STAT2001 and STAT6039 Final Exam – 1st Sem. 2015 – Solutions

(Note: The STAT2001 exam is identical to the STAT6039 exam except that it does not have Problem 8.)

Solution to Problem 1

(a) Let A be the event that Ann wins. Let 0 denote a number which is not 5 or 6, etc.

Then $\quad P(A) = P(0)P(A\,|\,0) + P(5)P(A\,|\,5) + P(6)P(A\,|\,6)$

$$= \frac{4}{6}(1 - P(A)) + \frac{1}{6}P(A\,|\,5) + \frac{1}{6}\times 1. \qquad (1)$$

Next, $\quad P(A\,|\,5) = P(50\,|\,0)P(A\,|\,50) + P(55)P(A\,|\,55) + P(56)P(A\,|\,56)$

$$= \frac{4}{6}P(A) + \frac{1}{6}P(A\,|\,55) + \frac{1}{6}\times 0. \qquad (2)$$

Also, $\quad P(A\,|\,55) = P(550\,|\,0)P(A\,|\,550) + P(555)P(A\,|\,555) + P(556)P(A\,|\,556)$

$$= \frac{4}{6}(1 - P(A)) + \frac{1}{6}\times 1 + \frac{1}{6}\times 1. \qquad (3)$$

With $a = P(A)$, $b = P(A\,|\,5)$ and $c = P(A\,|\,55)$, equations (1), (2) and (3) imply:

$$6a = 4 - 4a + b + 1, \quad 6b = 4a + c + 0, \quad 6c = 4 - 4a + 1 + 1.$$

Solving these equations yields the required probability, $a = P(A)$ = 93/170 = **<u>0.5471</u>**.

(Also, $b = P(A\,|\,5)$ = 24/51 = 0.4706 and $c = P(A\,|\,55)$ = 54/85 = 0.6353.)

(b) Let Y be the number of rolls. Then

$EY = P(0)E(Y\,|\,0) + P(5)E(Y\,|\,5) + P(6)E(Y\,|\,6)$

$$= \frac{4}{6}(1 + EY) + \frac{1}{6}E(Y\,|\,5) + \frac{1}{6}\times 1. \qquad (1)$$

Next, $\quad E(Y\,|\,5) = P(50\,|\,0)E(Y\,|\,50) + P(55)E(Y\,|\,55) + P(56)E(Y\,|\,56)$

$$= \frac{4}{6}(2 + EY) + \frac{1}{6}E(Y\,|\,55) + \frac{1}{6}\times 2. \qquad (2)$$

Also, $\quad E(Y\,|\,55) = P(550\,|\,0)E(Y\,|\,550) + P(555)E(Y\,|\,555) + P(556)E(Y\,|\,556)$

$$= \frac{4}{6}(3 + EY) + \frac{1}{6}\times 3 + \frac{1}{6}\times 3. \qquad (3)$$

With $a = EY$, $b = E(Y\,|\,5)$ and $c = E(Y\,|\,55)$, equations (1), (2) and (3) imply:

$$6a = 4 + 4a + b + 1, \quad 6b = 8 + 4a + c + 2, \quad 6c = 12 + 4a + 3 + 3.$$

Solving these equations yields the required expectation, $a = EY$ = 129/22 = **<u>5.864</u>**.

(Also, $b = E(Y\,|\,5)$ = 74/11 = 6.727 and $c = E(Y\,|\,55)$ = 76/11 = 6.909.)

Solution to Problem 2

(a) The distribution is exponential with mean λ but shifted to the right by λ. So its mean is $\lambda + \lambda = 2\lambda$. So we equate $2\lambda = \bar{y}$ and thereby obtain the MME,

$$\hat{\lambda} = \bar{y}/2 = 3/2 = \underline{\mathbf{1.5}}.$$

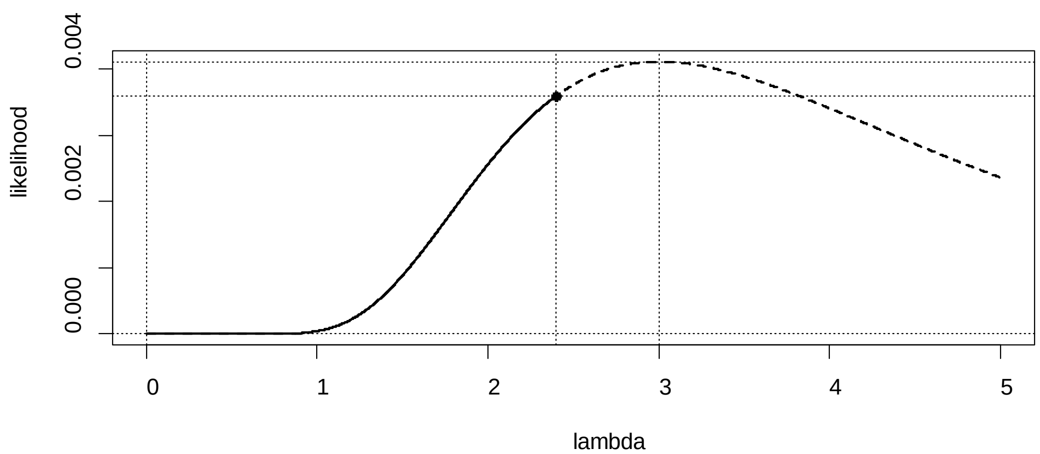
(b) The likelihood function is $L(\lambda) = \prod_{i=1}^{n} \frac{1}{\lambda} e^{-\frac{1}{\lambda}(y-\lambda)} = \lambda^{-n} e^{-\frac{1}{\lambda}(n\bar{y}-n\lambda)}$.

