THE ROYAL STATISTICAL SOCIETY

2005 EXAMINATIONS – SOLUTIONS

GRADUATE DIPLOMA

STATISTICAL THEORY AND METHODS

PAPER I

The Society provides these solutions to assist candidates preparing for the examinations in future years and for the information of any other persons using the examinations.

The solutions should NOT be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids.

Users of the solutions should always be aware that in many cases there are valid alternative methods. Also, in the many cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of these solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of these solutions.

Note. In accordance with the convention used in the Society's examination papers, the notation log denotes logarithm to base *e*. Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

© RSS 2005

(i) If $\{E_1, E_2, \ldots, E_n\}$ partition S, then $P(A) = \sum_{i=1}^{n} P(A \mid E_i)P(E_i)$. This is the law of total probability.

Since $P(A \cap E_j) = P(A | E_j)P(E_j) = P(E_j | A)P(A)$, we have (Bayes' Theorem)

$$P(E_j \mid A) = \frac{P(A \mid E_j)P(E_j)}{P(A)} = \frac{P(A \mid E_j)P(E_j)}{\sum_{i=1}^{n} P(A \mid E_i)P(E_i)}.$$

(ii) Let Y be the amount of time spent in the fitting room; Y is exponential with parameter $\frac{1}{3x}$.

(a) $P(Y < y \mid x) = \int_{t=0}^{y} \frac{1}{3x} e^{-t/3x} dt = \left[-e^{-t/3x}\right]_{t=0}^{y} = 1 - e^{-y/3x}$ (for $y > 0$).

Since x takes the values 1, 2, 3, 4 each with probability ¼, we therefore have

$$F(y) = \left(1 - e^{-y/3} + 1 - e^{-y/6} + 1 - e^{-y/9} + 1 - e^{-y/12}\right) \times \frac{1}{4} \quad \text{(using (i))}.$$

When $y = 5$ this is $1 - \frac{1}{4}\left(e^{-5/3} + e^{-5/6} + e^{-5/9} + e^{-5/12}\right)$

$$= 1 - \frac{1}{4}(0.1889 + 0.4346 + 0.5738 + 0.6592) = 1 - 0.464$$

and so $P(Y > 5) = 0.464$.

Solution continued on next page

(b) Let X be the number of garments taken to the room. Then

$$E(X) = \frac{1}{4}(1+2+3+4) = \frac{5}{2},$$

$$E(X^2) = \frac{1}{4}(1+4+9+16) = \frac{15}{2},$$

so $\text{Var}(X) = \dfrac{15}{2} - \dfrac{25}{4} = \dfrac{5}{4}$.

[These results may be quoted, as X has a discrete uniform distribution.]

Now, $E(Y|X) = 3X$. Also, because Y has an exponential distribution,

$$\text{Var}(Y|X) = (3X)^2 = 9X^2.$$

Thus

$$E(Y) = E\{E(Y|X)\} = E\{3X\} = 3E(X) = \frac{15}{2}.$$

Also,

$$\begin{aligned}
\text{Var}(Y) &= E\{\text{Var}(Y|X)\} + \text{Var}\{E(Y|X)\} \\
&= E\{9X^2\} + \text{Var}\{3X\} \\
&= 9E(X^2) + 9\text{Var}(X) \\
&= 9 \times \frac{15}{2} + 9 \times \frac{5}{4} \\
&= \frac{315}{4}.
\end{aligned}$$

<u>Graduate Diploma, Statistical Theory & Methods, Paper I, 2005. Question 2</u>

(i) $X + Y$ can take the values 0, 1, 2,…, $n + m$. For these values,

$$P(X+Y=z)=\sum_{x=0}^{z}P(X=x \text{ and } Y=z-x)$$

$$=\sum_{x=0}^{z}P(X=x)P(Y=z-x)$$

$$=\sum_{x=0}^{z}\left\{\binom{n}{x}\theta^{x}(1-\theta)^{n-x}\binom{m}{z-x}\theta^{z-x}(1-\theta)^{m-z+x}\right\}$$

$$=\theta^{z}(1-\theta)^{m+n-z}\sum_{x=0}^{z}\binom{n}{x}\binom{m}{z-x}$$

$$=\binom{m+n}{z}\theta^{z}(1-\theta)^{m+n-z}.$$

Thus $X + Y$ has the binomial distribution with parameters $m + n$ and θ.

