1 Basic concepts from probability theory

This chapter is devoted to some basic concepts from probability theory.

1.1 Random variable

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 $\sigma^2(X)$ where $\sigma(X)$ is the standard deviation of X.

An important quantity is the *coefficient of variation* of the positive random variable X defined as

$$c_X = \frac{\sigma(X)}{E(X)}.$$

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

1.2 Generating function

Let X be a nonnegative discrete random variable with $P(X = n) = p(n)$, $n = 0, 1, 2, \ldots$. Then the generating function $P_X(z)$ of X is defined as

$$P_X(z) = E(z^X) = \sum_{n=0}^{\infty} p(n)z^n.$$

Note that $|P_X(z)| \leq 1$ for all $|z| \leq 1$. Further

$$P_X(0) = p(0), \qquad P_X(1) = 1, \qquad P_X'(1) = E(X),$$

and, more general,

$$P_X^{(k)}(1) = E(X(X - 1) \cdots (X - k + 1)),$$

where the superscript (k) denotes the kth derivative. For the generating function of the sum $Z = X + Y$ of two *independent* discrete random variables X and Y, it holds that

$$P_Z(z) = P_X(z) \cdot P_Y(z).$$

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

$$P_Z(z) = qP_X(z) + (1 - q)P_Y(z).$$

1.3 Laplace-Stieltjes transform

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

$$\widetilde{X}(s) = E(e^{-sX}) = \int_{x=0}^{\infty} e^{-sx} dF(x), \qquad s \geq 0.$$

When the random variable X has a density $f(\cdot)$, then the transform simplifies to

$$\widetilde{X}(s) = \int_{x=0}^{\infty} e^{-sx} f(x) dx, \qquad s \geq 0.$$

Note that $|\widetilde{X}(s)| \leq 1$ for all $s \geq 0$. Further

$$\widetilde{X}(0) = 1, \qquad \widetilde{X}'(0) = -E(X), \qquad \widetilde{X}^{(k)}(0) = (-1)^k E(X^k).$$

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

$$\widetilde{Z}(s) = \widetilde{X}(s) \cdot \widetilde{Y}(s).$$

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

$$\widetilde{Z}(s) = q\widetilde{X}(s) + (1 - q)\widetilde{Y}(s).$$

1.4   Useful probability distributions

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

1.4.1   Geometric distribution

A geometric random variable X with parameter p has probability distribution

$$P(X = n) = (1 - p)p^n, \qquad n = 0, 1, 2, \ldots$$

For this distribution we have

$$P_X(z) = \frac{1-p}{1-pz}, \qquad E(X) = \frac{p}{1-p}, \qquad \sigma^2(X) = \frac{p}{(1-p)^2}, \qquad c_X^2 = \frac{1}{p}.$$

1.4.2   Poisson distribution

A Poisson random variable X with parameter μ has probability distribution

$$P(X = n) = \frac{\mu^n}{n!} e^{-\mu}, \qquad n = 0, 1, 2, \ldots$$

For the Poisson distribution it holds that

$$P_X(z) = e^{-\mu(1-z)}, \qquad E(X) = \sigma^2(X) = \mu, \qquad c_X^2 = \frac{1}{\mu}.$$

1.4.3 Exponential distribution

The density of an exponential distribution with parameter μ is given by

$$f(t) = \mu e^{-\mu t}, \qquad t > 0.$$

The distribution function equals

$$F(t) = 1 - e^{-\mu t}, \qquad t \geq 0.$$

For this distribution we have

$$\widetilde{X}(s) = \frac{\mu}{\mu + s}, \qquad E(X) = \frac{1}{\mu}, \qquad \sigma^2(X) = \frac{1}{\mu^2}, \qquad c_X = 1.$$

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

$$P(X > x + t | X > t) = P(X > x) = e^{-\mu x}.$$

So the remaining lifetime of X, given that X is still alive at time t, is again exponentially distributed with the same mean $1/\mu$. We often use the memoryless property in the form

$$P(X < t + \Delta t | X > t) = 1 - e^{-\mu \Delta t} = \mu \Delta t + o(\Delta t), \qquad (\Delta t \to 0), \tag{1}$$

where $o(\Delta t)$, $(\Delta t \to 0)$, is a shorthand notation for a function, $g(\Delta t)$ say, for which $g(\Delta t)/\Delta t$ tends to 0 when $\Delta t \to 0$ (see e.g. [1]).

If $X_1, \ldots, X_n$ are independent exponential random variables with parameters $\mu_1, \ldots, \mu_n$ respectively, then $\min(X_1, \ldots, X_n)$ is again an exponential random variable with parameter $\mu_1 + \cdots + \mu_n$ and the probability that X_i is the smallest one is given by $\mu_i/(\mu_1 + \cdots + \mu_n)$, $i = 1, \ldots, n$.

