STK2130 – Lecture 7

A. B. Huseby

Department of Mathematics
University of Oslo, Norway

Chapter 4.9 Markov Chain Monte Carlo Methods

Let Z be a discrete random variable with a state space $\mathcal{S}$, and assume that:

$$P(Z = i) = \pi_i = \frac{b_i}{B}, \quad i \in \mathcal{S}.$$

We assume that b_i is known for all $i \in \mathcal{S}$.

Since the probabilities must add up to 1, we obviously have:

$$\sum_{i \in \mathcal{S}} \frac{b_i}{B} = B^{-1} \sum_{i \in \mathcal{S}} b_i = 1,$$

Hence, it follows that the normalizing constant B is given by:

$$B = \sum_{i \in \mathcal{S}} b_i.$$

Thus, in principle B is known as well. However, if $|\mathcal{S}|$ is large, calculating B may be a time-consuming task.

MCMC Methods (cont.)

EXAMPLE: Let T and Z be two discrete random variables with state spaces $\mathcal{T}$ and $\mathcal{S}$ respectively. We assume that the marginal distribution of Z and the conditional distribution of T given Z are known.

The conditional distribution of Z given T is then:

$$P(Z = i|T = t) = \frac{P(Z = i)P(T = t|Z = i)}{\sum_{j \in \mathcal{S}} P(Z = j)P(T = t|Z = j)} = \frac{b_i(t)}{B(t)}, \quad i \in \mathcal{S},\ t \in \mathcal{T},$$

where we have introduced:

$$b_i(t) = P(Z = i)P(T = t|Z = i), \quad i \in \mathcal{S},\ t \in \mathcal{T},$$

$$B(t) = \sum_{j \in \mathcal{S}} P(Z = j)P(T = t|Z = j) = P(T = t), \quad t \in \mathcal{T}.$$

If $|\mathcal{S}|$ is large, we may want to avoid calculating $B(t)$.

MCMC Methods (cont.)

PROBLEM: Construct a Markov chain $\{X_n\}$ with state space $\mathcal{S}$ and stationary distribution equal to the distribution of Z.

SOLUTION (Hastings-Metropolis): Let $\boldsymbol{Q}$ be any given irreducible Markov chain transition probability matrix on $\mathcal{S}$, and define:

$$\alpha_{ij} = \min\left(\frac{b_j Q_{ji}}{b_i Q_{ij}}, 1\right), \quad i, j \in \mathcal{S}.$$

We then let the transition probability matrix of $\{X_n\}$, denoted $\boldsymbol{P}$, be defined as follows:

$$P_{ij} = Q_{ij}\alpha_{ij}, \quad i \neq j,$$

$$P_{ii} = 1 - \sum_{j \neq i} Q_{ij}\alpha_{ij}, \quad i \in \mathcal{S}$$

MCMC Methods (cont.)

We then claim that $\{X_n\}$ is time reversible and have a stationary distribution equal to the distribution of Z. That is, $\pi_i = b_i/B$, for all $i \in \mathcal{S}$, and:

$$\pi_i P_{ij} = \pi_j P_{ji}, \quad \text{for all } i, j \in \mathcal{S}. \tag{1}$$

Since (1) is trivially satisfied for $i = j$, we focus on the case where $i \neq j$, where (1) can be expressed as:

$$\frac{b_i}{B} Q_{ij}\alpha_{ij} = \frac{b_j}{B} Q_{ji}\alpha_{ji}, \quad i \neq j. \tag{2}$$

By eliminating B from these equations and inserting the expression for α_{ij} we get:

$$b_i Q_{ij} \cdot \min\left(\frac{b_j Q_{ji}}{b_i Q_{ij}}, 1\right) = b_j Q_{ji} \cdot \min\left(\frac{b_i Q_{ij}}{b_j Q_{ji}}, 1\right), \quad i \neq j. \tag{3}$$

MCMC Methods (cont.)

