

Chapter 2

Brief Introduction of Renewal Process

Length biased sampling is a fundamental problem in studying renewal processes. Suppose buses arrive at a stop after independent random intervals of T minutes, where T is uniformly distributed between 10 and 20. It is natural to wonder how long one is expected to wait from some random point in time t until next bus arrives. The next bus could arrive immediately, or one could be unlucky with time t just after the previous bus left and could wait as long as 20 minutes for the next bus. Interestingly this waiting time is no longer uniformly distributed. This is the so-called “inspection paradox”.

In order to understand this phenomenon we briefly introduce some basic concepts in the renewal process. More details can be found from renewal process textbooks, such as Cox and Isham (1980) and Feller (1965). Professor Sigman (2009)’s lecture notes are also very valuable. Some results discussed in this chapter will be used in Chap. 25.

2.1 Basic Concepts

Let X_1, X_2, \dots be i.i.d. positive random variables with a common distribution function F . Define a sequence of times by $T_0 = 0$, and $T_k = T_{k-1} + X_k$ for $k \geq 1$. A typical example is that X_i is the lifetime of a light bulb, and T_k is the time the k -th bulb burns out. Another example is T_k is the time of arrival of the k -th customer, or bus arrival time, etc. We assume $\mu = E(X_i)$ and $\sigma^2 = \text{Var}(X_i) < \infty$.

Let

$$N(t) = \max\{k : T_k \leq t\},$$

be the number of renewals in $(0, t]$. Immediately we have $N(t) \rightarrow \infty$ as $t \rightarrow \infty$, and

$$\{N(t) = n\} = \{T_n \leq t, T_{n+1} > t\}, \quad n \geq 1.$$

Let F_n be the distribution of T_n . Clearly we have a convolution relationship

$$F_{k+1}(x) = \int_0^x F_k(x-y) dF(y), \quad k \geq 1.$$

Also from the fact

$$\{N(t) = k\} = \{N(t) \geq k\} - \{N(t) \geq k+1\},$$

we have

$$P\{N(t) = k\} = F_k(t) - F_{k+1}(t).$$

The renewal function $m(t)$ is defined as $m(t) = E\{N(t)\}$. Noting

$$N(t) = \sum_{k=1}^{\infty} I_k, \quad I_k = I(T_k \leq t),$$

we have

$$m(t) = E[N(t)] = \sum_{k=1}^{\infty} E(I_k) = \sum_{k=1}^{\infty} F_k(t).$$

Conditioning on the first interval time X_1 ,

$$m(t) = E\{N(t)\} = E[E\{N(t)|X_1\}].$$

If $t < x$, then $E\{N(t)|X_1 = x\} = 0$ because the first arrival occurs after time t . On the other hand if $t \geq x$,

$$E\{N(t)|X_1 = x\} = 1 + E\{N(t-x)\}$$

since after the first arrival time x , the renewal process starts over again. Now we show the **renewal equation**

$$\begin{aligned} m(t) &= \int_0^{\infty} E\{N(t)|X_1 = x\} dF(x) \\ &= \int_0^t [1 + m(t-x)] dF(x) \\ &= F(t) + \int_0^t m(t-x) dF(x). \end{aligned}$$

Next we study the elementary renewal theorem.

Theorem 2.1 For the renewal process $N(t)$, $t \geq 0$ defined above, almost surely

$$\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{E(X)}, \quad \lim_{t \rightarrow \infty} \frac{E\{N(t)\}}{t} = \frac{1}{E(X)}.$$

Proof Noting $T_{N(t)} \leq t < T_{N(t)+1}$ and $T_{N(t)} = X_1 + \cdots + X_{N(t)}$, we have inequality

$$\frac{1}{N(t)} \sum_{j=1}^{N(t)} X_j \leq \frac{t}{N(t)} \leq \frac{1}{N(t)+1} \left\{ \sum_{j=1}^{N(t)+1} X_j \right\} \{1 + 1/N(t)\}.$$

Using the Law of Large Numbers we can show that both sides converge to $E(X)$ as $t \rightarrow \infty$.

The second part of proof needs $N(t)/t$ to be uniformly integrable. We leave this to readers as an exercise.

Clearly the number of renewals in the interval $(0, t]$ is inversely proportional to the length of inter-arrival time. Indeed the elementary renewal theorem tells us the average number of renewals in $(0, t]$ is reciprocal of the average inter-arrival time.

