Appendices

A Probability Refresher

A.1 Sample-space, events and probability measure

A *Probability space* is a triplet $(\Omega, \mathcal{F}, P)$ where

- Ω is the set of all *outcomes* associated with an experiment. Ω will be called the *sample-space*
- $\mathcal{F}$ is a set of subsets of Ω, called *events*, such that
  (i) $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$
  (ii) if $A \in \mathcal{F}$ then the complementary set A^c is in $\mathcal{F}$
  (iii) if $A_n \in \mathcal{F}$ for $n = 1, 2, \ldots$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$

  $\mathcal{F}$ is called a σ-algebra.
- P is a *probability measure* on $(\Omega, \mathcal{F})$, that is, P is a mapping from $\mathcal{F}$ into $[0, 1]$ such that
  (a) $P(\emptyset) = 0$ and $P(\Omega) = 1$
  (b) $P(\cup_{n \in I} A_n) = \sum_{n \in I} P(A_n)$ for any countable (finite or infinite) family $\{A_n,\, n \in I\}$ of *mutually exclusive* events (i.e., $A_i \cap A_j = \emptyset$ for $i \in I$, $j \in I$ such that $i \neq j$).

Axioms (ii) and (iii) imply that $\cap_{n \in I} A_n \in \mathcal{F}$ for any countable (finite or infinite) family $\{A_n,\, n \in I\}$ of events (write $\cap_{n \in I} A_n$ as $(\cup_{n \in I} A_n^c)^c$). The latter result implies, in particular, that $B - A \in \mathcal{F}$ (since $B - A = B \cap A^c$).

Axioms (a) and (b) imply that for any events A and B, $P(A) = 1 - P(A^c)$ (write Ω as $A \cup A^c$), $P(A) \leq P(B)$ if $A \subset B$ (write B as $A \cup (B - A)$), $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (write $A \cup B$ as $(A \cap B) \cup (A^c \cap B) \cup (A \cap B^c)$).

Example 9. The experiment consists in rolling a die. Then

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

$A = \{1, 3, 5\}$ is the event of rolling an odd number. Instances of σ-algebras on Ω are $\mathcal{F}_1 = \{\emptyset, \Omega\}$, $\mathcal{F}_2 = \{\emptyset, \Omega, A, A^c\}$, $\mathcal{F}_3 = \{\emptyset, \Omega, \{1, 2, 3, 5\}, \{4, 6\}\}$, $\mathcal{F}_4 = \mathcal{P}(\Omega)$ (the set of all subsets of Ω). Is $\{\emptyset, \Omega, \{1, 2, 3\}, \{3, 4, 6\}, \{5, 6\}\}$ a σ-algebra?

$\mathcal{F}_1$ and $\mathcal{F}_4$ are the smallest and the largest σ-algebras on Ω, respectively.

If the die is not biaised, the probability measure on, say, $(\Omega, \mathcal{F}_3)$, is defined by $P(\emptyset) = 0$, $P(\Omega) = 1$, $P(\{1, 2, 3, 5\}) = 4/6$ and $P(\{4, 6\}) = 2/6$. ♦

Example 10. The experiment consists in rolling two dice. Then

$$\Omega = \{(1, 1), (1, 2), \ldots, (1, 6), (2, 1), (2, 2), \ldots, (6, 6)\}.$$

$A = \{(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3)\}$ is the event of rolling a seven. ♦

Example 11. The experiment consists in tossing a fair coin until head appears. Then,

$$\Omega = \{H, TH, TTH, TTTH, \ldots\}.$$

$A = \{TTH, TTTH\}$ is the event that 3 or 4 tosses are required. ♦

Example 12. The experiment consists in measuring the time that elapses from the instant the last character of a request is typed on an inter-active terminal until the last character of the response from the computer has been received and displayed (referred to as *response time*). We assume that the response time is at least of 1 second. Then,

$$\Omega = \{\text{real } t \,:\, t \geq 1\}.$$

$A = \{10 \leq t \leq 20\}$ is the event that the response time is between 10 and 20 seconds. ♦

A.2 Combinatorial analysis

A *permutation* of order k of n elements is an ordered selection of k elements taken from the n elements.

A *combination* of order k of n elements is an unordered selection of k elements taken from the n elements.

Recall that $n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$ for any nonnegative integer n with $0! = 1$ by convention.