CASE 1: $b_i Q_{ij} \leq b_j Q_{ji}$

In this case $\alpha_{ij} = 1$ while $\alpha_{ji} = (b_i Q_{ij})/(b_j Q_{ji})$, and hence, (3) simplifies to:

$$b_i Q_{ij} = b_j Q_{ji} \cdot (b_i Q_{ij})/(b_j Q_{ji}), \quad i \neq j. \tag{4}$$

CASE 2: $b_i Q_{ij} \geq b_j Q_{ji}$

In this case $\alpha_{ij} = (b_j Q_{ji})/(b_i Q_{ij})$ while $\alpha_{ji} = 1$, and hence, (3) simplifies to:

$$b_i Q_{ij} \cdot (b_j Q_{ji})/(b_i Q_{ij}) = b_j Q_{ji}, \quad i \neq j. \tag{5}$$

Since obviously both (4) and (5) hold true, we conclude that (1) holds true as well.

MCMC Methods (cont.)

We recall that:

$$P_{ij} = Q_{ij}\alpha_{ij}, \quad i \neq j,$$

$$P_{ii} = 1 - \sum_{j \neq i} Q_{ij}\alpha_{ij}, \quad i \in \mathcal{S}$$

Assume that $X_n = i$. Then X_{n+1} can be generated using the following two-step Monte Carlo simulation procedure:

STEP 1. Generate $J \in \mathcal{S}$ such that $P(J = j) = Q_{ij}$, $j \in \mathcal{S}$.

STEP 2. Generate $K \in \{0, 1\}$ such that $P(K = 1 \mid J = j) = \alpha_{ij}$, and let:

$$X_{n+1} = K \cdot j + (1 - K) \cdot i$$

Thus, a transition from state i to state j where $i \neq j$ happens if and only if $J = j$ and $K = 1$. If not, the process stays in state i.

MCMC Methods (cont.)

The Monte Carlo simulation procedure can be used to estimate some unknown parameter in the distribution of Z, e.g.:

$$\theta = E[h(Z)] = \sum_{i \in \mathcal{S}} h(i) P(Z = i),$$

where h is some function of interest, and the normalizing constant B of the distribution of Z is too time-consuming to calculate.

By simulating the Markov chain $\{X_n\}$, having a stationary distribution which is equal to the distribution of Z, we may estimate θ by:

$$\hat{\theta}_n = \frac{1}{n} \sum_{m=1}^{n} h(X_m).$$

By the law of large numbers it follows that $\hat{\theta}_n \to \theta$ when $n \to \infty$.

MCMC Methods (cont.)

NOTE: $X_1, X_2, \ldots$ are not independent samples.

Moreover, the chain may converge slowly towards its stationary distribution.

Both these issues tend to have a negative effect on the convergence rate of the estimator $\hat{\theta}_n$.

If many of the α_{ij}-s are small, the Markov chain tends to get stuck for a long time before eventually transiting to another state. In such cases the estimator $\hat{\theta}_n$ will converge very slowly.

MCMC Methods (cont.)

For optimal performance, i.e., fast convergence, the matrix $\boldsymbol{Q}$ should ideally be chosen so that:

$$b_i Q_{ij} = b_j Q_{ji}, \quad \text{for all } i, j \in \mathcal{S}.$$

Then it follows that:

$$\alpha_{ij} = \min\left(\frac{b_j Q_{ji}}{b_i Q_{ij}}, 1\right) = 1, \quad \text{for all } i, j \in \mathcal{S}.$$

Hence, $\boldsymbol{Q} = \boldsymbol{P}$, i.e., $\boldsymbol{Q}$ is itself the transition probability matrix of $\{X_n\}$.

Finding the optimal matrix $\boldsymbol{Q}$ implies finding a transition probability matrix with a stationary distribution which is equal to the distribution of Z. In real-life applications, this can be difficult.

Instead we may think of $\boldsymbol{Q}$ as our best guess, while the α_{ij}-s are correction factors which are used to generate a Markov chain with the correct stationary distribution.

Gibbs sampling

Assume that $\boldsymbol{Z} = (Z_1, \ldots, Z_r)$ is a discrete random vector with values in $\mathcal{S}$ where:

$$P(\boldsymbol{Z} = \boldsymbol{z}) = p(\boldsymbol{z}) = g(\boldsymbol{z})/B, \quad \text{for all } \boldsymbol{z} \in \mathcal{S},$$

where the $g(\boldsymbol{z})$ is known for all $\boldsymbol{z} \in \mathcal{S}$ and B is an unknown normalizing constant.

