TPK4120 - Lecture summary

Jørn Vatn
eMail: jorn.vatn@ntnu.no

Updated 2022-01-22

Chapter 10 - Counting processes

This memo is based on the book: System Reliability Theory - Models, Statistical Methods, and Applications by Rausand, Barros and Hoyland (2021). John Wiley & Sons, and in particular on Chapter 10.

A stochastic process $\{X(t), t \in \Theta\}$ is a collection of random variables. The set Θ is called the *index set* of the process. For each index t in Θ is , $X(t)$ is called the *state* of the process at time t. In this chapter we only consider the situation where Θ is a continuum, that is we have a continuous-time stochastic process.

In Chapter 11 we say that a stochastic process has the Markov property if:

$$\Pr(X(t+s) = j | X(s) = i \cap \text{ some history up to time } s) = \Pr(X(t+s) = j | X(s) = i)$$

In Chapter 10 we basically consider a repairable system that is put into operation at time $t = 0$. Repair times are assumed to be small compared to failure times, and are considered negligible. In Chapter 6 and we also considered the repair times.

This means that we in Chapter 10 are considering a sequence of failure times. The main quantity of interest is $N(t)$ which is the number of failures in the time interval $(0, t]$. This means that we are primarily focusing on the number of failures in a time period rather than the actual failure times.

The process $\{N(t), t > 0\}$ is called a *counting process*. We use the notation T_i to denote the *time between failures* or *inter-arrival times*. The notation S_i is used to denote the (calendar) time of the individual failures. This means that $S_i = \sum_{j=1}^{i} T_i$, and $T_i = S_i - S_{i-1}$ where $S_0 = 0$. We use t to denote time, whether this is calendar time or local time.

Counting process

Definition: A stochastic process $\{N(t), t > 0\}$ is said to be a counting process if:

1. $N(t) \geq 0$

2. $N(t)$ is integer valued

3. If $s < t$ then $N(s) \le N(t)$

4. For $s < t$, $[N(t) - N(s)]$ is the number of failures in $(s, t]$.

Some basic concepts

The following concepts are introduced, some frequently used:

- *Independent increments*. A counting process has independent increments if the number of failures in non-overlapping intervals are stochastically independent.

- *Stationary increments*. A counting process has stationary increments if the number of failures in a time interval only depends on the length of the interval and not on where the interval is on the time axis. Such a process is said to be homogeneous.

- *Non-stationary process*. The number of failures in an interval depends both on the length of the interval and where the interval is on the time axis. Such a process is said to be non-homogeneous.

- *Regular process*. A counting process is said to be regular if the probability of two or more failures in a small interval, say Δt, is very small, i.e., $o(\Delta t)$.

- *Rate of the process*. The rate of a counting process at time t is defined as:

$$w(t) = W'(t) = \frac{d}{dt}\mathrm{E}[N(t)]$$

  where $W(t)$ is the expected number of failures in the interval $(0, t]$.

- ROCOF. $w(t)$ is denoted the rate of occurrence of failures. We have that for a small time interval, say Δt, $w(t)\Delta t$ is approximately equal to the probability of a failure in $(t, t + \Delta t]$.

Types of counting processes

Several types of counting processes exist, and we will elaborate on the following four:

1. Homogeneous Poisson processes

2. Renewal processes

3. Non-homogeneous Poisson processes

4. Imperfect repair processes.

Homogeneous Poisson Processes - HPP

A counting process is said to be an HHP with rate λ if:

1. $N(0) = 0$
2. The process has stationary and independent increments
3. $\Pr(N(\Delta t) = 1) = \lambda\Delta t + o(\Delta t)$
4. $\Pr(N(\Delta t) \geq 2) = o(\Delta t)$.