Proposition 20. *The number of permutations of order k of n elements is*

$$A(n,k) = \frac{n!}{(n-k)!} = n\,(n-1)\,(n-2)\cdots(n-k+1).$$

■

Proposition 21. *The number of combination of order k of n elements is*

$$C(n,k) = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

■

Example 13. Suppose that 5 terminals are connected to an on-line computer system via a single communication channel, so that only one terminal at a time may use the channel to send a message to the computer. At every instant, there maybe 0, 1, 2, 3, 4, or 5 terminals ready for transmission. One possible sample-space is

$$\Omega = \{(x_1, x_2, x_3, x_4, x_5) \,:\; \text{ each } x_i \text{ is either 0 or 1}\}.$$

$x_i = 1$ means that terminal i is ready to transmit a message, $x_i = 0$ that is it not ready. The number of points in the sample-space is 2^5 since each x_i of $(x_1, x_2, x_3, x_4, x_5)$ can be selected in two ways.

Assume that there are always 3 terminals in the ready state. Then,

$$\Omega = \{(x_1, x_2, x_3, x_4, x_5) : \text{ exactly 3 of the } x_i\text{'s are 1 and 2 are 0}\}.$$

In that case, the number n of points in the sample-space is the number of ways that 3 terminals that are ready can be chosen from the 5 terminals, that is from Proposition 21,

$$\binom{5}{3} = \frac{5!}{3!\,(5-3)!} = 10.$$

Assume that each terminal is equally likely to be in the ready condition.

If the terminals are polled sequentially (i.e., terminal 1 is polled first, then terminal 2 is polled, etc.) until a ready terminal is found, the number of polls required can be 1, 2 or 3. Let A_1, A_2, and A_3 be the events that the required number of polls is 1, 2, 3, respectively.

A_1 can only occur if $x_1 = 1$, and the other two 1's occur in the remaining four positions. The number n_1 of points favorable to A_1 is calculated as $n_1 = \binom{4}{2} = 6$ and therefore $P(A_1) = n_1/n = 6/10$.

A_2 can only occur if $x_1 = 0$, $x_2 = 0$, and the remaining two 1's occur in the remaining three positions. The number n_2 of points favorable to A_1 is calculated as $n_2 = \binom{3}{2} = 3$ and therefore $P(A_1) = 3/10$.

Similarly, $P(A_3) = 1/10$. ♦

A.3 Conditional probability

The probability that the event A occurs given the event B has occured is denoted by $P(A \,|\, B)$.

Proposition 22 (Bayes' formula)**.**

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

The conditional probability is not defined if $P(B) = 0$. It is easily checked that $P(\cdot \,|\, B)$ is a probability measure. ■

Interchanging the role of A and B in the above formula yields

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

provided that $P(A) > 0$.

Let A_i, $i = 1, 2, \ldots, n$ be n events. Assume that the events $A_1, \ldots, A_{n-1}$ are such that $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$. Then,

Proposition 23 (Generalized Bayes' formula)**.**

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)\, P(A_2 \,|\, A_1) \cdots \times P(A_n \,|\, A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

■

Proof. The proof is by induction on n. The result is true for $n = 2$. Assume that it is true for $n = 2, 3, \ldots, k$, and let us show that it is still true for $n = k + 1$.

Define $A = A_1 \cap A_2 \cap \cdots \cap A_k$. We have

$$\begin{aligned} P(A_1 \cap A_2 \cap \cdots \cap A_{k+1}) &= P(A \cap A_{k+1}) \\ &= P(A)\, P(A_{k+1} \,|\, A) \\ &= P(A_1)\, P(A_2 \,|\, A_1) \cdots P(A_k \,|\, A_1 \cap A_2 \cap \cdots \cap A_{k-1}) \\ &\quad \times P(A_{k+1} \,|\, A) \end{aligned}$$

from the induction assumption, which completes the proof. ★

Example 14. A survey of 100 computer installations in a city shows that 75 of them have at least one brand X computer. If 3 of these installations are chosen at *random*, what is the probability that each of them has at least one brand X machine?

Answer: let A_1, A_2, A_3 be the event that the first, second and third selection, respectively, has a brand X computer.

The required probability is

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1)\, P(A_2 \,|\, A_1)\, P(A_3 \,|\, A_1, A_2) \\ &= \frac{75}{>}100 \times \frac{74}{99} \times \frac{73}{98} \\ &= 0.418. \end{aligned}$$

♦

The following result will be extensively used throughout the course.

Proposition 24 (Law of total probability)**.** *Let $A_1, A_2, \ldots, A_n$ be events such that*

(a) $A_i \cap A_j = \emptyset$ *if* $i \neq j$ *(mutually exclusive events)*

(b) $P(A_i > 0)$ *for* $i = 1, 2, \ldots, n$

(c) $A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$.

