An Introduction to Probability and Queueing Theory

Andrew Klapper

1 Probability Spaces

Probability theory is concerned with events whose outcomes are not known ahead. Rather, the likelihoods of various outcomes are known. Formally, the set of outcomes is called a *sample space* S. For example, if the event is a coin flip, then $S = \{H, T\}$. If the event is the time of arrival of the next request for service at a server, then S is the nonnegative real numbers. We will be concerned with sample spaces of two types – *discrete* (meaning finite or countable) and *continuous* (ie the real numbers, or the nonnegative real numbers or an interval of real numbers). We need a way of specifying probabilities for sets of outcomes. The probabilities should satisfy various properties (we use $P(A)$ to denote the probability that some outcome in the set A occurs):

a. $P(A) \geq 0$ for any A.

b. If A is a subset of B, then $P(A) \leq P(B)$.

c. If A and B are disjoint, then $P(A \cup B) = P(A) + P(B)$.

d. $P(S) = 1, P(\emptyset) = 0$.

Note that (b) follows from (a) and (c). It also follows that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (exercise: prove this) and, $A \subset B$ implies $P(B - A) = P(B) - P(A)$ (so $P(\overline{A}) = 1 - P(A)$, where $\overline{A}$ is the complement of A). The pair (S, P) will be called a *probability space.*

In the case of a discrete sample space, this means that we must attach a number, $P(x)$, to each $x \in S$, such that $0 \leq P(x) \leq 1$, and the sum of all the $P(x)$s is 1. For example, in flipping a coin, the coin is "fair" if $P(H) = P(T) = \frac{1}{2}$.

In the case of a continuous sample space, say the nonnegative reals, we must specify, for each interval $[0, a]$, a nonnegative real number $P([0, a])$, such that $P([0, a]) \leq P([0, b])$ if $a < b$, and $\lim_{a\to\infty} P([0, a]) = 1$. Then for $0 \leq a \leq b, P((a, b]) = P([0, b]) - P([0, a])$. Probabilities of open intervals can be expressed as limits of probabilities of closed intervals (eg

$P((a,b)) = \lim_{c \to b^-} P((a,c]),)$. This allows us to specify probabilities for all sets which are unions of open and closed intervals. It is important to note that in a continuous probability space, not all sets have probabilities (this is difficult to prove). Moreover, the probability of a single point is often zero.

The function $F(x) = P([0,x])$ is called the *probability distribution function*, and its derivative, $f(x) = F'(x)$, if it exists, is called the *probability density function*. $F(x)$ is an antiderivative (indefinite integral) of $f(x)$, and $F(0) = 0$, so $F(a) = \int_0^a f(x)dx$. For example, the *exponential distribution* is defined by $f(x) = \lambda e^{-\lambda x}$, where λ is a positive constant. Then $F(a) = \int_0^a \lambda e^{-\lambda x} dx = 1 - e^{-\lambda a}$. We will often assume that arrival times at queues satisfy an exponential distribution. λ is called the arrival rate in this case. Exponential distributions have the property that as x increases, it is less and less likely to occur.

Another important example is the *uniform distribution*. Here $S = (0,1)$, and $f(x) = 1$ if $x \in S$, and $f(x) = 0$ if $x \notin S$. Thus $F(a) = a$ if $0 \leq a \leq 1$. More generally, $P([a,b]) = b - a$, $a \leq b$, $a, b \in S$. Thus all outcomes are equally likely.

As a third example, let a be a fixed point, and define $P(A) = 1$ if $a \in A$, $P(A) = 0$ if $a \notin A$.

From now on we will use the term *event* for a subset of a probability space for which a probability is defined.

2 Composite probability spaces and conditional probabilities

We often want to discuss probabilities of combinations of events. Let S_1, S_2 be two sample spaces, with probabilities P_1, P_2. We can define a new probability space $S = S_1 \times S_2$, with $P(A \times B) = P_1(A)P_2(B)$. (S, P) is the *composite* of (S_1, P_1) and (S_2, P_2). For example, flipping two coins is the composite of two coin flips. If the first coin satisfies $P_1(H) = p$, so $P_1(T) = 1 - p$, and the second $P_2(H) = q$, so $P_2(T) = 1 - q$, then $P(\text{ two heads}) = pq$, $P(\text{ a head and a tail}) = p(1-q) + q(1-p)$, $P(\text{ at least one head}) = p(1-q) + q(1-p) + qp = 1 - (1-p)(1-q)$, etc.

