

Chapter 2 - Random Variables

In this and the chapters that follow, we denote the *real line* as $\mathcal{R} = (-\infty < x < \infty)$, and the *extended real line* is denoted as $\mathcal{R}^+ \equiv \mathcal{R} \cup \{\pm\infty\}$. The extended real line is the real line with $\pm\infty$ thrown in.

Put simply (and incompletely), a *random variable* is a function that maps the sample space \mathcal{S} into the extended real line.

Example 2-1: In the die experiment, we assign to the six outcomes f_i the numbers $X(f_i) = 10i$. Thus, we have $X(f_1) = 10$, $X(f_2) = 20$, $X(f_3) = 30$, $X(f_4) = 40$, $X(f_5) = 50$, $X(f_6) = 60$.

For an arbitrary value x_0 , we must be able to answer questions like “what is the probability that random variable X is less than x_0 ?” Hence, the set $\{\rho \in \mathcal{S} : X(\rho) \leq x_0\}$ must be an event (*i.e.*, the set must belong to σ -algebra \mathcal{F}) for every x_0 (sometimes, the algebraic, non-random variable x_0 is said to be a *realization* of the random variable X). This leads to the more formal definition.

Definition: Given a probability space $(\mathcal{S}, \mathcal{F}, \mathbf{P})$, a *random variable* $X(\rho)$ is a function

$$X: \mathcal{S} \rightarrow \mathcal{R}^+. \quad (2-1)$$

That is, random variable X is a function that maps sample space \mathcal{S} into the extended real line \mathcal{R}^+ . In addition, random variable X must satisfy the two criteria discussed below.

- 1) Recall the Borel σ -algebra \mathcal{B} of subsets of \mathcal{R} that was discussed in Chapter 1 (see Example 1-9). For each $B \in \mathcal{B}$, we **must** have

$$X^{-1}(B) \equiv [\omega \in \mathcal{S} : X(\omega) \in B] \in \mathcal{F}, \quad B \in \mathcal{B}. \quad (2-2)$$

A function that satisfies this criteria is said to be *measurable*. A random variable X must be a *measurable function*.

- 2) $\mathbf{P}[\omega \in \mathcal{S} : X(\omega) = \pm\infty] = 0$. Random variable X is allowed to take on the values of $\pm\infty$;

however, it must take on the values of $\pm \infty$ with a probability of zero.

These two conditions hold for most elementary applications. Usually, they are treated as mere technicalities that impose no real limitations on real applications of random variables (usually, they are not given much thought).

However, good reasons exist for requiring that random variable X satisfy the conditions 1) and 2) listed above. In our experiment, recall that sample space \mathcal{S} describes the set of elementary outcomes. Now, it may happen that we cannot directly observe elementary outcomes $\omega \in \mathcal{S}$. Instead, we may be forced to use a measuring instrument (*i.e.*, random variable) that would provide us with measurements $X(\omega)$, $\omega \in \mathcal{S}$. Now, for *each* $\alpha \in \mathcal{R}$, we need to be able to compute the probability $\mathbf{P}[-\infty < X(\omega) \leq \alpha]$, because $[-\infty < X(\omega) \leq \alpha]$ is a meaningful, observable event in the context of our experiment/measurements. For probability $\mathbf{P}[-\infty < X(\omega) \leq \alpha]$ to exist, we must have $[-\infty < X(\omega) \leq \alpha]$ as an event; that is, we must have

$$[\omega \in \mathcal{S} : -\infty < X(\omega) \leq \alpha] \in \mathcal{F} \quad (2-3)$$

for *each* $\alpha \in \mathcal{R}$. It is possible to show that Conditions (2-2) and (2-3) are equivalent. So, while (2-2) (or the equivalent (2-3)) may be a *mere* technicality, it is an *important* technicality.