Then, for any event A,

$$P(A) = \sum_{i=1}^{n} P(A \,|\, A_i)\, P(A_i).$$

■

To prove this result, let $B_i = A \cap A_i$ for $i = 1, 2, \ldots, n$. Then, $B_i \cap B_j = \emptyset$ for $i \neq j$ (since $A_i \cap A_j = \emptyset$ for $i \neq j$) and $A = B_1 \cup B_2 \cup \cdots \cup B_n$. Hence,

$$P(A) = P(B_1) + P(B_2) + \cdots + P(B_n)$$

from axiom (b) of a probability measure. But $P(B_i) = P(A \cap A_i) = P(A \,|\, A_i)\, P(A_i)$ for $i = 1, 2, \ldots, n$ from Bayes' formula, and therefore $P(A) = \sum_{i=1}^{n} P(A \,|\, A_i)\, P(A_i)$, which concludes the proof. ★

Example 15. Requests to an on-line computer system arrive on 5 communication channels. The percentage of messages received from lines 1, 2, 3, 4, 5, are 20, 30, 10, 15, and 25, respectively. The corresponding probabilities that the length of a request will exceed 100 bits are 0.4, 0.6, 0.2, 0.8, and 0.9. What is the probability that a randomly selected request will be longer than 100 bits?

Answer: let A be the event that the selected message has more than 100 bits, and let A_i be the event that it was received on line i, $i = 1, 2, 3, 4, 5$. Then, by the law of total probability,

$$\begin{aligned} P(A) &= \sum_{i=1}^{5} P(A \mid A_i)\, P(A_i) \\ &= 0.2 \times 0.4 + 0.3 \times 0.6 + 0.1 \times 0.2 + 0.15 \times 0.8 + 0.25 \times 0.9 \\ &= 0.625. \end{aligned}$$

♦

Two events A and B are said to be *independent* if

$$P(A \cap B) = P(A)\, P(B).$$

This implies the usual meaning of independence; namely, that neither influences the occurence of the other. Indeed, if A and B are independent, then

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

and

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

The concept of independent should not be confused with the concept of their being mutually exclusive (i.e., $A \cap B = \emptyset$). In fact, if A and B are mutually exclusive then

$$0 = P(\emptyset) = P(A \cap B)$$

and thus $P(A \cap B)$ cannot be equal to $P(A)\, P(B)$ only if at least one event has the probability 0. Hence, mutually exclusive events *are not independent* except in the trivial case when at least one of them has zero probability.

A.4 Random variables

In many random experiments we are interested in some number associated with the experiment rather than the actual outcome (i.e., $\omega \in \Omega$). For instance, in Example 10 one may be interested in the sum of the numbers shown on the dice. We are thus interested in a function that associates a number with an experiment. Such function is called a *random variable* (rv).

More precisely, a *real-valued* rv X is a mapping from Ω into $\mathbf{R}$ such that

$$\{\omega \in \Omega \,:\, X(\omega) \le x\} \in \mathcal{F}$$

for all $x \in \mathbf{R}$.

As usual, we shall denote $X = x$ for the event $\{\omega \in \Omega : X(\omega) = x\}$, $X \le x$ for the event $\{\omega \in \Omega : X(\omega) \le x\}$, and $y \le X \le x$ for the event $\{\omega \in \Omega : y \le X(\omega) \le x\}$.

The requirement that $X \le x$ must be an event for X to be a rv is necessary so that probability calculations can be made.

For each rv X we define its *cumulative distribution function* (c.distribution function) F (also called the *probability distribution* of X or the *law* of X) as

$$F(x) = P(X \le x)$$

for each $x \in \mathbf{R}$.

F satisfies the following properties: $\lim_{x\to+\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$, and $F(x) \le F(y)$ if $x \le y$ (i.e., F is nondecreasing).

A rv is *discrete* if it takes only discrete values. The distribution function F of a discrete rv X with values in a countable (finite or infinite) set I (e.g. $I = \mathbf{N}$) is simply given by

$$F(x) = P(X = x)$$

for each $x \in I$. We have $\sum_{x\in I} F(x) = 1$.

Example 16 (The Bernoulli distribution)**.** Let $p \in (0, 1)$. A rv variable X taking values in the set $I = \{0, 1\}$ is said to be a *Bernoulli rv* with parameter p, or to have a *Bernoulli distribution* with parameter p if $P(X = 1) = p$ and $P(X = 0) = 1 - p$. ♦

A rv is *continuous* if $P(X = x) = 0$ for all x. The *density function* of a continuous rv is a function f such that

(a) $f(x) \ge 0$ for all real x

(b) f is integrable and $P(a \le X \le b) = \int_a^b f(x)\,dx$ if $a < b$

(c) $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

(d) $F(x) = \int_{-\infty}^{x} f(t)\,dt$ for all real x.