So the log-likelihood function is $l(\lambda) = -n\log\lambda - \frac{1}{\lambda}(n\bar{y} - n\lambda)$.

So $l'(\lambda) = -\frac{n}{\lambda} + \frac{n\bar{y}}{\lambda^2} = \frac{n}{\lambda^2}(\bar{y} - \lambda)$. Equating this to zero yields $\bar{y} = 3$. But this is not the MLE. Why? Because $\lambda \le y_i$ for all i = 1,...,n, and consequently, $\lambda \le m$, where $m = \min(y_1, ..., y_n)$. It is apparent that, since the 'extended' log-likelihood function is strictly increasing for all λ in the range (0, $\bar{y}$), the MLE is in fact $\hat{\lambda} = m$ = **<u>2.4</u>**.

Discussion

The following figure (not required) shows the likelihood, which is defined only over (0, 2.4). The dashed line shows the likelihood function extended beyond this range. The dot shows the maximum value of the likelihood functions at 2.4. This value is 0.003598. The maximum value of the 'extended' likelihood function is 0.004115, at 3.



(c) The bias of $\bar{y}$ is $B(\bar{y}) = E\bar{y} - \lambda = 2\lambda - \lambda = \lambda$. Also, $V\bar{y} = V(\bar{y} - \lambda) = \lambda^2/n$.

So $MSE(\bar{y}) = V\bar{y} + (B(\bar{y}))^2 = (\lambda^2/n) + \lambda^2 = \lambda^2(n+1)/n = 0.6^2 \times 6/5$ = **<u>0.432</u>**.

(d) The cdf of the MLE, $M = \min(Y_1,...,Y_n)$, is

$$F(m) = P(M \le m) = 1 - P(M > m) = 1 - P(Y_1 > m,...,Y_n > m)$$

$$= 1 - P(Y_1 > m)^n = 1 - e^{-n\left(\frac{m-\lambda}{\lambda}\right)},\ m \ge \lambda .$$

We see that the pdf of *M* is $f(m) = F'(m) = 0 - e^{-n\left(\frac{m-\lambda}{\lambda}\right)}\left(-\frac{n}{\lambda}\right) = \frac{n}{\lambda}e^{-\frac{n}{\lambda}(m-\lambda)}, m \ge \lambda$.

So *M* has an exponential distribution with mean λ / n but shifted to the right by λ.

So $EM = \frac{\lambda}{n} + \lambda = \frac{n+1}{n}\lambda$. It follows that an unbiased estimate of λ is

$$\hat{\lambda} = \frac{n}{n+1}m = \frac{5}{5+1} \times 2.4 = \underline{\mathbf{2}}.$$

(e) The variance of the MME in (a) is $V\left(\frac{\overline{y}}{2}\right) = \frac{1}{2^2} \times \frac{VY_i}{n} = \frac{1}{2^2} \times \frac{\lambda^2}{n} = \frac{\lambda^2}{4n}$.

The variance of the modified MLE in (d) is $V\left(\frac{n}{n+1}M\right) = \left(\frac{n}{n+1}\right)^2\left(\frac{\lambda}{n}\right)^2 = \frac{\lambda^2}{(n+1)^2}$.

So the required efficiency is $\frac{\lambda^2}{4n}\Big/\frac{\lambda^2}{(n+1)^2} = \frac{(n+1)^2}{4n} = \frac{(5+1)^2}{4 \times 5} = \frac{36}{20} = \underline{\mathbf{1.8}}$.

R Code for Problem 2 (not required)

```
# (b)
y=c(3.1, 2.8, 2.4, 3.0, 3.7); n=length(y); m=min(y); ybar=mean(y)
c(n,m,ybar) # 5.0 2.4 3.0
Lfun=function(lam,n,ybar){ lam^(-n)*exp(-(1/lam)*(n*ybar-n*lam)) }
lamvec=seq(0.001,5,0.001); Lvec=Lfun(lamvec,n=n,ybar=ybar); X11(w=8,h=4);

plot(lamvec,Lvec,type="l",xlab="lambda",ylab="likelihood",lwd=2,lty=2);
lines(lamvec[lamvec<=m],Lvec[lamvec<=m],lty=1,lwd=2)
maxL= Lfun(lam=m,n=n,ybar=ybar); maxL # 0.00359812
maxLextended=max(Lvec); maxLextended # 0.004115226
abline(v=c(0,m,ybar),lty=3); abline(h=c(0,maxL,maxLextended),lty=3)
points(m,maxL,pch=16)
```

Solution to Problem 3

(a) Equating y to $EY = k/(k+1)$ we get the MME, $\hat{k}_{MM} = \dfrac{y}{1-y}$, $0 < y < 1$.

So, if y = 0.7, then the MME is $\hat{k}_{MM} = \dfrac{0.7}{1-0.7}$ = **<u>2.333</u>**.

The log-likelihood is $l(k) = \log k + (k-1)\log y$. So $l'(k) = \dfrac{1}{k} + \log y$.

Equating this to zero, we get the MLE, $\hat{k}_{ML} = -1/\log y$.

So, if y = 0.7, then the MLE is $\hat{k}_{ML} = -1/\log 0.7$ = **<u>2.804</u>**.

(b) Here, $f(y) = ky^{k-1}, 0 < y < 1$, $x = 1/y^k$ is strictly decreasing, and $y = x^{-1/k}$.

