the example r = 16/9



In summary so far, $\quad F(r) = \begin{cases} r^{1/2} - (1/6)r^{3/2}, & 0 < r < 1 \\ 1 - (1/6)(2 - r^{1/2})^3, & 1 < r < 4 \end{cases}$

So $\quad f(r) = F'(r) = \begin{cases} (1/2)r^{-1/2} - (1/6)(3/2)r^{1/2} = (1/2)r^{-1/2} - (1/4)r^{1/2}, & 0 < r < 1 \\ 0 - (1/6)3(2 - r^{1/2})^2(-1/2)r^{-1/2} = (1/4)r^{-1/2}(2 - r^{1/2})^2, & 1 < r < 4 \end{cases}$

Thus, R has probability density function $\quad \boxed{f_R(r) = \begin{cases} \frac{1}{2}\left(\frac{1}{\sqrt{r}} - \frac{\sqrt{r}}{2} \right), & 0 \le r \le 1 \\ \frac{(2-\sqrt{r})^2}{4\sqrt{r}}, & 1 < r \le 4 \end{cases}}$

Next, $c = ER = \int r f_R(r)dr = \int_0^1 r \times \frac{1}{2}\left(\frac{1}{\sqrt{r}} - \frac{\sqrt{r}}{2}\right)dr + \int_1^4 r \times \frac{(2-\sqrt{r})^2}{4\sqrt{r}}dr$ = etc.

A simpler way to proceed is to first note that:

$EX = 1, \quad VX = \frac{(2-0)^2}{12} = \frac{1}{3}$ by properties of the uniform distribution

$EY = \frac{2}{2+1} = \frac{2}{3}, \quad VY = \frac{2\times 1}{(2+1)^2(2+1+1)} = \frac{1}{18}$

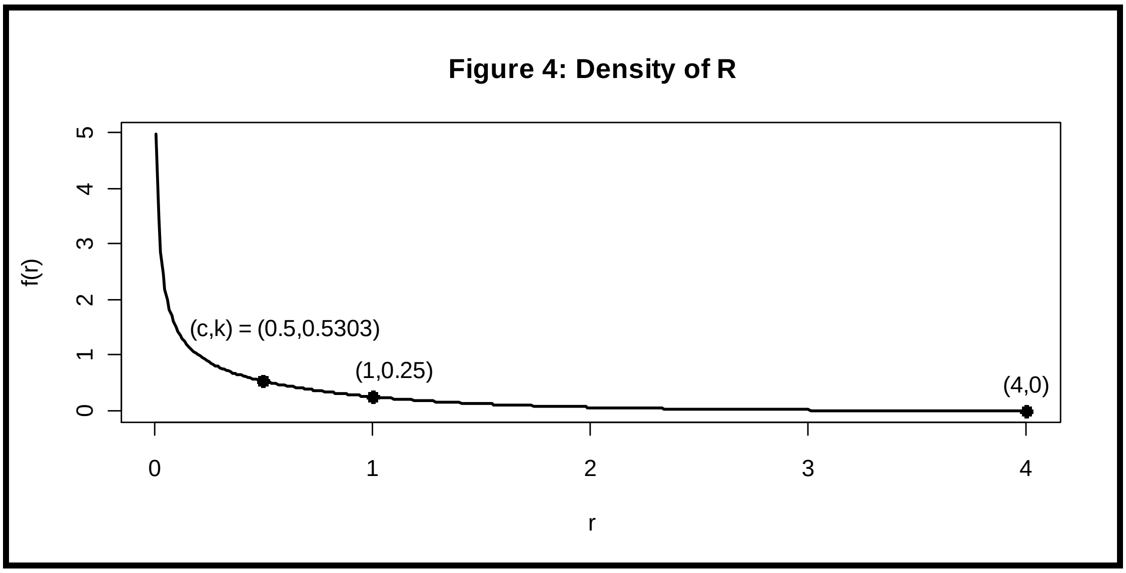
by properties of the beta distribution.