It may be shown that for an HPP the inter-arrival times are independent and exponentially distributed with parameter λ. Alternatively this fact could be used as a definition of the HPP. The following features apply for a HPP:

1. The ROCOF of the HPP is time independent i.e., $w(t) = \lambda$
2. The number of failures in the interval $(t, t+v)$ is Poisson distributed with mean λv
3. The mean number of failures in the interval $(t, t+v)$ is $W(t+v) - W(t) = \mathrm{E}[N(t+v) - N(t)] = \lambda v$
4. The inter-arrival times are independent and exponentially distributed with parameter λ The time of the nth failure $S_n = \sum_{i=1}^{n} T_i$ is gamma distributed with parameters (n, λ).

Compound HPPs

We consider a HPP where there is a cost, say V_i associated with each failure. Assume that all V_i's are independent and identically distributed with some distribution function $F_V(v)$. The cumulative cost at time t is given by

$$Z(t) = \sum_{i=1}^{N(t)} V_i$$

The process $\{Z(t), t > 0\}$ is denoted a *compound Poisson process*.

To find the expected value and variance of the accumulated cost in a time interval $(0, t]$ we may apply Wald's formula and the Blackwell–Girshick equation respectively.

Wald's formula states the following: Let V_i be independent an identically distributed stochastic variables with expected value and variance $\mathrm{E}(V)$ and Var(V) respectively. Further pick a random number, say $N(t)$, of V's. The expected sum of the V's picked is given by:

$$\mathrm{E}\left(\sum_{i=1}^{N(t)} V_i\right) = \mathrm{E}[N(t)]\,\mathrm{E}(V)$$

where $\mathrm{E}(N(t))$ is the expected number picked, i.e., the number of failures in the time interval.

The Blackwell-Girshick equation applies in the same situation, but gives the result for the variance:

$$\mathrm{Var}\left(\sum_{i=1}^{N(t)} V_i\right) = \mathrm{E}\left[N(t)\right]\mathrm{Var}(V) + \mathrm{E}^2(V)\mathrm{Var}\left[N(t)\right]$$

where $\mathrm{Var}(N(t))$ is the variance in the number picked, i.e., the variance in the number of failures in the time interval. Since $N(t)$ is Poisson distributed, $\mathrm{E}(N(t)) = \mathrm{Var}(N(t)) = \lambda t$.

The result is intuitive for the expected accumulated costs, whereas the result for the variance is not that obvious.

Renewal process - RP

A *renewal process* is a counting process $\{N(t), t > 0\}$ with independent and identically distributed inter-arrival times. The observed events are here the failure times (or more precisely the times of repairs after each failure). Let $F_T(t)$ be the distribution function of the inter-arrival times, i.e., the underlying distribution of the renewal process.

Of main interest will be to obtain the distribution function of $S_n = \sum_{i=1}^{n} T_i$ and the expected number of renewals, $W(t)$ at time t.

The distribution of S_n

Let $F^{(n)}(t)$ denote the distribution function of S_n. From the convolution theorem it follows:

$$F^{(n)}(t) = \int_0^t F^{(n-1)}(t-x) f_T(x) dx$$

For small values of n, say 2 and 3, we may use this result to find the distribution function for S_2 and S_3 by numerical integration. For large values of n this would be tedious.

The renewal function, $W(t) = \mathrm{E}(N(t))$

The renewal function, $W(t)$ is the expected number of failures in the time interval $(0, t]$. To obtain $W(t)$ we use that $= \mathrm{E}(N(t)) = \sum_{n=1}^{\infty} \Pr(N(t) \geq n)$. This result is similar to $\mathrm{E}(T) = \int_0^{\infty} t f(t) dt = \int_0^{\infty} R(t) dt$. We have:

$$W(t) = \mathrm{E}(N(t)) = \sum_{n=1}^{\infty} \Pr(N(t) \geq n) = \sum_{n=1}^{\infty} \Pr(S_n \leq t) = \sum_{n=1}^{\infty} F^{(n)}(t)$$

By some manipulation and inserting $F^{(n)}(t) = \int_0^t F^{(n-1)}(t-x) f_T(x) dx$ we obtain the following integral equation referred to as the fundamental renewal equation:

$$W(t) = F_T(t) + \int_0^t W(t-x) f_T(x) dx$$

It is not easy to solve the fundamental renewal equation. But if we have a reasonable approximation for $W(t)$, say $W_0(t)$ we may use the following iteration scheme:

$$W_i(t) = F_T(t) + \int_0^t W_{i-1}(t-x) f_T(x) dx$$

to obtain better and better solutions. This will be discussed in relation to Chapter 9.

