The Exponential Distribution

Recall that a continuous r.v X is *exponential* with parameter $\lambda > 0$ if its p.d.f is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

Properties related to exponentials:

• If $X \sim \text{expo}(\lambda)$, then $F_X(x) = 1 - e^{-\lambda x}$ for $x \geq 0$ and $\mathbb{E}\, X = 1/\lambda$, $\mathbb{E}\, X^2 = 2/\lambda^2$, $\text{Var}(X) = 1/\lambda^2$.

• The *memoryless* property: For all $s, t \geq 0$,

$$\mathbb{P}(X > s + t \,|\, X > t) = \mathbb{P}(X > s).$$

i.e. $\mathbb{P}(X > s + t) = \mathbb{P}(X > t) \cdot \mathbb{P}(X > s)$.

• The failure (or hazard) rate function

$$r(t) = \frac{f(t)}{1 - F(t)} \approx \frac{\mathbb{P}(X \in (t, t + \Delta t) \,|\, X > t)}{\Delta t}$$

of $X \sim \text{expo}(\lambda)$ is λ.

• If $X \sim$ expo(λ), then $\lambda X \sim$ expo(1) or $X \sim \frac{1}{\lambda}$expo(1).

• If $X_1 \sim$ expo(λ_1) and $X_2 \sim$ expo(λ_2) are ind. then $\mathbb{P}(X_1 < X_2) = \lambda_1/(\lambda_1 + \lambda_2)$.
Pf 1: Use joint density $\lambda_1 e^{-\lambda_1 x_1} \cdot \lambda_2 e^{-\lambda_1 x_2}$.
Pf 2: Conditioning, $\mathbb{P}(X_1 < X_2) = \int_0^\infty \mathbb{P}(X_1 < X_2 | X_1 = x)\lambda_1 e^{-\lambda_1 x} dx = \int_0^\infty e^{-\lambda_2 x} \cdot \lambda_1 e^{-\lambda_1 x} dx$

• If $X_1, \cdots,$ are i.i.d expo(λ) r.v.'s and $S_n = X_1 + \cdots + X_n$, then $S_n \sim$ gamma(n, λ), i.e.

$$f_{S_n}(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!}, \quad x \geq 0.$$

• If $X_1 \sim$ expo(λ_1), $\cdots, X_n \sim$ expo(λ_n) are ind. then min($X_1, \cdots, X_n$) $\sim$ expo($\lambda_1 + \cdots + \lambda_n$).
Pf: $\mathbb{P}(\min(X_1, \cdots, X_n) > x) = \prod_{i=1}^n \mathbb{P}(X_i > x) = \exp(-x \sum_{i=1}^n \lambda_i)$.

- Let $X_1 \sim \text{expo}(\lambda_1)$ and $X_2 \sim \text{expo}(\lambda_2)$ be ind. Define

$$\begin{aligned} N &= 1_{\{X_1 > X_2\}} \\ U &= \min(X_1, X_2) \\ V &= \max(X_1, X_2) \\ W &= V - U = |X_1 - X_2|. \end{aligned}$$

Then
(i). $U \sim \text{expo}(\lambda_1 + \lambda_2)$;
(ii). N and U are ind.
(iii). U and $W = V - U$ are ind.
(iv). $W|_{U=X_1} \sim \text{expo}(\lambda_2)$ and $W|_{U=X_2} \sim \text{expo}(\lambda_1)$.

