STAT2001/6039 Final Examination June 2014 Solutions

Solution to Problem 1

(a) Let X be number of times a single randomly selected Komto car breaks down in a given one year period, and let I be the indicator variable for the car being Type B.

Then: $\mu_1' = EX = P(I=0)E(X \mid I=0) + P(I=1)E(X \mid I=1)$

$$= (1-p)\times 0 + p\lambda = p\lambda$$

$$\mu_2' = EX^2 = P(I=0)E(X^2 \mid I=0) + P(I=1)E(X^2 \mid I=1)$$

$$= (1-p)\times 0^2 + p(\lambda+\lambda^2) = p\lambda(1+\lambda).$$

We now equate $\mu_1' = m_1'$ and $\mu_2' = m_2'$,

or equivalently $p\lambda = \bar{y}$ and $p\lambda(1+\lambda) = m$,

where $m_1' = \bar{y} = \frac{1}{n}(y_1 + ... + y_n)$ and $m = m_2' = \frac{1}{n}(y_1^2 + ... + y_n^2)$.

Solving, we arrive at the method of moments estimates, $\hat{\lambda} = \frac{m}{\bar{y}} - 1$ and $\hat{p} = \frac{\bar{y}}{\hat{\lambda}}$.

Numerically: $\bar{y} = \frac{1}{5}(1+0+5+1+3) = 2$, $m = \frac{1}{5}(1^2+0^2+5^2+1^2+3^2) = \frac{36}{5} = 7.2$

$$\hat{\lambda} = \frac{7.2}{2} - 1 = \boxed{2.6}, \qquad \hat{p} = \frac{2}{2.6} = \boxed{0.7692}.$$

(b) $P(X=0) = P(I=0)P(X=0 \mid I=0) + P(I=1)P(X=0 \mid I=1)$

$$= (1-p)\times 1 + p \times \frac{e^{-\lambda}\lambda^0}{0!} = 1 - p + pe^{-\lambda}.$$

Also, for $x > 0$, $P(X = x) = P(I = 0)P(X = x \mid I = 0) + P(I = 1)P(X = x \mid I = 1)$

$$= (1 - p) \times 0 + p \times \frac{e^{-\lambda}\lambda^x}{x!} = p\frac{e^{-\lambda}\lambda^x}{x!}.$$

Thus, X has density function $f(x) = \begin{cases} 1 - p + pe^{-\lambda}, & x = 0 \\ p\dfrac{e^{-\lambda}\lambda^x}{x!}, & x = 1, 2, 3, \ldots \end{cases}$

So $f(y) = f(y_1, \ldots, y_n) = \left\{\prod_{i:y_i=0} (1 - p + pe^{-\lambda})\right\} \prod_{i:y_i>0} p\frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$

$$= (1 - p + pe^{-\lambda})^{n-k} p^k \frac{e^{-k\lambda}\lambda^{y_T}}{y_1! \ldots y_n!} \qquad \text{(note that } 0! = 1\text{)},$$

where $k = \sum_{i=1}^{n} I(y_i > 0)$ is the number of non-zeros amongst $y_1, \ldots, y_n$.

So the likelihood function is

$$L(p, \lambda) = (1 - p + pe^{-\lambda})^{n-k} p^k e^{-k\lambda}\lambda^{y_T},$$

and the loglikelihood function is

$$l(p, \lambda) = (n - k)\log(1 - p + pe^{-\lambda}) + k\log p - k\lambda + y_T \log\lambda.$$

For the data $y = (1, 1, 4, 1, 3)$, we have that $k = n$, and so:

$$L(p, \lambda) = p^n e^{-n\lambda}\lambda^{y_T}$$

$$l(p, \lambda) = n\log p - n\lambda + y_T \log\lambda.$$

Taking the partial derivative with respect to λ, we have $\dfrac{\partial l(p, \lambda)}{\partial\lambda} = 0 - n + \dfrac{y_T}{\lambda}$.

Setting this to zero leads to the MLE, $\hat{\lambda} = \dfrac{y_T}{n} = \bar{y} = \dfrac{1}{5}(1 + 1 + 4 + 1 + 3) = \boxed{2}$.

Next, observe that for any fixed λ, $L(p, \lambda) = p^n e^{-n\lambda}\lambda^{y_T}$ is strictly increasing in p.

So the MLE of p is $\hat{p} = \boxed{1}$. (This makes sense, because all 5 y-values are positive).

(c) For a randomly selected Komto car with a *y*-value of zero, the probability of that car being Type B is

$$q = P(I=1 \mid X=0) = \frac{P(I=1)P(X=0 \mid I=1)}{P(I=0)P(X=0 \mid I=0) + P(I=1)P(X=0 \mid I=1)}$$

$$= \frac{pe^{-\lambda}}{(1-p)\times 1 + pe^{-\lambda}}.$$

Now, the cars in the sample with *y*-values 1 and 2 are definitely Type B. Let R be the number of the other three cars (those with a *y*-value of zero) which are Type B. Then

$$(R \mid y) \sim Bin(3, q).$$

So, given *y*, the mean of that number of cars is

$$E(R \mid y) = 3q = 3\times\frac{0.6e^{-0.2}}{1-0.6+0.6e^{-0.2}} = 3\times 0.551186 = 1.654.$$

So the mean of the total number of Type B cars in the sample is

$$E(2+R \mid y) = 2 + E(R \mid y) = 2 + 1.654 = \boxed{3.654}.$$

The variance of the total number of cars is

$$V(2+R \mid y) = V(R \mid y) = 3q(1-q) = 3\times 0.551186\times(1-0.551186) = \boxed{0.7421}.$$

Next, given *y*, the entire distribution of the number of cars amongst the three in the sample with *y*-value zero is given by

$$f(r \mid y) = \binom{3}{r} q^r (1-q)^{3-r} = \begin{cases} 0.0904, & r=0 \\ 0.3331, & r=1 \\ 0.4091, & r=2 \\ 0.1675, & r=3. \end{cases}$$

The largest of these four probabilities, 0.4091, corresponds to the value $r = 2$, and so the mode of the total number of Type B cars in the sample is $2 + 2 = \boxed{4}$.