Thus $c = ER = E\{(X-Y)^2\} = V(X-Y) + \{E(X-Y)\}^2$

$$= (VX + VY) + (EX - EY)^2 = \frac{1}{3} + \frac{1}{18} + \left(1 - \frac{2}{3}\right)^2 = \frac{1}{2} = \boxed{\mathbf{0.5}}.$$

Then also $k = f_R(c) = \frac{(2-\sqrt{0.5})^2}{4\sqrt{0.5}} = \boxed{\mathbf{0.5303}}$.

Some other observations regarding R's density are that $f_R(0) \to \infty$ as $r \to 0$, $f_R(4) = 0$ and $f_R(1)$ = 1/4. We may also derive $f_R'(r)$ and show that this is negative for all r > 0. These observations lead to the sketch of R's density shown in Figure 4.



R Code for Problem 3 (not required, only for interest)

```
# (a)
X11(w=8,h=4); wv=seq(1,10,0.01); fwv=2*(1/wv^2-1/wv^3)
plot(wv,fwv,type="l",xlab="w",ylab="f(w)",lwd=2,
        main="Figure 1: Density of W")
points(1.5, 2*(1/1.5^2-1/1.5^3), pch=16,cex=1.2)
2*(1/1.5^2-1/1.5^3) # 0.2962963
text(3, 0.28, "(m,h) = (1.5,0.2963)")
points(1,0, pch=16,cex=1.2)
text(1.5,0.02,"(1,0)")

# (b)
# Fig. 2
X11(w=8,h=4)
plot(c(-0.2,2.2),c(-0.2,1.2),type="n",xlab="x",ylab="y",
        main="Figure 2: The case 0 < r < 1 for deriving f(r) with the example r = 1/9")
lines(c(0,0,2,2,0),c(0,1,1,0,0),lwd=2); abline(v=0,h=0)
abline(-1/3,1,lty=2); abline(1/3,1,lty=2)
lines(c(0,0,2/3,4/3,1/3,0), c(0,1/3,1,1,0,0),lwd=2)
lines(c(0,2/3),c(1/3,1/3),lwd=2)
text(0.2,1/6,"A"); text(0.7,2/3,"B"); text(-0.1,1/3,"1/3"); text(1/3,-0.1,"1/3")

# Fig. 3
X11(w=8,h=4)
plot(c(-0.2,2.2),c(-0.2,1.2),type="n",xlab="x",ylab="y",
    main="Figure 3: The case 1 < r < 4 for deriving f(r) with the example r = 16/9")
lines(c(0,0,2,2,0),c(0,1,1,0,0),lwd=2); abline(v=0,h=0)
abline(-4/3,1,lty=2); abline(4/3,1,lty=2)
lines(c(4/3,2),c(0,2/3),lwd=2)
text(0.8,0.5,"C"); text(1.8,0.25,"D"); text(4/3,-0.1,"4/3")

# Fig. 4
X11(w=8,h=4)
rv1=seq(0.01,1,0.01); rv2=seq(1.01,4,0.01)
frv1=(1/(2*sqrt(rv1))) - sqrt(rv1)/4
frv2= (1/(4*sqrt(rv2)))*(2-sqrt(rv2))^2
rv=c(rv1,rv2); frv=c(frv1,frv2)
```

```
plot(rv,frv,type="l",xlab="r",ylab="f(r)",lwd=2,
        main="Figure 4: Density of R")
c=1/2; k=(1/(2*sqrt(c))) - sqrt(c)/4; k # 0.5303301
points(c,k,pch=16,cex=1.2)
text(0.6,1.5,"(c,k) = (0.5,0.5303)")
points(4,0,pch=16,cex=1.2)
text(4,0.5,"(4,0)")
points(1,0.25,pch=16,cex=1.2)
text(1.1,0.75,"(1,0.25)")

# Checking that ER = 0.5 in various ways

sum(rv*frv)/sum(frv) # 0.5392423

rv1=seq(0.0001,1,0.0001); rv2=seq(1.0001,4,0.0001)
frv1=(1/(2*sqrt(rv1))) - sqrt(rv1)/4
frv2= (1/(4*sqrt(rv2)))*(2-sqrt(rv2))^2
rv=c(rv1,rv2); frv=c(frv1,frv2)
sum(rv*frv)/sum(frv) # 0.5036776

rv1=seq(0.00001,1,0.00001); rv2=seq(1.00001,4,0.00001)
frv1=(1/(2*sqrt(rv1))) - sqrt(rv1)/4
frv2= (1/(4*sqrt(rv2)))*(2-sqrt(rv2))^2
rv=c(rv1,rv2); frv=c(frv1,frv2)
sum(rv*frv)/sum(frv) # 0.5011572

J=10000; ysamp=runif(J,0,2); xsamp=rbeta(J,2,1)
rsamp=(xsamp-ysamp)^2; est=mean(rsamp) #
ci=est+c(-1,1)*qnorm(0.975)*sd(rsamp)/sqrt(J)
c(est,ci) # 0.5034652 0.4918551 0.5150753

J=1000000; ysamp=runif(J,0,2); xsamp=rbeta(J,2,1)
rsamp=(xsamp-ysamp)^2; est=mean(rsamp) #
ci=est+c(-1,1)*qnorm(0.975)*sd(rsamp)/sqrt(J)
c(est,ci) # 0.5001565 0.4989960 0.5013170
```