We then consider the first step of the Hastings-Metropolis algorithm, and assume that $X_n = \boldsymbol{z} = (z_1, \ldots, z_r)$. The candidate for the next state, X_{n+1}, is generated as follows:

1. Generate $K = k$ uniformly from the set $\{1, \ldots, r\}$.
2. Generate $Z_k = z$ conditional on $Z_i = z_i$, $i = 1, \ldots, (k-1), (k+1) \ldots, r$.

The resulting candidate for the next state, denoted $\boldsymbol{y}$, is then:

$$\boldsymbol{y} = (z_1, \ldots, z_{k-1}, z, z_{k+1}, \ldots, z_r)$$

Gibbs sampling (cont.)

This implies that we have the following transition probabilities:

$$Q_{\boldsymbol{z},\boldsymbol{y}} = \tfrac{1}{r} P(Z_k = z \mid Z_i = z_i, i \neq k)$$

$$= \frac{g(\boldsymbol{y})/B}{r \cdot \sum_{z_k} g(\boldsymbol{y})/B} = \frac{g(\boldsymbol{y})}{r \cdot \sum_{z_k} g(\boldsymbol{y})}$$

By the same type of argument, we also have:

$$Q_{\boldsymbol{y},\boldsymbol{z}} = \frac{g(\boldsymbol{z})}{r \cdot \sum_{z_k} g(\boldsymbol{z})}.$$

However, since $\sum_{z_k} g(\boldsymbol{y}) = \sum_{z_k} g(\boldsymbol{z})$, this implies that:

$$g(\boldsymbol{z})Q_{\boldsymbol{z},\boldsymbol{y}} = g(\boldsymbol{y})Q_{\boldsymbol{y},\boldsymbol{z}}, \quad \text{for all } \boldsymbol{z}, \boldsymbol{y} \in \mathcal{S}.$$

Hence, $\alpha_{\boldsymbol{z},\boldsymbol{y}} = 1$ for all $\boldsymbol{z}, \boldsymbol{y} \in \mathcal{S}$, and thus, $\boldsymbol{Q}$ is an optimal transition probability matrix.

Chapter 5

The Exponential Distribution and the Poisson Process

Chapter 5.2 The Exponential Distribution

A continuous random variable X is said to have an exponential distribution with parameter $\lambda > 0$, denoted as $X \sim exp(\lambda)$, if its probability density function is given by:

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

If $X \sim exp(\lambda)$, then the cumulative distribution function of X is given by:

$$F(x) = P(X \leq x) = \int_0^x f(t)dt = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Moreover, the survival function of X is given by:

$$\bar{F}(x) = P(X > x) = 1 - F(x) = \begin{cases} e^{-\lambda x} & x \geq 0 \\ 1 & x < 0 \end{cases}$$

The Exponential Distribution (cont.)

The exponential distribution is a special case of the gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$, denoted as $X \sim gamma(\alpha, \lambda)$ with probability density function:

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where $\Gamma(\alpha)$, defined for all $\alpha > 0$, is the gamma function given by:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \qquad \Gamma(n) = (n-1)!, \quad n = 1, 2, \ldots$$

By substituting $u = \lambda x$ and $du = \lambda dx$, we find that:

$$\int_0^\infty f(x) dx = \frac{1}{\Gamma(\alpha)} \int_0^\infty u^{\alpha-1} e^{-u} du = 1.$$

Thus, $f(x)$ is indeed a proper probability density.

The Exponential Distribution (cont.)

Assume that $X \sim exp(\lambda)$, and let $p > -1$. We then have:

$$\begin{aligned} E[X^p] &= \int_0^\infty x^p f(x)dx = \int_0^\infty \lambda x^p e^{-\lambda x} dx \\ &= \frac{\Gamma(p+1)}{\lambda^p} \int_0^\infty \frac{\lambda^{p+1}}{\Gamma(p+1)} x^{(p+1)-1} e^{-\lambda x} dx \\ &= \frac{\Gamma(p+1)}{\lambda^p}. \end{aligned}$$

In particular:

$$E[X] = \frac{\Gamma(2)}{\lambda^1} = \frac{(2-1)!}{\lambda} = \frac{1}{\lambda}, \qquad E[X^2] = \frac{\Gamma(3)}{\lambda^2} = \frac{(3-1)!}{\lambda^2} = \frac{2}{\lambda^2},$$

$$\mathrm{Var}[X] = E[X^2] - (E[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

The Exponential Distribution (cont.)