[An alternative method is to use probability generating functions.]

(ii) $P\left(X=x \middle| X+Y=z\right)$

$$=\frac{P(X=x\cap X+Y=z)}{P(X+Y=z)}$$

$$=\frac{P(X=x\cap Y=z-x)}{P(X+Y=z)}$$

$$=\frac{\binom{n}{x}\theta^{x}(1-\theta)^{n-x}\binom{m}{z-x}\theta^{z-x}(1-\theta)^{m-z+x}}{\binom{m+n}{z}\theta^{z}(1-\theta)^{m+n-z}}$$

$$=\frac{\binom{n}{x}\binom{m}{z-x}}{\binom{m+n}{z}}$$

(i.e. a hypergeometric distribution).

Solution continued on next page

(iii) Let X and Y be the numbers of failed components in the two networks. We have $n = 20$, $m = 30$, $\theta = 0.1$, $z = 6$ in the above notation.

$$P(X = 3 \mid X + Y = 6) = \frac{\binom{20}{3}\binom{30}{3}}{\binom{50}{6}} = \frac{20.19.18.30.29.28.6.5.4.3.2.1}{3.2.1.3.2.1.50.49.48.47.46.45} = 0.2913.$$

$$f(x,y) = 12x^2 \qquad (0 < x < y < 1)$$

(i)

$$E\left(X^r Y^s\right) = \int_{x=0}^{1}\int_{y=x}^{1} 12x^{2+r} y^s \, dy \, dx$$

$$= 12\int_0^1 x^{2+r}\left[\frac{y^{s+1}}{s+1}\right]_{y=x}^{1} dx$$

$$= \frac{12}{s+1}\int_0^1 x^{2+r}(1 - x^{s+1})\, dx$$

$$= \frac{12}{s+1}\left[\frac{x^{r+3}}{r+3} - \frac{x^{r+s+4}}{r+s+4}\right]_0^1$$

$$= \frac{12}{s+1}\left(\frac{1}{r+3} - \frac{1}{r+s+4}\right)$$

$$= \frac{12(s+1)}{(s+1)(r+3)(r+s+4)} = \frac{12}{(r+3)(r+s+4)}.$$

Hence

$$E(X) = \frac{12}{4\times 5} = \frac{3}{5} \qquad \text{(put } r = 1,\ s = 0\text{; similarly for the others)}$$

$$E(Y) = \frac{12}{3\times 5} = \frac{4}{5}$$

$$E\left(X^2\right) = \frac{12}{5\times 6} = \frac{2}{5}, \quad \text{so } \operatorname{Var}(X) = \frac{2}{5} - \left(\frac{3}{5}\right)^2 = \frac{1}{25}$$

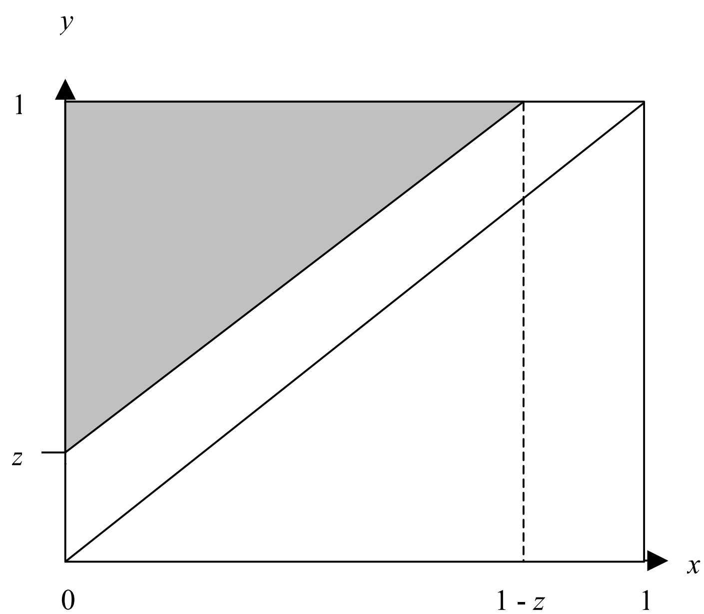
$$E\left(Y^2\right) = \frac{12}{3\times 6} = \frac{2}{3}, \quad \text{so } \operatorname{Var}(Y) = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{75}$$