Next we ask whether the central limit theorem holds true

$$\frac{N(t) - E\{N(t)\}}{\sqrt{\text{Var}\{N(t)\}}} \rightarrow N(0, 1)$$

as $t \rightarrow \infty$?

We already know that $E\{N(t)\} = m(t) \approx t/\mu$. How do we find the variance of $N(t)$?

For any real number x , denote r_t as the integer part of $t/\mu + x\sqrt{t\sigma^2/\mu^3}$. Then

$$r_t = t/\mu + x\sqrt{t\sigma^2/\mu^3} - \theta, \quad 0 \leq \theta < 1.$$

Note

$$\begin{aligned} P(T_{r_t} \geq t) &= P\left(\frac{T_{r_t} - r_t\mu}{\sigma\sqrt{r_t}} \geq \frac{t - r_t\mu}{\sigma\sqrt{r_t}}\right) \\ &= P\left(\frac{T_{r_t} - r_t\mu}{\sigma\sqrt{r_t}} \geq \frac{-x\sigma(t/\mu)^{1/2} + \mu\theta}{\sigma\sqrt{r_t}}\right). \end{aligned}$$

As $t \rightarrow \infty$,

$$\frac{r_t}{t/\mu} = \frac{t/\mu + x\sqrt{t\sigma^2/\mu^3} - \theta}{t/\mu} \rightarrow 1.$$

Using the Central Limit Theorem we can show that

$$P(N(t) \leq r_t) = P(T_{r_t} \geq t) \rightarrow 1 - \Phi(-x) = \Phi(x),$$

where Φ is the standard normal distribution function.

On the other hand

$$\begin{aligned} P(N(t) \leq r_t) &= P\left(\frac{N(t) - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \leq \frac{r_t - t/\mu}{\sqrt{t\sigma^2/\mu^3}}\right) \\ &= P\left(\frac{N(t) - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \leq \frac{x\sqrt{t\sigma^2/\mu^3} - \theta}{\sqrt{t\sigma^2/\mu^3}}\right) \\ &\approx P\left(\frac{N(t) - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \leq x\right), \end{aligned}$$

as $t \rightarrow \infty$. In conclusion we have proved the Central Limit Theorem for a renewal process.

Theorem 2.2 *As $t \rightarrow \infty$, in distribution*

$$\frac{N(t) - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \rightarrow N(0, 1). \quad (2.1.1)$$

From this with mild moment conditions we can find

$$\text{Var}\{N(t)\}/t \rightarrow \sigma^2/\mu^3.$$

2.2 Forward and Backward Recurrence Times

The forward recurrence time is a very important concept in renewal processes. It is the time between any given time t and the next epoch of the renewal process under consideration,

$$V(t) = T_{N(t)+1} - t, \quad t \geq 0.$$

It is also called residual lifetime or residual waiting time. If X_i has an exponential distribution with rate λ , then by the memoryless property of the exponential distribution, we know $V(t) \sim \exp(\lambda)$, $t \geq 0$. However for the general renewal process, the distribution of $V(t)$ is complicated and depends on the time t .

Theorem 2.3 *When the process $\{N(t), t \geq 0\}$ has run a long, it reaches the equilibrium, i.e.,*

$$\lim_{t \rightarrow \infty} \int_0^t V(s) ds = \frac{E(X^2)}{2E(X)},$$

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t I\{V(s) > x\} ds = \frac{E(X - x)^+}{\mu}, \quad \mu = E(X),$$

where $a^+ = \max(0, a)$.

The proof is not difficult. In fact from

$$T_{N(t)} \leq t < T_{N(t)+1},$$

we have

$$\int_0^t V(s) ds = \sum_{i=1}^{N(t)} \int_{t_i}^{t_{i+1}} (s - t_i) ds + \int_{T_{N(t)}}^t (s - T_{N(t)}) ds.$$

Immediately we have the inequality

$$\frac{1}{t} \sum_{j=1}^{N(t)} X_j^2 / 2 \leq \frac{1}{t} \int_0^t V(s) ds \leq \frac{1}{t} \sum_{j=1}^{N(t)+1} X_j^2 / 2.$$

Note that $t^{-1} = N^{-1}(t)N(t)/t$. Using the elementary renewal theorem and Strong Law of Large Numbers, we have the conclusion.