- If $X_1 \sim \text{expo}(\lambda_1), \cdots, X_n \sim \text{expo}(\lambda_n)$ are ind. then

$$\frac{\lambda_i X_i}{\lambda_1 X_1 + \cdots + \lambda_n X_n} \sim \text{Uniform}(0,1) \quad \forall i.$$

• **Representation formulas:** Assume that a_i and b_i are i.i.d. expo (1). Let $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ be the order statistic of $a_1, a_2, \cdots, a_n$. Then

$$X_{1:n} = \frac{b_1}{n}, \quad X_{2:n} = \frac{b_1}{n} + \frac{b_2}{n-1}, \quad \cdots$$

$$X_{i:n} = \sum_{j=1}^{i} \frac{b_j}{n-j+1}, \quad i = 1, 2, \cdots n$$

In particular, given $a_1 \le a_2 \le \cdots \le a_n$, then

$$(a_1, a_2 - a_1, \cdots, a_n - a_{n-1}) = (\frac{b_1}{n}, \frac{b_2}{n-1}, \cdots, b_n)$$

and

$$(a_1, a_2, \cdots, a_n) = (\frac{b_1}{n}, \frac{b_2}{n-1}, \cdots, \frac{b_2}{n-1} + \cdots + b_n)$$

Let $(X_1, X_2, \cdots, X_n)$ be ind. exponential r.v.'s with $X_j \sim expo(\lambda_j)$, of which the smallest is X_I. Then $\mathbb{P}(I = i) = \lambda_i / \sum_{j=1}^{n} \lambda_j$ for $1 \le i \le n$, X_I is exponentially distributed with intensity $\sum_{j=1}^{n} \lambda_j$, the differences $X_j - X_I$, $j \ne I$, are ind. expo(λ_j), and these differences, I, and X_I are all independent.

• **Ex. Post office problem.**

• **Ex. Analyzing greedy algorithms for the assignments problem.** We see that

$$\begin{aligned} & \mathbb{E}_A(\text{total cost under expo}(1)) \\ = \; & \mathbb{E}_B(\text{total cost under expo}(1)) = \sum_{j=1}^{n} \frac{1}{j} \sim \log n. \end{aligned}$$

We also know

$$\mathbb{E}_A(\text{total cost under } U) = \sum_{j=1}^{n} \frac{1}{j+1} \sim \log n$$

and kurtzburg (1962)

$$\mathbb{E}_B(\text{total cost under } U) \sim \log n.$$

Random Assignments Problems

When elements of an n-by-n matrix are random numbers, and n elements all in different rows and columns are chosen so as to minimize their total, what is the expectation of the total?

Let $A_e(n) = \min_\sigma \sum_{j=1}^n e_{j,\sigma(j)}$
and $A_u(n) = \min_\sigma \sum_{j=1}^n u_{j,\sigma(j)}$.
• Walkup(1979)

$$\mathbb{E}\, A_u(n) \leq 3 + O(\frac{\log n}{n}) = 3 + o(1)$$

• Karp (1983, 1987)

$$\mathbb{E}\, A_u(n) \leq \frac{2n-1}{n} < 2.$$

• Lazarus (1979), $\limsup_{n\to\infty} \mathbb{E}\, A_u(n) \geq 1 + e^{-1}$.
• Mézard and Parisi (1987) used the replica method from statistical physics to argue non-rigorously that $\mathbb{E}\, A_e(n) \to \zeta(2) = \pi^2/6$.
• Goemans and Kodialam (1993)

$$\limsup_{n\to\infty} \mathbb{E}\, A_u(n) \geq 1 + e^{-1} + \varepsilon_0$$

for some fixed $\varepsilon_0 > 0$ very small.

• Aldous (1992), $\lim_{n\to\infty} \mathbb{E}\, A_u(n) = \lim_{n\to\infty} \mathbb{E}\, A_e(n)$ exists.

• Conjecture:

(i). $\mathbb{E}\, A_u(n)$ and $\mathbb{E}\, A_e(n)$ are increasing.

(ii). $\lim_{n\to\infty} \mathbb{E}\, A_u(n) = \lim_{n\to\infty} \mathbb{E}\, A_e(n) = \pi^2/6$.