$$E(XY) = \frac{12}{4\times 6} = \frac{1}{2}, \quad \text{so } \operatorname{Cov}(X,Y) = \frac{1}{2} - \left(\frac{3}{5}\times\frac{4}{5}\right) = \frac{1}{50}$$

$$\therefore \rho_{XY} = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = \frac{\frac{1}{50}}{\sqrt{\frac{1}{25}\times\frac{2}{75}}} = \frac{\sqrt{3}}{2\sqrt{2}}$$

Solution continued on next page

(ii) $P(Y-X>z) = \int_{x=0}^{1-z}\left\{\int_{y=x+z}^{1} 12x^2 dy\right\} dx$, the evaluation being over the shaded region shown:



This is $12\int_0^{1-z} x^2 \left[y\right]_{x+z}^{1} dx = 12\int_0^{1-z} x^2(1-x-z) \quad dx$

$$= 12\int_0^{1-z}\left\{x^2(1-z)-x^3\right\} dx = 12\left[(1-z)\frac{x^3}{3}-\frac{x^4}{4}\right]_0^{1-z}$$

$$= 12\left\{\frac{(1-z)^4}{3}-\frac{(1-z)^4}{4}\right\} = (1-z)^4$$

Therefore $F(z) = 1-(1-z)^4 \qquad (\text{for } 0 \le z \le 1)$

and $f(z) = F'(z) = 4(1-z)^3 \qquad (\text{for } 0 \le z \le 1)$.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2005. Question 4

(i) $U = \dfrac{X}{X+Y}$, $V = X + Y$. So $X = UV$ and $Y = (1 - U)V$.

The Jacobian of the transformation is

$$\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ -v & 1-u \end{vmatrix} = v(1-u) + uv = v\,.$$

The joint pdf of X, Y is $f(x,y) = \dfrac{\theta^{\alpha+\beta} x^{\alpha-1} y^{\beta-1} e^{-\theta(x+y)}}{\Gamma(\alpha)\Gamma(\beta)}$, for $x > 0,\ y > 0$.

Hence the joint pdf of U, V is

$$g(u,v) = f(x,y)|J| = \frac{\theta^{\alpha+\beta}(uv)^{\alpha-1}(1-u)^{\beta-1}v^{\beta-1}e^{-\theta v}v}{\Gamma(\alpha)\Gamma(\beta)} \qquad (\text{for } v > 0,\ 0 < u < 1)$$

$$= \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)}\left\{u^{\alpha-1}(1-u)^{\beta-1}\right\}\left\{v^{\alpha+\beta-1}e^{-\theta v}\right\}.$$

This is of the form of a product

constant × function of u alone [$g(u)$, say] × function of v alone [$h(v)$, say]

and so U, V are independent. $g(u)$ is proportional to $u^{\alpha-1}(1 - u)^{\beta-1}$, the pdf of a beta distribution, and so U has a beta distribution. $h(v)$ is proportional to $v^{\alpha+\beta-1}e^{-\theta v}$, the pdf of a gamma distribution, and so V has a gamma distribution. The scale parameter of V is θ, as for X and Y.

(ii) $U = \dfrac{X}{X+Y}$ is the required distribution, where X and Y are the common exponential random variables. Taking $\alpha = \beta = 1$, $g(u) = u^0(1 - u)^0 = 1$ and so U has the uniform distribution on (0, 1).