Note: The median of that total number is also 4, since

$$0.0904 + 0.3331 + 0.4091 \ge 1/2 \text{ and } 0.4091 + 0.1675 \ge 1/2.$$

R Code (not required)

```
p=0.6; lam=0.2; q=p*exp(-lam)/(1-p+p*exp(-lam))
c(q,3*q,3*q*(1-q)) # 0.5511863 1.6535590 0.7421399
dbinom(0:3,3,q) # 0.0904062 0.3330825 0.4090574 0.1674539
```

Solution to Problem 2

(a) $C(W,X) = C(UY, U+Y) = E\{UY(U+Y)\} - \{E(UY)\}E(U+Y)$

$$= E(U^2Y) + E(UY^2) - (EU)(EY)(EU+EY)$$

$$= (EY)\{VU + (EU)^2\} + (EU)\{VY + (EY)^2\} - (EU)(EY)(EU+EY)$$

$$= 1\times\left\{\frac{1}{12} + \left(\frac{1}{2}\right)^2\right\} + \frac{1}{2}\times\left\{\frac{1}{2} + 1^2\right\} - \frac{1}{2}\times 1\times\left(\frac{1}{2}+1\right) = \frac{1}{3} + \frac{3}{4} - \frac{3}{4} = \boxed{1/3}.$$

Alternatively,

$$C(W,X) = C(UY, U+Y) = EC(UY, U+Y \mid Y) + C\{E(UY \mid Y), E(U+Y \mid Y)\}$$

$$= E\{YC(U,U \mid Y)\} + C\{YE(U \mid Y), Y + E(U \mid Y)\}$$

$$= E\{YVU\} + C\{YEU, Y + EU\}$$

$$= E\{Y(1/12)\} + C\{Y(1/2), Y + (1/2)\}$$

$$= \frac{1}{12}EY + \frac{1}{2}VY = \frac{1}{12}\times 1 + \frac{1}{2}\times\frac{1}{2} = \frac{1}{3}.$$

(b) The cdf of X is

$$F(x) = P(X \le x) = P(U+Y \le x) = P(U \le x - Y) = EP(U \le x - Y \mid Y)$$

$$= P(Y=0)P(U \le x - 0 \mid Y=0) + P(Y=1)P(U \le x-1 \mid Y=1)$$

$$+P(Y=2)P(U \le x-2 \mid Y=2)$$

$$= \frac{1}{4}P(U \le x) + \frac{1}{2}P(U \le x-1) + \frac{1}{4}P(U \le x-2),\quad 0 < x < 3.$$

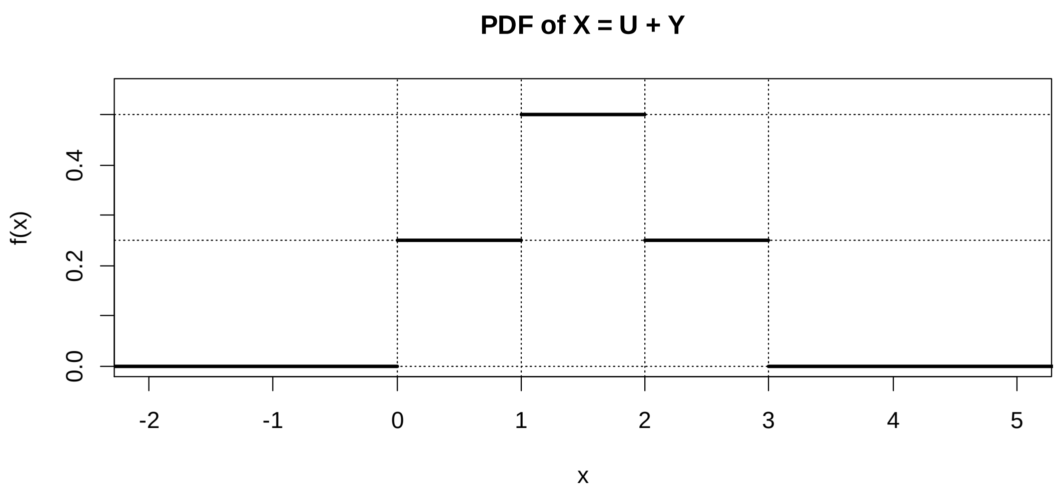
Considering three cases, we find that

$$F(x) = \begin{cases} \frac{1}{4}(x) + \frac{1}{2}(0) + \frac{1}{4}(0), & 0 < x < 1 \\ \frac{1}{4}(1) + \frac{1}{2}(x-1) + \frac{1}{4}(0), & 1 < x < 2 \\ \frac{1}{4}(1) + \frac{1}{2}(1) + \frac{1}{4}(x-2), & 2 < x < 3. \end{cases}$$

That is, $F(x) = \begin{cases} \frac{1}{4}x, & 0 < x < 1 \\ \frac{1}{4} + \frac{1}{2}(x-1), & 1 < x < 2 \\ \frac{3}{4} + \frac{1}{4}(x-2), & 2 < x < 3 \end{cases}$

Taking derivatives, we find that $\boxed{f(x) = \begin{cases} 1/4, & 0 < x < 1 \\ 1/2, & 1 < x < 2 \\ 1/4, & 2 < x < 3. \end{cases}}$

Note: We could also write $f(x) = \begin{cases} 1/4, & 0 < x < 1 \text{ and } 2 < x < 3 \\ 1/2, & 1 < x < 2 \\ 0, & x < 0 \text{ and } x > 3 \\ \text{undefined}, & x = 0, 1, 2, 3. \end{cases}$



