

Lesson 8. Arrival Counting Processes

1 Overview

- An **arrival counting process** is a stochastic process with one “arrival” system event type
- An “arrival” is broadly defined as any discrete unit that can be counted: for example,
 - customer arrivals
 - service requests
 - arrival of e-mail messages
 - accidents in a factory

2 The Case of the Reckless Beehunter

Citizens of Beehunter have complained that a busy intersection has recently become more dangerous, and they are demanding that the city council take action to make the intersection safer. There have been 103 accidents at the intersection since record keeping began. The city council agrees to undertake a study of the intersection to determine if the accident rate has actually increased above the 1 per week average that is (unfortunately) considered normal. It hires a traffic engineer from nearby Vincennes, to perform the study.

The traffic engineer recommends that the number of accidents at the intersection be recorded for a 24-week period. If the number of accidents is significantly larger than expected, then she will declare that the intersection has indeed become more dangerous. During the study period, 36 accidents were observed.

- How can we quantify “significantly larger”, or extreme behavior in general?
- Our approach:
 - Model the system as a stochastic process (algorithm that generates possible sample paths)
 - Define extreme behavior as a collection of sample paths that have relatively small probability of being generated by this model
 - e.g. According to the model we construct, what is the probability that 36 accidents occur in a 24 week period?
- Let’s assume the interarrival times between accidents are independent and time stationary random variables with common cdf F_G and $E[G] = 1$ week
 - Is this reasonable?

- State variables:

- System events:

- Simulation algorithm – just the general algorithm from before:

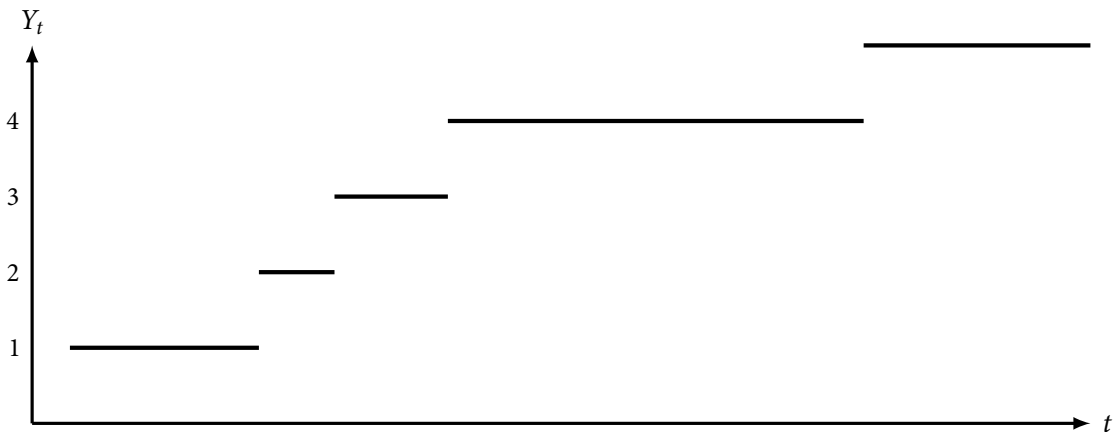
algorithm Simulation:

- | | |
|---|---|
| 1: $n \leftarrow 0$ | (initialize system event counter) |
| $T_0 \leftarrow 0$ | (initialize event epoch) |
| $e_0()$ | (execute initial system event) |
| 2: $T_{n+1} \leftarrow \min\{C_1, \dots, C_k\}$ | (advance time to next pending system event) |
| $I \leftarrow \arg \min\{C_1, \dots, C_k\}$ | (find index of next system event) |
| 3: $\mathbf{S}_{n+1} \leftarrow \mathbf{S}_n$ | (temporarily maintain previous state) |
| $C_I \leftarrow \infty$ | (event I no longer pending) |
| 4: $e_I()$ | (execute system event I) |
| $n \leftarrow n + 1$ | (update event counter) |
| 5: go to line 2 | |

- What does S_n equal for any n ?

- Output process: $Y_t \leftarrow S_n$ for all $T_n \leq t < T_{n+1}$, or in words,

- A stochastic process model defined in the way above is called a **renewal arrival counting process** or **renewal process** for short
 - In this case, “accidents” \leftrightarrow “arrivals”
- Suppose we begin observing at time a weeks (a is some fixed constant)
- The **forward-recurrence time** R_a at time a is the time that passes until the first accident after time a
- Define L_a as the length of the interarrival-time gap that contains time a
- An example of the output process for one sample path:



- We can modify algorithm Simulation to compute R_a and L_a for a given sample path:

- The number of accidents in the 24 week period we observe is

- What we're interested in:

- What does this depend on?

