EE5137 Lecture 4: Motivating the Poisson Process using Goals in World Cup Games

Vincent Y. F. Tan

Department of Electrical and Computer Engineering,
National University of Singapore

Feb 2022

The Poisson Process

- The inter-arrival times of a Poisson process is the i.i.d. process $\{X_i\}_{i=1}^{\infty}$, where X_i is an exponential rv with rate λ, i.e., $X_i \sim \mathrm{Exp}(\lambda)$ or

$$f_X(x) = \lambda \exp(-\lambda x), \qquad x \geq 0$$

The Poisson Process

- The inter-arrival times of a Poisson process is the i.i.d. process $\{X_i\}_{i=1}^{\infty}$, where X_i is an exponential rv with rate λ, i.e., $X_i \sim \mathrm{Exp}(\lambda)$ or

$$f_X(x) = \lambda \exp(-\lambda x), \qquad x \geq 0$$

- The arrivals of a Poisson process is the sequence of increasing random variables $0 < S_1 < S_2 < \ldots$. In particular,

$$S_n = \sum_{j=1}^{n} X_j, \qquad n \in \mathbb{N}.$$

It is known that S_n has an Erlang pdf with shape parameter n.

The Poisson Counting Process

- The Poisson counting process $\{N(t) : t > 0\}$ is the process in which $N(t)$ represents the number of arrivals up to and including time t.

The Poisson Counting Process

- The Poisson counting process $\{N(t) : t > 0\}$ is the process in which $N(t)$ represents the number of arrivals up to and including time t.
- $N(t)$ has the Poisson pmf

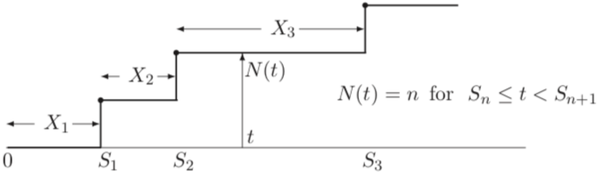
$$p_{N(t)}(n) = \frac{(\lambda t)^n \exp(-\lambda t)}{n!}, \qquad n \in \mathbb{N}$$

The Poisson Counting Process

- The Poisson counting process $\{N(t) : t > 0\}$ is the process in which $N(t)$ represents the number of arrivals up to and including time t.

- $N(t)$ has the Poisson pmf

$$p_{N(t)}(n) = \frac{(\lambda t)^n \exp(-\lambda t)}{n!}, \qquad n \in \mathbb{N}$$



Sample path of a Poisson process

Is the Poisson process a good model? An Application

- Focus on 232 World cup (soccer) games from 1990 to 2002

Is the Poisson process a good model? An Application

- Focus on 232 World cup (soccer) games from 1990 to 2002
- In the first game of 1990, Cameroon scored a single goal against Argentina at the 67th minute.

Is the Poisson process a good model? An Application

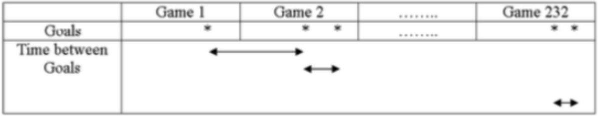
- Focus on 232 World cup (soccer) games from 1990 to 2002
- In the first game of 1990, Cameroon scored a single goal against Argentina at the 67th minute.
- In the second game, Romania scored 2 goals against the Soviet Union at the 42nd and 57th minutes.

Is the Poisson process a good model? An Application

- Focus on 232 World cup (soccer) games from 1990 to 2002
- In the first game of 1990, Cameroon scored a single goal against Argentina at the 67th minute.
- In the second game, Romania scored 2 goals against the Soviet Union at the 42nd and 57th minutes.
- Times to these goals are $65 = (90 - 67) + 42$ and $15 = 57 - 42$ min.