Sigma-Algebra Generated by Random Variable X

Suppose that we are given a probability space $(\mathcal{S}, \mathcal{F}, \mathbf{P})$ and a random variable X as described above. Random variable X *induces* a σ -algebra $\sigma(X)$ on \mathcal{S} . $\sigma(X)$ consists of all sets of the form $\{\omega \in \mathcal{S} : X(\omega) \in B, B \in \mathcal{B}\}$, where \mathcal{B} denotes the σ -algebra of Borel subsets of \mathcal{R} that was discussed in Chapter 1 (see Example 1-9). Note that $\sigma(X) \subset \mathcal{F}$; we say that $\sigma(X)$ is the *sub* σ -algebra of \mathcal{F} that is *generated* by random variable X .

Probability Space Induced by Random Variable X

Suppose that we are given a probability space $(\mathcal{S}, \mathcal{F}, \mathbf{P})$ and a random variable X as described above. Let \mathcal{B} be the Borel σ -algebra introduced by Example 1-9. By (2-2), for each $B \in \mathcal{B}$, we have $\{\omega \in \mathcal{S} : X(\omega) \in B\} \in \mathcal{F}$, so $\mathbf{P}[\{\omega \in \mathcal{S} : X(\omega) \in B\}]$ is well defined.

This allows us to use random variable X to define $(\mathcal{R}^+, \mathcal{B}, \mathbf{P}')$, a *probability space induced by X* . Probability measure \mathbf{P}' is defined as follows: for each $B \in \mathcal{B}$, we define $\mathbf{P}'(B) \equiv \mathbf{P}[\{\omega \in \mathcal{S} : X(\omega) \in B\}]$. We say that \mathbf{P} *induces* probability measure \mathbf{P}' .

Distribution and Density Functions

The *distribution function* of the random variable $X(\rho)$ is the function

$$F(x) = \mathbf{P}[X(\rho) \leq x] = \mathbf{P}[\rho \in \mathcal{S} : X(\rho) \leq x], \quad (2-4)$$

where $-\infty < x < \infty$.

Example 2-2: Consider the coin tossing experiment with $\mathbf{P}[\text{heads}] = p$ and $\mathbf{P}[\text{tails}] = q \equiv 1 - p$. Define the random variable

$$X(\text{head}) = 1$$

$$X(\text{tail}) = 0.$$

If $x \geq 1$, then both $X(\text{head}) = 1 \leq x$ and $X(\text{tail}) = 0 \leq x$ so that

$$F(x) = 1 \text{ for } x \geq 1.$$

If $0 \leq x < 1$, then $X(\text{head}) = 1 > x$ and $X(\text{tail}) = 0 \leq x$ so that

$$F(x) = \mathbf{P}[X \leq x] = q \text{ for } 0 \leq x < 1$$

Finally, if $x < 0$, then both $X(\text{head}) = 1 > x$ and $X(\text{tail}) = 0 > x$ so that

$$F(x) = \mathbf{P}[X \leq x] = 0 \text{ for } x < 0.$$

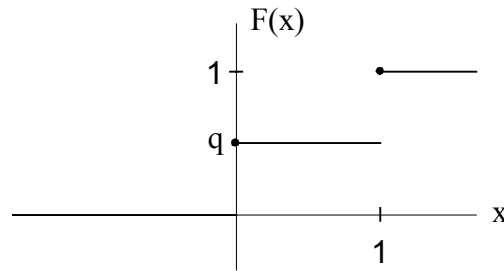


Figure 2-1: Distribution function for $X(\text{head}) = 1$, $X(\text{tail}) = 0$ random variable.

See Figure 2-1 for a graph of $F(x)$.

Properties of Distribution Functions

First, some standard notation:

$$F(x^+) \equiv \lim_{\rho \rightarrow x^+} F(\rho) \text{ and } F(x^-) \equiv \lim_{\rho \rightarrow x^-} F(\rho).$$

Some properties of distribution functions are listed below.

Claim #1: $F(+\infty) = 1$ and $F(-\infty) = 0$. (2-5)

Proof: $F(+\infty) = \lim_{x \rightarrow \infty} \mathbf{P}[X \leq x] = \mathbf{P}[\mathcal{S}] = 1$ and $F(-\infty) = \lim_{x \rightarrow -\infty} \mathbf{P}[X \leq x] = \mathbf{P}[\{\emptyset\}] = 0$.

Claim #2: The distribution function is a non-decreasing function of x . If $x_1 < x_2$, then $F(x_1) \leq F(x_2)$.

