Institute of Actuaries of India

Subject CT6 – Statistical Methods

November 2011 Examinations

INDICATIVE SOLUTIONS

Introduction

The indicative solution has been written by the Examiners with the aim of helping candidates. The solutions given are only indicative. It is realized that there could be other points as valid answers and examiner have given credit for any alternative approach or interpretation which they consider to be reasonable.

Question 1

Let A denote the event that the student answers correctly.
B_1: student knows the correct answer.
B_2: student does not know the correct answer.

By Bayes' theorem,

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} = \frac{1 \times 0.7}{1 \times 0.7 + 0.25 \times 0.3} = 0.9032.$$

(Note that the probability that a student answers correctly when he does not know the right answer is ¼, since there are four choices to each question, one out of which is the right answer).

The probability that the student who answered correctly did not actually know the answer is

$1 - P(B_1|A) = 0.0968$. **[2]**

[*Direct computation of* $P(B_1|A)$ *should also fetch full credit.*]

Question 2

(i) The distribution of claim amount in lacs is $f(x) = \lambda \exp(-\lambda x)$, $x > 0$.
Therefore, the likelihood function of λ for n claim amounts, $x_1, x_2, .., x_n$, is

$$\prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \lambda \exp(-\lambda x_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right).$$

When viewed as a function of λ, this function is proportional to the probability density function of a gamma distribution. It is well known that the product of two gamma density functions produces another gamma density function, and therefore the gamma distribution can serves as a conjugate prior for λ.

[*Only one mark should be awarded for correct specification of prior without justification.*]

(ii) Let the prior density be $g(\lambda) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} \lambda^{\alpha-1} \exp(-\lambda\beta)$.

From the given information, we have mean $\alpha / \beta = 1.95$ and variance $\alpha / \beta^2 = 0.01$.

By solving these equations, we have $\alpha = 380.25$ and $\beta = 195$.

The posterior probability density function is proportional to the likelihood times the prior density, i.e, it is proportional to

$$\lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right) \lambda^{\alpha-1} \exp(-\lambda\beta) = \lambda^{n+\alpha-1} \exp\left[-\lambda\left(\beta + \sum_{i=1}^{n} x_i\right)\right].$$

This expression shows that the posterior is gamma with parameters $n + \alpha$ and $\beta + \sum_{i=1}^{n} x_i$.

Further,

$n+\alpha = 100 + 380.25 = 480.25,$

$$\beta + \sum_{i=1}^{n} x_i = 195 + 100 \times 0.5 = 245.$$

The Bayes estimate of λ with respect to the squared error loss function is the posterior mean, namely,

$$\frac{n+\alpha}{\beta + \sum_{i=1}^{n} x_i} = \frac{480.25}{245} = 1.96.$$

[7]

Question 3

Let d_1 denote the decision to buy annual membership and d_2 denote the decision to borrow individual books. The total cost under each scenario is:

No. of books	d_1	d_2
1	850	175
2	850	350
3	850	525
4	850	700
5	850	875

Maximum cost under d_1 = 850
Maximum cost under d_2 = 875
Minimax decision is d_1 (buy annual membership).

Expected cost under d_1 = 850
Expected cost under d_2 = $175 \times 0.15 + 350 \times 0.2 + 525 \times 0.33 + 700 \times 0.22 + 875 \times 0.1 = 511$
Bayes' decision is d_2 (borrow books individually).

[3]

Question 4

(i) The unrestricted probability density function is

$$f(x) = \frac{a}{3} x^{-2/3} \exp\left(-ax^{1/3}\right).$$

Therefore, the truncated probability density function for the claims observed by the reinsurer is

$$\frac{f(x)}{1-F(R)} = \frac{\frac{a}{3} x^{-2/3} \exp\left(-ax^{1/3}\right)}{\exp\left(-aR^{1/3}\right)} = \frac{a}{3} x^{-2/3} \exp\left[-a\left(x^{1/3} - R^{1/3}\right)\right], \text{ where } R = 50{,}000.$$

The likelihood function of a for the 50 claims observed by the reinsurer is.