The formula $F(x) = \dfrac{\partial f(x)}{\partial x}$ that holds at each point x where f is continuous, provides a mean of computing the density function from the distribution function, and conversely.

Example 17 (The exponential distribution)**.** Let $\alpha > 0$. A rv X is said to be an *exponential rv* with parameter α or to have an *exponential distribution* with parameter α if

$$F(x) = \begin{cases} 1 - \exp(-\alpha x) & \text{if } x > 0 \\ 0 & \text{if } x \le 0. \end{cases}$$

The density function f is given by

$$f(x) = \begin{cases} \alpha \exp(-\alpha x) & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases}$$

Suppose that $\alpha = 2$ and we wish to calculate the probability that X lies in the interval $(1, 2]$. We have

$$\begin{aligned} P(1 < X \leq 2) &= P(X \leq 2) - P(X \leq 1) \\ &= F(2) - F(1) \\ &= (1 - \exp(-4)) - (1 - \exp(-2)) \\ &= 0.117019644. \end{aligned}$$

♦

Example 18 (The exponential distribution is memoryless)**.** Let us now derive a key feature of the exponential distribution: the fact that it is *memoryless*. Let X be an exponential rv with parameter α. We have

$$\begin{aligned} P(X > x + y \,|\, X > x) &= \frac{P(X > x + y,\, X > x)}{P(X > x)} \quad \text{from Bayes' formula} \\ &= \frac{P(X > x + y)}{P(X > x)} \\ &= e^{-\alpha y} \\ &= P(X > y) \end{aligned}$$

which does not depend on x! ♦

A.5 Parameters of a random variable

Let X be a discrete rv taking values in the set I.

The *mean* or the *expectation* of X, denoted as $E[X]$, is the number

$$E[X] = \sum_{x \in I} x\, P(X = x)$$

provided that $\sum_{x \in I} |x|\, P(X = x) < \infty$.

Example 19 (Expectation of a Bernoulli rv)**.** Let X be a Bernoulli rv with parameter p. Then,

$$\begin{aligned} E[X] &= 0 \times P(X = 0) + 1 \times P(X = 1) \\ &= p. \end{aligned}$$

♦

If X is a continuous rv with density function function f, we define the expectation or the mean of X as the number

$$E[X] = \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

provided that $\int_{-\infty}^{+\infty} |x|\, f(x)\, dx$.

Example 20 (Expectation of an exponential rv)**.** Let X be an exponential rv with parameter $\alpha > 0$. Then,

$$\begin{aligned} E[X] &= \int_{0}^{+\infty} x\, \alpha\, \exp(-\alpha x)\, dx \\ &= \frac{1}{\alpha}. \end{aligned}$$

by using an integration by parts (use the formula $\int udv = uv - \int vdu$ with $u = x$ and $dv = \alpha \exp(-\alpha x)dx$, together with the formula $\lim_{x\to+\infty} x \exp(-\alpha x) = 0$). ♦

Let us give some properties of the expectation operator $E[\cdot]$.

Proposition 25. *Suppose that X and Y are rvs. such that $E[X]$ and $E[Y]$ exist, and let c a real number. Then, $E[c] = c$, $E[X + Y] = E[X] + E[Y]$, and $E[cX] = c\, E[X]$.* ■

The *k-th moment* or the *moment of order* k ($k \geq 1$) of a discrete rv X taking values in the set I is given by

$$E[X^k] = \sum_{x\in I} x^k\, P(X = x)$$

provided that $\sum_{x\in I} |x^k|\, P(X = x) < \infty$.

The k-th moment or the moment of order k ($k \geq 1$) of a continuous rv X is given by

$$E[X^k] = \int_{-\infty}^{+\infty} x^k\, f(x)\, dx$$

provided that $\int_{-\infty}^{+\infty} |x^k|\, f(x)\, dx < \infty$.

The variance of a discrete or continuous rv X is defined to be

$$\text{var}\ (X) = E\, (X - E[X])^2 = E[X^2] - (E[X])^2.$$

Example 21 (Variance of the exponential distribution)**.** Let X be an exponential rv with parameter $\alpha > 0$. Then,

$$\begin{aligned} \text{var}\ (X) &= \int_{0}^{+\infty} x^2\, \alpha\, \exp(-\alpha x)\, dx - \frac{1}{\alpha^2} \\ &= \frac{2}{\alpha^2} - \frac{1}{\alpha^2} = \frac{1}{\alpha^2}. \end{aligned}$$

Hence, the variance of an exponential rv is the square of its mean. ♦

Example 22. Consider the situation described in Example 13 when the terminals are polled sequentially until one terminal is found ready to transmit. We assume that each terminal is ready to transmit with the probability p, $0 < p \leq 1$, when it is polled.