So $f(x) = f(y)\left|\dfrac{dy}{dx}\right| = \cancel{k}(x^{-1/k})^{k-1}\left|-\dfrac{1}{\cancel{k}}x^{-\frac{1}{k}-1}\right| = \dfrac{1}{x^2}$, $x > 1$, with $f(x=5) = 1/5^2$ = **<u>0.04</u>**.

(Note that k does not feature in the density of X, and so k = 3.8 is irrelevant.)

(c) We will use X as a pivot. First, $F(x) = \int_1^x t^{-2}dx = \left[\dfrac{t^{-1}}{-1}\Bigg|_{t=1}^{x}\right] = \dfrac{x^{-1}}{-1} - \dfrac{1^{-1}}{-1} = 1 - \dfrac{1}{x}$, $x > 1$.

Setting this to p, we obtain the quantile function of X, $F_X^{-1}(p) = \dfrac{1}{1-p}$, $0 < p < 1$.

Then, for any $\alpha \in (0,1)$ and $a \in [0, \alpha]$ (e.g. $a = \alpha/2$), it is true for all k that

$$\begin{aligned}
1-\alpha &= P(F_X^{-1}(a) < X < F_X^{-1}(a+1-\alpha)) \\
&= P\left(\frac{1}{1-a} < \frac{1}{Y^k} < \frac{1}{1-(a+1-\alpha)}\right) \\
&= P\left(-\log(1-a) < -k\log Y < -\log(\alpha-a)\right) \\
&= P\left(\frac{\log(1-a)}{\log Y} < k < \frac{\log(\alpha-a)}{\log Y}\right) \quad \text{(note that } \log Y \text{ is negative).}
\end{aligned}$$

So a $1-\alpha$ CI for k is given by $\left(\dfrac{\log(1-a)}{\log y}, \dfrac{\log(\alpha-a)}{\log y}\right)$.

Using y = 0.7, $\alpha = 0.05$ and $a = \alpha/2 = 0.025$, we obtain the central 95% CI

$$\left(\frac{\log(1-0.025)}{\log 0.7}, \frac{\log(0.05-0.025)}{\log 0.7}\right)$$ = **<u>(0.07098 10.34)</u>**.

Discussion

Using $a = 0$, we obtain the one-sided 95% CI

$$\left(\frac{\log(1-0)}{\log 0.7}, \frac{\log(0.05-0)}{\log 0.7} \right) = (0, 8.399).$$

Using $a = \alpha = 0.05$, we obtain the other one-sided 95% CI

$$\left(\frac{\log(1-0.05)}{\log 0.7}, \frac{\log(0.05-0.05)}{\log 0.7} \right) = (0.1438, \text{infinity}).$$

The following is a way to derive the above CIs without a pivot. First, set $F(y) = y^k$ to p to get the quantile function of Y, namely $F_Y^{-1}(p) = p^{1/k}$, $0 < p < 1$. Then

$$\begin{aligned} 1-\alpha &= P(F_Y^{-1}(a) < Y < F_Y^{-1}(a+1-\alpha)) \\ &= P(a^{1/k} < Y < (a+1-\alpha)^{1/k}) \\ &= P\left(\frac{1}{k}\log a < \log Y < \frac{1}{k}\log(a+1-\alpha) \right) \\ &= P\left(\frac{\log a}{\log Y} > k > \frac{\log(a+1-\alpha)}{\log Y} \right). \end{aligned}$$

So a $1-\alpha$ CI for k is given by $\left(\frac{\log(a+1-\alpha)}{\log y}, \frac{\log a}{\log y} \right)$.

This yields the same numerical results as previously.

R Code for Problem 3 (not required)

```
y=0.7; c(y/(1-y), -1/log(y)) # 2.333333 2.803673
y=0.7; alp=0.05; avec=c(0,alp/2,alp)
for(i in 1:3){ a=avec[i]; print(     c(log(1-a), log(alp-a)  )/log(y)     )   }
# [1]  0.000000       8.399054
# [1]  0.07098286    10.34241266
# [1]  0.1438096       Inf
y=0.7; alp=0.05; avec=c(0,alp/2,alp);
for(i in 1:3){ a=avec[i]; print(     c(log(a+1-alp), log(a)  )/log(y)     )   }
# [1]  0.1438096       Inf
# [1]  0.07098286    10.34241266
# [1]  0.000000        8.399054
```

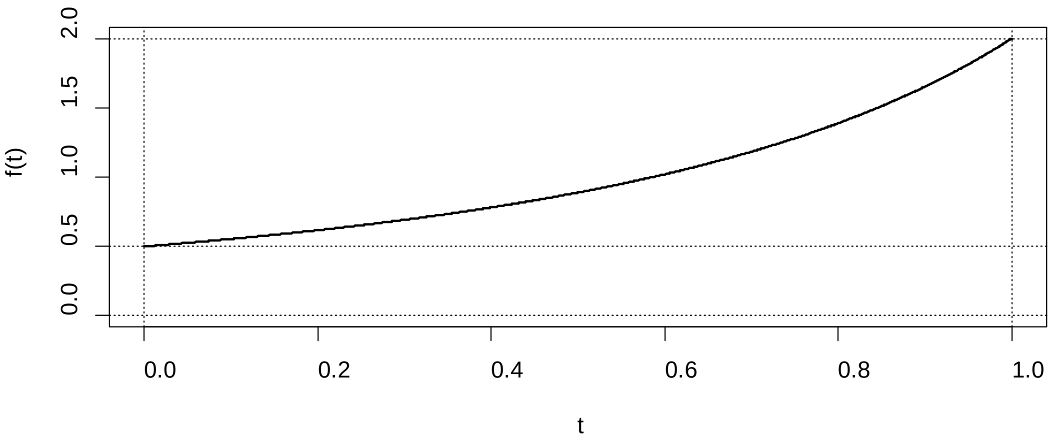
Solution to Problem 4

(a) First, $t = 2x/(x+1) \Rightarrow tx + t = 2x \Rightarrow x(2-t) = t \Rightarrow x = t(2-t)^{-1}$.

