TMA4265 Stochastic Modelling: Project 1

Kim-Iver Blindheimsvik, Max Pfisterer

Problem 1: Modelling an outbreak of Measles

We are considering a SIR-model of an infectious disease where infected individuals go through stages of susceptibility, infection, and a recovered state, immune from infection. The probabilities of transition from each state to the next during one time step is initially kept constant. The probability of infection for a susceptible individual is $0 < \beta < 1$. An infected individual has a probability $0 < \gamma < 1$ of recovering, and a recovered individual has a probability of $0 < \alpha < 1$ of becoming susceptible again. Each state can only stay in its current state or transition to the next, e.g. a susceptible individual has a $1 - \beta$ probability of of staying susceptible.

a)

Letting X_n be the state of an individual at time n with S, I, R corresponding to states 0, 1, and 2 respectively. $\{X_n : n = 0, 1, \ldots\}$ is a Markov chain:

$\sum_{i=0}^{2} P_{0i} = \sum_{i=0}^{2} P_{1i} = \sum_{i=0}^{2} P_{2i} = 1$, also $P\{X_{n+1} = x | X_n = y\} = P\{X_n = x | X_{n-1} = y\}$, i.e. the transition probabilities are only dependent on the current state.

By the previous description and explanation of the transition probabilities we can set up the following transition probability matrix:

$$\mathbf{P} = \begin{bmatrix} 1-\beta & \beta & 0 \\ 0 & 1-\gamma & \gamma \\ \alpha & 0 & 1-\alpha \end{bmatrix}$$

b)

Assuming $\beta = 0.01$, $\gamma = 0.10$, and $\alpha = 0.005$,

$$\mathbf{P} = \begin{bmatrix} 0.99 & 0.01 & 0 \\ 0 & 0.90 & 0.10 \\ 0.005 & 0 & 0.995 \end{bmatrix} \qquad \mathbf{P^2} = \begin{bmatrix} 0.980100 & 0.018900 & 0.001000 \\ 0.000500 & 0.810000 & 0.189500 \\ 0.009925 & 0.000050 & 0.990025 \end{bmatrix}$$

$\mathbf{P^2}$ is regular, so we can guarantee that it has a limiting distribution. We get the long-run mean number of days per year spent in each state by calculating the limiting distribution $\vec{\pi}$ of $\mathbf{P}$, which is given by the two conditions: $\pi_j = \sum_{i=0}^{2} \pi_i P_{ij}$ and $\sum_{i=0}^{2} \pi_i = 1$.

From these equations we get an over-determined system, which we solve by writing out π_0, π_1 and using the second condition, giving us three equations for three unknowns:

$$\text{I}: \quad \pi_0 = 0.99\pi_0 + 0.005\pi_2 \tag{1}$$

$$\text{II}: \quad \pi_1 = 0.01\pi_0 + 0.90\pi_1 \tag{2}$$

$$\text{III}: \quad \pi_0 + \pi_1 + \pi_2 = 1 \tag{3}$$

Giving us a system we can solve with Gaussian elimination (omitting the intermediary steps for brevity):

$$\begin{bmatrix} \pi_0 \\ \pi_1 \\ \sum_{i=0}^{2} \pi_i \end{bmatrix} = \left[\begin{array}{rrr|r} 0.01 & 0 & -0.005 & 0 \\ -0.01 & 0.10 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{10}{31} \\ 0 & 1 & 0 & \frac{1}{31} \\ 0 & 0 & 1 & \frac{20}{31} \end{array}\right]$$

From which we can calculate the long-run mean number of days per year spent in each state

$$365 \cdot \vec{\pi} = \begin{bmatrix} 117.7 \\ 11.8 \\ 235.5 \end{bmatrix}$$

c)

Assuming an individual is susceptible at time 0, i.e. $X_n = 0$, we first simulate the Markov chain for 7300 time steps (20 years), using the last 10 years for each run to estimate the long-run mean number of days per year spent in each state ($365 \cdot \vec{\pi}$). Then, using 30 simulations like the one just described we will compute an approximate 95% confidence interval (CI) for the long-run mean number of days per year spent in each state.

```
## Days spent in state 0: 107 , state 1: 12.5 , and state 2: 245.5
## , during an average year of the last 10 years

## Confidence interval for days spent in state 0: [ 55.59922 , 167.1941 ]

## Confidence interval for days spent in state 1: [ 2.90884 , 19.63782 ]

## Confidence interval for days spent in state 2: [ 187.4041 , 297.2559 ]
```