(c) The cdf of W is

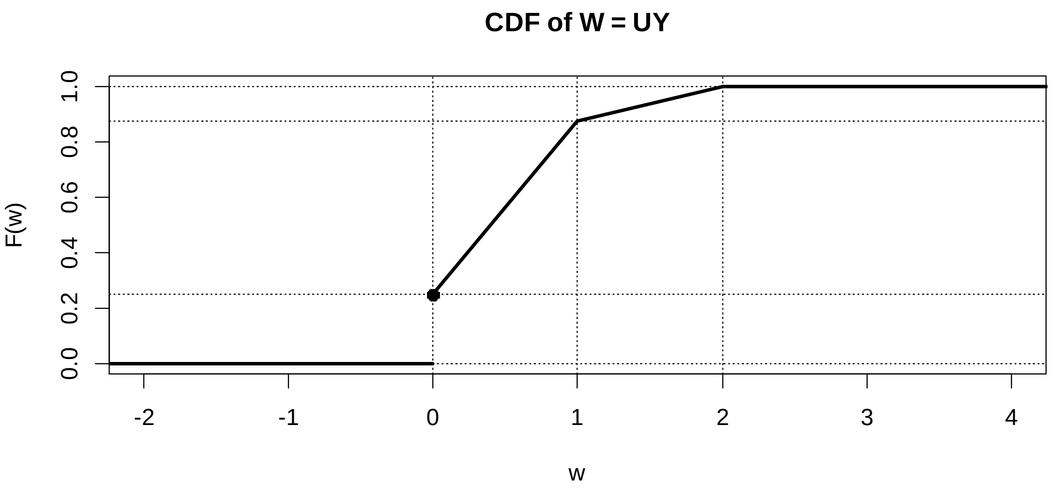
$$F(w) = P(W \le w) = P(UY \le w) = EP(UY \le x \mid Y)$$
$$= P(Y=0)P(U \times 0 \le w \mid Y=0) + P(Y=1)P(U \times 1 \le w \mid Y=1)$$
$$+P(Y=2)P(U \times 2 \le w \mid Y=2)$$
$$= \frac{1}{4}P(0 \le w) + \frac{1}{2}P(U \le w) + \frac{1}{4}P\left(U \le \frac{w}{2}\right), \quad 0 \le w < 2 .$$

Considering three cases, we find that

$$F(w) = \begin{cases} \frac{1}{4}(1) + \frac{1}{2}(0) + \frac{1}{4}(0), & w = 0 \\ \frac{1}{4}(1) + \frac{1}{2}(w) + \frac{1}{4}\left(\frac{w}{2}\right), & 0 < w < 1 \\ \frac{1}{4}(1) + \frac{1}{2}(1) + \frac{1}{4}\left(\frac{w}{2}\right), & 1 < w < 2. \end{cases}$$

Thus
$$\boxed{F(w) = \begin{cases} 0, & w < 0 \\ \frac{1}{4}, & w = 0 \\ \frac{1}{4} + \frac{5}{8}w, & 0 < w \le 1 \\ \frac{3}{4} + \frac{1}{8}w, & 1 < w \le 2 \\ 1, & w > 2. \end{cases}}$$

Note: This function is defined at every point on the real line.



Note: *W* has a *mixed* distribution which is continuous over (0,2) and discrete with probability 1/4 at 0 (thus, $P(W = 0) = 1/4$). We could write the density of *W* as

$$f(w) = \begin{cases} \frac{1}{4}, & w = 0 \quad \text{(discrete)} \\ \frac{5}{8}, & 0 < w < 1 \quad \text{(continuous)} \\ \frac{1}{8}, & 1 < w < 2 \quad \text{(continuous)} \\ 0, & w < 0 \text{ and } w > 2 \\ \text{undefined}, & w = 1, 2. \end{cases}$$

R Code (not required)

```
# (a)
X11(w=8,h=4)
plot(c(-2,5),c(0,0.55),type="n",xlab="x",ylab="f(x)",main="PDF of X = U + Y")
lines(c(0,1),c(1/4,1/4),lwd=3); lines(c(1,2),c(2/4,2/4),lwd=3)
lines(c(2,3),c(1/4,1/4),lwd=3); lines(c(-3,0),c(0,0),lwd=3); lines(c(3,6),c(0,0),lwd=3)
abline(h=c(0,1/4,1/2),lty=3); abline(v=c(0,1,2,3),lty=3)

# (b)
X11(w=8,h=4)
plot(c(-2,4),c(0,1),type="n",xlab="w",ylab="F(w)",main="CDF of W = UY")
points(0,1/4,pch=16,cex=1.2)
lines(c(0,1),c(1/4,7/8),lwd=3); lines(c(1,2),c(7/8,1),lwd=3)
lines(c(-3,0),c(0,0),lwd=3); lines(c(2,5),c(1,1),lwd=3)
abline(h=c(0,1/4,7/8,1),lty=3); abline(v=c(0,1,2),lty=3)
```

Solution to Problem 3

(a) $$\lambda = E\left(\frac{1}{Y+1}\right) = \sum_{y=0}^{n} \frac{1}{y+1} \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}$$

$$= \frac{1}{(n+1)p} \sum_{y=0}^{n} \frac{(n+1)!}{(y+1)!((n+1)-(y+1))!} p^{y+1} (1-p)^{(n+1)-(y+1)}$$

$$= \frac{1}{(n+1)p}\left\{ \sum_{x=0}^{m} \frac{m!}{x!(m-x)!} p^x (1-p)^{m-x} - \frac{m!}{0!(m-0)!} p^0 (1-p)^{m-0} \right\}$$

where $x = y + 1$ and $m = n + 1$

$$= \frac{1}{(n+1)p}\left\{1-(1-p)^{n+1}\right\} = \boxed{\frac{1-(1-p)^{n+1}}{(n+1)p}}.$$

If $n = 20$ and $p = 0.1$ then $\lambda = \dfrac{1-(1-0.1)^{20+1}}{(20+1)0.1} = \boxed{0.4241}$.

(b) Observe that

$$EY = P(Y=0)E(Y\,|\,Y=0) + P(Y=n)E(Y\,|\,Y=n) + P(0<Y<n)E(Y\,|\,0<Y<n).$$

Therefore,

$$\eta = E(Y\,|\,0<Y<n) = \frac{EY - P(Y=0)E(Y\,|\,Y=0) - P(Y=n)E(Y\,|\,Y=n)}{P(0<Y<n)}$$

$$= \frac{np-(1-p)^n \times 0 - p^n \times n}{1-(1-p)^n - p^n} = \boxed{\frac{np(1-p^{n-1})}{1-(1-p)^n-p^n}}.$$

If $n = 10$ and $p = 0.7$ then $\eta = \dfrac{10\times 0.7(1-0.7^{10-1})}{1-(1-0.7)^{10}-0.7^{10}} = \boxed{6.913}$.