We see that *t* is a strictly increasing function of *x* and has derivative

$$\frac{dx}{dt} = t(-1)(2-t)^{-2}(-1) + 1(2-t)^{-1} = \frac{1}{(2-t)^2}(t+2-t) = \frac{2}{(2-t)^2}.$$

So $f_T(t) = f(x)\left|\frac{dx}{dt}\right| = 1 \times \left|\frac{2}{(2-t)^2}\right| = \frac{2}{(2-t)^2}$, $0 < t < 1$, as shown in the figure below.



(b) The required expectation is $\quad ET = \int t f(t)dt = \int_0^1 t \frac{2}{(2-t)^2} dt$

$$= \int_{2-0}^{2-1} (2-w)\frac{2}{w^2}(-dw) \quad \text{(after substituting } w = 2-t\text{)}$$

$$= 2I,$$

where $I = \int_1^2 \left(\frac{2}{w^2} - \frac{1}{w}\right)dw = \left[\frac{-2}{w} - \log w \Big|_{w=1}^{2}\right] = \left(\frac{-2}{2} - \log 2\right) - \left(\frac{-2}{1} - \log 1\right) = 1 - \log 2$.

It follows that $ET = 2I = 2(1-\log 2)$ = **<u>0.6137</u>**.

Alternatively, $\qquad ET = E\left(\frac{2X}{X+1}\right) = \int \frac{2x}{x+1} f(x)dx = \int_0^1 \frac{2x}{x+1} \times 1dx$

$$= 2\int_1^2 \frac{w-1}{w} dw \quad \text{(after substituting } w = x + 1\text{)}$$

$$= 2\left\{\int_1^2 1dw - \int_1^2 \frac{1}{w}dw\right\} = 2\{1 - (\log 2 - \log 1)\} = 2(1-\log 2).$$

(c) The cdf of U is $F(u) = P(U \le u) = P((2Y-1)/X \le u) = P(Y \le (uX+1)/2)$.

We now need to consider four cases, and the results are as follows.

For $0 < u < 1$, $\quad F(u) = \dfrac{1}{2} + \dfrac{u}{4}$.

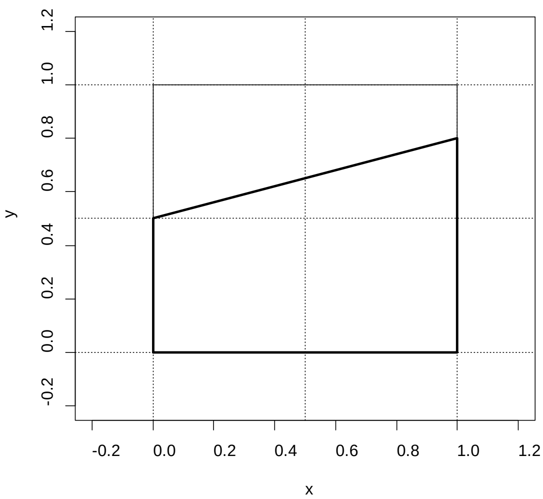
For $u > 1$, $\quad F(u) = 1 - \dfrac{1}{4u}$.

For $-1 < u < 0$, $\quad F(u) = \dfrac{1}{2} + \dfrac{u}{4}$.

For $u < -1$, $\quad F(u) = -\dfrac{1}{4u}$.

Each of these four results was obtained by drawing a unit square in the (x,y)-plane and determining the area under the function $y = (ux+1)/2$. The figure below illustrates the first case, $0 < u < 1$, in particular when $u = 0.6$. For that case, the relevant area is

$$\frac{1}{2} + \frac{1}{2} \times 1 \times \left(\frac{u \times 1 + 1}{2} - \frac{1}{2} \right) = \frac{1}{2} + \frac{u}{4} = \frac{1}{2} + \frac{0.6}{4} = 0.65.$$



In this figure, the co-ordinates of the four corners of the trapezoid are:

(0,0) (the origin, bottom left)

$(0.6 \times 0 + 1)/2$ = (0,0.5)

$(0.6 \times 1 + 1)/2$ = (0,0.8)

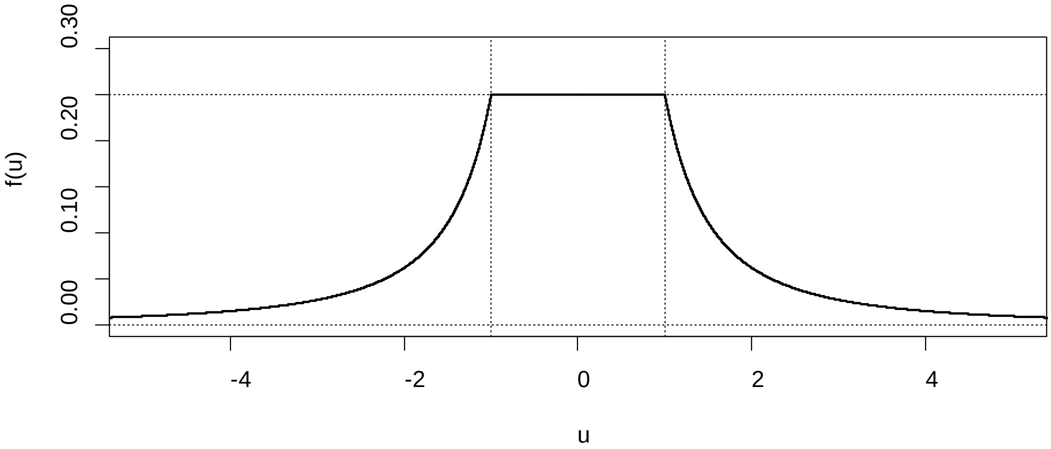
(1,0) (bottom right).