The confidence intervals were calculated using what we have leaned in a previous course about confidence intervals for normal distributions, using that we independently estimated our $\vec{\pi}$ 30 times, and that our distribution of the mean of our estimates for pi tend to a normal distribution due to the central limit theorem.

d)

$Y_n = (S_n, I_n, R_n),\ \ Z_n = (S_n, I_n),\ \ n \geq 0$. Given that we know $N = S_n + I_n + R_n$, i.e. the total number of individuals is constant, and that each individual has to be in one of the three states, $Z_n \sim Y_n$ as we can calculate $R_n = N - S_n - I_n$ and we know/can calculate all transition probabilities. $\{Y_n : n = 0, 1, \ldots\}$ is a Markov chain as it is a discrete time stochastic process with a discrete (finite) state space, and Y_{n+1} is completely independent of $Y_{n-k},\ \ k \geq 1$, so it satisfies the Markov property. Therefore both $\{Y_n : n = 0, 1, \ldots\}$ and $\{Z_n : n = 0, 1, \ldots\}$ are markov chains, as they are equivalent. $\{I_n : n = 0, 1, \ldots\}$ is not a markov chain as we cannot determine $\{I_{n+1}\}$ if $I_n = i < N$ (as then we neither know S_n or R_n). If we also know that $I_{n-1} = N$, however, we can estimate I_{n+1}, as then $R_n = N - i$ and $S_n = 0$. Therefore the Markov property is violated for $\{I_n : n = 0, 1, \ldots\}$ (I_n depends on more than the previous state), so it is not a Markov chain.

e)

We want to model a SIR model like the one described earlier, with probabilities of infection $\beta_n = 0.5\frac{I_n}{N}$, recovery $\gamma = 0.1$, and becoming susceptible to infection after recovery $\alpha = 0.005$. The total population is $N = 1000$ and we know that this is a Markov chain from task **1 d)**.

Plotting one simulated realization of the mentioned process Y_n, with the amount of susceptible individuals in red, the amount recovered in blue, and the amount infected in yellow on the next page, together with task 1g.

We see that we get a spike in infections after only a few timesteps, and that the vast majority of individuals are in either a susceptible or recovered state. As no-one is recovered in the initial state, and the vast majority are susceptible, we get a large increase in the infected population, as susceptible individuals can only stay susceptible or get infected. As the amount of infected people increases, so does the infection rate, creating an explosive outbreak. The probability of staying in the same state is highest for the recovered, 99.5%, but it varies for the susceptible, being 97.5% at the beginning, and decreasing during the outbreak. For the time interval 50-300 the outbreak has been slowed down by the slow transition from recovered to susceptible, and as it for the most part is more likely to stay in the recovered state we get a period of low infection rate, due to high amounts of recovered individuals, caused by the initial outbreak.

f)

Based on 1000 simulations of the measles outbreak as described in **e)**, we get these estimates:

```
## [1] 580.063
```

```
## [1] 11.93
```

Which are our $E[max\{I_0, ..., I_{300}\}]$, and $E[min\{arg \max_{n \leq 300} I_n\}]$ respectively.

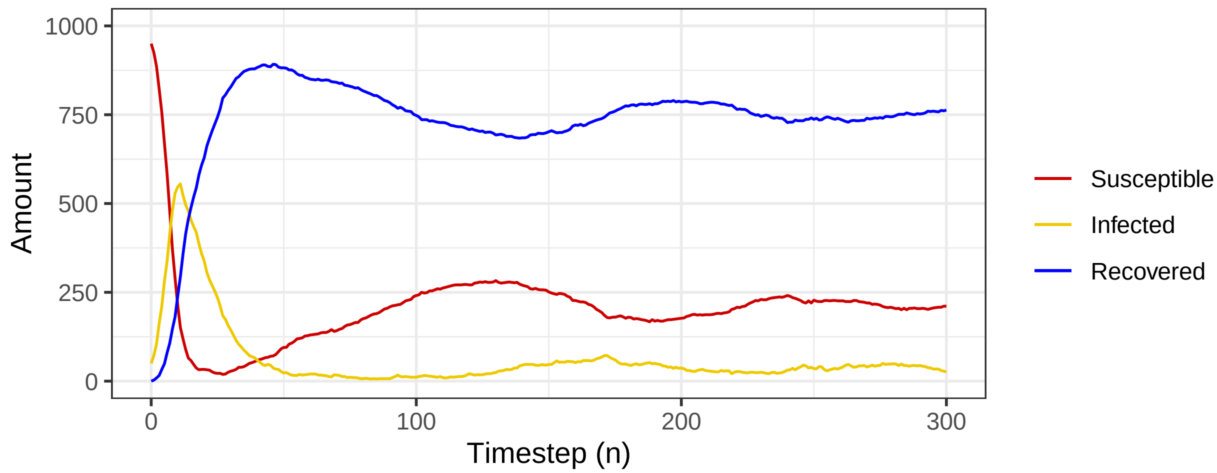
```
## Confidence interval for maximum infected individuals: [ 564.5645 , 595.5615 ]
```

```
## Confidence interval for expected first time of infection peak: [ 10.66964 , 13.19036 ]
```