$$\prod_{i=1}^{50} \frac{f(x_i)}{1-F(R)} = \left(\frac{a}{3}\right)^{50} \left(\prod_{i=1}^{50} x_i^{-2/3}\right) \exp\left\{-a\sum_{i=1}^{50}\left(x_i^{1/3} - R^{1/3}\right)\right\},$$

and the log-likelihood is

$$l(a) = 50\log a - a\sum_{i=1}^{50}\left(x_i^{1/3} - R^{1/3}\right)$$

plus terms that do not depend on a. By differentiating this above expression and setting it equal to zero, we have the likelihood equation

$$l'(a) = \frac{50}{a} - \sum_{i=1}^{50}\left(x_i^{1/3} - R^{1/3}\right) = 0.$$

The solution to this equation is

$$a = \frac{50}{\sum_{i=1}^{50}\left(x_i^{1/3} - R^{1/3}\right)} = \frac{50}{2600 - 50(50000)^{1/3}} = 0.06596.$$

Since the second derivative of the likelihood function,

$$l''(a) = -\frac{50}{a^2},$$

is always negative, the above location is indeed the *maximum* of the likelihood function.

(ii) The asymptotic variance of the maximum likelihood estimator is

$$\frac{1}{E[-l''(a)]} = \frac{a^2}{50}.$$

The corresponding standard deviation is $a/\sqrt{50}$, which may be estimated as $0.06596/\sqrt{50} = 0.009329$.

(iii) Under the revised scenario, the reinsurer has censored data (rather than truncated data). The likelihood function of a for the specified data is

$$[F(R)]^{600} \times \prod_{i=1}^{50} f(x_i) = \left[1 - \exp\left(-aR^{1/3}\right)\right]^{600} \left(\frac{a}{3}\right)^{50} \left(\prod_{i=1}^{50} x_i^{-2/3}\right) \exp\left(-a\sum_{i=1}^{50} x_i^{1/3}\right).$$

The log-likelihood is

$$l(a) = 600\log\left[1 - \exp\left(-aR^{1/3}\right)\right] + 50\log a - a\sum_{i=1}^{50} x_i^{1/3}.$$

Thus, the likelihood equation is

$$l'(a) = \frac{600R^{1/3}\exp\left(-aR^{1/3}\right)}{1-\exp\left(-aR^{1/3}\right)} + \frac{50}{a} - \sum_{i=1}^{50} x_i^{1/3} = 0,$$

i.e.,

$$\frac{600 \times (50000)^{1/3} \times \exp\left(-a(50000)^{1/3}\right)}{1-\exp\left(-a(50000)^{1/3}\right)} + \frac{50}{a} - 2600 = 0.$$

[10]

Question 5

Cumulative payments in monetary amounts

Accident Year	*Development Year* 0	1	2	3	4
2007	100	110	120	125	130
2008	110	120	125	130	
2009	105	115	125		
2010	95	105			
2011	120				

Incremental payments in monetary amounts

Accident Year	*Development Year* 0	1	2	3	4
2007	100	10	10	5	5
2008	110	10	5	5	
2009	105	10	10		
2010	95	10			
2011	120				

Incremental payments in real amounts (mid-2011 prices)

Accident Year	*Development Year* 0	1	2	3	4
2007	126.80	12.08	11.50	5.43	5.00
2008	132.84	11.50	5.43	5.00	
2009	120.76	10.85	10.00		
2010	103.08	10.00			
2011	120.00				

Cumulative payments in real amounts (mid-2011 prices)

Accident Year	*Development Year* 0	1	2	3	4
2007	126.80	138.87	150.38	155.80	160.80
2008	132.84	144.34	149.76	154.76	
2009	120.76	131.61	141.61		
2010	103.08	113.08			
2011	120.00				

Age-to-Age Factors	1.092	1.065	1.035	1.032	1.020

Note: 2007 is not fully run-off. Hence we need to calculate a tail factor.
Claims are assumed to be 98% paid by development year 4
and fully run-off by development year 5
Hence, the tail factor is calculated as 1 / 0.98 = 1.02041

Completing the triangle of future payments at mid-2011 prices

Accident Year	*Development Year* 0	1	2	3	4	5
2007					160.80	164.08
2008				154.76	159.73	162.99
2009			141.61	146.53	151.23	154.32
2010		113.08	120.41	124.60	128.60	131.22
2011	120.00	131.03	139.53	144.38	149.01	152.05