The area inside the trapezoid is 0.65.

By taking derivatives, we find that the density of U is

$$f(u) = \begin{cases} 1/4, & |u| \le 1 \\ 1/(4u^2), & |u| > 1. \end{cases}$$

The following figure illustrates.



R Code for Problem 4 (not required)

```
# (a)
tvec=seq(0,1,0.001); ftvec=2/(2-tvec)^2; X11(w=8,h=4)
plot(c(0,1),c(0,2),type="n", xlab="t",ylab="f(t)")
lines(tvec,ftvec,lwd=2)
abline(v=c(0,1),lty=3); abline(h=c(0,0.5,2),lty=3)

# (b)
X11(w=6,h=6)
plot(c(-0.2,1.2),c(-0.2,1.2),type="n",xlab="x",ylab="y")
abline(h=c(0,0.5,1),lty=3); abline(v=c(0,0.5,1),lty=3)
lines(c(0,0,1,1,0),c(0,1,1,0,0),lwd=1);
u=0.6; (c(0,1)*u+1)/2 # 0.5 0.8
lines(c(0,1),c(0.5,0.8), lwd=2)
lines(c(0,0,1,1,0),c(0,0.5,0.8,0,0),lwd=3);

uvec=seq(-6,6,0.001); fuvec=rep(0.25,length(uvec))
fuvec[abs(uvec)>1] = 0.25/uvec[abs(uvec)>1]^2;  X11(w=8,h=4)
plot(c(-5,5),c(0,0.3),type="n", xlab="u",ylab="f(u)")
lines(uvec,fuvec,lwd=2)
abline(v=c(-1,1),lty=3); abline(h=c(0,0.25),lty=3)
```

Solution to Problem 5

(a) By the central limit theorem, $\bar{Y} \sim N(\mu, \mu^2/n)$. So (approximately),

$$1-\alpha = P\left(-z < \frac{\bar{Y}-\mu}{\mu/\sqrt{n}} < z\right) \quad \text{where } z = z_{\alpha/2}$$

$$= P\left(-z\frac{\mu}{\sqrt{n}} < \bar{Y}-\mu, \bar{Y}-\mu < z\frac{\mu}{\sqrt{n}}\right)$$

$$= P\left(\mu\left(1-\frac{z}{\sqrt{n}}\right) < \bar{Y}, \bar{Y} < \mu\left(1+\frac{z}{\sqrt{n}}\right)\right)$$

$$= P\left(\mu < \frac{\bar{Y}}{1-z/\sqrt{n}}, \frac{\bar{Y}}{1+z/\sqrt{n}} < \mu\right)$$

$$= P\left(\frac{\bar{Y}}{1+z/\sqrt{n}} < \mu < \frac{\bar{Y}}{1-z/\sqrt{n}}\right).$$

So the CI is given generally by $\left(\frac{\bar{y}}{1+z_{\alpha/2}/\sqrt{n}}, \frac{\bar{y}}{1-z_{\alpha/2}/\sqrt{n}}\right)$.

This CI works out as $\left(\frac{5}{1+1.96/\sqrt{64}}, \frac{5}{1-1.96/\sqrt{64}}\right)$ = **<u>(4.016, 6.622)</u>**.

Discussion

Another solution can be obtained by estimating σ by $\bar{y}$ in $\left(\bar{y} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$. This yields $\left(\bar{y} \pm z_{\alpha/2}\frac{\bar{y}}{\sqrt{n}}\right)$ = (3.775, 6.225). Both solutions are correct but the first is better.

(b) By the central limit theorem, $\bar{Y} \sim N(\mu, \mu/n)$. So (approximately),

$$1-\alpha = P\left(-z < \frac{\bar{Y}-\mu}{\sqrt{\mu}/\sqrt{n}} < z\right) \quad \text{where } z = z_{\alpha/2}$$

$$= P\left(\left|\frac{\bar{Y}-\mu}{\sqrt{\mu/n}}\right| < z\right) = P\left(\frac{(\bar{Y}-\mu)^2}{\mu/n} < z^2\right).$$

So the CI is (a,b), where *a* and *b* are the solutions in μ of $\frac{(\bar{y}-\mu)^2}{\mu/n} = z^2$.