- The distribution of the interarrival times G ?
- Only the expected interarrival time $E[G]$?
- Does the time we start observing a matter?

3 Thought experiment: interarrival time as a degenerate random variable

- Suppose $\Pr\{G = 1\} = 1$

- Then $E[G] =$ and $\text{Var}(G) =$

- Such a random variable (takes a single value with probability 1) is called **degenerate**

- In this case, $\Pr\{Y_{a+24} - Y_a = 24\}$

- If G has some randomness, then $\Pr\{Y_{a+24} - Y_a = 24\}$

- Therefore,

- Let's assume G is degenerate again

- If we start observing at time $a = 0$, then the time until the first accident R_a

- If $a = 1/4$, then R_a

- Therefore,

4 Simulation experiments: different interarrival time distributions

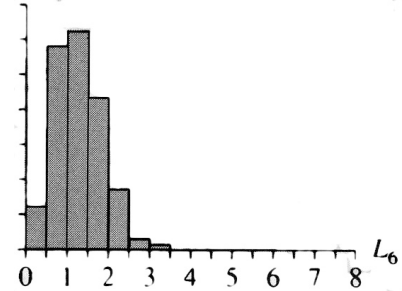
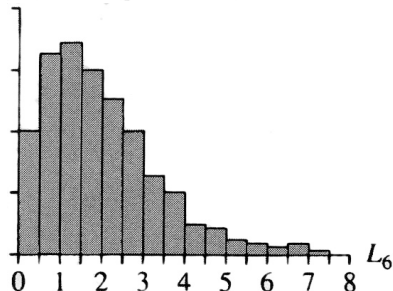
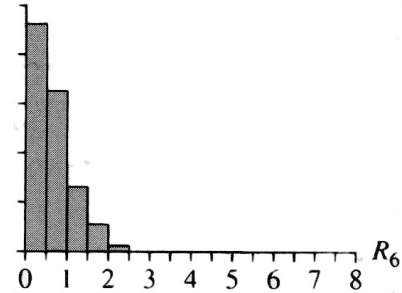
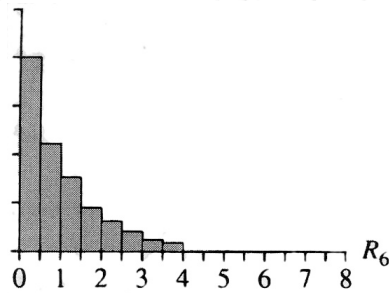
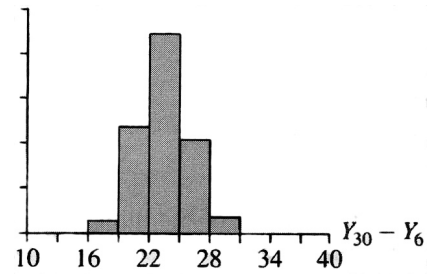
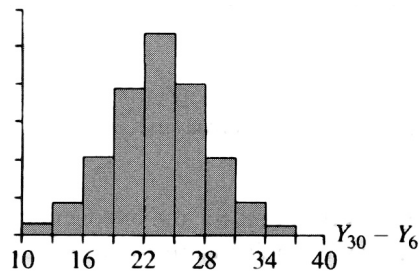
- Let's try two distributions for G :
 1. Exponential distribution with parameter $\lambda = 1 \Rightarrow E[G] = 1, \text{Var}(G) = 1$
 2. Weibull distribution with parameters $\alpha = 2, \beta = 2/\sqrt{\pi} \Rightarrow E[G] = 1, \text{Var}(G) = 0.25$
- For each distribution, run algorithm Simulation 1000 times with $a = 6$
 - Record $R_6, L_6,$ and $Y_{30} - Y_6$ for each replication
- Histograms and summary statistics:

$G \sim \text{Exponential}(\lambda = 1)$

- Sample average of $Y_{30} - Y_6 \approx 24$
- Sample average of $R_6 \approx 1$
- Sample average of $L_6 \approx 2$
- Estimated probability $\{Y_{30} - Y_6 > 30\} \approx 0.10$

$G \sim \text{Weibull}(\alpha = 2, \beta = 2/\sqrt{\pi})$

- Sample average of $Y_{30} - Y_6 \approx 24$
- Sample average of $R_6 \approx 0.6$
- Sample average of $L_t \approx 1.3$
- Estimated probability $\{Y_{30} - Y_6 > 30\} \approx 0.025$



- When $G \sim \text{Exp}(\lambda = 1), R_6$ looks like an exponential distribution and has sample average $\approx E[G]$
- Could it be that the distribution of R_a and G are the same in this case?
 - This would be a nice property – suggests choice of a doesn't matter