Future incremental payments at mid-2011 prices

Accident Year	*Development Year* 0	1	2	3	4	5
2007						3.28
2008					4.97	3.26
2009				4.92	4.70	3.09
2010			7.34	4.18	4.00	2.62
2011		11.03	8.50	4.85	4.63	3.04

Incremental payments adjusted for future inflation

Accident Year	*Development Year* 0	1	2	3	4	5
2007						3.61
2008					5.46	3.94
2009				5.41	5.69	4.11
2010			8.07	5.06	5.32	3.84
2011		12.13	10.29	6.45	6.78	4.90

(i) Reserve to be held at the end of 2011 (in monetary amounts) for future payments = 91.08.

(ii) Monetary amount of payments to be made in 2012 on this portfolio

= 12.13 + 8.07 + 5.41 + 5.46 + 3.61 = 34.69.

(iii) The key assumption underlying this method is that, for each origin year, the amount of claims paid, in real terms, in each development year is a constant proportion of the total claims, in real terms, from that origin year.
Explicit assumptions are made for both past and future claims inflation.

[14]

Question 6

(i) For the target distribution, the probability mass function is

$$f(x) = P(X = x) = \begin{cases} 0.70 & \text{for } x = 0, \\ 0.15 & \text{for } x = 1, \\ 0.08 & \text{for } x = 2, \\ 0.05 & \text{for } x = 3, \\ 0.02 & \text{for } x = 4. \end{cases}$$

On the other hand, for the discrete uniform distribution, the probability mass function is

$$h(x) = P(X = x) = \begin{cases} 0.2 & \text{for } x = 0, \\ 0.2 & \text{for } x = 1, \\ 0.2 & \text{for } x = 2, \\ 0.2 & \text{for } x = 3, \\ 0.2 & \text{for } x = 4. \end{cases}$$

Thus, $f(x) \le Ch(x)$ for all *x*, if $C = 0.7/0.2 = 3.5$, and in such a case, we can define the ratio

$$g(x) = \frac{f(x)}{Ch(x)} = \begin{cases} 1 & \text{for } x = 0, \\ 0.214 & \text{for } x = 1, \\ 0.114 & \text{for } x = 2, \\ 0.071 & \text{for } x = 3, \\ 0.029 & \text{for } x = 4. \end{cases}$$

Therefore, in order to generate random samples from target distribution, we would generate samples from the discrete uniform distribution, and accept the values with probability $g(x)$ as described above.

This acceptance/rejection decision can be made on the basis of a random sample from the Bernoulli distribution. Specifically, when the generated sample from the discrete uniform distribution is *x*, it is accepted if the sample from the Bernoulli distribution (with $p = g(x)$) is 1.

(ii) The overall acceptance probability is

$$E_h[g(X)] = \sum_{x=0}^{4} g(x)h(x) = \frac{1}{C}\sum_{x=0}^{4} f(x) = \frac{1}{C} = \frac{1}{3.5} = 0.2857,$$

i.e., on the average, 28.57% of the generated samples will be accepted.

(iii) Alternatively, we may use the inverse transformation method for the cumulative distribution function

$$F(x) = P(X \le x) = \begin{cases} 0.70 & \text{for } 0 \le x < 1, \\ 0.85 & \text{for } 1 \le x < 2, \\ 0.93 & \text{for } 2 \le x < 3, \\ 0.98 & \text{for } 3 \le x < 4, \\ 1 & \text{for } x \ge 4. \end{cases}$$

Thus, after generating a sample U from the uniform distribution over [0,1], one could choose the sample X from the target distribution as follows.

$$X = \begin{cases} 0 & \text{if } U \le 0.7, \\ 1 & \text{if } 0.70 < U \le 0.85, \\ 2 & \text{if } 0.85 < U \le 0.93, \\ 3 & \text{if } 0.93 < U \le 0.98, \\ 4 & \text{if } U > 0.98. \end{cases}$$

[9]

Question 7

(i) a. True.
b. True.
c. False.