Let X be the number of polls required before finding a terminal ready to transmit. Since $P(X = 1) = p$, $P(X = 2) = (1 - p)\, p$, and more generally since $P(X = n) = (1 - p)^{n-1}\, p$ for each n, we have

$$E[X] = \sum_{n=1}^{\infty} n\,(1 - p)^{n-1}\, p = 1/p.$$

Observe that $E[X] = 1$ if $p = 1$ and $E[X] \to \infty$ when $p \to 0$ which agrees with the intuition. ♦

A.6 Jointly distributed random variables

Sometimes it is of interest to investigate two or more rvs. If X and Y are defined on the same probability space, we define the *joint cumulative distribution function* (j.c.d.f.) of X and Y for all real x and y by

$$F(x, y) = P(X \leq x,\, Y \leq y) = P((X \leq x) \cap (Y \leq y)).$$

Define $F_X(x) = P(X \leq x)$ and $F_Y(y) = P(Y \leq y)$ for all real x and y. F_X and F_Y are called the *marginal cumulative distribution functions* of X and Y, respectively, corresponding to the joint distribution function F.

Note that $F_X(x) = \lim_{y \to +\infty} F(x, \infty)$ and $F_Y(y) = \lim_{x \to +\infty} F(\infty, y)$.

If there exists a nonnegative function f of two variables such that

$$P(X \leq x,\, Y \leq Y] = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, du\, dv$$

then f is called the *joint density function* of the rvs X and Y.

Suppose that g is a function of two variables and let f be the joint density function of X and Y. The expectation $E[g(X, Y)]$ is defined as

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)\, f(x, y)\, dx\, dy$$

provided that the integral exists.

Consider now the case when X and Y are discrete rvs taking values in some countable sets I and J, respectively. Then the *joint distribution function* of X and Y for all $x \in I$, $y \in J$, is given by

$$F(x, y) = P(X = x,\, Y = y) = P((X = x) \cap (Y = y)).$$

Define $F_X(x) = P(X = x)$ and $F_Y(y) = P(Y = y)$ for all $x \in I$ and $y \in J$ to be the marginal distribution functions of X and Y, respectively, corresponding to the joint distribution function F.

From the law of total probability, we see that $F_X(x) := \sum_{y \in J} F(x, y) = P(X = x)$ for all $x \in I$ and $F_Y(y) := \sum_{x \in I} F(x, y) = P(Y = y)$ for all $y \in J$.

Suppose that g is a nonnegative function of two variables. The expectation $E[g(X,Y)]$ is defined as

$$E[g(X,Y)] = \sum_{x\in I,\, y\in J} g(x,y)\, P(X=x,\, Y=y)$$

provided that the summation exists.

A.7 Independent random variables

Two rvs X and Y are said to be *independent* if

$$P(X \le x,\, Y \le y) = P(X \le x)\, P(Y \le y)$$

for all real x and y if X and Y are continuous rvs, and if

$$P(X = x,\, Y = y) = P(X = x)\, P(Y = y)$$

for all $x \in I$ and $y \in J$ if X and Y are discrete and take their values in I and J, respectively.

Proposition 26. *If X and Y are independent rvs such that $E[X]$ and $E[Y]$ exist, then*

$$E[X\,Y] = E[X]\,E[Y].$$

■

Proof. Let us prove this result when X and Y are discrete rvs taking values in the sets I and J, respectively. Let $g(x,y) = xy$ in the definition of $E[g(X,Y)]$ given in the previous section. Then,

$$\begin{aligned} E[X\,Y] &= \sum_{x\in I,\, y\in J} xyP(X=x,\, Y=y) \\ &= \sum_{x\in I,\, y\in J} xy\,P(X=x)\,P(Y=y) \quad \text{since } X \text{ and } Y \text{ are independent rvs} \\ &= \sum_{x\in I} x\,P(X=x) \left(\sum_{y\in I} y\,P(Y=y) \right) \\ &= E[X]\,E[Y]. \end{aligned}$$

The proof when X and Y are both continuous rvs is anologous and is therefore omitted. ★

A.8 Conditional expectation

Consider the situation in Example 13. Let X be the number of polls required to find a ready terminal and let Y be the number of ready terminals. The mean number of polls given that $Y = 1,2,3,4,5$ is the conditional expectation of X given Y (see the computation in Example 23).

Let X and Y be discrete rvs with values in the sets I and J, respectively.