• Parisi's Conjecture (1998):

$$\mathbb{E}\, A_e(n) = \sum_{j=1}^{n} \frac{1}{j^2}.$$

• Conjecture of Coppersmith and Sorkin (2000):

The expected cost of the minimum-cost matching of cardinality k in a complete bipartite graph K_{mn}, with independent random edge weights drawn from the exponential (1), is

$$\mathbb{E}\, A(k,m,n) = \sum_{i,j>0, i+j<k} \frac{1}{(m-i)(n-j)}$$

for $1 \le k \le m \le n$. In particular,

$$\mathbb{E}\, A(n,n,n) = \sum_{i+j<n} \frac{1}{(n-i)(n-j)} = \sum_{1\le j\le n} \frac{1}{n^2}.$$

Various very special cases were checked in Alm and Sorkin (2000), Coppersmith and Sorkin (2001).

•Papers of Coppersmith and Sorkin (2000, 2001) and Alm and Sorkin (2000) can be down-loaded from the web page of IBM research, Algorithms and Theory: http://www.research.ibm.com/TOC/
•Many problems in statistical physics deal with "frustrated" systems, in which not all optimization criteria can be satisfied simultaneously. Such problems make a natural bridge between this area of physics and mathematical optimization.
•The "random assignment problem" is one problem that has been studied by both computer scientists and physicists. When elements of an n-by-n matrix are random numbers, and n elements all in different rows and columns are chosen so as to minimize their total, what is the expectation of the total?

• Buck, Chan and Robbins (2002), Center for Communications Research.
Abstract: In this paper we describe our efforts to prove the Coppersmith Sorkin conjecture. We give evidence for a stronger conjecture, which generalizes theirs.

• Aldous (2002),

$$\lim_{n\to\infty} \mathbb{E}\, A_u(n) = \lim_{n\to\infty} \mathbb{E}\, A_e(n) = \pi^2/6$$

• Linusson and Wastlund (2002):
... There is a beautiful conjecture from 1998 by Parisi that says that if the a_{ij} are all exponentially distributed with mean one, then the expected value of this minimal assignment is $1+1/4+1/9+...+1/n^2$. Since the conjecture was formulated the problem has attracted a lot of attention from many different research groups in physics, combinatorics, probability and optimization.

This conjecture has been verified up to $n = 7$ using a computer aided proof. There is no explanation why such a simple answer comes up. Many partial results and some wrong proofs by physicists have been published in the last years. There have also been several generalizations of this conjecture which show surprisingly simple formulas as well. We have one of the most far reaching results so far, using tools from enumerative combinatorics. Very beautiful mathematics comes out of this, but the main conjecture is still open.

• Nair (2002): Towards the resolution of Coppersmith-Sorkin conjectures.

• Sharma and Prabhakar (2002): On Parisi's conjecture for the finite random assignment problem.

• The Finite Random Assignment Problem was solved simultaneously by (Linusson and Wastlund) and (Nair, Prabhkar and Sharma) who went about by resolving their own broader class of conjectures and thus leading to completely different solutions of Parisi's conjecture. Linusson and Wastlund have used their methods to establish the Buck et. al. conjectures as well.

• Linusson and Wastlund (2004): A proof of Parisi's conjecture.

• Nair, Prabhakar and Sharma (2004), Proofs of the Parisi and Coppersmith-Sorkin conjectures for the finite random assignment problem.

• Does $\sqrt{(n)}(A_e(n) - \mathbb{E}\, A_e(n))$ converge to to a Gaussian random variable G? Does $\mathbb{E}\, G^2 = 2$?