Moreover, if S is a composite space, then we can speak of the *conditional probability* of an event A given that some other event B occurs, denoted $P(A \mid B)$. We have

$$P(A \mid B) = \frac{P(AB)}{P(B)}.$$

A and B are called *independent events* if $P(A \mid B) = P(A)$, ie $P(A)P(B) = P(AB)$. For example, the probability of two heads given that the first coin is a head is

$$P(2 \text{ heads} \mid \text{first coin head}) \;=\; P(\text{two heads and first coin head})/P(\text{first head})$$

$$
\begin{aligned}
&= P(\text{two heads})/P(\text{first coin head}) \\
&= pq/p \\
&= q,
\end{aligned}
$$

and

$$P(2 \text{ heads}) = pq,$$

which is not equal to q (unless $p = 1$), so these are not independent events. But

$$
\begin{aligned}
P(\text{first coin head}|2nd \text{ coin head}) &= P(2 \text{ heads})/P(2nd \text{ coin head}) \\
&= pq/q \\
&= p, \\
P(\text{ first coin head}) &= p,
\end{aligned}
$$

so these are independent events.

Two useful formulas are the following, called Bayes' formulas:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)},$$

$$P(A) = P(A \mid B)P(B) + P(A \mid \bar{B})(1 - P(B))$$

where $\bar{B}$ is the complement of B. (exercise: prove them).

3 Random Variables

We are often concerned with some function of an event. E.g., in rolling dice, we are concerned with the total rolled, not the actual pair of numbers. Such a function is called a *random variable*. If X is a random variable, and T is a set of values in the range of X, then we define $P(X \in T)$ to be the probability of the set of points a such that $X(a) \in T$. If t is a single value in the range of T, we write $P(X = t) = P(X \in \{t\})$. For example, in the case of dice, if we assume the probability of each number on a die is $1/6$, and X is the random variable "the sum of the values on two dice" then $P(X = 5) = P((1, 4)) + P((2, 3)) + P((3, 2)) + P((4, 1)) = 4/36$.

As another example, suppose we toss a coin with $P(H) = p$ until a tail appears. Let N denote the number of heads before the first tail. Then $P(N = 0) = 1 - p$, $P(N = 1) = p(1 - p)$, $P(N = 2) = p^2(1 - p)$, and in general, $P(N = k) = p^k(1 - p)$. Note that $P(N = \text{ anything}) = \sum p^n(1 - p) = (1 - p)/(1 - p) = 1$, as it should. Random variables such as this and the previous example, whose ranges are discrete sets, are called *discrete* *random variables*, and all probabilistic questions can be answered by the probabilities that they equal single values, $f(a) = P(X = a)$ (the *probability mass function*).

The *distribution function* of the real valued (or continuous) random variable X is defined as $F(a) = P(X \leq a)$, and the *density function* of X (if it exists) is the derivative of F, $f(a) = F'(a)$. These functions are defined for all real values. $F(a)$ is nondecreasing and tends to 0 and 1 at positive and negative infinity, respectively. As with probability distributions, all probabilistic questions about X can be answered by $F(a)$ (or $f(a)$).

4 Examples of random variables

1. **Bernoulli trials:** If an experiment can be classified a success or failure, define $X = 0$ if success, $X = 1$ if failure. X is called a *Bernoulli trial*. The probability mass is defined by $p(0) = 1 - p, p(1) = p$, for some $p, 0 \leq p \leq 1$. The example of coin tossing is a Bernoulli trial. In the dice example, we might consider a 7 or 11 a success, anything else a failure, so $p = P(7) + P(11) = P(\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1),(5,6),(6,5)\}) = 8/36 = 2/9$. Delivery of a bit is a typical Bernoulli trial.