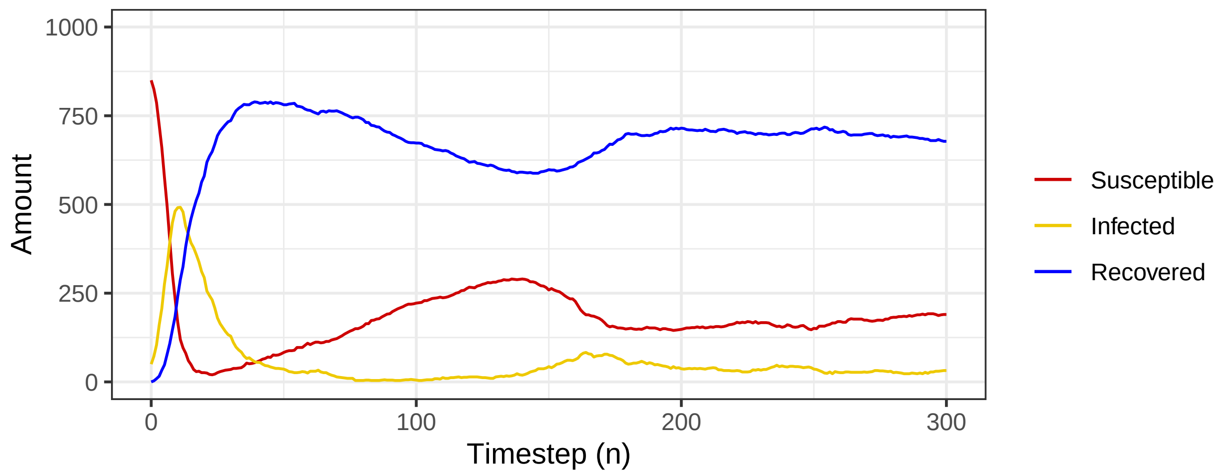
g

Here we are assuming that a vaccine completely immunizes, and set the probability of infection to 0. From the plots we can see that as the number of immune individuals increases, the proportion of infected to susceptible individuals increases, so the peak of the infection is reached earlier, while the maximal amount of infected individuals decreases (as expected).

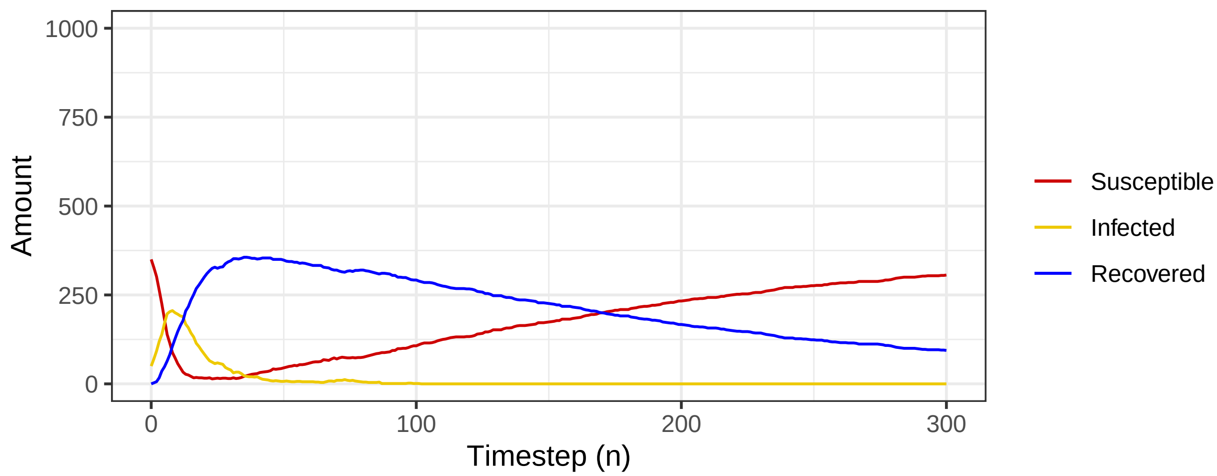
Simulation of Yn until n=300 for 1000 susceptible individuals