R Code (not required)

```
# (a)
n=20; p=0.1; (1-(1-p)^(n+1))/((n+1)*p) # 0.4240862
yv=0:n; sum(dbinom(yv,n,p)/(1+yv))  # 0.4240862 (check)

# (b)
n=10; p=0.7; (n*p*(1-p^(n-1)))/(1-(1-p)^n-p^n) # 6.912836
yv=1:(n-1); dv=dbinom(yv,n,p); sum(yv*dv)/sum(dv) # 6.912836  (check)
```

Solution to Problem 4

(a) Let $I_i = I(A_i)$. Then $X = I_1 + ... + I_n$, where $I_1,...,I_n \sim iid\ Bernoulli(p)$.

So $VX = VI_1 + ... + VI_n = p(1-p) + ... + p(1-p) = \boxed{np(1-p)} = \boxed{0.588}$.

(b) As in (a), $X = I_1 + ... + I_n$, where $I_1,...,I_n \sim Bernoulli(p)$ (not iid), with

$$C(I_i, I_j) = E(I_i I_j) - (EI_i)EI_j = P(A_i A_j) - P(A_i)P(A_j) = 0 - p^2 \text{ for all } i \neq j.$$

So $$VX = \sum_{i=1}^{n} VI_i + \sum_{i \neq j} C(I_i, I_j) = np(1-p) + (n^2 - n)(-p^2) = \boxed{np(1-np)} = \boxed{0.24}.$$

Alternative solution

Observe that $X = 0$ or 1 in this case. Therefore $X \sim Bernoulli(q)$, where

$$q = P(X = 1) = EX = EI_1 + ... + EI_n = p + ... + p = np.$$

It follows that $VX = q(1-q) = np(1-np)$, as before.

(c) In this case, observe that:

$$P(A_i) = P(A_i A_i) = \frac{p \times 1}{2-1} = p \quad \text{(consistent with line 1 of the problem)}$$

$$P(A_i A_{i+1}) = \frac{p \times 1}{2-0} = \frac{p}{2}$$

$$P(A_i A_{i+2}) = \frac{p \times 0}{2-0} = 0$$

$$P(A_i A_{i+3}) = \frac{p \times 0}{2-0} = 0, \text{ etc.}$$

So: $VI_i = p(1-p)$ (as previously)

$$C(I_i, I_{i+1}) = E(I_i I_{i+1}) - (EI_i)EI_{i+1} = P(A_i A_{i+1}) - P(A_i)P(A_{i+1}) = \frac{p}{2} - p^2$$

$$C(I_i, I_{i+2}) = E(I_i I_{i+2}) - (EI_i)EI_{i+2} = P(A_i A_{i+2}) - P(A_i)P(A_{i+2}) = 0 - p^2$$

$$C(I_i, I_{i+3}) = E(I_i I_{i+3}) - (EI_i)EI_{i+3} = P(A_i A_{i+3}) - P(A_i)P(A_{i+3}) = 0 - p^2, \text{ etc.}$$

Therefore

$$VX = \sum_{i=1}^{n} VI_i + \sum_{|i-j|=1} C(I_i, I_j) + \sum_{|i-j|>1} C(I_i, I_j)$$

$$= np(1-p) + 2(n-1)\left(\frac{p}{2} - p^2\right) + (n^2 - n - 2(n-1))(-p^2)$$

$$= \boxed{\{(2-np)n-1\}p} = \boxed{0.82}.$$

Alternative solution

Observe that:

$X = 0$ if $\bar{A}_1 \cdots \bar{A}_n$

$X = 2$ if $(A_1A_2) \cup (A_2A_3) \cup \cdots \cup (A_{n-1}A_n)$

$X = 1$ otherwise.

It follows that

$$P(X=0) = P\left(\bar{A}_1 \cdots \bar{A}_n\right) = P\left(\overline{A_1 \cup \cdots \cup A_n}\right) \text{ by de Morgan's laws}$$

$$= 1 - P\left(A_1 \cup \cdots \cup A_n\right)$$

$$= 1 - \{[P(A_1) + ... + P(A_n)] - [P(A_1A_2) + ... + P(A_{n-1}A_n)]$$

$$+ [P(A_1A_2A_3) + ...] - [P(A_1A_2A_3A_4) + ...] + ...\}$$

$$= 1 - \left\{[p + ... + p] - \left[\frac{p}{2} + ... + \frac{p}{2}\right] + 0 - 0 + ...\right\}$$

$$= 1 - np + (n-1)\frac{p}{2}.$$

Note: By the law of total probability,

$$P(A_1A_3) = P(A_1A_2A_3) + P(A_1\bar{A}_2A_3),$$

where $P(A_1A_3) = 0$, $P(A_1A_2A_3) \geq 0$ and $P(A_1\bar{A}_2A_3) \geq 0$.

Hence $P(A_1A_2A_3) = P(A_1\bar{A}_2A_3) = 0$.

Similarly, $P(A_iA_jA_k) = 0$ for all $i < j < k$.

Likewise, $P(A_iA_jA_kA_l) = 0$ for all $i < j < k < l$, etc.

Also, $P(X=2)=P\{(A_1A_2)\cup(A_2A_3)\cup\cdots\cup(A_{n-1}A_n)\}$

$$=P(A_1A_2)+P(A_2A_3)+...P(A_{n-1}A_n) \quad \text{(since } P((A_1A_2)(A_2A_3))=0\text{, etc.)}$$

$$=\frac{p}{2}+...+\frac{p}{2} \qquad (n-1 \text{ terms})$$

$$=(n-1)\frac{p}{2}.$$

Then also, $P(X=1)=1-P(X=0)-P(X=2)$

$$=1-\left\{1-np+(n-1)\frac{p}{2}\right\}-(n-1)\frac{p}{2}$$

$$=np-(n-1)p=p.$$

In summary so far,

$$f(x)=\begin{cases} p_0 \equiv 1-np+(n-1)p/2, & x=0 \\ p_1 \equiv p, & x=1 \\ p_2 \equiv (n-1)p/2, & x=2. \end{cases}$$