(ii) Under the assumption of a white noise process, the mean and the variance of the number of turning points (T) in a series of length N are

$$E(T) = \frac{2(N-2)}{3}, \quad V(T) = \frac{16N - 29}{90}.$$

In the present case, $N = 800$. Therefore,

$$E(T) = \frac{2 \times 798}{3} = 532, \quad V(T) = \frac{12800 - 29}{90} = \frac{12771}{90} = 141.9.$$

The two-sided p-value of the observed number of turning points (499) is

$$2 \times \left[1 - \Phi\left(\frac{|499 - 532|}{\sqrt{141.9}}\right)\right] = 2 \times [1 - \Phi(2.7703)] = 0.0056.$$

As the p-value is very small, the residuals cannot be said to have come from a white noise process.

[Alternatively, a 95% confidence interval for the number of turning points, under the white noise assumption, is $[532 - 1.96 \times \sqrt{141.9}, \quad 532 + 1.96 \times \sqrt{141.9}]$, i.e., [509,555]. Since the observed number lies outside this range, the hypothesis of white noise is rejected.]

[6]

Question 8

(i) Assume that claim sizes are independent and identically distributed,
that claim sizes are independent of the claims arrival process, and
that premiums are received continuously at constant rate.

(ii) $f(x) = 0.5e^{-x}(1+2e^{-x})$, x > 0.

Expected claim size is given by $\int_0^\infty x f(x)dx$

$$= \frac{1}{2}\int_0^\infty xe^{-x}(1+2e^{-x})dx$$

$$= \frac{1}{2}\int_0^\infty xe^{-x}dx + \int_0^\infty xe^{-2x}dx$$

$$= \frac{1}{2}(-xe^{-x} - e^{-x})\Big|_0^\infty + (-\frac{xe^{-2x}}{2} - \frac{e^{-2x}}{4})\Big|_0^\infty$$

= ½ + ¼ = ¾.

(iii) λ is the Poisson parameter.
θ = 0.25 is the premium loading factor.
$m_1 = E(X)$, X representing the claim size, and $m_2 = E(X^2)$.
Let μ = ¾ (calculated above) be the mean claim size.

Then the adjustment coefficient R is the positive solution of

$$M_X(r) = 1 + (1+\theta)\mu r\ .$$

$$M_X(r) = \int_0^\infty e^{rx} f(x)dx$$

$$= \frac{1}{2}\int_0^\infty e^{rx}e^{-x}(1+2e^{-x})dx$$

$$= \frac{1}{2}\int_0^\infty e^{-(1-r)x}dx + \int_0^\infty e^{-(2-r)x}dx$$

$$= \frac{1}{2(1-r)} + \frac{1}{(2-r)},$$

for $r < 1$.

Thus, R is obtained by solving

$$1 + 1.25\times\frac{3}{4}r = \frac{1}{2(1-r)} + \frac{1}{(2-r)}.$$

$$\Rightarrow 30r^3 - 58r^2 + 12r = 0$$

$$\Rightarrow 15r^2 - 29r + 6 = 0; r \neq 0$$

$$\Rightarrow r = \frac{29 \pm \sqrt{29^2 - 4\times15\times6}}{2\times15}$$

$$\Rightarrow R = 1.698,\quad 0.236.$$

Further, $R < 1$ implies $R = 0.236$.

Now, $m_1 = \frac{d}{dr} M_X(r)\Big|_{r=0} = E(X) = 0.75$;

$$m_2 = \frac{d^2}{dr^2} M_X(r)\Big|_{r=0} = \left((1-r)^{-3} + 2.(2-r)^{-3}\right)\Big|_{r=0} = 1.25 .$$

$$\therefore 2\theta \frac{m_1}{m_2} = \frac{2 \times 0.25 \times 0.75}{1.25} = 0.3,$$

$$\therefore R = 0.236 < 2\theta \frac{m_1}{m_2} = 0.3.$$

(iv) Lundberg's inequality gives us an upper bound for the probability of ruin:
$\psi(15) \le e^{-15R} = 0.029$.

(v) Now, if the claim sizes $X^{\#} \sim \exp(1/\mu)$,

$$M_{X^{\#}}(r) = \frac{1}{1-\mu r}, r < \frac{1}{\mu}.$$

Thus, $R^{\#}$ is obtained by solving

$$1 + 1.25 \times 0.75\text{r} = \frac{1}{1-0.75r}$$

$$\Rightarrow 0.703125r^2 - 0.1875r = 0$$

$$\Rightarrow r = \frac{0.1875}{0.703125}; r \ne 0$$

$$\Rightarrow R^{\#} = 0.267.$$

Lundberg's inequality gives us a modified upper bound for the probability of ruin
$\psi(15) \le e^{-15R^{\#}} = 0.018$.

