Basic concepts of probability theory

February 11, 2024

Contents

1 Introduction

- In general we do not like to wait. But reduction of the waiting time usually requires extra investments.
- So we need models and techniques to analyse such situations.
- In this course we treat a number of elementary queueing models.

2 Basic concepts of probability theory

2.1 Random variable

Definition 2.1. Assuming S represents the sample space of any random experiment, a random variable is defined as a function that assigns a real number to each element (outcome) of S; it can alternatively be represented by a capital letter (e.g., X,Y,...). Thus, any random variable X maps the outcomes of a random experiment to real numbers. This can be expressed as follows:

- A Random variables are denoted by capitals, X, Y , etc.
  - The expected value or mean of X is denoted by E(X) and its variance by $\sigma^2(X)$
  - $\sigma(X)$ is the standard deviation of X.
- An important quantity is the coefficient of variation of the positive random variable X defined as:

$$C_X = \frac{\sigma(X)}{E(X)} \tag{1}$$

- The coefficient of variation is a (dimensionless) measure of the variability of the random variable X.

3 Useful probability distributions

This section discusses a number of important distributions which have been found useful for describing random variables in many applications.

3.1 Geometric distribution

- A geometric random variable X with parameter p has probability distribution

$$P(X = n) = p\dot{(}1 - p) \tag{2}$$

Note that $|G_X(t)| \leq 1$ for all $|t| \leq 1$

$$G_X(0) = p(0),\ G_X(1) = 1,\ G_X^{'}(1) = E[X] \tag{3}$$

and, more general,

$$G_X^k(1) = E[X(X-1)...(X-k+1)] \tag{4}$$

where the superscript (k) denotes the k^{th} derivative.

3.2 Generating function

- Let X be a nonnegative discrete random variable with $P(X = x_i) = p_i, i = 0, 1, 2, ...$ Then the generating function $P_X(z)$ of X is defined as

$$G_X(t) = E[X^t] = \sum_{i=0}^{\infty} p(X = x_i)t^{x_i} \tag{5}$$

Note that $|G_X(t)| \leq 1$ for all $|t| \leq 1$

$$G_X(0) = p_0\dot{,}\ G_X(1) = 1,\ G_X^{'}(1) = E[X] \tag{6}$$

and, more general,

$$G_X^{(k)}(1) = E[X(X-1)...(X-k+1)] \tag{7}$$

where the superscript k denotes the k^{th} derivative.

- For the generating function of the sum $Z = X + Y$of two independent discrete random variables X and Y , it holds that

$$G_Z(t) = \ G_X(t)\dot{G}_Y(t) \tag{8}$$

3.3 Laplace-Stieltjes transform

- The Laplace-Stieltjes transform $\tilde{X}(s)$ of a nonnegative random variable X with distribution function $F\dot{()}$, is defined as:

$$\tilde{X}(s) = E(e^{-sX}) = \int_{x=0}^{\infty} e^{-sx} dF(x),\ s \geq 0 \tag{9}$$

- When the random variable X has a density $f\dot{()}$, then the transform simplifies to:

$$\tilde{X}(s) = E(e^{-sX}) = \int_{x=0}^{\infty} e^{-sx} f(x)dx,\ s \geq 0 \tag{10}$$

- Note that $\left|\tilde{X}(s)\right\| \leq 1$ for all $s \geq 0$. Further

$$\tilde{X}(0) = 1,\ \tilde{X}^{'}(0) = -E(X),\ \tilde{X}^{(k)}(0) = (-1)^k E(X^k) \tag{11}$$

- For the transform of the sum $Z = X + Y$ of two independent random variables X and Y , it holds that:

$$\tilde{Z}(s) = \tilde{X}(s)\dot{\tilde{Y}}(s) \tag{12}$$

.

- When Z is with probability q equal to X and with probability $1-q$ equal to Y , then:

$$\tilde{Z}(s) = q\tilde{X}(s) + (1-q)\tilde{Y}(s) \tag{13}$$

.

