MAST30001 Stochastic Modelling

Tutorial Sheet 10

1. The lifetime in hours of a saw blade in a lumber mill is a random variable with density proportional to $x(2-x)$ on $0 < x < 2$. When the mill opened a new blade was put in the saw. Every time a blade fails it is immediately replaced with a new one. Since the mill is automated, it operates 24 hours a day, seven days a week.
   (a) Model the number N_t of saw blades that have been replaced t hours after the mill opened as a renewal process and determine the density, mean, and variance of the random times between renewals.
   (b) On average, about how many replacement blades for the saw will the mill go through in the first 30 days of opening?
   (c) Give an interval around your estimate from (b) that will have a 95% chance of covering the true number of saws needed in the first 30 days.
   (d) If you show up at the mill at the end of day 30, about what is the mean and variance of the amount of time you'll have to wait for the current blade in use to be replaced?

2. A faculty course advisor is giving advice to students doing either course A or course B. Course A is rather structured and so advice is simple. Course B has more flexibility and so students need more time to select subjects. Moreover, in some complicated situations advising time can be more than 15 minutes and for these cases, the student is sent to talk to a faculty officer after 15 minutes. We assume that the distributions of the time (in hours) that the advisor spends talking to a given student have densities

$$\begin{array}{lll} \text{(A)} & f_A(x) = 32(1/4 - x), & x \in (0, 1/4), \\ \text{(B)} & f_B(x) = 32x, & x \in (0, 1/4). \end{array}$$

   On each day, either only the course A or course B students can come to the course advice session. We assume advisors work without breaks, there are always students waiting for advice, and that the time that advice sessions last are i.i.d. Denote by N_t the number of students an advisor has finished talking to t hours into the day. Answer the following questions for both cases (A) and (B).
   (a) About how many students can be advised in a typical 8 hour work day?
   (b) Determine an interval that contains N_8 with roughly 95% probability.
   (c) What is the density of the limiting (as $t \to \infty$) residual lifetime $T_{N_t+1} - t$?
   (d) What is the joint density of the limiting (as $t \to \infty$) residual lifetime $T_{N_t+1} - t$ and age $t - T_{N_t}$?

3. (Thinned Renewal Process) Consider a renewal process with mean interarrival time μ. Suppose that each event of this process is independently "counted" with probability p. Let N_t denote the number of counted events by time t; $t > 0$.
   (a) Is $(N_t)_{t \geq 0}$ a renewal process?
   (b) What is $\lim_{t \to \infty} N_t/t$?

4. (Delayed renewal process) A renewal process for which the time until the initial renewal, say τ_0, has a different distribution than the remaining interarrival times, say $\tau_1, \tau_2, \ldots$, is called a delayed (or a general) renewal process. If N_t is the number of renewals of a delayed renewal process in $[0, t]$, then show that $\lim_{t\to\infty} N_t/t \to 1/\mu$, where $\mu = E[\tau_1]$.

5. The phenomenon in a renewal process $(N_t)_{t\geq 0}$, that $T_{N_t+1} - T_{N_t}$, the length of a typical interval straddling a large value of t, is larger than the typical interval length, τ_i, is called *size biasing*. Intervals aren't selected at random to straddle t, but longer intervals are more likely. Size biasing arises frequently in statistical sampling. Consider a remote tourist destination that wants to estimate the average length of a tourist's visit. Two sampling schemes are up for discussion. The first scheme uses the estimate of the average of the lengths of tourists' visit from a randomly chosen sample of tourists leaving the airport of the destination while the second scheme uses the estimate of the average of the lengths of tourists' visit from a randomly chosen sample of tourists staying at hotels. These averages will not be the same. Why? Which is bigger? Which scheme is the most appropriate to estimate the average length of a tourist's visit?

read the unit $f_{Ti}(x) \propto x(2-x)$ $x \in (0,2)$.

1. The lifetime in hours of a saw blade in a lumber mill is a random variable with density proportional to $x(2-x)$ on $0 < x < 2$. When the mill opened a new blade was put in the saw. Every time a blade fails it is immediately replaced with a new one. Since the mill is automated, it operates 24 hours a day, seven days a week.