$R^{\#}$ is more than R. In other words, the probability of ruin reduces when the claim sizes follow the exponential distribution instead of the originally presumed distribution, even though there is no change in the mean claim amount and initial surplus.

[13]

Question 9

(i) X: Claim amount from the risk

$X|\theta \sim N(\theta, \sigma_1^2)$.

Prior distribution of θ is $N(\mu, \sigma_2^2)$

$$\Rightarrow f(\theta) \propto \exp\left[-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma_2}\right)^2\right].$$

Likelihood of $X_j|\theta$ is proportional to $\prod_{i=1}^{n} \exp\left[-\frac{1}{2}\left(\frac{x_i-\theta}{\sigma_1}\right)^2\right] = \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\theta}{\sigma_1}\right)^2\right]$

=> Posterior is proportional to

$$\exp\left[-\frac{1}{2}\left\{\left(\frac{\theta-\mu}{\sigma_2}\right)^2 + \sum_{i=1}^{n}\left(\frac{x_i-\theta}{\sigma_1}\right)^2\right\}\right] = \exp\left[-\frac{1}{2}\left(\frac{\theta-\mu^*}{\sigma^*}\right)^2\right] \times K .$$

Equating coefficients of θ and θ^2

$$\frac{\mu^*}{\sigma^{*2}} = \frac{\mu}{\sigma_2^2} + \left(\frac{1}{\sigma_1^2}\sum_{i=1}^{n} x_i\right) \text{ and } \frac{1}{\sigma^{*2}} = \frac{1}{\sigma_2^2} + \frac{n}{\sigma_1^2}$$

$$\Rightarrow \sigma^{*2} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + n\sigma_2^2}$$

and $\mu^* = \dfrac{\mu\sigma_1^2 + \sigma_2^2\sum_{i=1}^{n} x_i}{\sigma_1^2 + n\sigma_2^2}$

Under the quadratic loss function, the mean of the posterior is the Bayesian estimator of the unknown parameter θ

$$\Rightarrow \hat{\theta} = \frac{\mu\sigma_1^2 + \sigma_2^2.n.\bar{x}}{\sigma_1^2 + n\sigma_2^2} = \frac{n\sigma_2^2}{\sigma_1^2 + n\sigma_2^2}\bar{x} + \frac{\sigma_1^2}{\sigma_1^2 + n\sigma_2^2}\mu$$

$$= z\bar{x} + (1-z)\mu$$

where $z = \dfrac{n\sigma_2^2}{\sigma_1^2 + n\sigma_2^2} = \dfrac{n}{n + \dfrac{\sigma_1^2}{\sigma_2^2}}$ is the credibility factor

(ii) For $\sigma_1^2 = 100^2$, $\hat{\theta} = \dfrac{(500\times 100^2) + (250^2 \times 3195)}{100^2 + (10\times 250^2)} = 322.34$ and

for $\sigma_1^2 = 350^2$, $\hat{\theta} = \dfrac{(500\times 350^2) + (250^2 \times 3195)}{350^2 + (10\times 250^2)} = 349.08$

(iii) Risk premium is higher when $\sigma_1^2 = 350^2$ than when $\sigma_1^2 = 100^2$.

The credibility factor z is a decreasing function of σ_1^2. Therefore, as σ_1^2 increases, the value of z decreases and more emphasis is placed on the prior distribution's mean and lesser on the sample mean.

Prior mean = μ = 500 and sample mean = 319.5 in both cases.

The risk premium derived above in the case of $\sigma_1^2 = 350^2$ is closer to the prior mean μ than in the case of $\sigma_1^2 = 100^2$, as the data becomes less important in the latter case.