1.4.4 Erlang distribution

A random variable X has an *Erlang-k* $(k = 1, 2, \ldots)$ distribution with mean k/μ if X is the sum of k independent random variables $X_1, \ldots, X_k$ having a common exponential distribution with mean $1/\mu$. The common notation is $E_k(\mu)$ or briefly E_k. The density of an $E_k(\mu)$ distribution is given by

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

The distribution function equals

$$F(t) = 1 - \sum_{j=0}^{k-1} \frac{(\mu t)^j}{j!} e^{-\mu t}, \qquad t \geq 0.$$
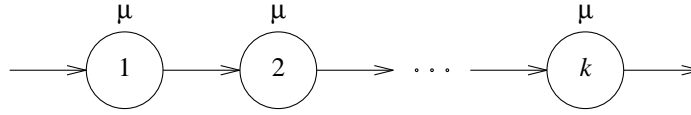
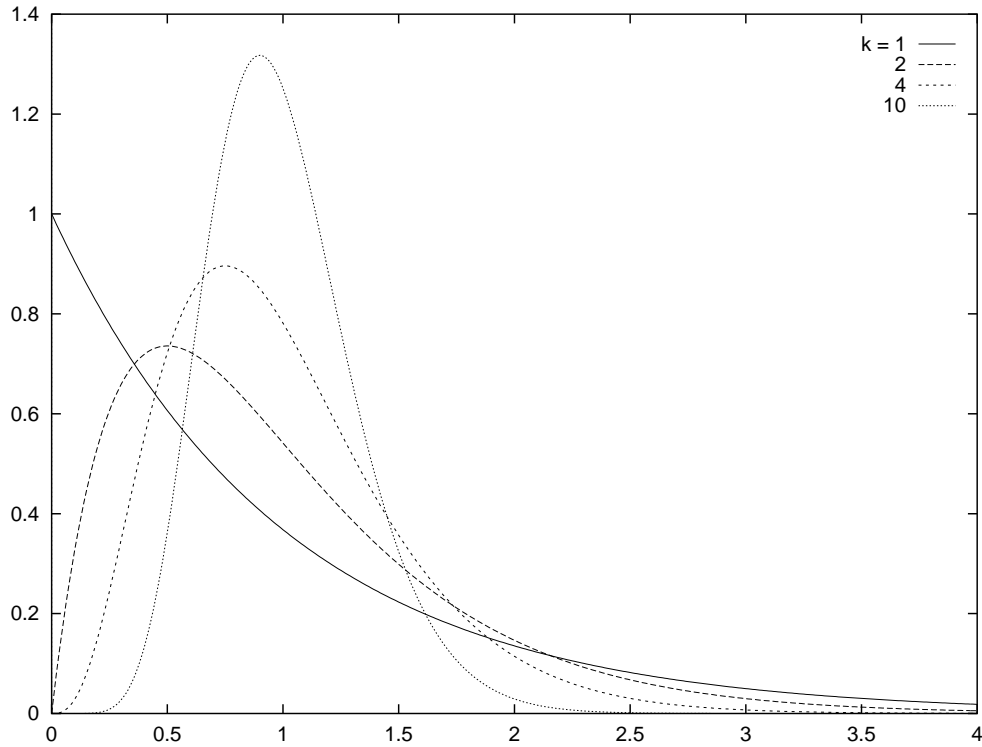Figure 1: Phase diagram for the Erlang-k distribution with scale parameter μ

Figure 2: The density of the Erlang-k distribution with mean 1 for various values of k

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 1.

In figure 2 we display the density of the Erlang-k distribution with mean 1 (so $\mu = k$) for various values of k.

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

$$E(X) = \frac{k}{\mu}, \qquad \sigma^2(X) = \frac{k}{\mu^2}, \qquad c_X^2 = \frac{1}{k}.$$

The Laplace-Stieltjes transform is given by

$$\widetilde{X}(s) = \left(\frac{\mu}{\mu + s}\right)^k.$$

A convenient distribution arises when we mix an E_{k-1} and E_k distribution with the same scale parameters. The notation used is $E_{k-1,k}$. A random variable X has an $E_{k-1,k}(\mu)$

4

distribution, if X is with probability p (resp. $1-p$) the sum of $k-1$ (resp. k) independent exponentials with common mean $1/\mu$. The density of this distribution has the form

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

where $0 \leq p \leq 1$. As p runs from 1 to 0, the squared coefficient of variation of the mixed Erlang distribution varies from $1/(k-1)$ to $1/k$. 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.