Solving, we get $\bar{y}^2 - 2\bar{y}\mu + \mu^2 = z^2\mu / n \quad \Rightarrow \mu^2 - \mu(2\bar{y} + z^2 / n) + \bar{y}^2 = 0$

$$\Rightarrow \mu = \frac{(2\bar{y} + z^2/n) \pm \sqrt{(2\bar{y} + z^2/n)^2 - 4\bar{y}^2}}{2}$$

$$= \left(\bar{y} + \frac{z^2}{2n}\right) \pm \sqrt{\left(\bar{y} + \frac{z^2}{2n}\right)^2 - \bar{y}^2}$$

$$= \bar{y} + \frac{z^2}{2n} \pm \sqrt{\bar{y}^2 + \frac{2\bar{y}z^2}{2n} + \frac{z^4}{4n^2} - \bar{y}^2} \qquad = \bar{y} + \frac{z^2}{2n} \pm z\sqrt{\frac{\bar{y}}{n} + \frac{z^2}{4n^2}}.$$

So the CI is given generally by $\left(\bar{y} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2}\sqrt{\frac{\bar{y}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}\right)$.

This CI works out as (5.03001 ± 0.54865) = **<u>(4.481, 5.579)</u>**.

Discussion

If n is very large then the above CI is approximately $\left(\bar{y} \pm z_{\alpha/2}\sqrt{\frac{\bar{y}}{n}}\right)$ = (4.452, 5.548).

This second solution can also be obtained by estimating σ by $\sqrt{\bar{y}}$ in $\left(\bar{y} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$.

Both solutions are correct but the first is better.

R Code for Problem 5 (not required)

```
# (a)
n=64; ybar=5; alp=0.05; z=qnorm(1-alp/2);
ybar/(1+c(1,-1)*z/sqrt(n)) # 4.016079 6.622477
ybar + c(-1,1)*z*abs(ybar)/sqrt(n) # 3.775023 6.224977

# (b)
term1 = ybar+z^2/(2*n); term2 = z*sqrt( ybar/n + z^2/(4*n^2) )
ci=term1 + c(-1,1)*term2
c(term1, term2, ci) # 5.030011 0.548648 4.481363 5.578659
ybar + c(-1,1)*z*sqrt(ybar/n) # 4.452173 5.547827
```

Solution to Problem 6

The expected value of *Y* is clearly a strictly increasing function of θ. So if *y* is large then this is evidence in favour of the alternative hypothesis. So a suitable *p*-value is $p = P(Y > y)$, where *y* is the observed value of *Y* and $\theta = k$. This *p*-value can be determined as follows. We first need to find the constant *c* as a function of θ. Now,

$$P(Y > y) = c\int_y^\theta \frac{1}{(t+1)^2}dt = c\int_{y+1}^{\theta+1} u^{-2}du \quad \text{where } u = t+1 \text{ (and } \theta = k\text{)}$$

$$= c\left[\frac{u^{-1}}{-1}\Bigg|_{u=y+1}^{\theta+1}\right] = -c\left(\frac{1}{\theta+1} - \frac{1}{y+1}\right) = c\left(\frac{1}{y+1} - \frac{1}{\theta+1}\right).$$

This function of *y* is defined for all $0 \le y \le \theta$, and so it must equal 1 at *y* = 0. Thus

$$1 = P(Y > 0) = c\left(\frac{1}{0+1} - \frac{1}{\theta+1}\right) = c\frac{\theta}{\theta+1}.$$

Therefore $c = \dfrac{\theta+1}{\theta}$. It follows that the *p*-value has a formula given by

$$p = P(Y > y) = \frac{\theta+1}{\theta}\left(\frac{1}{y+1} - \frac{1}{\theta+1}\right) = \frac{\theta+1}{\theta(y+1)} - \frac{1}{\theta} = \frac{k-y}{k(y+1)} \quad \text{(generally).}$$

So for the case *k* = 5 and *y* = 4, we obtain

$$p = q = \frac{5-4}{5(4+1)} = 1/25 = \underline{\mathbf{0.04}}.$$

The value of *y* which results in a *p*-value of *p* = 0.1 can be obtained by solving

$$p = \frac{k-y}{k(y+1)}$$

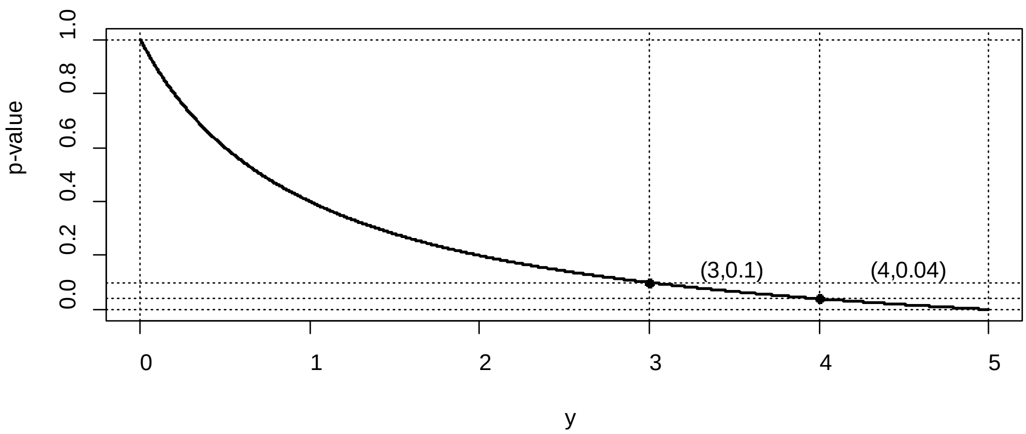
for *y*. The solution is

$$y = \frac{(1-p)k}{1+pk} \quad \text{(generally).}$$

So for the case *k* = 5 and *p* = 0.1, we obtain

$$y = z = \frac{(1-0.1)5}{1+0.1\times 5} = \underline{\mathbf{3.0}}.$$

The following figure is the required graph and also shows the above two (*y*,*p*) pairs. Note that the *p*-value appropriately decreases from 1 to 0 as *y* increases from 0 to *k*.