• W.V.Li (2002+): A completely different method is developed. In particular,

$$\mathbb{E}\,A(a_{ij}) = \left(1 + \frac{1}{n(n-1)}\right)\mathbb{E}\,A(e_{ij}) - \frac{1}{n-1}$$

and

$$\mathbb{E}\,A(b_{ij}) = \left(1 + \frac{n-2}{n(n-1)^2}\right)\mathbb{E}\,A(e_{ij}) - \frac{n-2}{(n-1)^2}$$

where

$$a_{11} = e_{11} + e_{12}, \quad a_{12} = 0, \quad a_{ij} = e_{ij} \quad \text{otherwise}$$

and

$$a_{11} = e_{11} + e_{12}, \quad a_{12} = 0, \quad a_{ij} = e_{ij} \quad \text{otherwise.}$$

Thus for $n \geq 2$,

$$\mathbb{E}\,A(a_{ij}) < \mathbb{E}\,A(b_{ij}) < \mathbb{E}\,A(e_{ij}).$$

Note that for any a,

$$\mathbb{E}\,A(a_{ij}) = \mathbb{E}\,a + \mathbb{E}\,A(e_{ij})$$

where for a fixed k, $a_{kj} = a + e_{kj}$, $1 \leq j \leq n$, and $a_{ij} = e_{ij}$, $i \neq k$, $1 \leq j \leq n$.

Uniform Distribution on Simplex

How does one sample uniformly from the unit simplex

$$S = \{(x_1, x_2, \cdots, x_n) | 0 \leq x_i \leq 1, x_1 + x_2 + \cdots + x_n = 1\}?$$

One way is to generate i.i.d random samples from an exponential distribution and take

$$X_i = \frac{e_i}{e_1 + \cdots + e_n}, \quad 1 \leq i \leq n.$$

Note that $e_i \sim -\log U_i([0,1])$ where $U_i([0,1])$ are i.i.d uniform on the interval $[0,1]$.

•For sampling from the unit sphere, a similar idea is normalizing i.i.d standard normal distribution.

•One can also generate $(n-1)$ samples from the uniform on $[0,1]$, sort them, and take the differences between the subsequent ones with zero and one as end points.

•Uniform sampling from the simplex is a special case of sampling from a Dirichlet distribution (whose main claim to fame is as the conjugate prior for a multinomial distribution). To sample from the Dirichlet distribution, you can take normalized samples from an appropriately constructed gamma distribution, which for the case of sampling from the simplex reduces to sampling from the exponential distribution as above.

Ref: 'Non-Uniform Random Variate Generation', Luc Devroye, **Free on Line**, http://cg.scs.carleton.ca/ luc/rnbookindex.html

•Let X be an exponential random variable. Without any computations, tell which one of the following is correct. Explain your answer.
a). $\mathbb{E}[X^2|X > 1] = \mathbb{E}[(X+1)^2]$.
b). $\mathbb{E}[X^2|X > 1] = \mathbb{E}[X^2] + 1$.
c) $\mathbb{E}[X^2|X > 1] = (1 + \mathbb{E}[X])^2$.

•Let $X \sim$ expo(λ). Then both $\lceil X \rceil$ and $1+[X]$ are geometric random variables with success probability $p = 1 - e^{-\lambda}$.

•Minimum of geometric random variables: Let $X_1 \sim Geom(p_1)$ and $X_2 \sim Geom(p_2)$ be independent. Then

$$X^{min} = \min(X_1, X_2) \sim Geom(p_1, p_2)$$

and

$$\mathbb{P}(X^{min} = X_i) = \frac{1 - p_i}{1 - p_1 p_2}, \quad i = 1, 2$$

Find $\mathbb{P}(X^{min} = X_1 = X_2)$.

Counting Processes

A stochastic process $\{N(t), t \geq 0\}$ is said to be a *counting* process if $N(t)$ represents the total number of "events" that have occured up to time t. A counting process $N(t)$ must satisfy:
(i). $N(t) \geq 0$.
(ii). $N(t)$ is integer valued.
(iii). If $s < t$, then $N(s) \leq N(t)$.
(iv). For $s < t$, $N(t) - N(s)$ equals the number of events that have occurred in the interval (s, t).