It follows that:

$$EX=0p_0+1p_1+2p_2=0+1\times p+2\times(n-1)\frac{p}{2}=np$$

$$EX^2=0^2p_0+1^2p_1+2^2p_2=0+1\times p+4\times(n-1)\frac{p}{2}=(2n-1)p$$

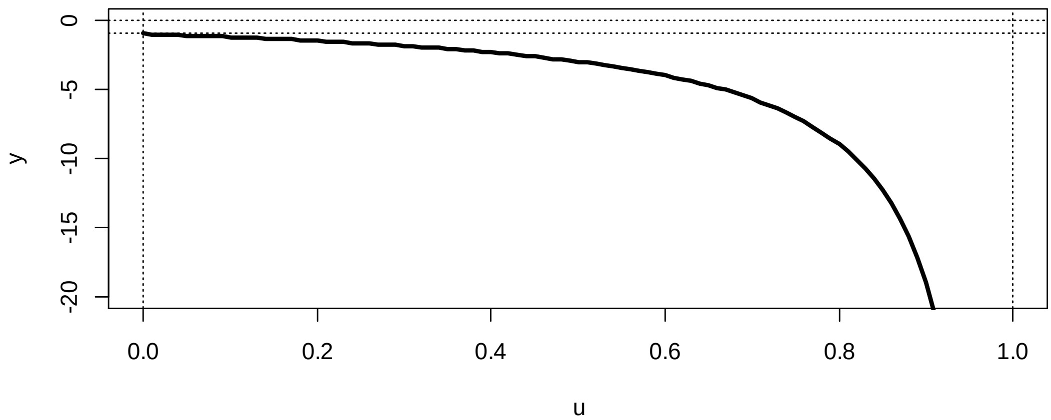
$$VX=EX^2-(EX)^2=(2n-1)p-(np)^2=\{(2-np)n-1\}p, \text{ as before.}$$

R Code (not required)

```
n=30; p = 0.02
n*p*(1-p) # 0.588
n*p*(1-n*p) # 0.24
((2-n*p)*n-1)*p # 0.82
n*p*(1-p)+2*(n-1)*(p/2-p^2)+(n^2-n-2*(n-1))*(-p^2)  # 0.82  (check)
```

Solution to Problem 5

(a) The function $y = \dfrac{u+1}{u-1}$ is depicted in the next figure.



We see that the inverse of this function is defined for $y < -1$ and given by

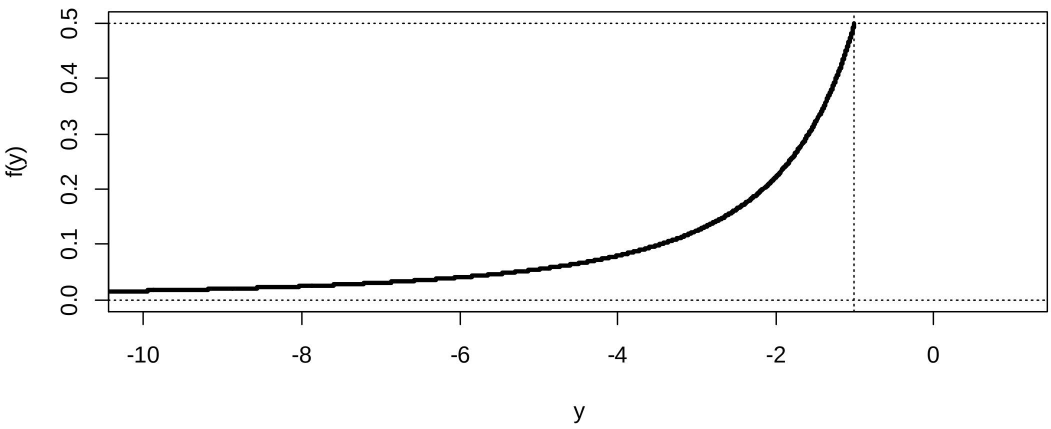
$$yu - y = u + 1 \;\Rightarrow\; u(y-1) = y+1 \;\Rightarrow\; u = \frac{y+1}{y-1}.$$

Then, $F(y) = P(Y \le y) = P\left(\dfrac{y+1}{y-1} < U < 1\right) = 1 - \dfrac{y+1}{y-1} = \dfrac{2}{1-y} = 2(1-y)^{-1}, \quad y < -1$.

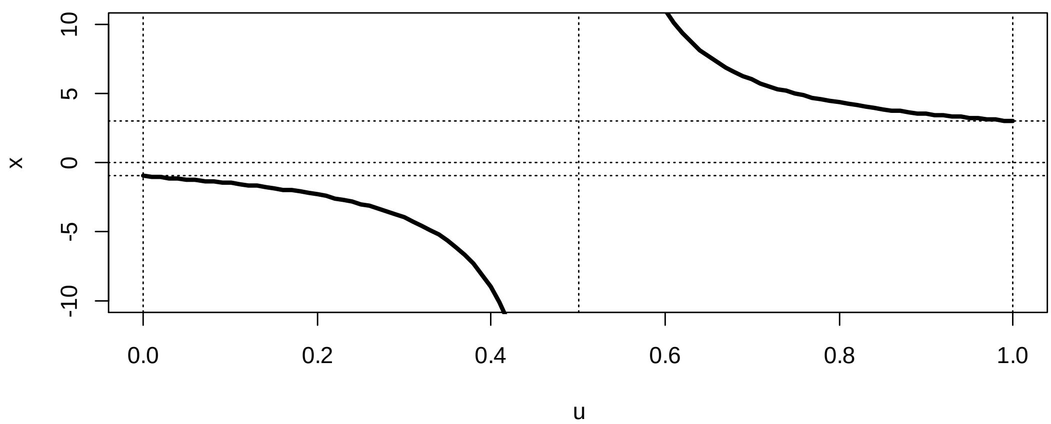
So $f(y) = F'(y) = 2(-1)(1-y)^{-2}(-1)$. That is, $\boxed{f(y) = \dfrac{2}{(1-y)^2}, \quad y < -1}$.