2. **Binomial random variable:** Suppose n independent trials are performed, each with probability p of success. Let X be the number of successes in the n trials. X is the *binomial* random variable with parameters (n, p). We have $p(i) = \binom{n}{i} p^i (1 - p)^{n-i}, i = 0, 1, \ldots, n$. For example, if four fair coins are flipped, the probability of two heads is $p(2) = \binom{4}{2} (\frac{1}{2})^2 (\frac{1}{2})^2 = \frac{3}{8}$. If an eight bit byte is transmitted, and the probability of a single bit being damaged is .1, then the probability of at least two errors is $1 - \binom{8}{0}(.9)^8 - \binom{8}{1}(.9)^7(.1)^1 = .18689527$.

3. **Poisson random variable:** X, whose range is the nonnegative integers, is *Poisson* with parameter λ if $p(i) = e^{-\lambda}(\lambda^i)/(i!), i = 0, 1, \ldots$ A binomial random variable with parameters (n, p) can be approximated by a Poisson random variable with $\lambda = np$, if n is large (Poisson random variables are generally easier to compute with).

4. **Uniform random variable:** X is *uniform* over $(0, 1)$ if $f(x) = 1$ for $0 < x < 1, f(x) = 0$ otherwise. Thus $P(a \leq X \leq b) = \int_a^b 1 dx = b - a$, if $0 \leq a \leq b \leq 1$.

5. **Exponential random variable:** X is *exponential* with parameter $\lambda > 0$ if $f(x) = \lambda e^{-\lambda x}$ for nonnegative x, and $f(x) = 0$ for negative x. Thus $F(a) = 1 - e^{-\lambda x}$. The time between consecutive arrivals at a queue and service times at a resource are frequently assumed to be exponential random variables.

5 Expectations

We want a notion of average value of a random variable. In the discrete case, if X has mass $p(x)$, we define $E[X] = \sum xp(x)$. In the continuous case, if X has density $f(x)$, we define $E[X] = \int_{-\infty}^{\infty} xf(x)dx$. $E[X]$ is called the *expected value* of X. Examples:

1. Bernoulli trials with parameter p: $E[X] = 0 \times (1-p) + 1 \times p = p$.

2. Binomial with parameter (n, p):

$$\begin{aligned}
E[X] &= \sum_{i=0}^{n} iP(i) \\
&= \sum_{i=0}^{n} i\binom{n}{i} p^i(1-p)^{n-i} \\
&= \sum_{i=0}^{n} i\frac{n!}{i!(n-i)!} p^i(1-p)^{n-i} \\
&= \sum_{i=1}^{n} \frac{n!}{(i-1)!(n-i)!} p^i(1-p)^{n-i} \\
&= np\sum_{i=1}^{n} \frac{(n-1)!}{(i-1)!(n-i)!} p^{i-1}(1-p)^{n-i} \\
&= np\sum_{k=0}^{n-1} \binom{n-1}{k} p^k(1-p)^{n-1-k} \\
&= np(p+(1-p))^{n-1} \\
&= np
\end{aligned}$$

3. Poisson random variable with parameter λ:

$$\begin{aligned}
E[X] &= \sum_{i=0}^{\infty} i\frac{e^{-\lambda}\lambda^i}{i!} \\
&= \sum_{i=1}^{\infty} \frac{e^{-\lambda}\lambda^i}{(i-1)!} \\
&= \lambda e^{-\lambda}\sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda}\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda}e^{\lambda} \\
&= \lambda
\end{aligned}$$

4. Uniform random variable over (a,b): $f(x) = 1/(b-a)$ for $a < x < b$.

$$\begin{aligned} E[X] &= \int_a^b \frac{x}{b-a} \\ &= \frac{b^2-a^2}{2(b-a)} \\ &= \frac{b+a}{2} \end{aligned}$$

5. Exponential random variable with parameter λ:

$$\begin{aligned} E[X] &= \int_0^\infty x\lambda e^{-\lambda x} dx \\ &= -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx \\ &= 0 - \frac{e^{-\lambda x}}{\lambda}\Big|_0^\infty \\ &= \frac{1}{\lambda} \end{aligned}$$

If $g(x)$ is a function, and X is a random variable, we can define a new random variable $Y = g(X)$ and ask about its expectation. One way to do this is to compute the distribution of Y from that of X. A second way is the following, known as the law of the unconscious statistician:

Proposition 5.1 *If X is discrete with mass $p_X(x)$, then $E[g(X)] = \sum g(x)p_X(x)$. If X is continuous with density $f_X(x)$ then $E[g(X)] = \int g(x)f_X(x)dx$.*

6 Joint distributions

We are often interested in probability statements concerning two random variables X and Y at once. We define the *joint probability distribution* of X and Y to be $F(a,b) = P(X \leq a, Y \leq b)$, any real a, b. We can derive the distribution functions F_X and F_Y of X and Y from F by $F_X(a) = P(X \leq a) = P(X \leq a, Y \leq \infty) = F(a, \infty)$ and similarly $F_Y(b) = F(\infty, b)$.