   (a) Model the number N_t of saw blades that have been replaced t hours after the mill opened as a renewal process and determine the density, mean, and variance of the random times between renewals.

   (b) On average, about how many replacement blades for the saw will the mill go through in the first 30 days of opening?

   (c) Give an interval around your estimate from (b) that will have a 95% chance of covering the true number of saws needed in the first 30 days.

   (d) If you show up at the mill at the end of day 30, about what is the mean and variance of the amount of time you'll have to wait for the current blade in use to be replaced?

(a).

$\int_0^2 Ax(2-x)\,dx$

$= A\int_0^2 2x - x^2\,dx$

$= A\left[x^2 - \frac{1}{3}x^3\right]_0^2$

$= A\left[4 - \frac{8}{3}\right] = 1$

$\Rightarrow A\ \frac{4}{3} = 1 \Rightarrow A = \frac{3}{4}.$

$\therefore f_{Ti}(x) = \frac{3}{4}x(2-x).$ $x \in (0,2)$

$E(X) = \frac{3}{4}\int_0^2 x^2(2-x)\,dx$

$= \frac{3}{4}\left[\frac{2}{3}x^3 - \frac{1}{4}x^4\right]_0^2$

$= \frac{3}{4}\left[\frac{16}{3} - \frac{16}{4}\right]$

$= \cancel{\frac{3}{4}} \times \cancel{16}^{4} \times \frac{1}{\cancel{12}\,3} = 1.$

$var(X) = E(X^2) - E(X)^2$

$E(X^2) = \frac{3}{4}\int_0^2 x^2 x(2-x)\,dx$

$= \frac{3}{4}\left[\frac{1}{2}x^4 - \frac{1}{5}x^5\right]_0^2$

$= \frac{3}{4}\left[8 - \frac{32}{5}\right]$

$= \frac{3}{\cancel{4}} \times \frac{\cancel{8}\,2}{5} = \frac{6}{5}$

$\therefore var(X) = \frac{6}{5} - 1 = \frac{1}{5}.$

(b). $N_{30\times 24} = \frac{720}{\mu} = 720$ $\frac{t\sigma^2}{\mu^3}$ $t=720$ $\sigma^2 = \frac{1}{5}$

(c). $N_{720} \approx N(720, 144).$

$\therefore \frac{N_{720} - 720}{12} \sim N(0,1)$

95% CI $720 \pm 1.96 \times 12$

720 ± 23.52

(d) about residual time Y_t

Y_t roughly have a distribution $\frac{1}{\mu}\left[1 - F(x)\right]$

$= \frac{1}{1}\left[1 - \frac{3}{4}t^2 + \frac{1}{4}t^3\right]$

$= 1 - \frac{3}{4}t^2 + \frac{1}{4}t^3.$

$F_X = \int_0^t \frac{3}{4}x(2-x)\,dx$

$= \frac{3}{4}\int_0^t 2x - x^2\,dx$

$= \frac{3}{4}\left[x^2 - \frac{1}{3}x^3\right]_0^t$

$= \frac{3}{4}t^2 - \frac{1}{4}t^3.$

$E(Y_t) = \int_0^2 x\left(1 - \frac{3}{4}x^2 + \frac{1}{4}x^3\right)dx$

$= \int_0^2 x - \frac{3}{4}x^3 + \frac{1}{4}x^4\,dx$

$= \left[\frac{1}{2}x^2 - \frac{3}{16}x^4 + \frac{1}{20}x^5\right]_0^2$

$= 2 - 3 + \frac{\cancel{32}\,8}{\cancel{20}\,5}$

$= \frac{3}{5}$

$E(Y_t^2) = \int_0^2 x^2 - \frac{3}{4}x^4 + \frac{1}{4}x^5\,dx$

$= \left[\frac{1}{3}x^3 - \frac{3}{20}x^5 + \frac{1}{24}x^6\right]_0^2$

$= \frac{8}{15}$

$$var(Y_t) = \frac{8}{15} - \frac{9}{25}$$
$$= \frac{13}{75}$$

2. A faculty course advisor is giving advice to students doing either course A or course B. Course A is rather structured and so advice is simple. Course B has more flexibility and so students need more time to select subjects. Moreover, in some complicated situations advising time can be more than 15 minutes and for these cases, the student is sent to talk to a faculty officer after 15 minutes. We assume that the distributions of the time (in hours) that the advisor spends talking to a given student have densities

$$\begin{array}{lll} \text{(A)} & f_A(x) = 32(1/4 - x), & x \in (0, 1/4), \\ \text{(B)} & f_B(x) = 32x, & x \in (0, 1/4). \end{array}$$