Proof: $x_1 < x_2$ implies that $\{X(\rho) \leq x_1\} \subset \{X(\rho) \leq x_2\}$. But this means that $\mathbf{P}[\{X(\rho) \leq x_1\}] \leq \mathbf{P}[\{X(\rho) \leq x_2\}]$ and $F(x_1) \leq F(x_2)$.

Claim #3: $\mathbf{P}[X > x] = 1 - F(x)$. (2-6)

Proof: $\{X(\rho) \leq x_1\}$ and $\{X(\rho) > x_1\}$ are mutually exclusive. Also, $\{X(\rho) \leq x_1\} \cup \{X(\rho) > x_1\} =$

S. Hence, $\mathbf{P}\{X \leq x_1\} + \mathbf{P}\{X > x_1\} = \mathbf{P}\{S\} = 1$.

Claim #4: Function $F(x)$ may have jump discontinuities. It can be shown that a jump is the only type of discontinuity that is possible for distribution $F(x)$ (and, $F(x)$ may have a countable number of jumps, at most). $F(x)$ must be right continuous; that is, we must have $F(x^+) = F(x)$. At a “jump”, take F to be the larger value; see Figure 2-2.

Claim #5: $\mathbf{P}[x_1 < X \leq x_2] = F(x_2) - F(x_1)$ (2-7)

Proof: $\{X(\rho) \leq x_1\}$ and $\{x_1 < X(\rho) \leq x_2\}$ are mutually exclusive. Also, $\{X(\rho) \leq x_2\} = \{X(\rho) \leq x_1\} \cup \{x_1 < X(\rho) \leq x_2\}$. Hence, $\mathbf{P}[X(\rho) \leq x_2] = \mathbf{P}[X(\rho) \leq x_1] + \mathbf{P}[x_1 < X(\rho) \leq x_2]$ and $\mathbf{P}[x_1 < X(\rho) \leq x_2] = F(x_2) - F(x_1)$.

Claim #6: $\mathbf{P}[X = x] = F(x) - F(x^-)$. (2-8)

Proof: $\mathbf{P}[x - \varepsilon < X \leq x] = F(x) - F(x - \varepsilon)$. Now, take limit as $\varepsilon \rightarrow 0^+$ to obtain the desired result.

Claim #7: $\mathbf{P}[x_1 \leq X \leq x_2] = F(x_2) - F(x_1^-)$. (2-9)

Proof: $\{x_1 \leq X \leq x_2\} = \{x_1 < X \leq x_2\} \cup \{X = x_1\}$ so that

$$\mathbf{P}[x_1 \leq X \leq x_2] = (F(x_2) - F(x_1)) + (F(x_1) - F(x_1^-)) = F(x_2) - F(x_1^-).$$

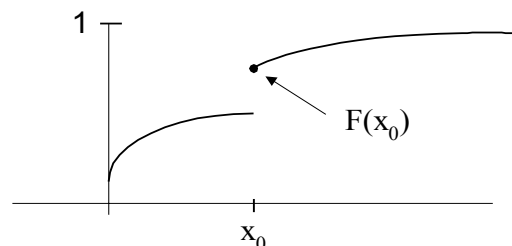


Figure 2-2: Distributions are right continuous.

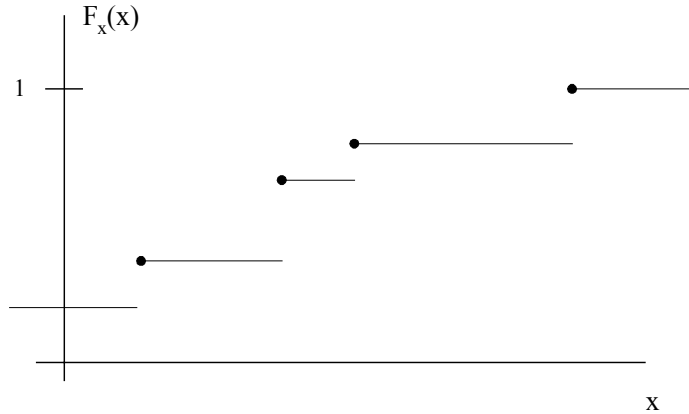


Figure 2-3: Distribution function for a discrete random variable.