Let $P_{X|Y}(x, y) := P(X = x \,|\, Y = y)$ be the conditional probability of the event $(X = x)$ given the event $(Y = y)$. From Proposition 22 we have

$$P_{X|Y}(x, y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

for each $x \in I$, $y \in J$, provided that $P(Y = y) > 0$.

$P_{X|Y}(\cdot \,|\, y)$ is called the *conditional distribution function* of X given $Y = y$.

The *conditional expectation* of X given $Y = y$, denoted as $E[X \,|\, Y = y]$, is defined for all $y \in J$ such that $P(Y = y) > 0$, by

$$E[X \,|\, Y = y] = \sum_{x \in I} x \, P_{X|Y}(x, y)$$

Example 23. Consider Example 13. Let $X \in \{1, 2, 3, 4, 5\}$ be the number of polls required to find a terminal in the ready state and let $Y \in \{1, 2, 3, 4, 5\}$ be the number of ready terminals. We want to compute $E[X \,|\, Y = 3]$, the mean number of polls required given that $Y = 3$.

We have

$$\begin{aligned} E[X \,|\, Y = 3] &= 1 \times P_{X|Y}(1, 3) + 2 \times P_{X|Y}(2, 3) + 3 \times P_{X|Y}(3, 3) \\ &= 1 \times P(A_1) + 2 \times P(A_2) + 3 \times P(A_3) \\ &= \frac{6}{10} + 2 \times \frac{3}{10} + 3 \times \frac{1}{10} = \frac{15}{10} = 1.5\,. \end{aligned}$$

The result should not be surprising since each terminal is equally likely to be in the ready state. ♦

Consider now the case when X and Y are both continuous rvs with density functions f_X and f_Y, respectively, and with joint density function f. The *conditional probability density function* of X given $Y = y$, denoted as $f_{X|Y}(x, y)$, is defined for all real y such that $f_Y(y) > 0$, by

$$f_{X|Y}(x, y) = \frac{f(x, y)}{f_Y(y)}.$$

The conditional expectation of X given $Y = y$ is defined for all real y such that $f_Y(y) > 0$, by

$$E[X \,|\, Y = y] = \int_{-\infty}^{+\infty} x \, f_{X|Y}(x, y) \, dx.$$

Below is a very useful result on conditional expectation. This is the version of the law of total probability for the expectation.

Proposition 27 (Law of conditional expectation)**.** *For any rvs X and Y,*

$$E[X] = \sum_{y \in J} E[X \,|\, Y = y] \, P(Y = y)$$

if X is a discrete rv, and

$$E[X] = \int_{-\infty}^{\infty} E[X \,|\, Y = y] \, f_Y(y) \, dy$$

if X is a continuous rv ■

We prove the result in the case when X and Y are both discrete rvs. Since $E[X \mid Y = y] = \sum_{x \in I} x\, P(X = x \mid Y = y)$ we have from the definition of the expectation that

$$
\begin{aligned}
\sum_{y \in J} E[X \mid Y = y]]\, P(Y = y) &= \sum_{y \in J} \left(\sum_{x \in I} x\, P(X = x \mid Y = y) \right) P(Y = y) \\
&= \sum_{x \in I} x \left(\sum_{y \in J} P(X = x \mid Y = y)\, P(Y = y) \right) \\
&= \sum_{x \in I} x\, P(X = x) \quad \text{by using the law of total probability} \\
&= E[X].
\end{aligned}
$$

The proof in the case when X and Y are continuous rvs is analogous and is therefore omitted. ★

B Stochastic Process

B.1 Definitions

All the rvs considered from now on are supposed to be constructed on a common probability space $(\Omega, \mathcal{F}, P)$.

Notation: We shall denote by $\mathbf{N}$ the set of all nonnegative integers and by $\mathbf{R}$ the set of all real numbers.

A collection of rvs $\mathbf{X} = (X(t),\, t \in T)$ is called a *stochastic process*. In other word, for each $t \in T$, $X(t)$ is a mapping from Ω into some set E where $E = \mathbf{R}$ or $E \subset \mathbf{R}$ (e.g., $E = [0, \infty)$, $E = \mathbf{N}$) with the interpretation that $X(t)(\omega)$ (also written as $X(t, \omega)$) is the value of the stochastic process $\mathbf{X}$ at time t on the outcome (or path) ω.

The set T is the *index set* of the stochastic process.

If T is countable (e.g., $T = \mathbf{N}$, $T = \{\ldots, -2, -1, 0, 1, 2 \ldots\}$) then $\mathbf{X}$ is called a *discrete-time* stochastic process; if T is continuous (e.g., $T = \mathbf{R}$, $T = [0, \infty)$) then $\mathbf{X}$ is called a *continuous-time* stochastic process. When T is countable one will in general substitute the notation $X(t)$ for $X(n)$ (or $X(n)$, t_n, etc.).