Similarly we can show that

$$\begin{aligned} \int_0^t I\{V(s) > x\} ds &= \sum_{j=1}^{N(t)} \int_{t_{j-1}}^{t_j} I\{V(s) > x\} ds + \int_{N(t)}^t I\{V(s) > x\} ds \\ &= \sum_{j=1}^{N(t)} \int_{t_{j-1}}^{t_j} I\{x_j - x > s - t_{j-1}\} ds \\ &= \sum_{j=1}^{N(t)} (x_j - x)^+. \end{aligned}$$

Note that if $x_j - x > 0$, then the integral becomes $\int_{t_{j-1}}^{t_j} 1 ds = x_j - x$. On the other hand if $x_j - x < 0$, the indicator function gives zero, and the integral becomes 0.

Note that

$$E(X - x)^+ = \int_x^\infty (t - x) dF(t) = \int_x^\infty \bar{F}(t) dt.$$

Intuitively the limiting value of $P(V(t) > x)$ as $t \rightarrow \infty$ should be equal to the average $t^{-1} \int_0^t P\{V(s) > x\}ds$. Therefore the equilibrium distribution as $t \rightarrow \infty$ has a distribution

$$\frac{1}{t} \int_0^t I\{V(s) > x\}ds = \frac{N(t)}{t} \frac{1}{N(t)} \sum_{j=1}^{N(t)} (x_j - x)^+ \rightarrow \int_x^\infty \bar{F}(t)dt/\mu, \quad x \geq 0.$$

This is also called a residual density

$$f_R(x) = \frac{\bar{F}(x)}{\mu}, \quad x \geq 0.$$

Note that $f_R(x)$ is always non-increasing.

Similarly we can define backward time. For any given time t , the backward time is the time between the last epoch of the renewal process to t ,

$$A(t) = t - T_{N(t)}, \quad t \geq 0.$$

It can be shown that the backward time A and forward time V have the same limiting distribution. Intuitively, when a renewal process reaches the equilibrium status, it becomes reversible, i.e., the statistical properties of this process are the same as the statistical properties for time-reversed data from the same process. The backward time can also be treated as the forward time if time periods are reversible.

The inter-arrival interval is defined as

$$Y(t) = T_{N(t)+1} - T_{N(t)} = A(t) + V(t).$$

Using the results from backward time and forward time, we immediately have

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t Y(s)ds = \frac{E(X^2)}{E(X)} \geq E(X).$$

This is the so called “inspection paradox”. When the renewal process reaches the equilibrium status, the mean inter-arrival time is longer than $E(X)$.

Moreover

$$\lim_{t \rightarrow \infty} \int_0^t I\{Y(s) > x\}ds = \frac{E\{XI(X > x)\}}{\mu}.$$

This can be shown by noting

$$\frac{1}{t} \int_0^t I\{Y(s) > x\}ds = \frac{1}{t} \sum_{j=1}^{N(t)} X_j I(X_j > x),$$

the length of time during the j -th inter-arrival time that $T(s) > x$ is precisely X_j if $X_j > x$ and 0 otherwise.

Observe

$$t^{-1} \int_0^t I\{V(s) > y, A(s) > x\} ds = t^{-1} \sum_{j=1}^{N(t)} \{X_j - (y + x)\}^+.$$

The length of time during the j -th inter-arrival time that $A(s) > y$ and $V(s) > x$ is precisely $\{X_j - (y + x)\}^+$.

Consider the first inter-arrival time, so that $A(s) = s$ and $V(s) = X_1 - s$, $s \in [0, X_1)$. Then

$$I\{A(s) > y, V(s) > x\} = I\{s > y, X_1 - s > x\} = I\{y < s < X_1 - x\}.$$

If $X_1 - (y + x) > 0$, then this indicator is non-zero and

$$\int_0^{X_1} I(y < s < X_1 - x) ds = \int_y^{X_1 - x} ds = X_1 - (y + x).$$

If $X_1 - (y + x) \leq 0$, then the indicator is 0 and

$$\int_0^{X_1} I(y < s < X_1 - x) ds = 0.$$

Therefore we obtain $\{X_1 - (y + x)\}^+$.