1.4.5   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 $1/\mu_i$. For this random variable we use the notation $H_k(p_1,\ldots,p_k;\mu_1,\ldots,\mu_k)$, or simply H_k. The density is given by

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

and the mean is equal to

$$E(X) = \sum_{i=1}^{k} \frac{p_i}{\mu_i}.$$

The Laplace-Stieltjes transform satisfies

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

The coefficient of variation c_X of this distribution is always greater than or equal to 1. A phase diagram of the H_k distribution is shown in figure 3.
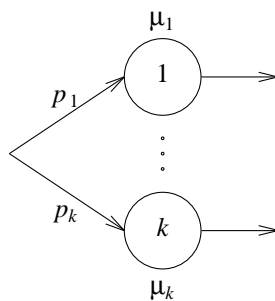

Figure 3: Phase diagram for the hyperexponential distribution

1.4.6 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 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 $1/\mu_i$, and the Markov chain is entered with probability p_i in state i, $i = 1, \ldots, k$. Then the random 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 $F(\cdot)$ a sequence of phase-type distributions can be found which pointwise converges at the points of continuity of $F(\cdot)$. The denseness of the two classes makes them very useful as a practical modelling tool. A proof of the denseness can be found in [6, 7]. The first class is the class of *Coxian distributions*, notation C_k, 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 the figures 4 and 5.
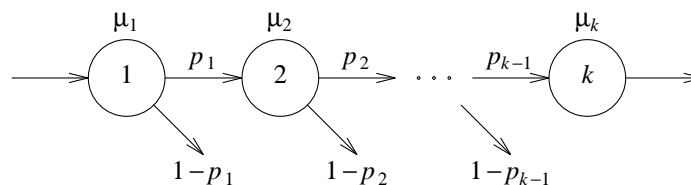

Figure 4: Phase diagram for the Coxian distribution

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 $1/\mu_n$, $n = 1, \ldots, k$. 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.

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 $1/\mu$, $n = 1, \ldots, k$.

1.5 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 fit 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.
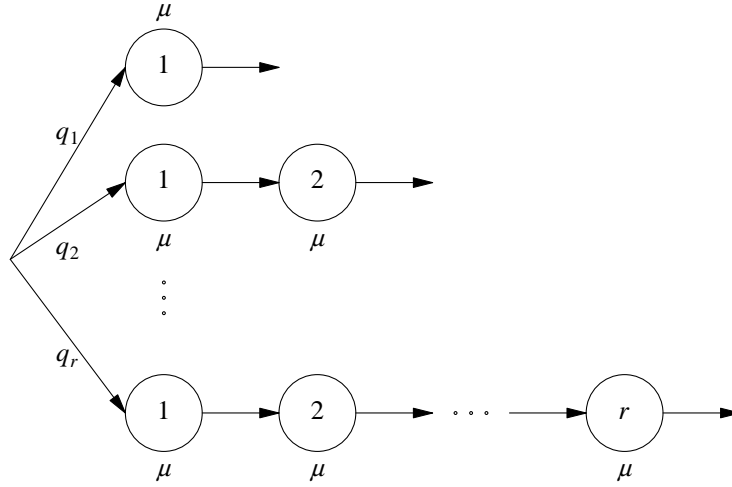
Figure 5: Phase diagram for the mixed Erlang distribution

In case $0 < c_X < 1$ one fits an $E_{k-1,k}$ distribution (see subsection 1.4.4). More specifically, if

$$\frac{1}{k} \leq c_X^2 \leq \frac{1}{k-1},$$

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 $1/\mu$. By choosing (see e.g. [8])

$$p = \frac{1}{1+c_X^2}[kc_X^2 - \{k(1+c_X^2) - k^2c_X^2\}^{1/2}], \qquad \mu = \frac{k-p}{E(X)},$$

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

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

is used. The parameters of the H_2 distribution with balanced means and fitting $E(X)$ and c_X (≥ 1) are given by

$$p_1 = \frac{1}{2}\left(1 + \sqrt{\frac{c_X^2 - 1}{c_X^2 + 1}}\right), \qquad p_2 = 1 - p_1,$$

$$\mu_1 = \frac{2p_1}{E(X)}, \qquad \mu_1 = \frac{2p_2}{E(X)}.$$

In case $c_X^2 \geq 0.5$ one can also use a Coxian-2 distribution for a two-moment fit. The following set is suggested by [4],

$$\mu_1 = 2/E(X), \quad p_1 = 0.5/c_X^2, \quad \mu_2 = \mu_1 p_1.$$

It also possible to make a more sophisticated use of phase-type distributions by, e.g., trying to match the first three (or even more) moments of X or to approximate the shape of X (see e.g. [9, 2, 3]).

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.