Assume that $X \sim gamma(\alpha, \lambda)$. Then the moment generating function of X is given by:

$$M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx$$

$$= \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda-t)x} dx$$

$$= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} \int_0^\infty \frac{(\lambda-t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda-t)x} dx$$

$$= \frac{\lambda^\alpha}{(\lambda-t)^\alpha}, \quad \text{for all } t < \lambda.$$

In particular, if $X \sim exp(\lambda)$, we have:

$$M_X(t) = \frac{\lambda}{\lambda - t}, \quad \text{for all } t < \lambda.$$

The Exponential Distribution (cont.)

Proposition (5.1)

Assume that $X_1, \ldots, X_n$ are independent and $X_i \sim exp(\lambda)$, $i = 1, \ldots, n$, and let:

$$Y = X_1 + \cdots + X_n$$

Then $Y \sim gamma(n, \lambda)$.

PROOF: Using moment generating functions we get:

$$M_Y(t) = E[e^{tY}] = E[e^{tX_1 + \cdots + tX_n}] = M_{X_1}(t) \cdots M_{X_n}(t)$$

$$= \frac{\lambda}{\lambda - t} \cdots \frac{\lambda}{\lambda - t} = \frac{\lambda^n}{(\lambda - t)^n}$$

Hence, $Y \sim gamma(n, \lambda)$.

The Exponential Distribution (cont.)

Proposition 5.1 is a special case of the following more general result:

Proposition (5.1b)

Assume that $X_1, \ldots, X_n$ *are independent and* $X_i \sim gamma(\alpha_i, \lambda)$, $i = 1, \ldots, n$, *and let:*

$$Y = X_1 + \cdots + X_n$$

Then $Y \sim gamma(\alpha, \lambda)$, *where* $\alpha = \sum_{i=1}^{n} \alpha_i$.

PROOF: Using moment generating functions we get:

$$M_Y(t) = E[e^{tY}] = E[e^{tX_1 + \cdots + tX_n}] = M_{X_1}(t) \cdots M_{X_n}(t)$$

$$= \frac{\lambda^{\alpha_1}}{(\lambda - t)^{\alpha_1}} \cdots \frac{\lambda^{\alpha_n}}{(\lambda - t)^{\alpha_n}} = \frac{\lambda^{\alpha}}{(\lambda - t)^{\alpha}}$$

Hence, $Y \sim gamma(\alpha, \lambda)$.

Memoryless stochastic variables

A random variable X is said to be memoryless if:

$$P(X > s + t | X > t) = P(X - t > s | X > t) = P(X > s), \quad \text{for all } s, t \geq 0.$$

Thus, X is memoryless if $(X - t) | (X > t)$ has the same distribution as X.

Note that if X is the lifetime of some unit, $(X - t)$ is the remaining lifetime given that the unit has survived up to the time t.

If $X \sim exp(\lambda)$, we have:

$$P(X > s + t | X > t) = \frac{P(X > s + t \cap X > t)}{P(X > t)} = \frac{P(X > s + t)}{P(X > t)}$$

$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda(t)}} = e^{-\lambda s} = P(X > s)$$

Hence, we conclude that X is memoryless.

Memoryless stochastic variables (cont.)

The memoryless property:

$$P(X > s + t | X > t) = P(X > s), \quad \text{for all } s, t \geq 0.$$

is equivalent to the following:

$$P(X > s + t) = P(X > s)P(X > t), \quad \text{for all } s, t \geq 0.$$

Since $\bar{F}(x) = P(X > x)$, this property can also be written as:

$$\bar{F}(s + t) = \bar{F}(s)\bar{F}(t), \quad \text{for all } s, t \geq 0.$$

We now show that the exponential distribution is essentially the only distribution with this property.

Memoryless stochastic variables (cont.)

Proposition

Let X be a random variable and let $\bar{F}(x) = P(X > x)$ be such that:

$$\bar{F}(x+y) = \bar{F}(x) \cdot \bar{F}(y), \qquad \textit{for all } x, y \geq 0. \tag{6}$$

$$\lambda = -\log(\bar{F}(1)) > 0. \tag{7}$$

Then $X \sim exp(\lambda)$.