Solution to Problem 4

(a) Let X be the number of defectives in the sample of 100, and let $p = p_1$.

Then $P(X=0\,|\,p) = (1-p)^{100}$ and $f(p) = \dfrac{p^{1-1}(1-p)^{19-1}}{\Gamma(1)\Gamma(19)/\Gamma(20)} = 19(1-p)^{18},\ 0<p<1$.

So $P(X=0) = EP(X=0\,|\,p) = \int_0^1 P(X=0\,|\,p)f(p)dp = \int_0^1 (1-p)^{100}19(1-p)^{18}dp$

$$= \frac{19}{119}\int_0^1 119(1-p)^{118}dp = \frac{19}{119}\times 1 = \boxed{\mathbf{0.1597}}.$$

(b) Let Y_i be the number of defectives on day i, i = 1,...,100.

Then $(Y_i\,|\,p_i) \sim Bin(2, p_i)$, with $E(Y_i\,|\,p_i) = 2p_i$ and $V(Y_i\,|\,p_i) = 2p_i(1-p_i)$.

Thus $Y_1,...,Y_{100} \sim iid$, with each random variable having mean and variance given by:

$$\mu = EY_i = EE(Y_i\,|\,\mu_i) = E(2p_i) = 2Ep_i$$

$$\sigma^2 = VY_i = EV(Y_i\,|\,\mu_i) + VE(Y_i\,|\,\mu_i) = E\{2p_i(1-p_i)\} + V(2p_i)$$

$$= 2Ep_i - 2Ep_i^2 + 4Vp_i = 2Ep_i - 2\{Vp_i + (Ep_i)^2\} + 4Vp_i$$

$$= 2\{Vp_i + (Ep_i)(1-Ep_i)\}.$$

Now: $Ep_i = \dfrac{1}{1+19} = 0.05$

$$Vp_i = \frac{1\times 19}{(1+19)^2(1+19+1)} = 0.002261905.$$

So: $\mu = 2\times 0.05 = 0.1$

$$\sigma^2 = 2\left\{0.002261905 + 0.05(1-0.05)\right\} = 0.09952381.$$

Thus by the central limit theorem,

$$T = Y_1 + ... + Y_{100} \overset{\sim}{\sim} N(100\mu, 100\sigma^2) = N(10, 9.952381).$$

Hence $q = P(T \geq 20) \cong P\left(Z > \dfrac{19.5-10}{\sqrt{9.952381}}\right) = P(Z > 3.01)$

$\cong P(Z > 3.0) = \boxed{\mathbf{0.00135}}$ (where $Z \sim N(0,1)$).

Note: The above answer makes use of a continuity correction and normal tables. Marks will be deducted for any answer which fails to apply the continuity correction or uses a Poisson approximation in place of the exact binomial distribution for $(Y_i\,|\,p_i)$. No marks will be lost for writing more accurately

$$q = P(T \geq 20) \cong P\left(Z > \frac{19.5-10}{\sqrt{9.952381}}\right) = P(Z > 3.01134) = 0.0013005.$$

(c) To find an exact upper bound for q we write

$$q = P(T \geq 20) = P(|T - 10| \geq 10) - P(T = 0),$$

where $P(T = 0) = P(Y_1 = 0, ..., Y_{100} = 0) = P(Y_1 = 0)...P(Y_{100} = 0) = \{P(Y_1 = 0)\}^{100}$.