Bounds and limiting values for $W(t)$

It may be shown that a general renewal process is bounded by:

$$\frac{t}{\mu} - 1 \le W(t) \le \frac{t}{\mu} + \frac{\sigma^2}{\mu^2}$$

where μ and σ^2 are the expected value and variance of the individual inter-arrival times. For a distribution of type "new better than used" somewhat stricter bound by be obtained:

$$\frac{t}{\mu} - 1 \le W(t) \le \frac{t}{\mu} + \frac{\sigma^2}{\mu^2}$$

When t becomes large we have the obvious result:

$$\lim_{t \to \infty} W(t) = \frac{t}{\mu}$$

When using the iteration scheme $W_i(t) = F_T(t) + \int_0^t W_{i-1}(t-x) f_T(x) dx$ we need an initial approximation for $W_0(t)$. For small t values $F(t)$ would be a good approximation. For larger values of t we could use some of the bounds described above.

Numerical methods for $W(t)$

The `NumLibTPK4120.xlsm` MS-Excel file available on Blackboard contains some Visual Basic code (VBA) for an iterative solution of the renewal equation. The basic code is given below. On the W(t) sheet an example is given.

```
' Calculate initial W(), F1=CDF and f=PDF
For i = 0 To MaxDim
    W(i) = 1# - Exp(-(lambda * i * dt) ^ alpha) ' W_0 = CDF
```

```
    F1(i) = 1# - Exp(-(lambda * i * dt) ^ alpha) ' CDF
    f(i) = alpha * lambda * (lambda * i * dt) ^ (alpha - 1) * (1 - F1(i)) ' PDF
Next
prev = W(MaxDim)
Do
    cnt = cnt + 1
    prev = W(MaxDim)
    For i = MaxDim To 1 Step -1
        W(i) = F1(i) + TrapezConv(0, i, dt, W, f) 'integrate W(t-x)*f(x)
    Next
Loop While Abs(prev - W(MaxDim)) / prev > eps Or cnt < 2
RenewalWeibull = W(MaxDim)
```

Monte Carlo methods for $W(t)$

The `NumLibTPK4120.xlsm` MS-Excel file available on Blackboard contains some Visual Basic code (VBA) for Monte Carlo simulation for obtaining the renewal equation. The basic code is given below. On the W(t) sheet an example is given.

```
For n = 1 To nSim
  nFail = 0#
  i = 0 ' Current pointer to W-array
  t = 0# ' Current time
  Do While t < maxT
    t = t + rndWeibull(alpha, lambda) ' Next failure time
    Do While i * step < t
      If i > nTimes Then Exit Do
      W(i) = W(i) + nFail / nSim ' W(i) will hold avg # failures at time i*step
      i = i + 1
    Loop
    nFail = nFail + 1
  Loop
Next n
```

Non-Homogeneous Poisson processes - NHHP

A counting process is said to be an NHHP with rate function $w(t)$ if:

1. $N(0) = 0$
2. The process has independent increments
3. $\Pr(N(t+\Delta t) - N(t) = 1) = w(t)\Delta t + o(\Delta t)$

4. $\Pr(N(t+\Delta t) - N(t) \geq 2) = o(\Delta t)$.

The basic "parameter" of the NHPP is the ROCOF function, $w(t)$. There are many similarities with the HPP, but we have to use $w(t)$ rather than λ. We have:

1. $W(t) = \mathrm{E}(N(t)) = \int_0^t w(u)du$
2. The number of failures in a time interval $(v, t+v)$ is Poisson distributed with mean value $W(t+v) - W(v)$
3. The expected number of failures in a time interval $(v, t+v)$ is given by $W(t+v) - W(v) = \int_v^{t+v} w(u)du$.