As $t \rightarrow \infty$, the joint density of the backward time $A(t)$ and forward time $V(t)$ has a survival function

$$t^{-1} \int_0^t I\{V(s) > y, A(s) > x\} ds = \frac{N(t)}{t} \frac{1}{N(t)} \sum_{j=1}^{N(t)} \{X_j - (y + x)\}^+ \rightarrow \frac{\int_{x+y}^{\infty} \bar{F}(u) du}{\mu}.$$

Thus we have shown that

Theorem 2.4 *At the equilibrium, the backward time A and forward time V have a joint density*

$$\frac{f(a + v)}{\mu}, \quad a, v > 0. \quad (2.2.2)$$

Let $Y = A + V$ be the inter-arrival interval, then

$$(Y, A) \sim \frac{f(y)}{\mu}, \quad y > a > 0. \quad (2.2.3)$$

The marginal density of Y is

$$\int_0^y f(y) da / \mu = yf(y)/\mu, \quad y > 0, \quad (2.2.4)$$

which is a length biased version of $f(y)$. The conditional density of A given Y is

$$\frac{f(y)/\mu}{yf(y)/\mu} = 1/y, \quad 0 \leq a \leq y. \quad (2.2.5)$$

In other words, given the length of the inter-arrival time, the backward time is a uniform distribution between $(0, y)$, irrelevant of the density f . By symmetry, this result also applies to the forward time. From a statistical inference point of view the inter-arrival length has captured all information on the underlying distribution F and the backward time or forward time is not informative for F any more as long as the inter-arrival length is given. However, the backward time indeed contains information about F if the inter-arrival length is not available, which is the case in cross sectional studies where the forward time is not available or partially available (with right censoring). We will use this result in Chap. 25 for length biased sampling data with right censoring.

In contrast to the length biased sampling problem, Rao (1965) discussed a reciprocal length biased sampling problem in aerial survey of low density traffic streams. Assume that vehicles enter the highway according to a nonhomogeneous Poisson process. The vehicles choose velocities at random from a distribution of $V \sim f(v)$. However the recorded velocity of a vehicle in the interval $[a, b]$ at time t is

$$f_w(v) = \frac{v^{-1} f(v)}{\int v^{-1} f(v) dv},$$

as $t \rightarrow \infty$, i.e., at the equilibrium. In this example, slower vehicles take longer time to transverse $[a, b]$. As a result, it would be easier to be recorded.

We conclude this chapter with some important results on the Poisson process which may be useful in later chapters.

2.3 Basic Results on Poisson Process

A special case of a renewal process is the Poisson process. It is defined as follows.

- (1) $N(0) = 0$.
- (2) The process has stationary and independent increments.
- (3) The number of events in an interval of length t is Poisson distributed with mean λt , i.e., for all $s, t > 0$,

$$P(N(t+s) - N(s) = k) = \frac{(\lambda t)^k \exp(-\lambda t)}{k!}, \quad k = 0, 1, 2, \dots,$$

Note that the first inter-arrival time $X_1 > t$ implies $N(t) = 0$.

$$P(X_1 > t) = P(N(t) = 0) = \exp(-\lambda t).$$

Hence X_1 has an exponential distribution with mean $1/\lambda$.

$$\begin{aligned} P(X_2 > t | X_1 = s) &= P\{0 \text{ event in } (s, s+t] | X_1 = s\} \\ &= P\{0 \text{ event in } (s, s+t]\} \\ &= \exp(-\lambda t), \end{aligned}$$

where the second equality has used the independent increments assumption and the third one has used the stationary increments argument.

Therefore the inter-arrival times X_1, X_2, \dots are i.i.d. exponential random variables.

Exercise Given $N(t) = n$, the n arrival times T_1, \dots, T_n have the same distribution as the order statistics corresponding to n independent variables uniformly distributed on the interval $(0, t)$.

Nonhomogeneous Poisson Process

In practical applications the requirement that the arrival rate at any time t is a constant λ is too strong. To relax this assumption, we consider a nonhomogeneous Poisson process.

The counting process $\{N(t), t \geq 0\}$ is said to be a non-stationary or nonhomogeneous Poisson process with intensity function $\lambda(t)$, $t \geq 0$ if

- (1) $N(t) = 0$.
- (2) $\{N(t), t \geq 0\}$ has independent increments.
- (3) $P\{N(t+\delta) - N(t) \geq 2\} = o(\delta)$, as $\delta \rightarrow 0$.
- (4) $P\{N(t+\delta) - N(t) = 1\} = \lambda(t)\delta + o(\delta)$.