(i) $M_X(t) = E\left(e^{tX}\right) = \int_0^\infty e^{tx} \frac{1}{\sqrt{2\pi x}} e^{-x/2} dx = \frac{1}{\sqrt{2\pi}} \int_0^\infty x^{-1/2} e^{-x\left(\frac{1}{2}-t\right)} dx$.

$t < ½$ is used in what follows to ensure convergence of the integral.

Write $u = x\left(\frac{1}{2}-t\right)$, so that $du = \left(\frac{1}{2}-t\right) dx$.

Then $M_X(t) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \left(\frac{u}{\frac{1}{2}-t}\right)^{-1/2} e^{-u} \frac{1}{\frac{1}{2}-t} du$

$$= \frac{2}{\sqrt{2\pi}\sqrt{2}\sqrt{1-2t}} \int_0^\infty u^{-1/2} e^{-u} du$$

The integral here should be recognised as $\Gamma(½) = \sqrt{\pi}$; alternatively, refer back to the original pdf

$$= \frac{1}{\sqrt{1-2t}}.$$

$M_X(t) = (1-2t)^{-1/2}$, so $M'_X(t) = -\frac{1}{2}(1-2t)^{-3/2}(-2) = (1-2t)^{-3/2}$

$\therefore E(X) = M'_X(0) = 1$

$M''_X(t) = -\frac{3}{2}(1-2t)^{-5/2}(-2) = 3(1-2t)^{-5/2}$

$\therefore E\left(X^2\right) = M''_X(0) = 3$

$\therefore \text{Var}(X) = E\left(X^2\right) - \{E(X)\}^2 = 3 - 1^2 = 2$.

Solution continued on next page

(ii) $M_{X_1+\ldots+X_n}(t)=\{M_X(t)\}^n$

$$=(1-2t)^{-n/2}.$$

Now using $M_{aY+b}(t)=e^{bt}M_Y(at)$,

$$M_z(t)=e^{-\sqrt{\frac{n}{2}}\,t}\left(1-2\frac{t}{\sqrt{2n}}\right)^{-n/2}.$$

To find the limiting form of $M_Z(t)$, we take logs:

$$\log M_Z(t)=-\sqrt{\frac{n}{2}}\,t\;-\;\frac{n}{2}\log\left(1-t\sqrt{\frac{2}{n}}\right)$$

$$=-t\sqrt{\frac{n}{2}}\;-\;\frac{n}{2}\left\{-t\sqrt{\frac{2}{n}}-\frac{1}{2}\left(t\sqrt{\frac{2}{n}}\right)^2-\frac{1}{3}\left(t\sqrt{\frac{2}{n}}\right)^3-\ldots\right\}$$

$$=\frac{1}{2}t^2\;+\;\frac{1}{3}t^3\sqrt{\frac{2}{n}}\;+\;\ldots$$

$$\to\frac{t^2}{2}\quad\text{as } n\to\infty.$$

So $M_Z(t)\to e^{t^2/2}$ as $n\to\infty$, which is the mgf of N(0, 1).

Therefore in the limit *Z* becomes N(0, 1), i.e. the standard Normal distribution.

<u>Graduate Diploma, Statistical Theory & Methods, Paper I, 2005. Question 6</u>

(i) In a $U(-\theta, \theta)$ distribution, $f(x) = \dfrac{1}{2\theta}$ and $F(x) = \dfrac{x}{2\theta} + \dfrac{1}{2}$, for $-\theta < x < \theta$.