PROOF: We first note that by (7), it follows that:

$$0 < \bar{F}(1) = e^{-\lambda} < 1. \tag{8}$$

Secondly we note that since cumulative distribution functions always are right-continuous, it follows that $\bar{F} = 1 - F$ is righ-continuous as well.

Memoryless stochastic variables (cont.)

By repeated use of (6) it follows that for $n, m \in \mathbb{N}^+$, we have:

$$\bar{F}(\tfrac{m}{n}) = \bar{F}(\tfrac{1}{n} + \cdots + \tfrac{1}{n}) = \bar{F}^m(\tfrac{1}{n}), \tag{9}$$

where the sum contains m terms. In particular, by letting $m = n$, we get:

$$\bar{F}(1) = \bar{F}(\tfrac{n}{n}) = \bar{F}^n(\tfrac{1}{n}). \tag{10}$$

Alternatively, (10) can be written as:

$$\bar{F}(\tfrac{1}{n}) = [\bar{F}(1)]^{1/n}. \tag{11}$$

By (8) and that $\bar{F}$ is right-continuous, (11) implies that:

$$\bar{F}(0) = \lim_{n\to\infty} \bar{F}(\tfrac{1}{n}) = \lim_{n\to\infty} [\bar{F}(1)]^{1/n} = 1.$$

Hence, since $\bar{F}$ must be non-increasing, $\bar{F}(x) = 1$ for all $x \leq 0$.

Memoryless stochastic variables (cont.)

We now combine (9) and (11), and get:

$$\bar{F}(\tfrac{m}{n}) = \bar{F}^m(\tfrac{1}{n}) = \bar{F}(1)^{m/n}, \quad \text{for all } m, n \in \mathbb{N}^+.$$

Thus, since $\bar{F}(1) = e^{-\lambda}$, we have proved that:

$$\bar{F}(q) = \bar{F}(1)^q = e^{-\lambda q}, \quad \text{for all } q \in \mathbb{Q}^+.$$

Now, let $x \in \mathbb{R}^+$. Since the set $\mathbb{Q}^+$ is dense in $\mathbb{R}^+$, there exists a decreasing sequence $\{q_r\} \subset \mathbb{Q}^+$ such that:

$$\lim_{r \to \infty} q_r = x$$

Since $\bar{F}$ is right-continuous, this implies that:

$$\bar{F}(x) = \lim_{r \to \infty} \bar{F}(q_r) = \lim_{r \to \infty} e^{-\lambda q_r} = e^{-\lambda x}.$$

Hence, we conclude that $X \sim exp(\lambda)$ ∎

Example 5.2

The amount of time one spends in a bank, denoted X, is exponentially distributed with mean ten minutes. That is, $X \sim exp(\lambda) = exp(\frac{1}{10})$.

PROBLEM 1: What is the probability that a customer will spend more than fifteen minutes in the bank?

SOLUTION:

$$P(X > 15) = e^{-15\lambda} = e^{-15/10} \approx 0.223$$

PROBLEM 2: What is the probability that a customer will spend more than fifteen minutes in the bank given that she is still in the bank after ten minutes?

SOLUTION:

$$P(X > 15 \mid X > 10) = e^{-(15-10)\lambda} = e^{-5/10} \approx 0.607$$

The Exponential Distribution (cont.)

Assume that X_1, X_2 are independent and that $X_i \sim exp(\lambda_i)$, $i = 1, 2$. We want to calculate the probability of the event that $X_1 < X_2$.

$$\begin{aligned}
P(X_1 < X_2) &= \int_0^\infty P(X_1 < X_2 | X_1 = x) \lambda_1 e^{-\lambda_1 x} dx \\
&= \int_0^\infty P(X_2 > x) \lambda_1 e^{-\lambda_1 x} dx \\
&= \int_0^\infty e^{-\lambda_2 x} \lambda_1 e^{-\lambda_1 x} dx \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2} \int_0^\infty (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda_2) x} dx \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2}.
\end{aligned}$$

The Exponential Distribution (cont.)

Assume that $X_1, \ldots, X_n$ are independent and that $X_i \sim exp(\lambda_i)$, $i = 1, \ldots, n$.

$$
\begin{aligned}
P(\min_{1 \leq i \leq n} X_i > x) &= P(\bigcap_{i=1}^{n} X_i > x) \\
&= \prod_{i=1}^{n} P(X_i > x) \qquad \text{(by independence)} \\
&= \prod_{i=1}^{n} e^{-\lambda_i x} \\
&= e^{-(\sum_{i=1}^{n} \lambda_i)x}
\end{aligned}
$$

Thus, we have shown that $\min_{1 \leq i \leq n} X_i \sim exp(\sum_{i=1}^{n} \lambda_i)$.