Def: $N(t)$ has *independent increment* if $N(t_4) - N(t_3)$ is independent of $N(t_2) - N(t_1)$ for any $t_1 < t_2 \leq t_3 < t_4$.

Def: $N(t)$ has *stationary increment* if $N(t_2 + s) - N(t_1 + s) \stackrel{d}{=} N(t_2) - N(t_1)$ for any $t_2 > t_1$, $s \geq 0$.

Homework: Ch5: 9th edition [8th](7th) . **3**, **4**(2), **6**, **7*** (5*), **10*** (8*), **12**, **18***(13*), **23***(17*), **34** [29] (19), **40***, **42**, **57*** (54*), **60*** (57*), **64***, **74**, **84*** **91***.

Poisson Process

One of the most important counting processes is the Poisson process which is defined as follows:

Def. 1: The counting process $\{N(t), t \geq 0\}$ is said to be a *Poisson process* having rate λ, $\lambda > 0$, if
(i). $N(0) = 0$.
(ii). The process has independent increments.
(iii). The number of events in any interval of length t is Poisson distributed with mean λt. That is, for all $s, t \geq 0$

$$\mathbb{P}(N(t+s) - N(s) = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n}, \quad n = 0, 1, \ldots$$

i.e. $(N(t+s) - N(s)) \sim \text{Poisson}(\lambda t)$. Note that it follows from condition (iii) that a Poisson process has stationary increments and also that $\mathbb{E}\,N(t) = \lambda t$ which explains why λ is called the rate of the process.

Alternative Definitions of a Poisson Process

Def. 2: The counting process $\{N(t), t \geq 0\}$ is a Poisson process with rate $\lambda, \lambda > 0$, if
(i). $N(0) = 0$.
(ii). The process has stationary and independent increments.
(iii). $\mathbb{P}(N(h) = 1) = \lambda h + o(h)$.
(iv). $\mathbb{P}(N(h) \geq 2) = o(h)$.

Here a function $f(\cdot)$ is said to be $o(h)$ if $\lim_{h \to 0} f(h)/h = 0$ and we write $f(h) = o(h)$ as $h \to 0$. Note that (iii) and (iv) imply that $\mathbb{P}(N(h) = 0) = 1 - \lambda h + o(h)$

Def. 3: The counting process $\{N(t), t \geq 0\}$ is a Poisson process with rate $\lambda, \lambda > 0$, if
(i). $N(0) = 0$.
(ii). The process has independent increments.
(iii). $\mathbb{P}(N(t+h) - N(t) = 1) = \lambda h + o(h)$.
(iv). $\mathbb{P}(N(t+h) - N(t) \geq 2) = o(h)$.

Equivalence of various definitions

The main part is to show, from Def. 2,

$$P_n(t) = \mathbb{P}(N(t) = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n = 0, 1, \ldots.$$

The ideas in our proofs are useful.

Pf 1: As given in class via $P_n'(t) = -\lambda P_n(t) + \lambda P_{n-1}(t)$, $P_n(0) = 0$, and inductively solve the system.

Pf 2: See later handout.

Properties of $N(t) \sim$ Poisson(λ)

- For any $0 = t_0 < t_1 < \cdots < t_m$ and $0 = k_0 \le k_1 \le \cdots \le k_m$,

$$
\begin{aligned}
& \mathbb{P}(N(t_1) = k_1, N(t_2) = k_2, \cdots, N(t_m) = k_m) \\
= \; & \mathbb{P}(N(t_1) = k_1, N(t_2) - N(t_1) = k_2 - k_1, \cdots, \\
& \qquad N(t_m) - N(t_{m-1}) = k_m - k_{m-1}) \\
= \; & \prod_{i=1}^{m} \mathbb{P}(N(t_i) - N(t_{i-1}) = k_i - k_{i-1}) \\
= \; & \prod_{i=1}^{m} \mathbb{P}(N(t_i - t_{i-1}) = k_i - k_{i-1}) \\
= \; & \prod_{i=1}^{m} e^{-\lambda(t_i - t_{i-1})} \frac{(\lambda(t_i - t_{i-1}))^{k_i - k_{i-1}}}{(k_i - k_{i-1})!} = e^{-\lambda t_m} \lambda^{k_m} \prod_{i=1}^{m} \frac{(t_i - t_{i-1})^{k_i - k_{i-1}}}{(k_i - k_{i-1})!}
\end{aligned}
$$