$$F(u_{(1)}, u_{(n)}) = P(U_{(n)} \le u_{(n)}) - P(U_{(1)} > u_{(1)} \text{ and } U_{(n)} \le u_{(n)})$$

$$= P(\text{all data} \le u_{(n)}) - P(\text{all data between } u_{(1)} \text{ and } u_{(n)})$$

$$= \{F(u_{(n)})\}^n - \{F(u_{(n)}) - F(u_{(1)})\}^n$$

$$= \left\{\frac{u_{(n)}}{2\theta}\right\}^n - \left\{\frac{u_{(n)} - u_{(1)}}{2\theta}\right\}^n, \quad \text{for } -\theta < u_{(1)} < u_{(n)} < \theta.$$

$$f(u_{(1)}, u_{(n)}) = \frac{\partial^2}{\partial u_{(1)} \partial u_{(n)}} F(u_{(1)}, u_{(n)})$$

$$= \frac{n(n-1)(u_{(n)} - u_{(1)})^{n-2}}{(2\theta)^n}.$$

[An argument using the multinomial distribution with one observation at each of $u_{(1)}$ and $u_{(n)}$ and with $n - 2$ in between is also acceptable.]

(ii) Transforming to $R = U_{(n)} - U_{(1)}$ and $T = U_{(1)}$ (so that $U_{(n)} = R + T$), we have the Jacobian

$$J = \frac{\partial\left(u_{(1)}, u_{(n)}\right)}{\partial(r, t)} = \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix}, \quad \text{so } |J| = 1.$$

Hence $f(r, t) = \dfrac{n(n-1)r^{n-2}}{(2\theta)^n}$ (for $-\theta < t < \theta - r$, $0 < r < 2\theta$).

$$\therefore f(r) = \int_{-\theta}^{\theta - r} \frac{n(n-1)r^{n-2}}{(2\theta)^n} dt$$

$$= \frac{n(n-1)r^{n-2}(2\theta - r)}{(2\theta)^n}, \quad \text{for } 0 < r < 2\theta.$$

Solution continued on next page

(iii)

$$
\begin{aligned}
E(R) &= \frac{n(n-1)}{(2\theta)^n}\int_0^{2\theta} r^{n-1}(2\theta - r)\,dr \\
&= \frac{n(n-1)}{(2\theta)^n}\int_0^{2\theta} (2\theta r^{n-1} - r^n)\,dr \\
&= \frac{n(n-1)}{(2\theta)^n}\left[\frac{2\theta r^n}{n} - \frac{r^{n+1}}{n+1}\right]_0^{2\theta} \\
&= \frac{n(n-1)}{(2\theta)^n}\times\frac{(2\theta)^{n+1}}{n(n+1)} \\
&= 2\theta\left(\frac{n-1}{n+1}\right).
\end{aligned}
$$

Hence $\frac{1}{2}R$ is a biased estimator of θ (but asymptotically unbiased as $n \to \infty$).

Graduate Diploma, Statistical Theory & Methods, Paper I, 2005. Question 7

(i) (a) The inverse cumulative distribution function method can be used with tables of the standard Normal cdf $\Phi(x)$. The values of z are such that $\Phi(z) = u$, and for the four values of u the corresponding values of z are $-1.07, -0.42, +0.46, +1.40$.

(b) These can be transformed to N(–2, 0.81) by $w = \mu + \sigma z$ or $w = -2 + 0.9z$, to give $-2.963, -2.378, -1.586, -0.740$.

(c) The chi-squared distribution with one degree of freedom is the square of N(0, 1), so take values of z^2 from (i): 1.14, 0.18, 0.21, 1.96.

(ii) The probabilities and cumulative probabilities for a Poisson distribution with mean 2 are:

r	0	1	2	3	4	5	…
$P(r)$	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	
$F(r)$	0.1353	0.4060	0.6767	0.8571	0.9473	0.9834	

Taxis: 0.553 corresponds to $r = 2$ (it is between 0.4060 and 0.6767) etc, giving 2, 3, 1, 5, 1.

Similarly for customers: 3, 1, 1, 2, 2.