The Exponential Distribution (cont.)

The following result combines the two previous results:

Assume that $X_1, \ldots, X_n$ are independent and that $X_i \sim exp(\lambda_i)$, $i = 1, \ldots, n$. We want to calculate the probability that X_i is the smallest of all the variables, i.e., that $X_i < X_j$ for all $j \neq i$.

$$P(\bigcap_{j \neq i} [X_i < X_j]) = P(X_i < \min_{j \neq i} X_j)$$

$$= \frac{\lambda_i}{\lambda_i + \sum_{j \neq i} \lambda_j}, \qquad \text{since } \min_{j \neq i} X_j \sim exp(\textstyle\sum_{j \neq i} \lambda_j)$$

$$= \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}$$

The Exponential Distribution (cont.)

Proposition (5.2)

Assume that $X_1, \ldots, X_n$ are independent and that $X_i \sim exp(\lambda_i)$, $i = 1, \ldots, n$. Then $\min_i X_i \sim exp(\sum_{i=1}^n \lambda_i)$. Moreover, $\min_i X_i$ and the rank order of $X_1, \ldots, X_n$ are independent.

PROOF: Since the exponential distribution is memoryless, we get that:

$$P(X_{i_1} < \cdots < X_{i_n} \mid \min_{1 \leq i \leq n} X_i > t)$$

$$= P(X_{i_1} < \cdots < X_{i_n} \mid \bigcap_{i=1}^{n} X_i > t)$$

$$= P(X_{i_1} - t < \cdots < X_{i_n} - t \mid \bigcap_{i=1}^{n} X_i > t)$$

$$= P(X_{i_1} < \cdots < X_{i_n})$$

Example 5.8

A post office with two busy clerks. No one is waiting in line except you.

$$R_i = \text{Time until for clerk } i \text{ becomes available}, \quad i = 1,2$$

$$S = \text{Your service time}$$

$$T = \text{The total time spent in the post office}$$

We assume that R_1, R_2 are independent and $R_i \sim exp(\lambda_i)$, $i = 1, 2$.

$$E[T] = E[T|R_1 < R_2]P(R_1 < R_2) + E[T|R_2 \leq R_1]P(R_2 < R_1)$$
$$= E[R_1 + S|R_1 < R_2]\frac{\lambda_1}{\lambda_1 + \lambda_2} + E[R_2 + S|R_2 < R_1]\frac{\lambda_2}{\lambda_1 + \lambda_2}$$

We get:

$$E[R_i|R_i < R_{3-i}] = E[\min(R_1, R_2)] = \frac{1}{\lambda_1 + \lambda_2}, \quad i = 1,2$$

Moreover, we assume that $S|R_i < R_{3-i} \sim exp(\lambda_i)$, $i = 1, 2$, and get:

$$E[S|R_i < R_{3-i}] = \frac{1}{\lambda_i}, \quad i = 1,2$$

Example 5.8 (cont.)

Combining these results we get:

$$
\begin{aligned}
E[T] &= E[R_1 + S | R_1 < R_2] \frac{\lambda_1}{\lambda_1 + \lambda_2} + E[R_2 + S | R_2 < R_1] \frac{\lambda_2}{\lambda_1 + \lambda_2} \\
&= \left(\frac{1}{\lambda_1 + \lambda_2} + \frac{1}{\lambda_1} \right) \frac{\lambda_1}{\lambda_1 + \lambda_2} + \left(\frac{1}{\lambda_1 + \lambda_2} + \frac{1}{\lambda_2} \right) \frac{\lambda_2}{\lambda_1 + \lambda_2} \\
&= \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} + 1 \right) \frac{1}{\lambda_1 + \lambda_2} + \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} + 1 \right) \frac{1}{\lambda_1 + \lambda_2} \\
&= \left(\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2} + 1 + 1 \right) \frac{1}{\lambda_1 + \lambda_2} \\
&= \frac{3}{\lambda_1 + \lambda_2}.
\end{aligned}
$$