On each day, either only the course A or course B students can come to the course advice session. We assume advisors work without breaks, there are always students waiting for advice, and that the time that advice sessions last are i.i.d. Denote by N_t the number of students an advisor has finished talking to t hours into the day. Answer the following questions for both cases (A) and (B).

(a) About how many students can be advised in a typical 8 hour work day?

(b) Determine an interval that contains N_8 with roughly 95% probability.

(c) What is the density of the limiting (as $t \to \infty$) residual lifetime $T_{N_t+1} - t$?

(d) What is the joint density of the limiting (as $t \to \infty$) residual lifetime $T_{N_t+1} - t$ and age $t - T_{N_t}$?

(a). $E(A) = \int_0^{\frac{1}{4}} 32x(\frac{1}{4} - x)\,dx = 32\int_0^{\frac{1}{4}} \frac{1}{4}x - x^2\,dx$

$= 32\left[\frac{1}{8}x^2 - \frac{1}{3}x^3\right]_0^{\frac{1}{4}} = 32\left[\frac{1}{8}\times\frac{1}{16} - \frac{1}{3}\times\frac{1}{64}\right]$

$= \frac{1}{12}$ / hour

$N_{8A} = \frac{8}{\frac{1}{12}} = 96.$

$\frac{t\sigma^2}{\mu^3}$

$E(A^2) = \int_0^{\frac{1}{4}} 32x^2(\frac{1}{4} - x)\,dx$

$= 32\int_0^{\frac{1}{4}} \frac{1}{4}x^2 - x^3\,dx$

(b). $N_{8A} \approx N(96, 48).$

$= 32\left[\frac{1}{12}x^3 - \frac{1}{4}x^4\right]_0^{\frac{1}{4}}$

$\frac{N_{8A} - 96}{\sqrt{48}} \approx N(0, 1)$

$= 32\left[\frac{1}{12}\cdot\frac{1}{64} - \frac{1}{4}\cdot\frac{1}{4^4}\right) = \frac{1}{96}$

95% CI $96 \pm 1.96\cdot\sqrt{48}$

96 ± 13.58 $\quad$ $\mathrm{var}(A) = \frac{1}{96} - \frac{1}{12^2} = \frac{1}{288}$ $\quad$ σ^2.

(c). Y_t roughly has density $\frac{1}{\mu}[1-F(x)]$

$$F(x) = \int_0^x 32\left(\tfrac{1}{4}-x\right)dx = 32\int_0^x \tfrac{1}{4}-x\,dx$$
$$= 32\left[\tfrac{1}{4}x - \tfrac{1}{2}x^2\right]$$
$$= 8x - 16x^2$$

$$f_{Y_t}(t) = 12\left(1-8x-16x^2\right)$$
$$= 12\,(1-4x)^2 \qquad x \in (0, \tfrac{1}{4}).$$

(d). Joint density of $(Y_t, A_t) = \frac{1}{\mu} f(x+y)$

$$= 12 \times 32\left(\tfrac{1}{4}-x-y\right).$$
$$x+y \in (0, \tfrac{1}{4}).$$

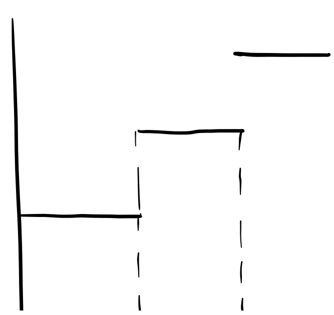
3. (Thinned Renewal Process) Consider a renewal process with mean interarrival time μ. Suppose that each event of this process is independently "counted" with probability p. Let N_t denote the number of counted events by time t; $t > 0$.

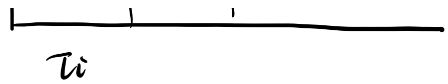
(a) Is $(N_t)_{t\geq 0}$ a renewal process?