3.4 Poisson distribution

- A Poisson random variable X with parameter λ has probability distribution:

$$P(X = n) = \frac{\lambda^n}{n!}e^{-\lambda} \tag{14}$$

For the Poisson distribution it holds that:

$$P_X(z) == e^{-\lambda(1-z)},\ E[X] = \sigma^2(X) = \lambda,\ C_X^2 = \frac{1}{\lambda} \tag{15}$$

3.5 Exponential distribution

- The density of an exponential distribution with parameter λ is given by:

$$f(t) = \lambda e^{-\lambda t},\ t > 0 \tag{16}$$

- The distribution function equals:

$$F(t) = 1 - e^{-\lambda t},\ t \geq 0 \tag{17}$$

- For this distribution we have:

$$\tilde{X}(s) = \frac{\lambda}{\lambda + s},\ E[X] = \frac{1}{\lambda},\ \sigma^2(X) = \frac{!}{\lambda^2},\ C_X = 1 \tag{18}$$

- An important property of an exponential random variable X with parameter λ is the memoryless property. This property states that for all $x \geq 0$ and $t \geq 0$,

$$P(X > x + t/X > t) = P(X > x) = e^{-\lambda x} \tag{19}$$

3.6 Erlang distribution

- A random variable X has an Erlang-k ($k = 1, 2, ...$) distribution with mean $\frac{k}{\mu}$ if X is the sum of k independent random variables $X_1, ..., X_n$ having a common exponential distribution with mean $\frac{1}{\mu}$. The common notation is $E_k(\mu)$ or briefly Ek. The density of an $E_k(\mu)$ distribution is given by:

$$f(t) = \mu\frac{(\mu t)^{k-1}}{(k-1)!}e^{-\mu t}, \quad t > 0 \tag{20}$$

- The distribution function equals:

$$F(t) = 1 - \sum_{i=0}^{k-1}\frac{(\mu t)^i}{(i)!}e^{-\mu t}, \quad t \geq 0 \tag{21}$$

- The parameter μ is called the scale parameter, k is the shape parameter. A phase diagram of the E_k distribution is shown in
figure 2.1
- In figure 2.2 we display the density of the Erlang-k distribution with mean 1 (so $\mu = k$) for various

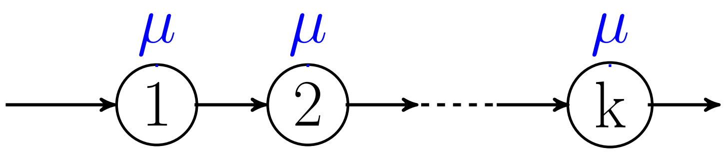

Figure 1: Phase diagram for the Erlang-k distribution with scale parameter μ

values of k.

- The mean, variance and squared coefficient of variation are equal to:

$$E(X) = \frac{k}{\mu}, \qquad \sigma^2(X) = \frac{k}{\mu^2}, \quad C_X^2 = \frac{1}{k}. \tag{22}$$

- The Laplace-Stieltjes transform is given by:

$$\tilde{X}(s) = \left(\frac{\mu}{\mu + s}\right)^2. \tag{23}$$

- A convenient distribution arises when we mix an E_{k-1} and E_k distribution with thesame scale parameters. The notation used is $E_{k-1,k}$. A random variable X has an $E_{k-1,k}(\mu)$ distribution, if X is with probability p (resp. $1-p$) the sum of $k-1$ (resp. k) independent exponentials with common mean $\frac{1}{\mu}$.
- The density of this distribution has the form:

$$f(t) = p\mu\frac{(\mu t)^{k-2}}{(k-2)!}e^{-\mu t} + (1-p)\mu\frac{(\mu t)^{k-1}}{(k-1)!}e^{-\mu t}, \quad t > 0 \tag{24}$$

- where $0 \leq p \leq 1$. As p runs from 1 to 0, the squared coeffcient of variation of the mixed Erlang distribution varies from $\frac{1}{(k-1)}$ to $\frac{1}{k}$. It will appear (later on) that this distribution is useful for fitting a distribution if only the first two moments of a random variable are known.