[10]

Question 10

$\bar{X}_A = (5+7+5+6)/4 = 5.75,$

$\bar{X}_B = 30,$

$\bar{X}_C = 110.$

$$\bar{X} = \frac{1}{3}\sum_{i=1}^{3}\bar{X}_i = 48.58$$

$$\therefore E[m(\theta)] = \bar{X} = 48.58.$$

$$E[s^2(\theta)] = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{n-1}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2\right\}$$

$$= \frac{1}{3}\sum_{i=1}^{3}\left\{\frac{1}{3}\sum_{j=1}^{3}(X_{ij} - \bar{X}_i)^2\right\} = 55.86.$$

[$\because$ for I = Fleet A, $\frac{1}{n-1}\sum_{j=1}^{n}(X_{ij}-\overline{X}_i)^2$

$=\frac{1}{3}\left[(5-5.75)^2+(7-5.75)^2+(5-5.75)^2+(6-5.75)^2\right]$

$=0.92.$

Similarly for Fleet B, 50 and for Fleet C, 116.67]

$$Var[m(\theta)]=\frac{1}{N-1}\sum_{i=1}^{N}(\overline{X}_i-\overline{X})^2-\frac{1}{Nn}\sum_{i=1}^{N}\left\{\frac{1}{n-1}\sum_{j=1}^{n}\left(X_{ij}-\overline{X}_i\right)^2\right\}$$

$$=\frac{1}{2}\left[(5.75-48.58)^2+(30-48.58)^2+(110-48.58)^2\right]-\frac{1}{4}\times 55.86$$

$$=2976.02-13.97=2962.06.$$

So the credibility factor is given by: $z=\frac{n}{\left(n+\frac{E(s^2(\theta))}{Var(m(\theta))}\right)}=\frac{4}{4+\frac{55.86}{2902.06}}=0.99531.$

EBCT premium (in Rs.'000s) for

Fleet A = $(5.75\times 0.99531)+(1-0.99531)\times 48.58\approx 5.95$;

Fleet B ≈ 30.09;

Fleet C ≈ 109.71. **[6]**

Question 11

(i) $f(x)=\frac{1}{\mu}e^{-\frac{x}{\mu}}=\exp\left(-\frac{x}{\mu}-\ln(\mu)\right)$, which is in the form of the exponential family of distributions.

(ii) The canonical parameter θ is $-\frac{1}{\mu}$;

Dispersion parameter $\phi=1$; $a(\phi)=1$.

$b(\theta)=\ln(\mu)=\ln\left(-\frac{1}{\theta}\right)=-\ln(-\theta)$,

$b'(\theta)=-1/\theta$,

$b''(\theta)=1/\theta^2=\mu^2$.

The variance function $V(\mu)=\mu^2$; the variance of X is $a(\phi)V(\mu)=\mu^2$.

Alternatively, the canonical parameter θ is $\frac{1}{\mu}$;

Dispersion parameter $\phi=1$; $a(\phi)=-1$.

$b(\theta)=-\ln(\mu)=-\ln\left(\frac{1}{\theta}\right)=\ln\theta$,

$b'(\theta)=1/\theta$,

$b''(\theta)=-1/\theta^2=-\mu^2$.

The variance function $V(\mu)=-\mu^2$ and the variance of X is $a(\phi)V(\mu)=\mu^2$.

[4]

Question 12

(i) For a random variable X having the geometric distribution

$P(X = x) = p(1-p)^x$, $x = 0,1,2,\ldots,$

the MGF is

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} p(1-p)^x = \frac{p}{1-(1-p)e^t},$$

so that

$$M'_X(t) = \frac{p(1-p)e^t}{\left[1-(1-p)e^t\right]^2};\quad M''_X(t) = \frac{p(1-p)e^t\left[1-(1-p)e^t\right]^2 + 2\left[1-(1-p)e^t\right]p(1-p)^2 e^{2t}}{\left[1-(1-p)e^t\right]^4};$$

and

$$E(X) = M'(0) = \frac{1-p}{p};\quad E(X^2) = M''(0) = \frac{(1-p)(2-p)}{p^2};$$

$$V(X) = E(X^2) - E(X)^2 = \frac{(1-p)}{p^2}.$$

Since N has the distribution of X with $p = 0.25$, it follows that

$$M_X(t) = \frac{0.25}{1-0.75e^t};\quad E(N) = \frac{0.75}{0.25} = 3;\qquad V(N) = \frac{0.75}{0.25^2} = 12.$$