Continuous Random Variables

Random variable X is of *continuous type* if $F_x(x)$ is continuous. In this case, $\mathbf{P}[X = x] = 0$; the probability is zero that X takes on a given value x .

Discrete Random Variables

Random variable X is of *discrete type* if $F_x(x)$ is piece-wise constant. The distribution should look like a staircase. Denote by x_i the points where $F_x(x)$ is discontinuous. Then $F_x(x_i) - F_x(x_i^-) = \mathbf{P}[X = x_i] = p_i$. See Figure 2-3.

Mixed Random Variables

Random variable X is said to be of *mixed type* if $F_x(x)$ is discontinuous but not a staircase.

Density Function

The derivative

$$f_x(x) \equiv \frac{dF_x(x)}{dx} \quad (2-10)$$

is called the *density function* of the random variable X . Suppose F_x has a jump discontinuity at a point x_0 . Then $f(x)$ contains the term

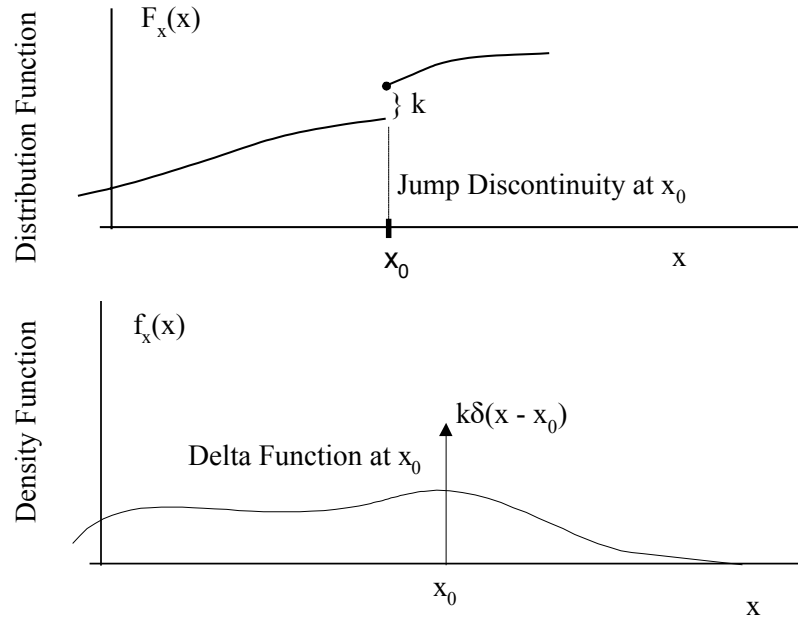


Figure 2-4: "Jumps" in distribution function causes delta function in density function. The distribution jumps by the value k at $x = x_0$.

$$\left[F_x(x_0^+) - F_x(x_0^-) \right] \delta(x - x_0) = \left[F_x(x_0) - F_x(x_0^-) \right] \delta(x - x_0). \quad (2-11)$$

See Figure 2-4.

Suppose that X is of a discrete type taking values x_i , $i \in \mathcal{J}$. The density can be written as

$$f_x(x) = \sum_{i \in \mathcal{J}} \mathbf{P}[X = x_i] \delta(x - x_i),$$

where \mathcal{J} is an index set. Figures 2-5 and 2-6 illustrate the distribution and density, respectively, of a discrete random variable.

Properties of f_x

The monotonicity of F_x implies that $f_x(x) \geq 0$ for all x . Furthermore, we have

$$f_x(x) = \frac{dF_x(x)}{dx} \Leftrightarrow F_x(x) = \int_{-\infty}^x f_x(\rho) d\rho \quad (2-12)$$

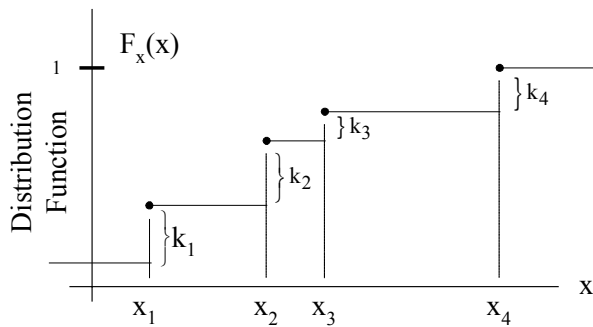


Figure 2-5: Distribution function for a discrete random variable.