Alternatively, as $y = \dfrac{u+1}{u-1}$ is strictly decreasing, we may use the transformation rule:

$$f(y) = f(u)\left|\frac{du}{dy}\right| = 1 \times \left|(y+1)(-1)(y-1)^{-2} + (1)(y-1)^{-1}\right| = \frac{2}{(1-y)^2}, \quad y < -1.$$



(b) The function $x = \dfrac{u+1/2}{u-1/2}$ is depicted in the next figure.



We see that the inverse of this function is defined for $x < -1$ and $x > 3$ and given by

$$xu - x/2 = u + 1/2 \;\Rightarrow\; u(x-1) = (x+1)/2 \;\Rightarrow\; u = \frac{(x+1)/2}{x-1}.$$

So for $x < -1$:

$$F(x) = P(X \le x) = P\left(\frac{(x+1)/2}{x-1} < U < \frac{1}{2}\right) = \frac{1}{2} - \frac{(x+1)/2}{x-1} = \frac{1}{1-x}.$$

Also, for $x > 3$:

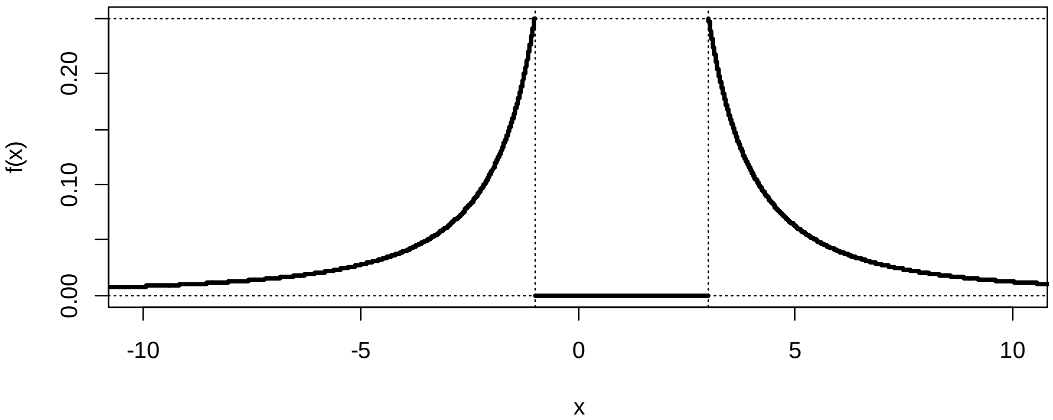
$$F(x) = P(X \le x) = \frac{1}{2} + P\left(\frac{(x+1)/2}{x-1} < U < 1\right) = \frac{1}{2} + 1 - \frac{(x+1)/2}{x-1} = \frac{x-2}{x-1}.$$

And for $-1 < x < 3$: $F(x) = P(X \le x) = P(U < 1/2) = 1/2$.

In summary so far, $F(x) = \begin{cases} \dfrac{1}{1-x} = (1-x)^{-1}, & x < -1 \\ 1/2, & -1 \le x \le 3 \\ \dfrac{x-2}{x-1} = (x-2)(x-1)^{-1}, & x > 3. \end{cases}$

So $\quad f(x) = F'(x) = \begin{cases} -(1-x)^{-2}(-1) = (1-x)^{-2}, & x < -1 \\ 0, & -1 < x < 3 \\ (x-2)(-1)(x-1)^{-2} + (1)(x-1)^{-1} = (x-1)^{-2}, & x > 3. \end{cases}$

Thus, $\boxed{f(x) = \begin{cases} 1/(x-1)^2, & x < -1 \text{ and } x > 3 \\ 0, & -1 < x < 3. \end{cases}}$



R Code (not required)

```
# (a)

X11(w=8,h=4)
plot(c(0,1),c(-20,0),type="n",xlab="u",ylab="y")
uvec=seq(0,1,0.01); yvec=(uvec+1)/(uvec-1)
lines(uvec,yvec,lwd=3)
abline(h=c(-1,0),lty=3); abline(v=c(0,1),lty=3)

yvec=seq(-15,-1,0.001); fvec=2/(1-yvec)^2
plot(c(-10,1),c(0,0.5),type="n", xlab="y",ylab="f(y)")
lines(yvec,fvec,lwd=3)
abline(h=c(0,0.5),lty=3); abline(v=-1,lty=3)

# (b)
X11(w=8,h=4)
plot(c(0,1),c(-10,10),type="n",xlab="u",ylab="x")
uvec=seq(0,0.49,0.01); xvec=(uvec+1/2)/(uvec-1/2)
lines(uvec,xvec,lwd=3)
uvec=seq(0.51,1,0.01); xvec=(uvec+1/2)/(uvec-1/2)
lines(uvec,xvec,lwd=3)
abline(h=c(-1,0,3),lty=3); abline(v=c(0,0.5,1),lty=3)

plot(c(-10,10),c(0,0.25),type="n", xlab="x",ylab="f(x)")
lines(seq(-15,-1,0.001), 1/( seq(-15,-1,0.001) - 1)^2, lwd=3)
lines(seq(3,15,0.001), 1/( seq(3,15,0.001) - 1)^2, lwd=3)
lines(c(-1,3),c(0,0),lwd=3)
abline(h=c(0,0.25),lty=3); abline(v=c(-1,3),lty=3)
```

Solution to Problem 6

(a) Let Y_i be the number coming up on the *i*th roll (i = 1,2,3,...) and let n be the number of rolls. Then, by the central limit theorem, $Y_T \equiv Y_1 + ... + Y_n \sim N(n\mu, n\sigma^2)$,

where: $\mu = EY_i = \frac{1}{6}(1 + ... + 6) = 3.5$, $\mu_2' = EY_i^2 = \frac{1}{6}(1^2 + ... + 6^2) = \frac{91}{6}$ = 15.166667

$$\sigma^2 = \mu_2' - \mu^2 = \frac{35}{12} = 2.916667, \qquad \sigma = \sqrt{\frac{35}{12}} = 1.707825.$$

We wish to solve $p = 0.975$ for n (with a view to rounding at the end), where

$$p = P(Y_T > c) \approx P\left(Z > \frac{c - n\mu}{\sigma\sqrt{n}}\right), \quad Z \sim N(0,1) \quad \text{and } c = 5000.$$

Now $0.975 = P(Z > -z)$, where $z = z_{0.025} = 1.96$. So we equate $\frac{c - n\mu}{\sigma\sqrt{n}} = -z$.