Now, $P(Y_1 = 0) = EP(Y_1 = 0 \mid p_1) = E\{(1 - p_1)^2\}$

$$= V(1 - p_1) + \{E(1 - p_1)\}^2$$

$$= Vp_1 + (1 - Ep_1)^2$$

$$= 0.002261905 + (1 - 0.05)^2$$

$$= 0.9047619.$$

Thus $P(T = 0) = 0.9047619^{100}$ = 0.00004502261.

Next, $P(|T - 10| \geq 10) = P(|T - ET| \geq kSD(T)) \leq \dfrac{1}{k^2} = \dfrac{VT}{100}$ = 0.0995238.

(This follows by Chebyshev's theorem with $k = 10 / SD(T)$.)

It follows that $q = P(T \geq 20) \leq 0.0995238 - 0.00004502261$ = $\boxed{\mathbf{0.09948}}$.

R Code for Problem 4 (not required, only for interest)

```
# (b)
Ep = 1/20; mu = 2*Ep; Vp = 19/(20^2 *21); sig2=2*(Vp+Ep*(1-Ep))
c(Ep,Vp,mu,sig2,sqrt(sig2))
    # 0.050000000 0.002261905 0.100000000 0.099523810 0.315473944
z = (19.5-10)/sqrt(100*sig2); z # 3.011342
1-pnorm(z) # 0.001300478


# (c)
PY1is0 = Vp+(1-Ep)^2; PY1is0 # 0.9047619
PTis0 = PY1is0^100; PTis0 # 4.502261e-05
VT=100*sig2; VT # 9.95238
inversek2 = VT/10^2; inversek2 # 0.0995238
inversek2- PTis0 # 0.09947879
```

Solution to Problem 5

(a) Here: $n = \sum_{k=0}^{5} f_k = 186+42+13+5+3+1 = 250$ (check)

$$\sum_{i=1}^{n} x_i = \sum_{k=0}^{5} f_k n_k = 186\times0+42\times1+13\times2+5\times3+3\times4+1\times5 = 100$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{250}(100) = 0.4$$

$$\sum_{i=1}^{n} x_i^2 = \sum_{k=0}^{5} f_k n_k^2 = 186\times0^2+42\times1^2+13\times2^2+5\times3^2+3\times4^2+1\times5^2 = 212$$

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = \frac{1}{249}\left(212-250\times0.4^2\right) = \frac{172}{249} = 0.6907631$$

$$s = \sqrt{s^2} = 0.8311216, \quad \alpha = 0.2, \quad z_{\alpha/2} = z_{0.1} = 1.281552.$$

So the required 80% confidence interval is

$$\left(\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}\right) = \left(0.4 \pm 1.281552\times\frac{0.8311216}{\sqrt{250}}\right) = \boxed{\mathbf{(0.3474,\ 0.4526)}}.$$

(b) We effectively have a random sample of m = 105 from the adults in Urbania. Of these, y = 42 + 13 + 5 + 3 + 1 = 64 own at least one house each, this being a proportion of $\hat{p} = y/m$ = 64/105 = 0.6095238, which estimates p, the true proportion of all adults in Urbania who own at least one house each. So an appropriate hypothesis test is as follows:

$$H_0 : p = 0.5; \quad H_1 : p \geq 0.5$$

$$TS : Z = \frac{\hat{p}-0.5}{\sqrt{0.5(1-0.5)/m}} \sim N(0,1) \text{ if } H_0 \text{ is true}$$

$$RR : Z > z_{0.05} = 1.645 \quad \text{(from normal tables)}$$

$$z = \frac{0.6095238-0.5}{\sqrt{0.5(1-0.5)/105}} = 2.24457 \in RR.$$

So we reject the null hypothesis at the 5% level and conclude in favour of the alternative that more than 50% of adults in Urbania own at least one house each.

The p-value associated with this test is $P(Z > z) = P(Z > 2.24) = \boxed{\mathbf{0.0125}}$.