If X and Y are discrete, we define the *joint probability mass* $p(x,y) = P(X = x, Y = y)$ and $p_X(x) = \sum_y p(x,y), p_Y(y) = \sum_x p(x,y)$. If X and Y are continuous, we say they are *jointly continuous* if there is a function $f(x,y)$ (the *joint density function*) such that for any sets A and $B, P(X \in A, Y \in B) = \int_A \int_B f(x,y)dydx$. The density functions $f_X(x)$ and $f_Y(y)$ can be derived from $f(x,y)$ by $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$, $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$.

If $g(x,y)$ is a function of two variables, then $E[g(X,Y)] = \sum_x \sum_y g(x,y)p(x,y)$, in the discrete case, $E[g(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dydx$, in the continuous case.For example, if $g(x,y) = x+y$, then, in the continuous case,

$$\begin{aligned} E[X+Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)dydx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dydx + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dydx \\ &= \int_{-\infty}^{\infty} xf_X(x)dx + \int_{-\infty}^{\infty} yf(x,y)dxdy \\ &= E[X] + \int_{-\infty}^{\infty} yf_Y(y)dy \\ &= E[X] + E[Y] \end{aligned}$$

that is, the expected sum of two random variables is the sum of their expected values. This result generalizes: if $X_1, \ldots, X_n$ are random variables, and $a_1, \ldots, a_n$ are real numbers, then $E[a_1X_1 + \ldots + a_nX_n] = a_1E[X_1] + \ldots + a_nE[X_n]$. Exercise: use this to compute the expectation of a binomial random variable.

X and Y are said to be *independent* if $P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b)$. It follows that $F(a,b) = F_X(a)F_Y(b), p(x,y) = p_X(x)p_Y(y)$ if X and Y are discrete, and $f(x,y) = f_X(x)f_Y(y)$ if X and Y are continuous. Exercise: if X and Y are independent continuous random variables, prove that $E[XY] = E[X]E[Y]$.

7 Higher moments

The expected value $E[X]$ of a random variable X is also referred to as the first moment or mean. $E[X^n]$, n a positive integer, is referred to as the *nth* moment, and provides additional information on average behavior. Another interesting quantity is the variance $Var(X) = E[(X - E[X])^2]$. This measures the expected square of the deviation of X from its expected value, ie, how scattered the values of X are likely to be. Note that $Var(X) = E[X^2 - 2XE[X] + E[X]^2] = E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2$.

For example, if X is Poisson with parameter λ, we know $E[X] = \lambda$, so $Var(X) = E[X^2] - \lambda^2$. Now

$$\begin{aligned} E[X^2] &= \sum_0^{\infty} i^2 \frac{e^{-\lambda}\lambda^i}{i!} \\ &= e^{-\lambda}\sum_1^{\infty} i\frac{\lambda^i}{(i-1)!} \end{aligned}$$

$$
\begin{aligned}
&= e^{-\lambda}(\lambda \sum_{1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} + \lambda^2 \sum_{2}^{\infty} \frac{\lambda^{i-2}}{(i-2)!} \\
&= e^{-\lambda}(\lambda + \lambda^2)e^{\lambda} \\
&= \lambda + \lambda^2
\end{aligned}
$$

so $Var(X) = \lambda$.

Now suppose X is exponentially distributed with parameter λ. Then $E[X] = 1/\lambda$, so $Var(X) = E[X^2] - 1/\lambda^2$.

$$
\begin{aligned}
E[X^2] &= \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx \\
&= -x^2 e^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} 2x e^{-\lambda x} dx \\
&= 0 + \frac{2}{\lambda} E[X] \\
&= \frac{2}{\lambda^2}
\end{aligned}
$$

so $Var(X) = 1/\lambda^2$.