The space E is called the *state-space* of the stochastic process $\mathbf{X}$. If the set E is countable then $\mathbf{X}$ is called a *discrete-space* stochastic process; if the set E is continuous then $\mathbf{X}$ is called a *continuous-space* stochastic process.

When speaking of "the process $X(t)$" one should understand the process $\mathbf{X}$. This is a common abuse of language.

Example 24 (Discrete-time discrete-space stochastic process)**.** $X(n) =$ number of jobs processed during the n-th hour of the day. The stochastic process $(X(n),\, n = 1, 2, \ldots, 24)$ is a discrete-time discrete-space stochastic process. ♦

Example 25 (Discrete-time continuous-space stochastic process)**.** $X(n) =$ response time of the n-th inquiry to a central processing system of an interactive computer system. The stochastic process $(X(n),\, n = 1, 2, \ldots)$ is a discrete-time continuous-space stochastic process. ♦

Example 26 (Continuous-time discrete-space stochastic process)**.** $X(t) =$ number of messages that have arrived at a given node of a communication network in the time period $(0, t)$. The stochastic process $\{X(t),\, t \geq 0\}$ is a continuous-time discrete-space stochastic process. ♦

Example 27 (Continuous-time continuous-space stochastic process)**.** $X(t) =$ waiting time of an inquiry received at time t. The stochastic process $\{X(t),\, t \geq 0\}$ is a continuous-time continuous-space stochastic process. ♦

Introduce the following notation: a function f is $o(h)$ if

$$\lim_{h \to 0} \frac{f(h)}{h} = 0.$$

For instance, $f(h) = h^2$ is $o(h)$, $f(h) = h$ is not, $f(h) = h^r$, $r > 1$, is $o(h)$, $f(h) = \sin(h)$ is not. Any linear combinaison of $o(h)$ functions is also $o(h)$.

Example 28. Let X be an exponential rv with parameter λ. In other words, $P(X \leq x) = 1 - \exp(-\lambda x)$ for $x \geq 0$ and $P(X \leq x) = 0$ for $x < 0$. Then, $P(X \leq h) = \lambda\, h + o(h)$.

Similarly, $P(X \leq t + h \,|\, X > t) = \lambda\, h + o(h)$ since $P(X \leq t + h \,|\, X > t) = P(X \leq h)$ from the memoryless property of the exponential distribution. ♦

C Poisson Process

A *Poisson process* is one of the simplest interesting stochastic processes.

Consider a *point process* $(t_n)_n$, that is a collection of random points in time such that $0 \leq t_1 < t_2 < \cdots < t_n < t_{n+1} < \cdots$, where t_n records the occurence time of the n-th event in some experiment. For instance, t_n will be the arrival time of the n-th request to a database.

For any interval $[s,t)$ define the integer rv $N([s,t)) = \sum_{n\geq 1} \mathbf{1}(t_n \in [s,t))$. In words, $N([s,t))$ is the number of occurences (or events) of the point process $(t_n)_n$ in $[s,t)$.

We say that $(N([s,t)), 0 \leq s < t)$ is a *Poisson process* if

(1) $\{(t_{n+1} - t_n), n = 0, 1, \ldots\}$ is a collection of independent and identically distributed rvs (with $t_0 = 0$ by convention);

(2) $t_{n+1} - t_n$ is exponentially distributed with rate $\lambda > 0$, namely,

$$P(t_{n+1} - t_n < x) = 1 - \exp(-\lambda x), \quad x > 0.$$

It is a common abuse of terminology to say that $(t_n)_n$ is a Poisson process if (1)-(2) hold. We will also use this terminology.

One of the original applications of the Poisson process in communications was to model the arrivals of calls to a telephone exchange (the work of A. K. Erlang in 1919). The use of each telephone, at least in a first analysis, can be modeled as a Poisson process.

Below are important consequences of the definition of a Poisson process.

Proposition 28 (Stationary and independent increments)**.** *(i) Stationary increments: The number of occurences of a Poisson process in a given time interval only depends on the length of this interval.*

(ii) Independent increments: The number of occurences of a Poisson process in two disjoint time intervals are independent rvs.

■

Both results immediately follow from the memoryless property of the exponential distribution and from the assumption that the inter-event times $(t_{n+1} - t_n)_n$ are iid rvs.

Proposition 29 (Poisson distribution)**.** *Let $P_n(t)$ be the probability that exactly n events occur in an interval of length t. We have, for each $n \in \mathbf{N}$, $t \geq 0$,*

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}. \tag{104}$$

■

The probability distribution in (104) is called a Poisson distribution.