- Let the sequence $\{T_n, n = 1, 2, \cdots\}$ be *interarrival times*, i.e. $T_1 = \inf\{t : N(t) = 1\}$ and

$$
\begin{aligned}
T_n &= \inf\{t : N(T_1 + \cdots + T_{n-1} + t) - N(T_1 + \cdots + T_{n-1}) = 1\} \\
&= \inf\{t : N(T_1 + \cdots + T_{n-1} + t) = n\}
\end{aligned}
$$

Then T_n are i.i.d expo(λ) r.v's.

• Let S_n be the *waiting time* until the n^{th} events (the arrival time of the n^{th} events). Then

$$S_n = T_1 + \cdots + T_n, \quad \text{and} \quad S_n \sim \text{Gamma}(n, \lambda).$$

Note that the event $\{N(t) \geq n\} = \{S_n \leq t\}$ and $\{S_n > t\} = \{N(t) < n\}$ but $\{S_n \geq t\} \subset \{N(t) \leq n\}$, $\{S_n \geq t\} \neq \{N(t) \leq n\}$.

Three ways to obtain the density of S_n:
M1: Calculations based S_n is the sum of i.i.d. expo(λ) r.v's.
M2: Use $f_{S_n}(t) = F'_{S_n}(t)$ and

$$\begin{aligned} F_{S_n}(t) &= \mathbb{P}(S_n \leq t) = \mathbb{P}(N(t) \geq n) \\ &= 1 - \mathbb{P}(N(t) < n) = 1 - \sum_{k=0}^{n-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!}. \end{aligned}$$

M3: Use the fact that

$$\begin{aligned} f_{S_n}(t) \cdot \Delta t &\approx \mathbb{P}(t < S_n < t + \Delta t) \\ &\approx \mathbb{P}(N(t) = n-1, N(t+\Delta t) - N(t) = 1) \\ &= \mathbb{P}(N(t) = n-1) \cdot \mathbb{P}(N(\Delta t) = 1) \\ &\approx e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \cdot \lambda \Delta t \end{aligned}$$

and thus $f_{S_n}(t) = e^{-\lambda t} \frac{\lambda^n t^{n-1}}{(n-1)!}$.

• Given i.i.d sequence $\{T_n, n = 1, 2, \cdots\}$ with $T_i \sim \text{expo}(\lambda)$ and $S_n = T_1 + \cdots + T_n$, $S_0 = 0$, then

$$N(t) = \max\{n : S_n \le t\}$$

is a Poisson process.

• The joint distribution of $S_1, S_2, \cdots, S_n$ can be found. In the case $n = 2$, we have for $t_1 < t_2$,

$$\begin{aligned}
\mathbb{P}(S_1 > t_1, S_2 > t_2) &= \mathbb{P}(N(t_1) = 0, N(t_2) - N(t_1) \le 1) \\
&= \mathbb{P}(N(t_1) = 0) \cdot \mathbb{P}(N(t_2) - N(t_1) \le 1) \\
&= e^{-\lambda t_1} \cdot \left(e^{-\lambda(t_2 - t_1)} + e^{-\lambda(t_2 - t_1)} \lambda(t_2 - t_1)\right) \\
&= e^{-\lambda t_2} \cdot (1 + \lambda(t_2 - t_1))
\end{aligned}$$

and thus $f_{S_1, S_2}(t_1, t_2) = \lambda^2 e^{-\lambda t_2}$ for $0 < t_1 < t_2$.