3.7 Hyperexponential distribution

- A random variable X is hyperexponentially distributed if X is with probability p_i, $i = 1; \ldots; k$ an exponential random variable X_i with mean $\frac{1}{\mu_i}$. For this random variable we use the notation $H_k(p_1; \ldots; p_k; \frac{1}{\mu_1}; \ldots; \frac{1}{\mu_k})$, or simply H_k. The density is given by:

$$f(t) = \sum_{i=1}^{k} p_i \mu_i e^{-\mu_i t}, \quad t > 0 \tag{25}$$

- and the mean is equal to:

$$E[X] = \sum_{i=1}^{k} \frac{p_i}{\mu_i} \tag{26}$$

- The Laplace-Stieltjes transform satisfies:

$$\tilde{X}(s) = \sum_{i=1}^{k} \frac{p_i \mu_i}{\mu_i + s}. \tag{27}$$

- The cofficient of variation C_x of this distribution is always greater than or equal to 1

3.8 Phase-type distribution

The preceding distributions are all special cases of the phase-type distribution. The notation is PH. This distribution is characterized by a Markov chain with states $1; \ldots; k$ (the so called phases) and a

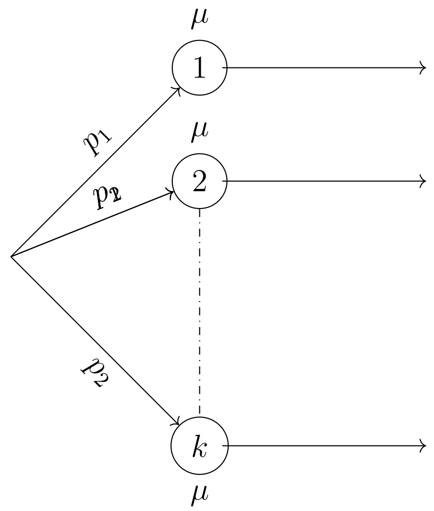

Figure 2: Phase diagram for the hyperexponential distribution

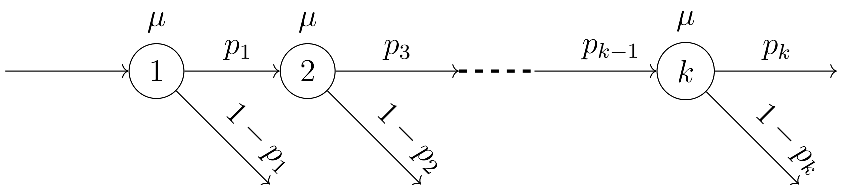

Figure 3: Phase diagram for the hyperexponential distribution

transition probability matrix P which is transient. This means that P^n tends to zero as n tends to infinity. In words, eventually you will always leave the Markov chain. The residence time in state i is exponentially distributed with mean $\frac{1}{\mu_i}$, and the Markov chain is entered with probability pi in state i, $i = 1; \ldots; k$. Then therandom variable X has a phase-type distribution if X is the total residence time in the preceding Markov chain, i.e. X is the total time elapsing from start in the Markov chain till departure from the Markov chain. We mention two important classes of phase-type distributions which are dense in the class of all non-negative distribution functions. This is meant in the sense that for any non-negative distribution function F(.) a sequence of phase-type distributions can be found which pointwise converges at the points of continuity of F(.). The denseness of the two classes makes them very useful as a practical modelling tool. A proof of the denseness can be found in [23, 24]. The first class is the class of Coxian distributions, notation Ck, and the other class consists of mixtures of Erlang distributions with the same scale parameters. The phase representations of these two classes are shown in figures 4 and 5. A random variable X has a Coxian distribution of order k if it has to go through up to at most k exponential phases. The mean length of phase n is $\frac{1}{\mu_n}, n = 1; \ldots; k$. It starts in phase 1. After phase n it comes to an end with probability $1 - p_n$ and it enters the next phase with probability p_n. Obviously $p_k = 0$. For the Coxian-2 distribution it holds that the squared coefficient of variation is greater than or equal to 0.5 (see exercise 8). A random variable X has a mixed Erlang distribution of order k if it is with probability p_n the sum of n exponentials with the same mean $\frac{1}{\mu_n}, n = 1; \ldots; k$.