Proof. The proof uses the fact that the sum of n iid exponential defines an Erlang-n rv whose probability density function is known in explicit form. More specifically, if $Y_1, \ldots, Y_n$ are iid rvs with common (exponential) probability distribution $P(Y_i < x) = 1 - \exp(-\lambda x)$ then the probability density function of $Y_1 + \cdots + Y_n$ is given by

$$\frac{d}{dx} P(Y_1 + \cdots + Y_n < x) = \frac{\lambda^n x^{n-1} \exp(-\lambda x)}{(n-1)!}. \quad (105)$$

The proof goes as follows. First observe that

$$t_k = \tau_1 + \cdots + \tau_k \quad (106)$$

with $\tau_k := t_k - t_{k-1}$ (recall that $t_0 = 0$ by convention).

We know from Proposition 28 that $P_n(t) = P(N([0,t)) = n)$ for any t. For this reason we will only focus on the probability distribution of $N([0,t))$. We have

$$\begin{aligned}
&P(N([0,t)) = n) = P(t_n > t, t_{n+1} < t) \\
&= P(\tau_1 + \cdots + \tau_n > t, \tau_1 + \cdots + \tau_{n+1} < t) \quad \text{from (106)} \\
&= \int_t^\infty P(\tau_1 + \cdots + \tau_{n+1} < t \,|\, \tau_1 + \cdots + \tau_n = y) \frac{\lambda^n y^{n-1} e^{-\lambda y}}{(n-1)!} \, dy \quad \text{from (105)} \\
&= \int_t^\infty P(\tau_{n+1} < t - y \,|\, \tau_1 + \cdots + \tau_n = y) \frac{\lambda^n y^{n-1} e^{-\lambda y}}{(n-1)!} \, dy \\
&= \int_t^\infty P(\tau_{n+1} < t - y) \frac{\lambda^n y^{n-1} e^{-\lambda y}}{(n-1)!} \, dy \quad \text{by using (1) in def. of a Poisson process} \\
&= \int_t^\infty \left(1 - e^{-\lambda(t-y)}\right) \frac{\lambda^n y^{n-1} e^{-\lambda y}}{(n-1)!} \, dy \quad \text{by using (2) in def. of a Poisson process} \\
&= \frac{(\lambda t)^n}{n!} e^{-\lambda t}
\end{aligned}$$

which is obtained by performing an integration by part. ★

A function f is $o(h)$ if $f(h)/h \to 0$ as $h \to 0$. From Proposition 29 we conclude that

Corollary 1.

Prob(a single event in an interval of duration h) $= \lambda h + o(h)$

and

Prob(more than one event in an interval of duration h) $= o(h)$.

Let us now compute $E[N(t)]$, the expected number of events of a Poisson process with rate λ in an interval of length t.

Proposition 30 (Expected number of events in an interval of length t). *For each $t \geq 0$*

$$E[N(t)] = \lambda\, t.$$

■

Proposition 30 says that the expected number of events per unit of time, or equivalently, the rate at which the events occur, is given by $E[N(t)]/t = \lambda$. This is why λ is called the rate of the Poisson process.

Proof of Proposition 30. We have by using Proposition 29

$$\begin{aligned} E[N(t)] &= \sum_{k=0}^{\infty} k\, P_k(t) = \sum_{k=1}^{\infty} k\, P_k(t) \\ &= \left(\sum_{k=1}^{\infty} k\, \frac{(\lambda t)^k}{k!} \right) e^{-\lambda t} \\ &= \lambda t \left(\sum_{k=1}^{\infty} \frac{(\lambda t)^{k-1}}{(k-1)!} \right) e^{-\lambda t} \\ &= \lambda t. \end{aligned}$$

★

Proposition 31. *The superposition of two independent Poisson processes with rates λ_1 and λ_2 is a Poisson process with rate $\lambda_1 + \lambda_2$.* ■

The proof is omitted.

Example 29. Consider the failures of a link in a communication network. Failures occur according to a Poisson process with rate 2.4 per day. We have:

(i) $P(\text{time between failures } \le T \text{ days }) = 1 - e^{-2.4T}$

(ii) $P(k \text{ failures in } T \text{ days }) = \frac{(2.4T)^k}{k!}\, e^{-2.4T}$

(iii) Expected time between two consecutive failures = 10 hours

(iv) $P(0 \text{ failure in next day}) = e^{-2.4}$

(v) Suppose 10 hours have elapsed since the last failure. Then,
Expected time to next failure = 10 hours (memoryless property).

♦