Is the Poisson process a good model? An Application

- Focus on 232 World cup (soccer) games from 1990 to 2002
- In the first game of 1990, Cameroon scored a single goal against Argentina at the 67th minute.
- In the second game, Romania scored 2 goals against the Soviet Union at the 42nd and 57th minutes.
- Times to these goals are $65 = (90 - 67) + 42$ and $15 = 57 - 42$ min.



Process of the arrival of goals

Property 1: Mean and Standard Deviation Coincide

- For $X \sim \mathrm{Exp}(\lambda)$, we have
$$\mathbb{E}[X] = \frac{1}{\lambda}, \qquad \mathrm{var}(X) = \frac{1}{\lambda^2}$$
so the mean and standard deviation coincide.

Property 1: Mean and Standard Deviation Coincide

- For $X \sim \mathrm{Exp}(\lambda)$, we have
$$\mathbb{E}[X] = \frac{1}{\lambda}, \qquad \mathrm{var}(X) = \frac{1}{\lambda^2}$$
so the mean and standard deviation coincide.

- Empirically,
$$\frac{1}{n}\sum_{i=1}^{n} X_i = 36.25 \text{ min}, \quad \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\right)^2} = 36.68 \text{ min}$$

Property 1: Mean and Standard Deviation Coincide

- For $X \sim \mathrm{Exp}(\lambda)$, we have

$$\mathbb{E}[X] = \frac{1}{\lambda}, \qquad \mathrm{var}(X) = \frac{1}{\lambda^2}$$

so the mean and standard deviation coincide.

- Empirically,

$$\frac{1}{n}\sum_{i=1}^{n} X_i = 36.25 \text{ min}, \quad \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\right)^2} = 36.68 \text{ min}$$

- There were 574 goals in the 232 games. Thus, $\lambda = 574/232$ goals per game.

Property 1: Mean and Standard Deviation Coincide

- For $X \sim \mathrm{Exp}(\lambda)$, we have
$$\mathbb{E}[X] = \frac{1}{\lambda}, \qquad \mathrm{var}(X) = \frac{1}{\lambda^2}$$
so the mean and standard deviation coincide.

- Empirically,
$$\frac{1}{n}\sum_{i=1}^{n} X_i = 36.25 \text{ min}, \quad \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\right)^2} = 36.68 \text{ min}$$

- There were 574 goals in the 232 games. Thus, $\lambda = 574/232$ goals per game.

- These compare well to the theoretical expectation
$$\frac{90}{\lambda} \approx 36.31 \text{ min}$$

Property 2: Memorylessness

Inter-goal Duration (minutes)	Actual	Empirical Probability	Theoretical Probability	Expected
0-10	144	0.2504	0.2407	138
10-20	106	0.1843	0.1828	105
20-30	86	0.1496	0.1388	80
30-40	52	0.0904	0.1054	60
40-50	46	0.0800	0.0800	46
50-60	27	0.0470	0.0607	35
60-70	35	0.0626	0.0461	26
70-80	16	0.0278	0.0350	20
80-90	22	0.0383	0.0266	15
90-100	12	0.0209	0.0202	12
100-110	3	0.0052	0.0153	9
110-120	3	0.0052	0.0116	7
120-130	6	0.0104	0.0088	5
130 or more	16	0.0278	0.0279	16
Total	574	1	1	574