(ii) Let us use, for simplicity, the notation U in lieu of U_i. Since $U - 1$ has the distribution of X with $p = 0.75$, it follows that

$$E(U-1) = \frac{0.25}{0.75} = \frac{1}{3},\ \text{i.e., } E(U) = \frac{4}{3},$$

$$V(U) = V(U-1) = \frac{0.25}{0.75^2} = \frac{4}{9}.$$

(iii) $$E(S) = E(U)\times E(N) = \frac{4}{3}\times 3 = 4.$$

$$V(S) = [E(U)]^2 \times V(N) + V(U)\times E(N) = \left(\frac{4}{3}\right)^2 \times 12 + \frac{4}{9}\times 3 = \frac{68}{3}.$$

(iv) For a given accident, probability that the number of fatalities is more than 2 is

$$P(U > 2) = 1 - P(U = 1) - P(U = 2) = 1 - 0.75 - 0.75\times 0.25 = \frac{1}{16}.$$

Let $J_i = \begin{cases} 1 & \text{if } i^{\text{th}} \text{ fatal accident involves more than 2 fatalities,} \\ 0 & \text{if } i^{\text{th}} \text{ fatal accident involves one or 2 fatalities.} \end{cases}$

Then, $P(J_i = 1) = 1/16$.

Let $N^{\#}$ be the number of fatal accidents involving more than 2 fatalities. Given N, the distribution of $N^{\#}$ is binomial with parameters N and 1/16. Therefore, for $i = 0,1,2,\ldots,$

$$
\begin{aligned}
P\left(N^{\#}=i\right) &= \sum_{n=0}^{\infty} P\left(N^{\#}=i \middle| N=n\right) \times P(N=n) \\
&= \sum_{n=i}^{\infty} \binom{n}{i} \left(\frac{1}{16}\right)^{i} \left(\frac{15}{16}\right)^{n-i} \left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^{n} \\
&= \left(\frac{1}{4}\right)\left(\frac{3}{64}\right)^{i} \sum_{n=i}^{\infty} \binom{n}{n-i} \left(\frac{45}{64}\right)^{n-i} \\
&= \left(\frac{1}{4}\right)\left(\frac{3}{64}\right)^{i} \sum_{n=0}^{\infty} \binom{n+i}{n} \left(\frac{45}{64}\right)^{n} \\
&= \left(\frac{1}{4}\right)\left(\frac{3}{64}\right)^{i} \left(\frac{64}{19}\right)^{i+1} \sum_{n=0}^{\infty} \binom{n+i}{n} \left(\frac{45}{64}\right)^{n} \left(\frac{19}{64}\right)^{i+1} \\
&= \left(\frac{16}{19}\right)\left(\frac{3}{19}\right)^{i},
\end{aligned}
$$

which shows that $N^{\#}$ has another geometric distribution with parameter 16/19. Therefore,

$$
E\left(N^{\#}\right) = \frac{3/19}{16/19} = 3/16;
$$

$$
V\left(N^{\#}\right) = \frac{3/19}{\left(16/19\right)^{2}} = 57/256;
$$

Let $U^{\#}$ be the number of fatalities in excess of 2 in a particular accident involving more than 2 fatalities. Then, for $u = 0,1,2,\ldots$,

$$
P\left(U^{\#}=u\right) = P\left(U=u+2 \middle| U>2\right) = \frac{P(U=u+2)}{P(U>2)} = \frac{\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^{u+1}}{\left(\frac{1}{16}\right)} = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^{u-1},
$$

which is the same as the distribution of U, i.e., $E\left(U^{\#}\right) = \frac{4}{3}$, $V\left(U^{\#}\right) = \frac{4}{9}$.

Hence, $E(Y) = E\left(U^{\#}\right) \times E\left(N^{\#}\right) = \frac{4}{3} \times \frac{3}{16} = \frac{1}{4}$.

$$
V(Y) = \left[E\left(U^{\#}\right)\right]^{2} \times V\left(N^{\#}\right) + V\left(U^{\#}\right) \times E\left(N^{\#}\right) = \left(\frac{4}{3}\right)^{2} \times \frac{57}{256} + \frac{4}{9} \times \frac{3}{16} = \frac{23}{48}.
$$

[16]

****