Note: If the null hypothesis is true, then $Y \sim Bin(105, 0.5)$.

So we may report the p-value more accurately as

$$P(Y \geq 64) \cong P\left(R > \frac{64 - 0.5 - 105 \times 0.5}{\sqrt{105 \times 0.5 \times (1 - 0.5)}} \right) \text{ where } R \sim N(0,1)$$

and where "-0.5" is a suitable continuity correction

$$= P(R > 2.15) = 0.0158,$$

or even more accurately, with the aid of a computer, as

$$P(Y \geq 64) = \sum_{y=64}^{105} \binom{105}{y} \left(\frac{1}{2}\right)^y \left(1 - \frac{1}{2}\right)^{105-y} = 0.01565.$$

R Code for Problem 5 (not required, only for interest)

```
# (a)
nv=c(0,1,2,3,4,5);  fv=c(186,42,13,5,3,1); n = sum(fv); n # 250 (check)
sumxi=sum(nv*fv); sumxi2=sum(nv^2 * fv); c(sumxi,sumxi2) # 100 212
xbar = sumxi/n; s2=(1/(n-1))*(sumxi2 - n*xbar^2); s1 = sqrt(s2)
c(xbar,s2,s1) # 0.4000000 0.6907631 0.8311216
z=qnorm(0.9); z # 1.281552
ci=xbar + c(-1,1)*z*s1/sqrt(n); ci # 0.3326356 0.4673644

# (b)
1-pnorm(2.15) # 0.01577761
1-pbinom(63,105,0.5) # 0.01565089
```

Solution to Problem 6

Let X be the number of sixes that come up. Then $X \sim Bin(2, 1/6)$,
and consequently $EX = 2 \times \frac{1}{6} = \frac{1}{3}$ and $VX = 2 \times \frac{1}{6}\left(1 - \frac{1}{6}\right) = \frac{5}{18} \left(= \frac{10}{36}\right)$.

Also, let Y be the total of the two numbers that come up. Then $Y = Y_1 + Y_2$,
where Y_1 and Y_2 are independent random variables, each with density

$f(y_i) = 1/6, \; y_i = 1, \ldots, 6$.

Thus $EY_i = \frac{1}{6}(1 + \ldots + 6) = \frac{7}{2}$, $EY_i^2 = \frac{1}{6}(1^2 + \ldots + 6^2) = \frac{91}{6}$ and $VY_i = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$.

Therefore $EY = EY_1 + EY_2 = 2\times\frac{7}{2} = 7$ and $VY = VY_1 + VY_2 = 2\times\frac{35}{12} = \frac{35}{6}\left(=\frac{210}{36}\right)$.

Next, $E(XY) = EE(XY \mid X) = \sum_{x=0}^{2} P(X = x)\times E(XY \mid X = x)$

$$= P(X=0)E(XY \mid X=0) + P(X=1)E(XY \mid X=1) + P(X=2)E(XY \mid X=2)$$
$$= \left(\frac{5}{6}\right)^2 E(0\times Y \mid X=0) + 2\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)E(1\times Y \mid X=1) + \left(\frac{1}{6}\right)^2 E(2\times Y \mid X=2)$$
$$= \frac{35}{36}\times 0 + \frac{10}{36}\times E(Y \mid X=1) + \frac{1}{36}\times 2\times E(Y \mid X=2)$$
$$= 0 + \frac{10}{36}\times 9 + \frac{1}{36}\times 2\times 12 = \frac{114}{36} = \frac{19}{6}.$$

Note: If X = 1 then exactly one six comes up. In that case the other number is equally likely to be 1, 2, 3, 4 or 5, and so its mean is $(1+2+3+4+5)/5 = 3$. Thus $E(Y \mid X=1) = 6+3 = 9$. Likewise, but more simply, $E(Y \mid X=2) = 2\times 6 = 12$.

It follows that $C(X,Y) = E(XY) - (EX)EY = \frac{19}{6} - \frac{1}{3}\times 7 = \frac{5}{6}\left(=\frac{30}{36}\right)$.