This equation may also be written as $c - t^2\mu = -z\sigma t$, where $t = \sqrt{n}$,

or equivalently $\mu t^2 - z\sigma t - c = 0$. Solving this quadratic equation we get

$$t = \frac{z\sigma + \sqrt{z^2\sigma^2 + 4\mu c}}{2\mu} \qquad \text{(this is the only positive solution)}$$

$$= 38.27766,$$

and thereby $n = t^2$ = 1465.18, which after rounding to the nearest integer is $\boxed{1465}$.

Note 1: If we repeat this logic using a suitable continuity correction (+ 0.5), we get

$$t = \frac{z\sigma + \sqrt{z^2\sigma^2 + 4\mu(c + 0.5)}}{2\mu} = 38.27955,$$

and thereby $n = t^2$ = 1465.32 (i.e. exactly the same answer after rounding).

Note 2: If the question had stated "at least 5000" (rather than "greater than 5000"), then the appropriate continuity correction above would have been -0.5 (not + 0.5).

Note 3: The answer n = 1465 is checked via Monte Carlo in the R code below. It is shown that this value of n implies a value of p slightly *smaller* than 0.975. On the other hand, using n = 1466 (obtained by *rounding up*, rather than *rounding*) implies a value of p slightly *larger* than 0.975, but one which is also slightly further away from 0.975. These calculations are for interest only and not required as part of the solution. Both 1465 and 1466 are acceptable solutions.

Note 4: The following is another approach to solving $\dfrac{c-n\mu}{\sigma\sqrt{n}}=-z$ for n.

This equation implies $c-n\mu=-z\sigma\sqrt{n}$

$$\Rightarrow \quad c^2-2cn\mu+n^2\mu^2=z^2\sigma^2 n$$

$$\Rightarrow \quad n^2\mu^2-n(2c\mu+z^2\sigma^2)+c^2=0$$

$$\Rightarrow n=\frac{2c\mu+z^2\sigma^2\pm\sqrt{4c^2\mu^2+4c\mu z^2\sigma^2+z^4\sigma^4-4\mu^2c^2}}{2\mu^2}$$

$$=\frac{2c\mu+z^2\sigma^2\pm z\sigma\sqrt{4c\mu+z^2\sigma^2}}{2\mu^2} \quad = \; 1392.88 \; \text{ or } \; 1465.18.$$

Now $\dfrac{c-1392.88\mu}{\sigma\sqrt{1392.88}}=1.96=+z$ (wrong) and $\dfrac{c-1465.18\mu}{\sigma\sqrt{1465.18}}=-1.96=-z$ (as required).

So the answer is again 1465.

(b) We wish to find the smallest integer n such that $P(Y_T>c)\geq 0.975$,

or equivalently, such that: $1-P(Y_T\leq c)\geq 0.975 \quad \Leftrightarrow \quad P(Y_T\leq c)\leq 0.025$

$$\Leftrightarrow \quad \frac{1}{2}P\left(\left|Y_T-n\mu\right|\geq n\mu-c\right)\leq 0.025$$

$$\Leftrightarrow \quad P\left(\left|Y_T-n\mu\right|\geq n\mu-c\right)\leq 0.05\,.$$

To clarify this argument, note that the distribution of Y_T is discrete and symmetric about $n\mu$. To illustrate further, draw a symmetric bell-shaped discrete probability density function centred at $n\mu$ (the mean of Y_T), mark a point c to the left of $n\mu$, and consider that we require the sum of probabilities at c and to the left to be 0.025 or less.

Now, Chebyshev's theorem implies that $P\left(|Y_T - n\mu| \geq k\sigma\sqrt{n}\right) \leq \frac{1}{k^2}$.

Therefore we equate $0.05 = 1/k^2$ and $n\mu - c = k\sigma\sqrt{n}$.

This yields $k = \sqrt{20}$ = 4.472136 and

$$t^2\mu - c = k\sigma t \quad \text{(where } t = \sqrt{n}\text{, as in (a))}$$

$$\Leftrightarrow \quad \mu t^2 - k\sigma t - c = 0$$

$$\Leftrightarrow \quad t = \frac{k\sigma + \sqrt{k^2\sigma^2 + 4\mu c}}{2\mu} \quad \text{(this is the only positive solution)}$$

= 38.90328.

We thereby obtain n = 1513.467, which rounded up gives the answer, n = $\boxed{1514}$.

As a check, we may reverse the above logic and write:

$$P(Y_T > c) = 1 - P(Y_T \leq c)$$

$$= 1 - \frac{1}{2} P\left(|Y_T - n\mu| \geq n\mu - c\right) \quad \text{(by symmetry)}$$

$$\geq 1 - \frac{1}{2} \times \frac{1}{k^2}, \quad \text{where } n\mu - c = k\sigma\sqrt{n} \quad \text{(by Chebyshev's theroem).}$$

With n = 1513, this implies that k = 4.448306 and

$$P(Y_T > c) \geq 1 - \frac{1}{2} \times \frac{1}{k^2} = 0.9747314 \quad \text{(which is less than the required 0.975).}$$

However, with n = 1514, this implies that k = 4.499507 and

$$P(Y_T > c) \geq 1 - \frac{1}{2} \times \frac{1}{k^2} = 0.9753032 \quad \text{(which is at least the required 0.975).}$$