Note the difference between the failure rate function, $z(t)$ and the ROCOF, $w(t)$. $z(t)$ is the conditional probability of a failure in a small time interval given that no failures has occurred up to current time, whereas $w(t)$ is an unconditional probability of failure in a small time interval.

Further note that since the probability of failure in a small time interval only depends on $w(t)$, we have that the probability of a failure just after a failure is the same as it was just prior to that failure. This corresponds to a so-called *minimal repair*, or an *as-bad-as-old* situation. This is in contrast to the renewal process where we assume perfect repair, or an *as-good-as-new* situation after repair.

The Nelson-Aalen Estimator

The Nelson-Aalen estimator is appropriate if we have observed one ore more NHPP and we would like to look for trends in the data. Rather than visualizing $w(t)$ we estimate $W(t)$. The procedure is as follows:

1. We observe data for n processes, and for system i the observation period is $(a_i, b_i]$ relative to the global age of the process.
2. Let T_{ij} denote the (calendar) time of the jth failure of process i
3. Merge all T_{ij}'s and sort them in increasing order. Denote the result T_k, $k = 1, 2, \ldots$
4. For each k,let O_k denote the number of processes observed immediately before T_k
5. Let $\hat{W}_0 = 0$
6. Calculate $\hat{W}_k = \hat{W}_{k-1} + 1/O_k$, $k = 1, 2, \ldots$
7. Plot $(t_k, \hat{W}_k)$, i.e., the Nelson-Aalen plot.

What actually is taken place by this procedure is that the value on the y-axis is incremented by $1/O_k$ for each failure time. If we only have one process the procedure simplifies, we just increment the plot by 1 for each failure time. The interpretation of the plot is as follows:

- If the plot is *concave* this indicates that we have a set of improving processes where the failures become more and more seldom
- If the plot is *convex* this indicates that we have a set of deteriorating systems where the failures become more and more frequent
- If the plot is more or less a straight line there is no indication of trend, this means that a renewal process could be more appropriate than a NHPP
- If the plot is concave in the beginning and then turns to be convex we have a bathtub like ROCOF.

Generally a cumulative plot such as the Nelson-Aalen plot is more appropriate for trend identification than a plot estimating the underlying rate function $w(t)$.

Parametric NHPP models

The most used parametric NHPP models are:

1. The power law model with ROCOF function: $w(t) = \lambda\beta t^{\beta-1}$
2. The log-linear model with ROCOF function: $w(t) = e^{\alpha+\beta t}$.

The power law process is deteriorating for $\beta > 1$ whereas the log-linear process is deteriorating for $\beta > 0$. It may be proven that the first inter-arrival time, T_1 in the power law model is Weibull distributed. This is not surprising since $z(t) = w(t)$ for the power law process.

Imperfect repair processes

We have seen that the renewal process corresponds to an as-good-as-new situation after a repair (renewal), whereas the non-homogeneous Poisson process corresponds to an as-bad-as-old situation after a repair. In many situations it is natural to consider something between. A wide range of models exist for modelling imperfect repair. Some of these are listed in Section 7.5 in the textbook. Most of the models may be classified in two main groups: (i) models where the repair actions reduce the ROCOF, and (ii) models where the repair actions reduce the (virtual) age of the system.

Learning objectives, Chapter 10

- Understand what is a counting process, and basic terms like the number of failures in a time period, the expected number of failures in a time period, $W(t)$, and the ROCOF = $w(t)$
- Understand the relation between the inter-arrival times and the number of failures at a given point of time
- Understand that we basically focus on the number of failures, and not the point of time of the individual failure times
- Understand the difference between the HPP, RP and NHPP
- Use the Nelson-Aalen estimator to check a dataset for trend
- For the RP know that the renewal function, $W(t)$ could in principle be found by a the fundamental renewal equation, but that it is not easy to use in order to calculate $W(t)$
- For the HPP and the NHPP know that the number of failures in a time interval is Poisson distributed with mean value equal to the integral of the ROCOF function over that interval. For the RP is in not easy to find the distribution of number of failures in a time interval.