Times Between Goals

Property 2: Memorylessness

- The interarrival times are also memoryless, i.e., that

$$\Pr(X > t + x) = \Pr(X > x) \Pr(X > t), \qquad \forall\, t, x \geq 0.$$

or

$$\Pr(X > t + x | X > t) = \Pr(X > x)$$

Property 2: Memorylessness

- The interarrival times are also memoryless, i.e., that

$$\Pr(X > t + x) = \Pr(X > x)\Pr(X > t), \qquad \forall\, t, x \geq 0.$$

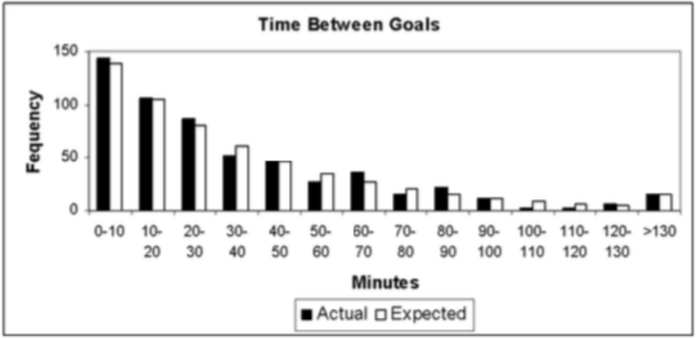
or

$$\Pr(X > t + x | X > t) = \Pr(X > x)$$

- We calculate empirically

$$\begin{aligned}
\Pr(X > 10) &= 1 - \frac{144}{574} \approx 0.7491, \\
\Pr(X > 20 | X > 10) &= \frac{574 - 144 - 106}{574 - 144} \approx 0.7534, \\
\Pr(X > 30 | X > 20) &\approx 0.7346
\end{aligned}$$

Property 2: Memorylessness



Times Between Goals

Property 3: Independence of Inter-arrival Times

- The X_i's (Inter-Goal or Inter-Arrival Times) are independent.

Property 3: Independence of Inter-arrival Times

- The X_i's (Inter-Goal or Inter-Arrival Times) are independent.
- Autocorrelations of times between goals

Lag	1	2	3	4	5	6	7	8	9
Autocorrelation	-0.0122	-0.0086	0.0021	-0.0033	0.0128	-0.0051	0.0003	0.0092	0.0033

Autocorrelation Between Goals

- All these times lie inside a band of plus or minus twice the standard error i.e., $2/\sqrt{574}$ or 0.0835.

Property 3: Independence of Inter-arrival Times

- The X_i's (Inter-Goal or Inter-Arrival Times) are independent.
- Autocorrelations of times between goals

Lag	1	2	3	4	5	6	7	8	9
Autocorrelation	-0.0122	-0.0086	0.0021	-0.0033	0.0128	-0.0051	0.0003	0.0092	0.0033

Autocorrelation Between Goals

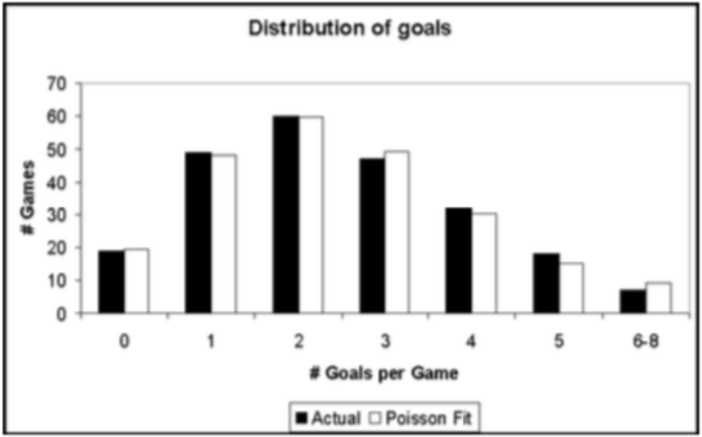
- All these times lie inside a band of plus or minus twice the standard error i.e., $2/\sqrt{574}$ or 0.0835.
- So empirically inter-arrival times also appear approximately independent.

Property 4: Poisson Counts $N(t)$

- We will prove in class that $N(t)$ is Poisson.

Property 4: Poisson Counts $N(t)$

- We will prove in class that $N(t)$ is Poisson.



Distribution of Goals per Game

Reference

- Singfat Chu, (2003) Using Soccer Goals to Motivate the Poisson Process. INFORMS Transactions on Education 3(2):64-70. https://doi.org/10.1287/ited.3.2.64