Problem Sheet 4

1. Show that
\[
\int_0^\infty \int_0^y (x + y) e^{-(x+y)} dx dy = \int_0^\infty \int_0^\infty I(x, y) (x + y) e^{-(x+y)} dx dy,
\]
where
\[
I(x, y) = \begin{cases} 
1, & x < y; \\
0, & x \geq y.
\end{cases}
\]
Use simulation to evaluate the following integrals, carefully explaining your procedure and comparing your answer with the exact answer.

(i) \[\int_0^{\pi/2} \frac{d\theta}{1 + \frac{2}{3} \cos \theta},\]
(ii) \[\int_0^\infty \int_0^y (x + y) e^{-(x+y)} dx dy.\]

2. Let \( U_1 \) and \( U_2 \) each be independent uniform random variables on \((0, 1)\). What is the joint distribution of \((X, Y)\) where \(X = 2U_1 - 1, Y = 2U_2 - 1?\)

Let \( I \) be the random variable
\[
I = \begin{cases} 
1, & X^2 + Y^2 \leq 1; \\
0, & \text{otherwise}.
\end{cases}
\]
Find \( E(I) \).

Use your result to derive a simulation method of estimating \( \pi \). Generate 1,000 pairs of uniform random numbers and use them to obtain an estimate.

3. In 1969 a number of married women from a district in Guatemala, all born during the period 1935 – 44, were interviewed and asked the age at which they were married (Scholl, T.O., Odell, M.E. and Johnston, F.E. (1976) Biological correlates of modernization in a Guatemalan highland municipio. *Annals of Human Biology*, 3, 23 – 32). The data are given below.

<table>
<thead>
<tr>
<th>Age</th>
<th>9, 10</th>
<th>11, 12</th>
<th>13, 14</th>
<th>15, 16</th>
<th>17, 18</th>
<th>19, 20</th>
<th>21, 22</th>
<th>23, 24</th>
<th>25, 26</th>
<th>( \geq 27 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>( \geq 9 )</td>
</tr>
<tr>
<td>Number</td>
<td>5</td>
<td>11</td>
<td>18</td>
<td>28</td>
<td>8</td>
<td>7</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Test the goodness of fit of a Poisson distribution to these data by carrying out a \( \chi^2 \) test (a) without pooling, (b) with pooling. Comment on the plausibility of your results.

Simulate 5,000 of these data sets from your Poisson distribution. Show how these can be used to carry out your goodness-of-fit test.
4. Let $X$ have a geometric distribution with probability mass function

$$p_X(x) = \theta (1 - \theta)^{x-1}, \quad x = 1, 2, \ldots \quad 0 < \theta < 1.$$ 

Show that $Y = \sum_{i=1}^{n} X_i$ has a negative binomial distribution with probability mass function

$$p_Y(y) = \binom{y-1}{n-1} \theta^n (1 - \theta)^{y-n}, \quad y = n, n+1, \ldots$$

(a) Show how to simulate from the distribution of $X$.
(b) Hence obtain an algorithm for simulating from $Y$.
(c) Obtain another algorithm for simulating from $Y$ by using a recurrence relation.
(d) The negative binomial distribution may be regarded as the number of independent Bernoulli trials, each with probability of success $\theta$, needed to obtain a total of $n$ successes. Use this idea to produce a third algorithm for simulating from $Y$.

Compare your three algorithms. Which do you prefer?