R Code for Problem 6 (not required)

```
yvec=seq(0,5,0.001); pvec=(5-yvec)/(5*(yvec+1)); X11(w=8,h=4)
plot(yvec,pvec,type="l",xlab="y",ylab="p-value", lwd=2)
(1-0.1)*5/(1+0.1*5) # 3
abline(v=c(0,3,4,5),lty=3); abline(h=c(0,0.04,0.1,1),lty=3);
points(c(3,4),c(0.1,0.04),pch=16)
text(3.3,0.15,"(3,0.1)"); text(4.3,0.15,"(4,0.04)")
```

Solution to Problem 7

(a) Let m = 9, w = 13, $N = m + w$ = 22 and c = 6. Also let A be the event that the oldest man is on the committee, and let B be the event that the oldest woman is *not* on the committee. Then the required probability is

$$P(A\cup B)=1-P(\overline{A\cup B})=1-P(\bar{A}\bar{B})=1-\{P(\bar{B})-P(A\bar{B})\}$$

$$=1-\frac{\binom{N-1}{c-1}}{\binom{N}{c}}+\frac{\binom{N-2}{c-2}}{\binom{N}{c}}=1-\frac{\binom{21}{5}}{\binom{22}{6}}+\frac{\binom{20}{4}}{\binom{22}{6}}=1-\frac{6}{22}+\frac{6\times 5}{22\times 21}\ = 61/77 = \underline{\mathbf{0.7922}}.$$

Alternatively,

$$P(A\cup B)=P(A)+P(B)-P(AB)=\frac{\binom{21}{5}}{\binom{22}{6}}+\frac{\binom{21}{6}}{\binom{22}{6}}-\frac{\binom{20}{5}}{\binom{22}{6}}\ = 0.7922.$$

(b) Let L be the event that the committee has at least one man, and let F be the event that the first person selected is a man. Then the required probability is

$$P(F \mid L) = \frac{P(F)P(L \mid F)}{P(L)} = \frac{(m/N)\times 1}{1-\binom{m}{0}\binom{w}{c}\Big/\binom{N}{c}} = \frac{9/22}{1-\binom{13}{6}\Big/\binom{22}{6}} = \underline{\mathbf{0.4187}}.$$

(c) Let X be the number of men on the first committee, and consider the distribution of this random variable jointly with Y, the number of men on both committees. Also denote the number of men on the second committee by k = 4. Then:

$$X \sim Hyp(N,m,c)$$

$$(Y \mid X = x) \sim Hyp(N,x,k).$$

So: $EX = c\frac{m}{N}$ = 2.45455, $VX = c\frac{m}{N}\left(1-\frac{m}{N}\right)\frac{N-c}{N-1}$ = 1.10508

$$EX^2 = VX + (EX)^2 = 7.12987$$

$$E(Y \mid X = x) = k\frac{x}{N}, \qquad V(Y \mid X = x) = k\frac{x}{N}\left(1-\frac{x}{N}\right)\frac{N-k}{N-1}.$$

It follows that: $EY = EE(Y \mid X) = E\left\{k\frac{X}{N}\right\} = \frac{k}{N}EX = \frac{ckm}{N^2} = \frac{54}{121}$ = $\underline{\mathbf{0.4463}}$

$$VY = VE(Y \mid X) + EV(Y \mid X) = V\left\{k\frac{X}{N}\right\} + E\left\{k\frac{X}{N}\left(1-\frac{X}{N}\right)\frac{N-k}{N-1}\right\}$$

$$= \left(\frac{k}{N}\right)^2 VX + \frac{k}{N}\left(\frac{N-k}{N-1}\right)\left(EX - \frac{1}{N}EX^2\right) = \underline{\mathbf{0.3686}}.$$

An alternative solution to (c) is as follows. Let X_i be the indicator variable for the ith man being on both committees (i = 1,…,m). (Thus X_i = 1 if the ith man is on both committees, and X_i = 0 otherwise.) Then we wish to find the mean

$$e = E(X_1 + ... + X_m) = mEX_1,$$

where: $X_1 \sim Bern(p)$

$p = EX_1 = \frac{6}{22}\times\frac{4}{22} = \frac{6}{121}$ (the probability that man 1 is on both committees).

It follows that $e = mp = 9 \times \frac{6}{121} = \frac{54}{121}$ = 0.4463 (as previously).

We also wish to find the variance

$$v = V(X_1 + ... + X_m) = mVX_1 + m(m-1)C(X_1, X_2)\,,$$

where:

$$VX_1 = p(1-p) = \frac{6}{121}\left(1 - \frac{6}{121}\right) = \frac{690}{121^2} = 0.0471279$$

$$C(X_1, X_2) = E(X_1X_2) - EX_1EX_2$$

$X_1X_2 \sim Bern(q)$ (noting that the product X_1X_2 equals 0 or 1)

$$q = E(X_1X_2) = P(X_1X_2 = 1) = P(X_1 = 1, X_2 = 1)$$

= probability that man 1 and man 2 are both on both committees

$$= \frac{\binom{22-2}{6-2}}{\binom{22}{6}} \times \frac{\binom{22-2}{4-2}}{\binom{22}{4}} = \frac{6\times 5}{22\times 21} \times \frac{4\times 3}{22\times 21} = \frac{10}{77^2} = 0.001686625.$$

We see that

$$C(X_1, X_2) = q - p^2 = \frac{10}{77^2} - \left(\frac{6}{121}\right)^2 = -0.000772223,$$

and so the required variance is

$$v = 9\times 0.0471279 + 9(9-1)(-0.000772223)$$

= 0.4241511 – 0.05560 = 0.36855 (as previously).