Denote

$$p_j(t) = P(j \text{ event before time } t),$$

then

$$\begin{aligned} p_j(t + \Delta t) &= P\{j \text{ events before } t \text{ and no events in } (t, t + \Delta t)\} \\ &\quad + P\{(j-1) \text{ events before } t \text{ and no events in } (t, t + \Delta t)\} + (\Delta t)^2, \end{aligned}$$

or

$$p_j(t + \Delta t) = p_j(t)(1 - \lambda(t)\Delta t) + p_{j-1}(t)\lambda(t)\Delta t + (\Delta t)^2.$$

As a consequence we have a differential equation

$$p_0'(t) = -\lambda(t)p_0(t).$$

Clearly

$$p_0(t) = \exp \left\{ - \int_0^t \lambda(u) du \right\}$$

is the solution. In general we can derive (exercise)

$$P\{N(t) - N(s) = k\} = \frac{1}{k!} \{\Lambda(t) - \Lambda(s)\}^k \exp[-\{\Lambda(t) - \Lambda(s)\}], \quad t > s > 0.$$

where

$$\Lambda(t) = \int_0^t \lambda(u) du.$$

Define a time transformed process

$$N^*(s) = N(\Lambda(t)),$$

$$P(T_1 > t) = P(N(t) = 0) = \exp\{-\Lambda(t)\},$$

$$P\{T_2 > t_2 | T_1 = t_1\} = P\{N(t_2) - N(t_1) = 0\} = \exp[-\{\Lambda(t_2) - \Lambda(t_1)\}].$$

Therefore

$$f_2(t_2|t_1) = \lambda(t_2) \exp[-\{\Lambda(t_2) - \Lambda(t_1)\}].$$

In general if a Poisson process is observed in the interval $(0, \tau]$ with event times at $t_1 < t_2 < \dots < t_n < \tau$, then the likelihood is

$$\begin{aligned} L &= f_1(t_1) f_2(t_2|t_1) \cdots f_n(t_n|t_{n-1}) P(T_{n+1} > \tau) \\ &= \lambda(t_1) \exp\{-\Lambda(t_1)\} \cdots \lambda(t_n) \exp[-\{\Lambda(t_n) - \Lambda(t_{n-1})\}] \exp[-\{\Lambda(\tau) - \Lambda(t_n)\}] \\ &= \prod_{i=1}^n \lambda(t_i) \exp\{-\Lambda(\tau)\}. \end{aligned} \tag{2.3.6}$$

Theorem 2.5 *Given that $N(\tau) = n$, the pdf of the arrival times T_1, \dots, T_n is*

$$f(t_1, \dots, t_n) = \frac{n! \prod_{i=1}^n \lambda(t_i)}{\Lambda^n(\tau)}. \tag{2.3.7}$$

Define

$$h(t) = \lambda(t)/\Lambda(\tau), \quad 0 < t < \tau, \quad (2.3.8)$$

then $h(t)$ is a density function in $[0, \tau]$. As a consequence, $f(t_1, \dots, t_n)$ is the joint density of n order statistics from g .

In medical applications, due to correlations among recurrent events, frequently the intensity is assumed to be a random variable given by $\xi\lambda(t)$, where ξ (called frailty) has a gamma distribution. Then conditional on ξ , $N(t)$ still has independent increments. However, it is not the case unconditionally.

The joint likelihood is

$$\int \xi^n \prod_{i=1}^n \lambda(t_i) \exp\{-\xi \Lambda(\tau)\} dG(\xi), \quad \xi \sim G(\cdot).$$

The conditional density

$$(T_1, \dots, T_n) | N(\tau) = n \sim \frac{n! \prod_{i=1}^n \xi \lambda(t_i)}{\xi^n \Lambda^n(\tau)} = \frac{n! \prod_{i=1}^n \lambda(t_i)}{\Lambda^n(\tau)}, \quad 0 \leq t_1, \dots, t_n \leq \tau,$$

however, is free of the frailty ξ . In other words, conditional on $N(\tau) = n$, again T_1, \dots, T_n can be treated as order statistics generated from $h(t)$ defined in (2.3.8). Note that the frailty ξ is cancelled out.

Biased Sampling, Over-identified Parameter Problems
and Beyond

Qin, J.

2017, XVI, 624 p. 5 illus., 1 illus. in color., Hardcover

ISBN: 978-981-10-4854-8