• If $N_1(t) \sim \text{Poisson}(\lambda_1)$ and $N_2(t) \sim \text{Poisson}(\lambda_2)$ are ind. processes, then $N_1(t) + N_2(t) \sim \text{Poisson}(\lambda_1 + \lambda_2)$

• If $N(t) \sim \text{Poisson}(\lambda)$ and $M(t) \sim \text{Binomial}(N(t), p)$, then $M(t)$ and $N(t) - M(t)$ are ind. processes, and

$$M(t) \sim \text{Poisson}(p\lambda), \quad N(t) - M(t) \sim \text{Poisson}((1 - p)\lambda).$$

• Let $N_1(t) \sim$ Poisson(λ_1) and $N_2(t) \sim$ Poisson(λ_2) be ind. processes. Set

$$S_n^{(i)} = n^{th}\text{arrival time of } N_i(t),\ i = 1, 2.$$

Then

$$\mathbb{P}(S_n^{(1)} < S_m^{(2)}) = \sum_{k=n}^{n+m-1} \binom{n+m-1}{k} \cdot \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n+m-1-k}$$

Note that

$$\mathbb{P}(\text{an event is from } N_i(t)) = \frac{\lambda_i}{\lambda_1+\lambda_2}, \quad i = 1, 2.$$

and it is ind. of what happened before. Thus

$$\begin{aligned} \mathbb{P}(S_n^{(1)} < S_m^{(2)}) &= \mathbb{P}(\ n \text{ heads before } m \text{ tails with } p = \tfrac{\lambda_1}{\lambda_1+\lambda_2}) \\ &= \mathbb{P}(\ n \text{ or more heads in the first } n+m-1). \end{aligned}$$

and we are dealing with the *problem of points* which can be solved in at least three ways. The above approach is due to Fermat and the first step conditioning approach is due to Pascal.

Below is the disjoint partitioning approach:

$$\begin{aligned}\mathbb{P}(S_n^{(1)} < S_m^{(2)}) &= \sum_{l=n}^{n+m-1} \mathbb{P}(\ n \text{ heads in exactly } l \text{ tosses, not less}) \\ &= \sum_{l=n}^{n+m-1} \binom{l-1}{n-1} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^n \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{l-n}\end{aligned}$$

The special cases such as

$$\mathbb{P}(S_1^{(1)} < S_1^{(2)}) = \frac{\lambda_1}{\lambda_1+\lambda_2}, \quad \mathbb{P}(S_2^{(1)} < S_1^{(2)}) = \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^2$$

can be seen easily by drawing appropriate diagram.

- Let $N(t) \sim$ Poisson(λ). Then for $0 < u < t$ and $0 \le k \le n$,

$$\mathbb{P}(N(u) = k \,|\, N(t) = n) = \binom{n}{k} \left(\frac{u}{t}\right)^k \left(1 - \frac{u}{t}\right)^{n-k}$$

i.e.

$$N(u)\Big|_{N(t)=n} \sim \text{Binomial}(n, \frac{u}{t}).$$

- **Campbell's Thm:** Let $U_1(t), \cdots, U_n(t)$ be i.i.d uniform on interval $(0, t)$ and $U_{(1)}(t) \le U_{(2)}(t) \le \cdots \le U_{(n)}(t)$ be the $U_i(t)$, $1 \le i \le n$, arranged in increasing order, i.e. order statistic. Then

$$(S_1, S_2 \cdots, S_n)\Big|_{N(t)=n} \stackrel{d}{=} \Big(U_{(1)}(t), \cdots, U_{(n)}(t)\Big)$$

and

$$\left(\frac{S_1}{S_{n+1}}, \cdots, \frac{S_n}{S_{n+1}}\right) \stackrel{d}{=} \Big(U_{(1)}(1), U_{(2)}(1), \cdots, U_{(n)}(1)\Big)$$