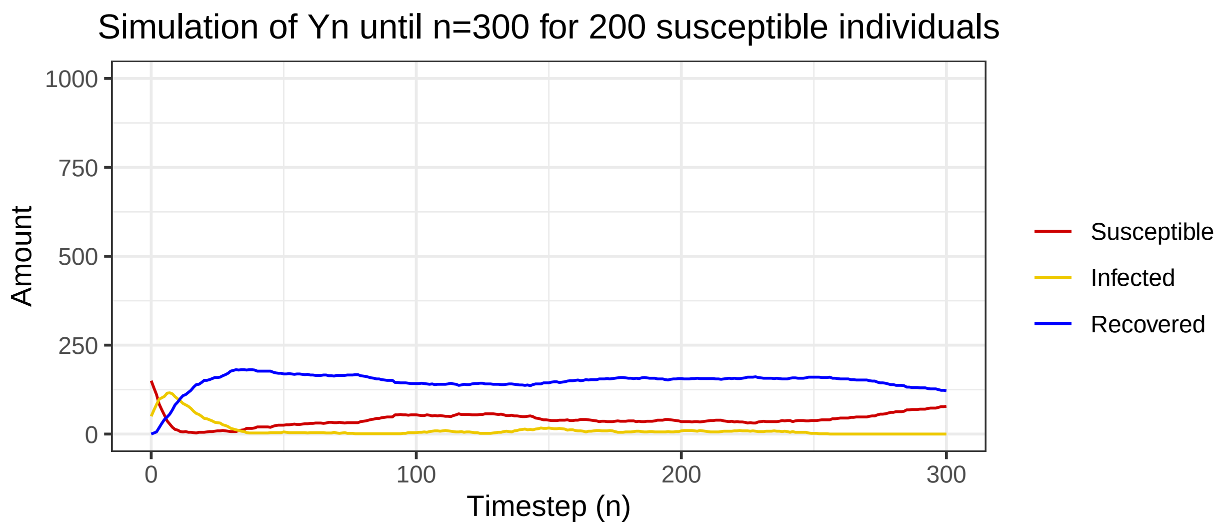


Simulation of Yn until n=300 for 900 susceptible individuals



Simulation of Yn until n=300 for 400 susceptible individuals





Problem 2: Insurance claims

In this problem we examine a Poisson process $\{X(t) : t \geq 0\}$, where $X(t)$ is the number of claims received in the interval $[0, t]$, and t is measured in days from the start of January 1st. We assume it has rate $\lambda(t) = 1.5, t \geq 0$.

a)