R Code for Problem 7 (not required)

```
# (a)

m = 9; w = 13; N = m + w; c = 6; k = 4

1-choose(N-1,c-1)/choose(N,c) +choose(N-2,c-2)/choose(N,c) # 0.7922078

p=(c/N)*(k/N); p # 0.04958678

m*p # 0.446281

(choose(21,5)+choose(21,6)-choose(20,5))/choose(22,6)  # 0.7922078


# (b)

(9/22) / (1 - choose(13,6)/choose(22,6)) # 0.4187209
```

```
# (c)
EX=c*m/N; VX=c*(m/N)*(1-m/N)*((N-c)/(N-1)); EX2 = VX + EX^2
c(EX, VX, EX2) # 2.454545 1.105077 7.129870
EY=(k/N)*EX
VY = (k/N)^2 * VX + (k/N) * ((N-k)/(N-1)) * (   EX -  (1/N)*EX2  )
c(EY,VY)  #  0.4462810 0.3685513

# Alternative solution
p=c*k/N^2; e=m*p; c(p,e) # 0.04958678 0.44628099
VarX1 = p*(1-p); VarX1 # 0.04712793
q=(choose(N-2,c-2)/choose(N,c))*choose(N-2,k-2)/choose(N,k); q # 0.001686625
CovX1X2=q-p^2; CovX1X2 # -0.0007722234
v=m*VarX1+m*(m-1)*CovX1X2; v   # 0.3685513
```

Solution to Problem 8

Here: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ = 2.5321, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ = –0.0109

$S_{xx} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$ = 2187.8, $S_{yy} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$ = 1775.9

$S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$ = 1747.8, $SSE = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$ = 379.51.

So the required point estimates of the intercept α, slope β and variance σ^2 are:

$b = \frac{S_{xy}}{S_{xx}} =$ **<u>0.79892</u>**, $a = \bar{y} - b\bar{x} =$ **<u>–2.0338</u>**, $s^2 = \frac{SSE}{n-2}$ = **<u>0.38027</u>**.

Further, $\hat{V}b = \frac{s^2}{S_{xx}}$ = 0.00017382, and $t_{0.025}(n-2) \approx z_{0.025} = 1.96$. So a 95% CI for β is

$\left(b \pm z_{\alpha/2}\sqrt{\hat{V}b}\right) \approx \left(0.79892 \pm 1.96\sqrt{0.00017382}\right)$ = **<u>(0.7731, 0.8248)</u>**.

We may predict *v* (a new *y*-value with *x*-value *u* = 0.6) by

$\hat{v} = a + bu$ = **<u>–1.5545</u>**.

A 95% prediction interval for v is

$$\left(\hat{v} \pm z_{\alpha/2} s \sqrt{1 + \frac{1}{n} + \frac{(u-\overline{x})^2}{S_{xx}}}\right) = \underline{\mathbf{(-2.7648,\ -0.3442)}},$$

and a 95% confidence interval for the *mean* of v is

$$\left(\hat{v} \pm z_{\alpha/2} s \sqrt{0 + \frac{1}{n} + \frac{(u-\overline{x})^2}{S_{xx}}}\right) = \underline{\mathbf{(-1.6174,\ -1.4916)}}.$$

R Code for Problem 8 (not required)

```
options(digits=5); n=1000; set.seed(442); x = round(runif(n,0,5),1)
alp=-2; bet=0.8; sig=0.6; y=round(rnorm(n,alp+bet*x,sig), 1); plot(x,y) # Looks OK
rbind(c(1,2,n), x[c(1,2,n)], y[c(1,2,n)]) # Data look OK

sumx=sum(x); xbar=sumx/n; sumx2=sum(x^2)
sumy=sum(y); ybar=sumy/n; sumy2=sum(y^2)
sumxy = sum(x*y);  Sxy=sumxy-n*xbar*ybar
Sxx=sumx2-n*xbar^2; Syy=sumy2-n*ybar^2
c(sumx,xbar,sumy,ybar,sumx2,sumy2,sumxy)
  # 2532.1000    2.5321  -10.9000   -0.0109 8599.2900 1776.0100 1720.2400

c(Sxx,Syy,Sxy) # 2187.8 1775.9 1747.8
SSE=Syy-Sxy^2 / Sxx; SSE # 379.51
b=Sxy/Sxx; a=ybar-xbar*b; s2=SSE/(n-2)
c(b,a,s2) #  0.79892 -2.03384  0.38027
abline(a,b,lwd=2) # Regression line looks OK
Vhatb=s2/Sxx; Vhatb # 0.00017382
b+c(-1,1)*1.96*sqrt(Vhatb) # 0.77308 0.82476 (CI using CLT and normal dsn)
b+c(-1,1)*qt(0.975,n-2)*sqrt(Vhatb) # 0.77305 0.82479 (CI using t dsn)

u=0.6; vhat=a+b*u
predint= vhat +c(-1,1)*1.96*sqrt(s2*(1+1/n+(u-xbar)^2 / Sxx))
confint= vhat +c(-1,1)*1.96*sqrt(s2*(0+1/n+(u-xbar)^2 / Sxx))
c(vhat,predint,confint) # -1.5545 -2.7648 -0.3442 -1.6174 -1.4916
```