Time	*Taxis*	*Arrivals*	*Customers*	*Arrivals*
3.00	0	2	0	3
3.01	0	3	1	1
3.02	1	1	0	1
3.03	1	5	0	2
3.04	4	1	0	2
3.05	3		0	

(i) (a)

$$\text{If } \mathbf{C} = \begin{bmatrix} 1 & -\alpha \\ 1 & \beta \end{bmatrix} \text{ then } \mathbf{C}^{-1} = \frac{1}{\alpha+\beta}\begin{bmatrix} \beta & \alpha \\ -1 & 1 \end{bmatrix}$$

$$\mathbf{CDC}^{-1} = \frac{1}{\alpha+\beta}\begin{bmatrix} 1 & -\alpha \\ 1 & \beta \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1-\alpha-\beta \end{bmatrix}\begin{bmatrix} \beta & \alpha \\ -1 & 1 \end{bmatrix}$$

$$= \frac{1}{\alpha+\beta}\begin{bmatrix} 1 & \alpha(\alpha+\beta-1) \\ 1 & \beta(1-\alpha-\beta) \end{bmatrix}\begin{bmatrix} \beta & \alpha \\ -1 & 1 \end{bmatrix}$$

$$= \frac{1}{\alpha+\beta}\begin{bmatrix} \alpha+\beta-\alpha^2-\alpha\beta & \alpha^2+\alpha\beta \\ \alpha\beta+\beta^2 & \alpha+\beta-\alpha\beta-\beta^2 \end{bmatrix}$$

$$= \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$$

$$= \mathbf{P}$$

(b) The n-step transition matrix is $\mathbf{P}^n$, which can be written $\left(\mathbf{CDC}^{-1}\right)\left(\mathbf{CDC}^{-1}\right)\left(\mathbf{CDC}^{-1}\right)\cdots\left(\mathbf{CDC}^{-1}\right)$ and every pair $\mathbf{C}^{-1}\mathbf{C}$ is replaced by $\mathbf{I}$ to give $\mathbf{CD}^n\mathbf{C}^{-1}$.

$\mathbf{D}^n$ is simply $\begin{bmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^n \end{bmatrix}$, i.e. $\begin{bmatrix} 1 & 0 \\ 0 & \lambda^n \end{bmatrix}$ in the given notation.

Since $0 < \alpha < 1$ and $0 < \beta < 1$, we have $-1 < \lambda < 1$, i.e. $|\lambda| < 1$; therefore $\lambda^n \to 0$.

Thus $\mathbf{P}^n \to \mathbf{C}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\mathbf{C}^{-1}$ which is

$$\begin{bmatrix} 1 & -\alpha \\ 1 & \beta \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \beta & \alpha \\ -1 & 1 \end{bmatrix} \times \frac{1}{\alpha+\beta} = \frac{1}{\alpha+\beta}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \beta & \alpha \\ -1 & 1 \end{bmatrix} = \frac{1}{\alpha+\beta}\begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix}.$$

Solution continued on next page

(ii) Let state 0 be no rain and state 1 be rain. The transition matrix is

$$\mathbf{P} = \begin{bmatrix} 0.8 & 0.2 \\ 0.9 & 0.1 \end{bmatrix}.$$

This is the matrix in (i) with $\alpha = 0.2$ and $\beta = 0.9$.

As there is no rain on the first visit, $\begin{bmatrix} 1 & 0 \end{bmatrix}\mathbf{P}^n$ gives the probabilities for the two states on the next visit in n days' time. As n is large, $\mathbf{P}^n$ can be taken as approximately equal to the limiting value in (i)(b), i.e. here

$$\tfrac{1}{1.1}\begin{bmatrix} 0.9 & 0.2 \\ 0.9 & 0.2 \end{bmatrix}.$$

This gives

$$\begin{bmatrix} 1 & 0 \end{bmatrix}\mathbf{P}^n = \begin{bmatrix} \frac{0.9}{1.1} & \frac{0.2}{1.1} \end{bmatrix},$$

i.e. $P(\text{rain}) = \frac{0.2}{1.1} = \frac{2}{11}$.

Replacing $\begin{bmatrix} 1 & 0 \end{bmatrix}$ with $\begin{bmatrix} 0 & 1 \end{bmatrix}$ for the first visit gives the same answer because of the form of $\mathbf{P}^n$. In the long run there are about 9 days without rain for every 2 days with rain.