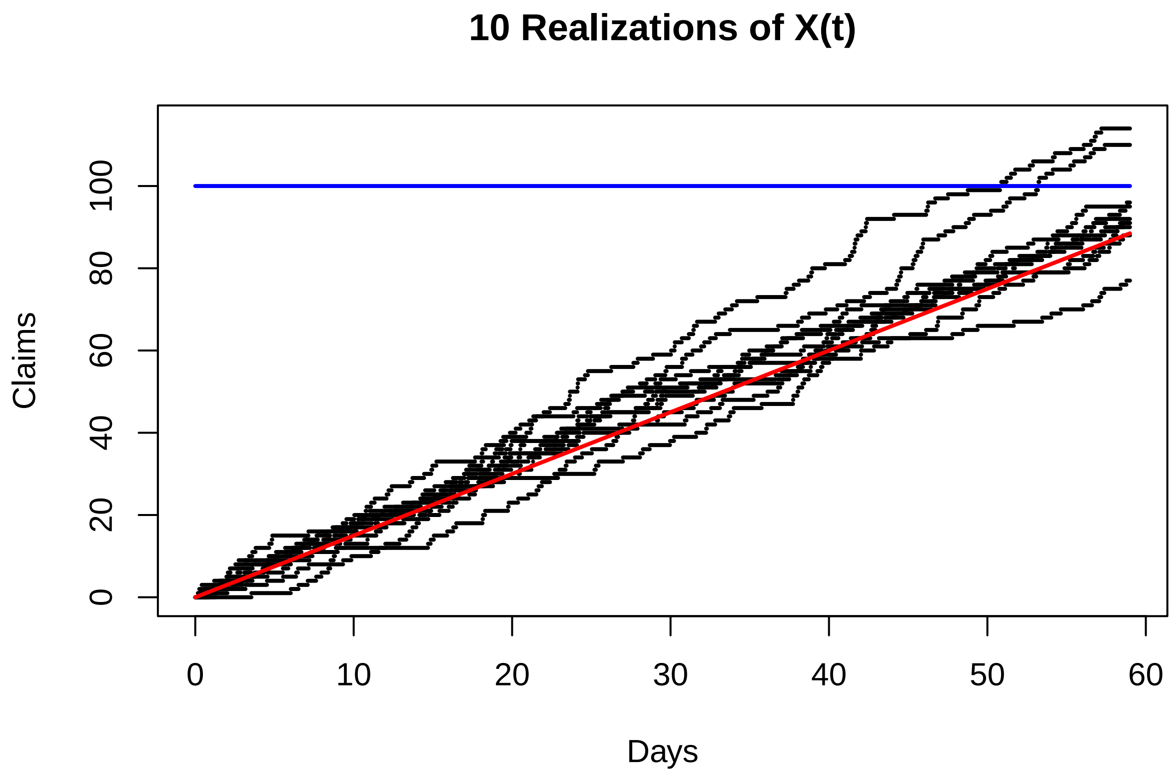
We want to find $P\{X(59) > 100\} = 1 - P\{X(59) \leq 100\} = 1 - \sum_{s=0}^{x} \frac{(\lambda t)^s}{s!} \exp\{-\lambda t\} =$

```
## [1] 0.1028222
```

Giving us a 10.28% probability of having more than 100 claims at March 1st.

Simulating 1000 realizations of the Poisson process we got an estimated probability of receiving over 100 claims by March 1st of 10.9%, corresponding well with our exact calculations.

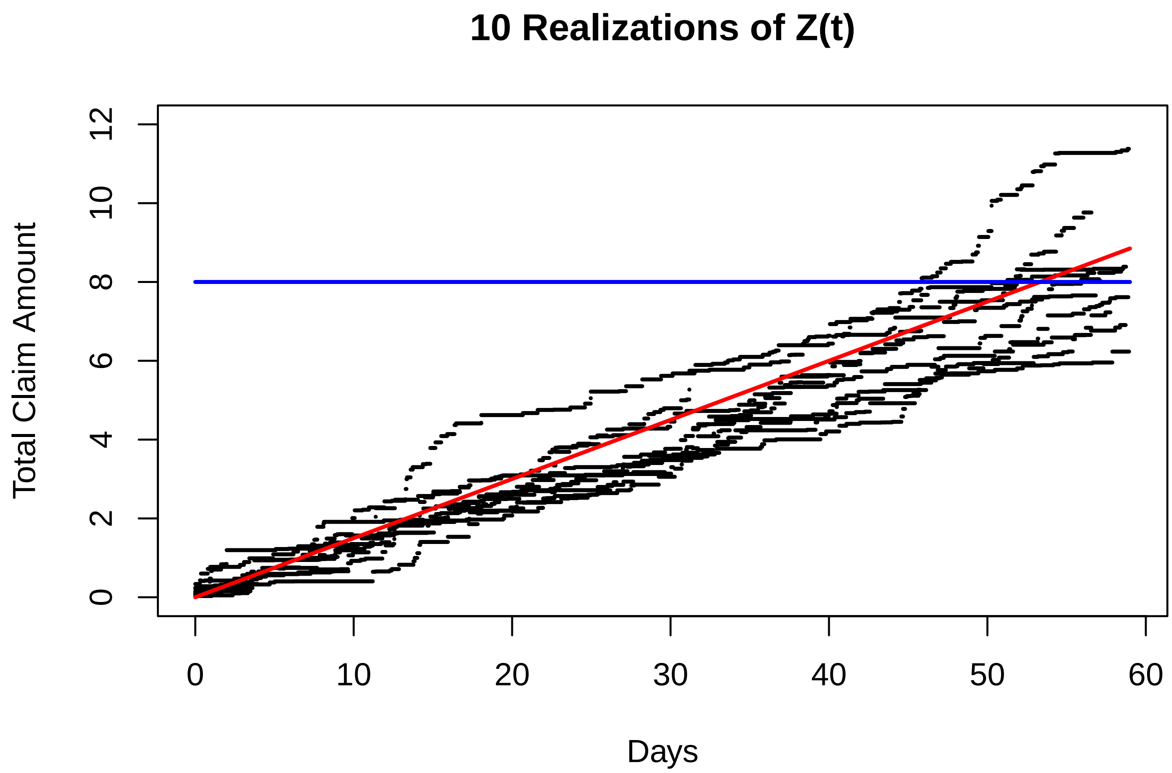
Plotting 10 realizations until t=59, with the mean in red and X(t)=100 in blue:



We assume the monetary claims $C_i \sim Exp(\gamma), i \geq 1$ with rate parameter $\gamma = 10$, are independent of each other and of the arrival times. We get that the total claim amount at time t is given by $Z(t) = \sum_{i=1}^{X(t)} C_i$.

b)

Simulating 1000 realizations of our Poisson process and computing the realized claim amounts, we get an estimated probability of 72.5% that the claim amounts will exceed 8 (million kroner) at March 1st.



c)

Call the number of claims exceeding 250000 (1/4 million) kr. by time t, Y_t. As the insurance company has a policy of investigating these claims, we want want to find the distribution of $\{Y(t) : t \geq 0\}$.

Finding the probability density function of $Y(t)$: We know that $X(t) \sim \text{Poisson}(\lambda)$, and that $X(t) = Y(t) + \overline{Y}(t)$, as those are the only two outcomes for each occurrence of $X(t)$, and they are mutually exclusive. Using the law of total probability and conditional probability, we obtain: $P\{Y(t) = y\} = \sum_{n=0}^{\infty} P\{Y(t) = y|X(t) = n\}P\{X(t) = n\}$

As the claim amounts for each occurrence are independent, and that the probability p of a claim amount exceeding $\frac{1}{4}$ million is $p = P\{C_i > \frac{1}{4}\} = 1 - P\{C_i \leq \frac{1}{4}\} = \exp\{-\frac{\gamma}{4}\} = \exp\{-\frac{5}{2}\}$, we know that the amount of claims exceeding $\frac{1}{4}$ million given that n amounts occurred is Binomially distributed. $P\{Y(t) = y|X(t) = n\} = \binom{n}{y} p^y (1-p)^{n-y}$

From earlier we know that $P\{X(t) = n\} = \frac{(\lambda t)^n}{n!} \exp\{-\lambda t\} = \frac{(\lambda t)^{(n-y)} (\lambda t)^y}{n!} \exp\{-\lambda p t\} \exp\{-\lambda(1-p)t\}$

This gives us that $P\{Y(t) = y\} = \sum_{n=0}^{\infty} P\{Y(t) = y|X(t) = n\}P\{X(t) = n\}$
$= \sum_{n=y}^{\infty} \frac{n!}{(n-y)!y!} \frac{(\lambda p t)^y}{n!} \exp\{-\lambda p t\}(\lambda(1-p)t)^{n-y} \exp\{-\lambda(1-p)t\}$
$= \frac{(\lambda p t)^y}{y!} \exp\{-\lambda p t\} \exp\{-\lambda(1-p)t\} \sum_{n=y}^{\infty} \frac{(\lambda(1-p)t)^{n-y}}{(n-y)!}$
$= \frac{(\lambda p t)^y}{y!} \exp\{-\lambda p t\}$

This is the density function of a Poisson distribution, which fulfills all necessary conditions of a poisson process: It has independent increments that are of distribution $\text{Poisson}(\lambda p)$, with rate $\lambda p = 1.5 \exp\{-\frac{5}{2}\} \approx 0.123$, about a tenth of the original rate. Also $Y(0) = 0$