8 Exponential distribution

A random variable X is said to be *memoryless* if for all $s, t \geq 0, P(X > s + t \mid X > t) = P(X > s)$. If X is the time until the next request for service in seconds, it says the probability that the next request doesn't come in $s + t$ seconds given that it has not come by t seconds is the same as the initial probability that it does not come in s seconds. In other words, if no request arrives by time t, then the distribution of the remaining time until an arrival is the same as the distribution of the original time until an arrival, i.e. the arrivals forget that nothing has arrived in t seconds. The condition is equivalent to $P(X > s + t) = P(X > s)P(X > t)$. It is simple to check that the exponential random variable is memoryless. It can also be proven that any memoryless random variable is exponential.

If X represents the interarrival time as above, we can define the *arrival rate function* $r(t) = f(t)/(1 - F(t))$. To interpret $r(t)$, suppose no arrival has occurred in t units, and we want the probability that an arrival occurs within dt units. This is $P(X \in (t, t + dt) \mid X > t) = P(X \in (t, t + dt), X > t)/P(X > t) = P(X \in (t, t + dt))/P(X > t)$, which is approximately $f(t)dt/(1 - F(t)) = r(t)dt$. So $r(t)$ is the probability density of an arrival after t units of waiting. If X is exponential, $f(t) = \lambda e^{-\lambda t}$, and $F(t) = 1 - e^{-\lambda t}$, so $r(t) = \lambda$. Thus the arrival rate is constant. λ is often called the arrival rate of X.

9 Stochastic processes: the Poisson process

A *stochastic process* is a collection $R = \{X(t), t \in T\}$ of random variables, for some index set T. T may be discrete or continuous and we call R *discrete* or *continuous* accordingly. T is often interpreted as time, and $X(t)$ as the state of a system at time t. The range of the $X(t)$s is then referred to as the state space of the system. For example, we can model a queue by a stochastic process $\{X(t)\}$, with index set the nonnegative reals, where $X(t)$ is the number of tasks in the queue at time t. A special case of a stochastic process is a *counting process*. This is a stochastic process $\{N(t)\}$ in which $N(t)$ is the number of occurrences of some event. Hence in a counting process, $N(t) \geq 0$, and $N(t)$ is integer valued and nondecreasing. For example, if $N(t)$ is the number of arrivals at a queue up to time t, then $N(t)$ is a counting process, as is the process $M(t) =$ the number of requests that have been granted service up to time t. Note that $X(t) = N(t) - M(t)$ is the number of requests in the waiting queue at time t. Two processes $\{X_1(t)\}$ and $\{X_2(t)\}$ are *independent* if for each $t, X_1(t)$ and $X_2(t)$ are independent random variables.

A process has *independent increments* if the numbers of events occuring in disjoint intervals are independent. This is typically assumed in the case of queueing systems. It would mean for example that the number of arrivals between the tenth and twentieth seconds is independent of the number of arrivals in the first five seconds. This assumption is really not quite valid (since most computing systems are closed - ie have a bounded number of tasks - the number of tasks awaiting service will affect the number that are likely to arrive), but makes the analysis tractable, and is often close to being valid.

If N is a counting process and $I = [a, b]$ is any interval, we can define a random variable $X_I = N(b) - N(a) =$ the number of events in the interval I. N is said to have *stationary increments* if X_I depends only on the length of I, ie if J is another interval with the same length as I, then X_J and X_I have the same distribution. This assumption is reasonable in a system whose behavior does not vary, say, with the time of day. For example, it is not reasonable to assume stationary increments for numbers of arrivals at a bank (nobody arrives at night). This is another simplifying assumption that makes analysis tractable.

A counting process $\{N(t), t \geq 0\}$ is a *Poisson process* with rate $\lambda > 0$ if

a. $N(0) = 0$,

b. N has independent increments, and

c. the number of events in any interval of length t is Poisson distributed with mean λt.