1.6 Poisson process

Let $N(t)$ be the number of arrivals in $[0, t]$ for a *Poisson process* with rate λ, i.e. the time between successive arrivals is exponentially distributed with parameter λ and independent of the past. Then $N(t)$ has a *Poisson distribution* with parameter λt, so

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \qquad k = 0, 1, 2, \ldots$$

The mean, variance and coefficient of variation of $N(t)$ are equal to (see subsection 1.4.2)

$$E(N(t)) = \lambda t, \qquad \sigma^2(N(t)) = \lambda t, \qquad c_{N(t)}^2 = \frac{1}{\lambda t}.$$

From (1) it is easily verified that

$$P(\text{arrival in } (t, t + \Delta t]) = \lambda \Delta t + o(\Delta t), \qquad (\Delta t \to 0).$$

Hence, for small Δt,

$$P(\text{arrival in } (t, t + \Delta t]) \approx \lambda \Delta t. \tag{2}$$

So in each small time interval of length Δt the occurence of an arrival is equally likely. In other words, Poisson arrivals occur completely random in time. In figure 6 we show a realization of a Poisson process and an arrival process with Erlang-10 interarrival times. Both processes have rate 1. The figure illustrates that Erlang arrivals are much more equally spread out over time than Poisson arrivals.

The Poisson process is an extremely useful process for modelling purposes in many practical applications, such as, e.g. to model arrival processes for queueing models or demand processes for inventory systems. It is empirically found that in many circumstances the arising stochastic processes can be well approximated by a Poisson process.

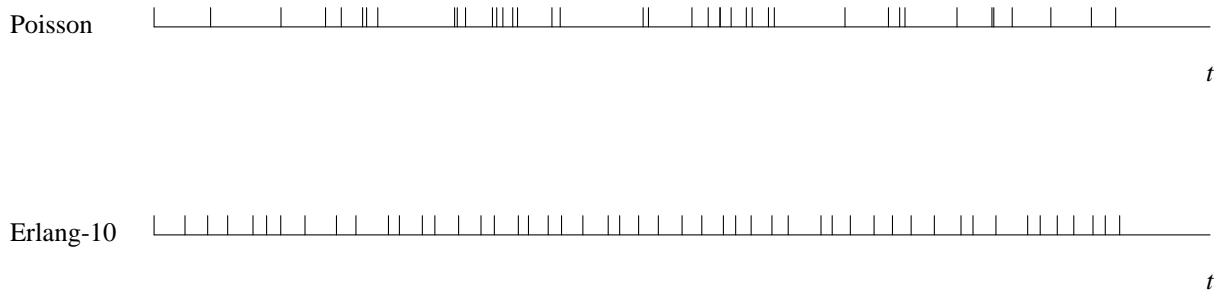Next we mention two important properties of a Poisson process (see e.g. [5]).

Figure 6: A realization of Poisson arrivals and Erlang-10 arrivals, both with rate 1

(i) *Merging.*
Suppose that $N_1(t)$ and $N_2(t)$ are two independent Poisson processes with respective rates λ_1 and λ_2. Then the sum $N_1(t) + N_2(t)$ of the two processes is again a Poisson process with rate $\lambda_1 + \lambda_2$.

(ii) *Splitting.*
Suppose that $N(t)$ is a Poisson process with rate λ and that each arrival is marked with probability p independent of all other arrivals. Let $N_1(t)$ and $N_2(t)$ denote respectively the number of marked and unmarked arrivals in $[0, t]$. Then $N_1(t)$ and $N_2(t)$ are both Poisson processes with respective rates λp and $\lambda(1 - p)$. And these two processes are independent.

So Poisson processes remain Poisson processes under merging and splitting.

References

[1] N.G. DE BRUIJN, *Asymptotic methods*, Dover, 1981.

[2] M.C. VAN DER HEIJDEN, *Performance analysis for reliability and inventory models*, Thesis, Vrije Universiteit, Amsterdam, 1993.

[3] M.A. JOHNSON, *An emperical study of queueing approximations based on phase-type approximations*, Stochastic Models, 9 (1993), pp. 531–561.

[4] R.A. MARIE, *Calculating equilibrium probabilities for $\lambda(n)/C_k/1/N$ queue*, in: Proceedings Performance' 80, Toronto, (May 28–30, 1980), pp. 117–125.

[5] S.M. ROSS, *Introduction to probability models*, 6th ed., Academic Press, London, 1997.

[6] R.S. SCHASSBERGER, *On the waiting time in the queueing system $GI/G/1$*, Ann. Math. Statist., 41 (1970), pp. 182–187.

[7] R.S. SCHASSBERGER, *Warteschlangen,* Springer-Verlag, Berlin, 1973.

[8] H.C. Tijms, *Stochastic models: an algorithmic approach*, John Wiley & Sons, Chichester, 1994.

[9] W. Whitt, *Approximating a point process by a renewal process I: two basic methods*, Opns. Res., 30 (1986), pp. 125–147.