(b) What is $\lim_{t\to\infty} N_t/t$?

(a) ~~The mean interarrival time of $(N_t)_{t\geq 0}$ process is $p\mu$.~~

The process is a renewal process





with interarrival times distributed as $\sum_{i=1}^{M} \tau_i$, where τ_i is the original renewal process interarrival distribution and M is geometric distribution with parameter p start from one and independent of the τ_i.

(b). $\lim_{t \to \infty} \frac{N_t}{t} = \frac{1}{\mu'} = \frac{p}{\mu}$

$$E\left(\sum_{i=1}^{M} \tau_i\right) = E\left[E\left(\sum_{i=1}^{M} \tau_i \,\middle|\, M\right)\right]$$

$$= E\left[M E(\tau_i)\right]$$

$$= E(M) E(\tau_i)$$

$$= \mu \cdot \frac{1}{p}$$

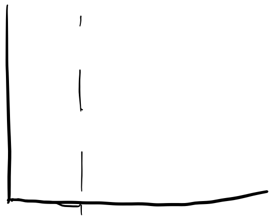
$$= \frac{\mu}{p}$$

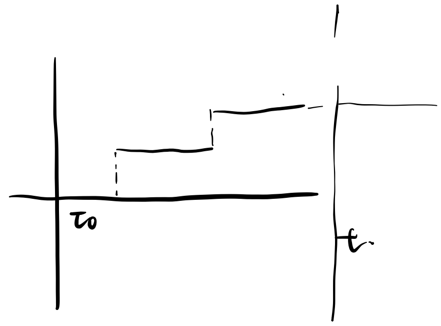
4. (Delayed renewal process) A renewal process for which the time until the initial renewal, say τ_0, has a different distribution than the remaining interarrival times, say $\tau_1, \tau_2, \ldots$, is called a delayed (or a general) renewal process. If N_t is the number of renewals of a delayed renewal process in $[0, t]$, then show that $\lim_{t\to\infty} N_t/t \to 1/\mu$, where $\mu = E[\tau_1]$.

? why $t-\tau_0$ will be a new renewal process

↓

memoryless property ?





$$N_t = N_{t-\tau_0} + 1$$

$$\lim_{t\to\infty} \frac{N_{t-\tau_0}}{t} = \frac{1}{\mu}$$

$$\lim_{t\to\infty} \frac{N_t}{t} = \lim_{t\to\infty} \frac{N_{t-\tau_0}+1}{t} = \lim_{t\to\infty} \frac{N_{t-\tau_0}}{t} + \lim_{t\to\infty} \frac{1}{t} = \frac{1}{\mu} + 0 = \frac{1}{\mu}$$

$$\frac{N_t}{t} = \frac{N_{t-\tau_0}}{t-\tau_0} \cdot \frac{t-\tau_0}{t} + \frac{1}{t}$$

$$\frac{N_{t-\tau_0}}{t-\tau_0} \to \frac{1}{\mu} \qquad \frac{t-\tau_0}{t} \to 1 \qquad \frac{1}{t} \to 0$$

5. The phenomenon in a renewal process $(N_t)_{t\geq 0}$, that $T_{N_t+1} - T_{N_t}$, the length of a typical interval straddling a large value of t, is larger than the typical interval length, τ_i, is called *size biasing.* Intervals aren't selected at random to straddle t, but longer intervals are more likely. Size biasing arises frequently in statistical sampling. Consider a remote tourist destination that wants to estimate the average length of a tourist's visit. Two sampling schemes are up for discussion. The first scheme uses the estimate of the average of the lengths of tourists' visit from a randomly chosen sample of tourists leaving the airport of the destination while the second scheme uses the estimate of the average of the lengths of tourists' visit from a randomly chosen sample of tourists staying at hotels. These averages will not be the same. Why? Which is bigger? Which scheme is the most appropriate to estimate the average length of a tourist's visit?

The average of tourists staying at hotel will be longer

The chance that a given person is sampled in a hotel is proportional to the length of their stay in that hotel

Sampling from airport gives a better estimate since each person flies out of the airport exactly once and the length of time spent in the airport doesn't vary greatly from person to person.