I.e., for all $s, t \geq 0$, and $n \geq 0$,

$$P(N(t+s) - N(s) = n) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}.$$

Note that this implies that a Poisson process has stationary increments. The first two conditions are usually easy to verify by the nature of our assumptions about N, but the third is difficult to verify. In order to simplify this, we have

Theorem 9.1 *N is a Poisson process if and only if*

i. $N(0) = 0$,

ii. *N has stationary, independent increments,*

iii. $P(N(h) = 1) = \lambda h + o(h)$,

iv. $P(N(h) \geq 2) = o(h)$.

(Recall that a function f is $o(h)$ if $\lim_{h\to 0}(f(h)/h) = 0$.)

Thus the probability of one event in a small interval is approximately proportional to the length of the interval, with proportion λ, and the probability of two events is negligible.

Now suppose N is a Poisson process, and let T_1 be the time of occurence of the first event, and T_n be the time between the occurences of the $(n-1)$th and nth events. The sequence $\{T_n, n = 1, 2, \ldots\}$ is called the sequence of *interarrival times*. We wish to determine the distributions of the T_n. First note that $T_1 > t$ iff no events occur in $[0, t]$, so $P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t}$. So T_1 is exponentially distributed with parameter λ. Moreover,

$$P(T_2 > t) = \int_0^\infty P(T_2 > t \mid T_1 = s)ds = E[P(T_2 > t|T_1)].$$

But

$$\begin{aligned} P(T_2 > t \mid T_1 = s) &= P(\text{ no events in } (s, s+t]|T_1 = s) \\ &= P(\text{ no events in } (s, s+t]), \text{ by independent increments,} \\ &= e^{-\lambda t}. \end{aligned}$$

Thus T_2 is exponentially distributed with parameter λ, and also is independent of T_1. Repeating this gives

Proposition 9.2 *$T_n, n = 1, 2, \ldots$ are independent identically distributed exponential random variables with parameter λ (ie, mean $= 1/\lambda$).*

This makes sense. The assumption of stationary, independent increments says the system "restarts" itself at each instant, ie, has no memory. Note that we can recover $N(t)$ from $\{T_i\} : N(t) = n$ if and only if n is the least integer such that $\sum_{i=1}^{n} T_i \leq t$. In fact, if we start with $\{T_i\}$, independent identically distributed exponential random variables, then the stochastic process defined as above will be Poisson. Thus this is an equivalent definition of a Poisson process.

We are also interested in the waiting time S_n until the nth event, which is $\sum_1^n T_i$. It can be shown that the density function f_{S_n} of S_n is $\lambda e^{-\lambda t}(\lambda t)^{n-1}/(n-1)!$. This distribution is called the *gamma distribution*, and has mean n/λ, and variance n/λ^2 (exercise: prove this).

Suppose $\{N_1(t)\}$ and $\{N_2(t)\}$ are independent Poisson processes, with rate λ_1 and λ_2 respectively. We can define $N(t) = N_1(t) + N_2(t)$. We claim that $N(t)$ is also a Poisson process, with rate $\lambda_1 + \lambda_2$. This can be seen most easily from the theorem above, since

$$\begin{aligned} P(N(h)=1) &= P(N_1(h)=1)P(N_2(h)=0) + P(N_1(h)=0)P(N_2(h)=1) \\ &= \lambda_1 h(1-\lambda_2 h) + (1-\lambda_1 h)\lambda_2 h + o(h) \\ &= (\lambda_1+\lambda_2)h + o(h), \text{ and} \end{aligned}$$

$$\begin{aligned} P(N(h)>1) &= P(N_1(h)>1) + P(N_2(h)>1) + P(N_1(h)=1)P(N_2(h)=1) \\ &= o(h) + o(h) + \lambda_1 h \lambda_2 h + o(h) \\ &= o(h). \end{aligned}$$

On the other hand, suppose that $N(t)$ is a Poisson process with rate λ, and the events seperate into two classes, such that the probability that an event is of the first class is p. Let $N_1(t)$ and $N_2(t)$ be the number of class 1 and class 2 arrivals up to time t. Then it can be shown that $N_1(t)$ and $N_2(t)$ are independent Poisson processes with rates λp and $\lambda(1-p)$, respectively.

10 References

There are many introductory books on probability and stochastic processes. Look for them in the mathematics or statistics sections of the library or bookstores. One good reference is:

Sheldon Ross, “Introduction to Probability Models”, Academic Press, New York, 1972.