3.9 Fitting distributions

In practice it often occurs that the only information of random variables that is available is their mean and standard deviation, or if one is lucky, some real data. To obtain an approximating distribution it is common to t a phase-type distribution on the mean, $E(X)$, and the coefficient of variation, C_X, of a given positive random variable X, by using the following simple approach.

- In case $0 < cX < 1$ one fits an $E_{k-1;k}$ distribution (see subsection 2.4.4). More specically, if

$$\frac{1}{k} \leq C_X^2 \leq \frac{1}{k-1} \tag{28}$$

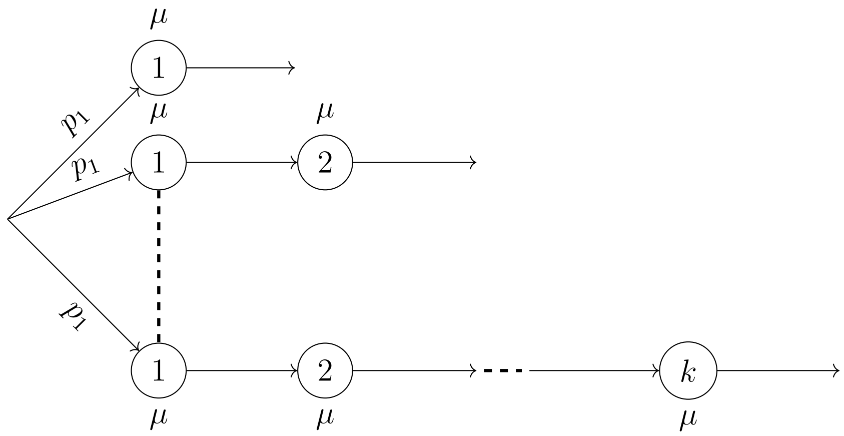

Figure 4: Phase diagram for the mixed Erlang distribution

- for certain $k = 2; 3; \ldots$, then the approximating distribution is with probability p (resp. $1-p$) the sum of $k-1$ (resp. k) independent exponentials with common mean $\frac{1}{\mu}$. By choosing

$$p = \frac{1}{1+C_X^2}\left[kC_X^2 - \left(k(1+C_X^2) - K^2C_X^2\right)\right] \tag{29}$$

the $E_{k-1;k}$ distribution matches $E(X)$ and C_X.

- In case $C_X \geq 1$ one fits a $H_2(p_1; p_2; \mu_1; \mu_2)$ distribution. The hyperexponential distribution however is not uniquely determined by its first two moments. In applications, the H_2 distribution with balanced means is often used. This means that the normalization

$$\frac{p_1}{\mu_1} = \frac{p_2}{\mu_2} \tag{30}$$

is used,

- The parameters of the H_2 distribution with balanced means and fitting $E(X)$ and C_X (≥ 1)vare given by

$$p_1 = \frac{1}{2}\left(1 + \sqrt{\frac{C_X^2 - 1}{C_X^2 + 1}}\right), \; p_2 = 1 - p_1, \; \mu_1 = \frac{p_1}{E[X]}, \; \mu_2 = \frac{p_2}{E[X]}$$

the $E_{k-1;k}$ distribution matches $E(X)$ and C_X.

- In case $C_X \geq 1$ one fits a $H_2(p_1; p_2; \mu_1; \mu_2)$ distribution. The hyperexponential distribution however is not uniquely determined by its first two moments. In applications, the H_2 distribution with balanced means is often used. This means that the normalization

$$\frac{p_1}{\mu_1} = \frac{p_2}{\mu_2} \tag{31}$$

is used,

- In case $C_X^2 \geq 0.5$ one can also use a Coxian-2 distribution for a two-moment fit.

$$\mu_1 = \frac{2}{E[X]}, \; p_1 = \frac{0,5}{C_X^2}, \; \mu_2 = \mu_2 p_1$$

It also possible to make a more sophisticated use of phase-type distributions by, e.g., trying to match the
rst three (or even more) moments of X or to approximate the shape of X.
Phase-type distributions may of course also naturally arise in practical applications. For example, if the processing of a job involves performing several tasks, where each task takes an exponential amount of time, then the processing time can be described by an Erlang distribution.