So the required correlation is

$$\rho = Corr(X,Y) = \frac{C(X,Y)}{SD(X)SD(Y)} = \frac{30/\cancel{36}}{\sqrt{10/\cancel{36}}\times\sqrt{210/\cancel{36}}} = \sqrt{\frac{3}{7}} = \boxed{\mathbf{0.6547}}.$$

Alternative working

We have already seen that: $E(Y \mid X=1) = 6+3 = 9$ $(= 6+3\times 1)$

$E(Y \mid X=2) = 2\times 6 = 12$ $(= 6+3\times 2)$.

Also, by the same logic: $E(Y \mid X=0) = 3+3 = 6$ $(= 6+3\times 0)$.

Thus we may generally write $E(Y \mid X=x) = 6+3x$
(for all possible values x of X, namely x = 0,1,2).

Consequently, $E(Y \mid X) = 6+3X$, and so an alternative working for the covariance term is: $C(X,Y) = EC(X,Y \mid X) + C\{E(X \mid X), E(Y \mid X)\}$

$$= E0 + C(X, 6+3X)$$
$$= 0 + 3VX = 3\times\frac{5}{18} = \frac{5}{6} \text{ (as before).}$$

An alternative working for the entire problem is as follows. Consider all 36 equiprobable outcomes of the experiment, and for each one write the corresponding values of X (= number of 6s) and Y (= total), separated by a comma, as follows:

	6	1,7	1,8	1,9	1,10	1,11	2,12
	5	0,6	0,7	0,8	0,9	0,10	1,11
Die 2	**4**	0,5	0,6	0,7	0,8	0,9	1,10
	3	0,4	0,5	0,6	0,7	0,8	1,9
	2	0,3	0,4	0,5	0,6	0,7	1,8
	1	0,2	0,3	0,4	0,5	0,6	1,7
		1	**2**	**3**	**4**	**5**	**6**
				Die 1			

From this table we see that:

$$EX = \frac{0\times 25+1\times 10+2\times 1}{36} = \frac{1}{3}, \qquad EX^2 = \frac{0^2\times 25+1^2\times 10+2^2\times 1}{36} = \frac{7}{18}$$

$$EY = \frac{2\times 1+3\times 2+4\times 3+5\times 4+6\times 5+7\times 6+8\times 5+9\times 4+10\times 3+11\times 2+12\times 1}{36} = 7$$

$$EY^2 = \frac{2^2\times 1+3^2\times 2+4^2\times 3+5^2\times 4+6^2\times 5+7^2\times 6+8^2\times 5+9^2\times 4+10^2\times 3+11^2\times 2+12^2\times 1}{36}$$

= 329/6

$$E(XY) = \frac{2\times(1\times 7+1\times 8+1\times 9+1\times 10+1\times 11)+2\times 12}{36} = \frac{19}{6}.$$

Thus: $VX = \frac{7}{18} - \left(\frac{1}{3}\right)^2 = \frac{5}{18}$, $VY = \frac{329}{6} - 7^2 = \frac{35}{6}$ and $C(X,Y) = \frac{19}{6} - \frac{1}{3}\times 7 = \frac{5}{6}$.

Consequently, $\rho = \frac{C(X,Y)}{SD(X)SD(Y)} = \frac{5/6}{\sqrt{(5/18)(35/6)}} = \sqrt{\frac{3}{7}}$ = 0.6547 (as before).

R Code for Problem 6 (not required, only for interest)

```
J=10000; nv=rep(NA,J); tv=rep(NA,J); set.seed(217)
for(j in 1:J){    x=sample(1:6,2,T);     nv[j]=length(x[x==6]); tv[j]=sum(x)    }
est = cor(nv,tv); est   # 0.6555782    (A Monte Carlo estimate)
z=qnorm(0.975)
a=0.5*log((1+est)/(1-est))-z/sqrt(J-3); b=0.5*log((1+est)/(1-est))+z/sqrt(J-3)
L=(exp(2*a)-1)/(exp(2*a)+1); U =(exp(2*b)-1)/(exp(2*b)+1)
c(L,U) # 0.6442564 0.6666127    (Monte Carlo 95% confidence interval)
# For more information (all non-assessable) you may look up "Pearson product-moment correlation coefficient" on Wikipedia, or elsewhere.
```