R Code (not required)

```
# (a)
mu=3.5; mu2p=mean((1:6)^2); sig2=mu2p-mu^2; sig=sqrt(sig2)
c(mu2p,sig2,sig) # 15.166667  2.916667  1.707825
```

```
z=1.96; c = 5000; t=(z*sig+sqrt(z^2*sig2+4*mu*c))/(2*mu); n=t^2
c(t,n) # 38.27766 1465.17950
z=1.96; t=(z*sig+sqrt(z^2*sig2+4*mu*(c+0.5)))/(2*mu); n=t^2
c(t,n) # 38.27955 1465.32416    Solution using the continuity correction +0.5

# Check via Monte Carlo (Note 3)
n = 1465; set.seed(532)
tot = sum(    sample(1:6,n,rep=T)     ); tot # 5197       (simulation of 1465 die rolls)
ct=0; J=100000; for(j in 1:J){             tot = sum(    sample(1:6,n,rep=T)   )
                                           if(tot>5000) ct=ct+1    }
phat=ct/J; ci=phat+c(-1,1)*qnorm(0.975)*sqrt(phat*(1-phat)/J); c(phat,ci)
# 0.9736800 0.9726878 0.9746722    (p  is slightly smaller than 0.975)                 (1)

n = 1466; set.seed(142)   # Test n as obtained by rounding up (rather than rounding)
ct=0; J=100000; for(j in 1:J){           tot = sum(    sample(1:6,n,rep=T)   )
                                           if(tot>5000) ct=ct+1    }
phat=ct/J; ci=phat+c(-1,1)*qnorm(0.975)*sqrt(phat*(1-phat)/J); c(phat,ci)
# 0.9770800 0.9761525 0.9780075    (p is slightly larger than 0.975,
#                                  p is also slightly more distant from 0.975 than p at (1) above

# Alternative approach (Note 4)
nvals = (0.5/mu^2) * ( 2*c*mu+sig2*z^2+c(-1,1)*z*sig*sqrt(4*c*mu+sig2*z^2)  )
nvals  # 1392.878 1465.179
(c-nvals*mu)/(sig*sqrt(nvals))  # 1.96 -1.96

# (b) Continues on from (a)
k=sqrt(20); t = (k*sig+sqrt(k^2*sig2+4*mu*c))/(2*mu); n=t^2
c(t,n) # 38.90328 1513.46535

# Check:
k=(1513*3.5-5000)/(1.707825*sqrt(1513)); k # 4.448306
        1-0.5/k^2 # 0.9747314
k=(1514*3.5-5000)/(1.707825*sqrt(1514)); k # 4.499507
        1-0.5/k^2 # 0.9753032
```

Solution to Problem 7

(a) Here: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{4}(0+0+2+2) = 1$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{4}(0+1+0+3) = 1$

$\sum_{i=1}^{n} x_i^2 = 0^2 + 0^2 + 2^2 + 2^2 = 8$, $\sum_{i=1}^{n} y_i^2 = 0^2 + 1^2 + 0^2 + 3^2 = 10$

$\sum_{i=1}^{n} x_i y_i = 0+0+0+2\times 3 = 6$

$S_{xx} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = 4$, $S_{yy} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = 6$

$S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} = 2$, $SSE = S_{yy} - \frac{S_{xy}^2}{S_{xx}} = 5.$

So the required point estimates of the intercept α, slope β and normal variance σ^2 are: $b = \frac{S_{xy}}{S_{xx}} = \boxed{0.5}$, $a = \bar{y} - b\bar{x} = \boxed{0.5}$ and $s^2 = \frac{SSE}{n-2} = \boxed{2.5}$ (respectively).

(b) We may write the average of interest as

$$q = \frac{1}{2}\{(\alpha + 1\beta + w_1) + (\alpha + 2\beta + w_2)\} = \alpha + 1.5\beta + \left(\frac{w_1 + w_2}{2}\right),$$

where $w_1, w_2 \sim iid\ N(0, \sigma^2)$, independently of the sample errors, $e_1, ..., e_n$.

A point predictor of q is $\hat{q} = a + 1.5b = \boxed{1.25}$.

Next, the error of prediction has variance

$$V(\hat{q} - q) = V\left\{[a + 1.5b] - \left[\alpha + 1.5\beta + \left(\frac{w_1 + w_2}{2}\right)\right]\right\}$$

$$= V\left\{a + 1.5b - \left(\frac{w_1 + w_2}{2}\right)\right\}$$

$$= Va + 1.5^2 Vb + 2\times 1.5 C(a,b) + \frac{\sigma^2}{2}$$

$$= \left(\frac{\sigma^2 \sum x_i^2}{nS_{xx}} \right) + 2.25 \left(\frac{\sigma^2}{S_{xx}} \right) + 3 \left(\frac{-\bar{x}\sigma^2}{S_{xx}} \right) + \frac{\sigma^2}{2}$$

$$= \sigma^2 \left\{ \frac{1}{2} + \frac{1}{S_{xx}} \left[\frac{\sum x_i^2}{n} + 2.25 - 3\bar{x} \right] \right\}.$$

This may be estimated by

$$\hat{V}(\hat{q} - q) = s^2 \left\{ \frac{1}{2} + \frac{1}{S_{xx}} \left[\frac{\sum x_i^2}{n} + 2.25 - 3\bar{x} \right] \right\} = 2.031.$$

It follows that a 95% PI for q is

$$\left(\hat{q} \pm t_{\tau/2}(n-2)\sqrt{\hat{V}(\hat{q} - q)} \right) = \left(1.25 \pm 4.303\sqrt{2.031} \right) = \boxed{(-4.882, 7.382)}.$$

R Code (not required)

```
# (a)
options(digits=4); x = c(0,0,2,2); y = c(0,1,0,3)
xbar=mean(x); sumx2=sum(x^2);   ybar=mean(y); sumy2=sum(y^2)
sumxy = sum(x*y);  n=length(y); Sxy=sumxy-n*xbar*ybar
Sxx=sumx2-n*xbar^2; Syy=sumy2-n*ybar^2
c(xbar,ybar,sumx2,sumy2,sumxy, Sxx,Syy,Sxy) # 1  1  8 10  6  4  6  2
SSE=Syy-Sxy^2 / Sxx; SSE # 5
b=Sxy/Sxx; a=ybar-xbar*b; s2=SSE/(n-2)
c(b,a,s2) #  0.5 0.5 2.5

# (b)
qhat=a + 1.5*b
Vhat=s2   *   (   0.5 + (1/Sxx) * ( (1/n)*sumx2 + 2.25 - 3*xbar )   )
PI = qhat + c(-1,1)*qt(0.975,n-2)*sqrt(Vhat)
c(qhat,Vhat, qt(0.975,n-2), PI) # 1.250  2.031  4.303 -4.882  7.382
```