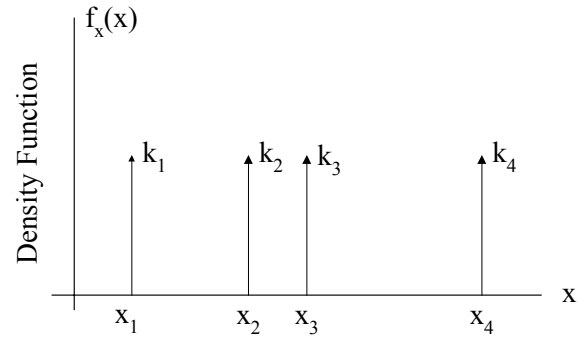


Figure 2-6: Density function for a discrete random variable.

$$\mathbf{P}[x_1 < X \leq x_2] = F_x(x_2) - F_x(x_1) = \int_{x_1}^{x_2} f_x(\rho) d\rho. \quad (2-13)$$

If X is of continuous type, then $F_x(x) = F_x(x^-)$, and

$$\mathbf{P}[x_1 \leq X \leq x_2] = F_x(x_2) - F_x(x_1) = \int_{x_1}^{x_2} f_x(\rho) d\rho.$$

For continuous random variables $\mathbf{P}[x < X \leq x + \Delta x] \approx f_x(x)\Delta x$ for small Δx .

Normal/Gaussian Random Variables

Let η , $-\infty < \eta < \infty$, and σ , $\sigma > 0$, be constants. Then

$$f_x(x) = \frac{1}{\sigma} g\left(\frac{x-\eta}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\eta}{\sigma}\right)^2\right] \quad (2-14)$$

is a Gaussian density function with parameters η and σ . These parameters have special

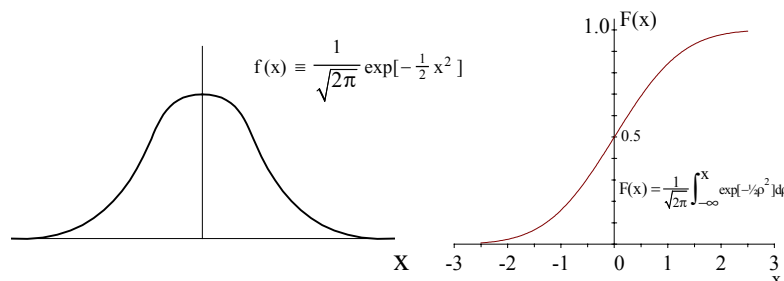


Figure 2-7: Density and distribution functions for a Gaussian random variable with $\eta = 0$ and $\sigma = 1$.

meanings that will be discussed below. The notation $N(\eta;\sigma)$ is used to indicate a Gaussian random variable with parameters η and σ . Figure 2-7 illustrates Gaussian density and distribution functions.

Random variable X is said to be *Gaussian* if its distribution function is given by

$$F(x) = \mathbf{P}[X \leq x] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp\left[-\frac{(x-\eta)^2}{2\sigma^2}\right] dx \quad (2-15)$$

for given η , $-\infty < \eta < \infty$, and σ , $\sigma > 0$. Numerical values for $F(x)$ can be determined with the aid of a table. To accomplish this, make the change of variable $y = (x - \eta)/\sigma$ in (2-15) and obtain

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\eta)/\sigma} \exp\left[-\frac{y^2}{2}\right] dy = G\left(\frac{x-\eta}{\sigma}\right). \quad (2-16)$$

Function $G(x)$ is tabulated in many reference books, and it is built in to many popular computer math packages (*i.e.*, Matlab, Mathcad, etc.).

Uniform

Random variable X is *uniform* between x_1 and x_2 if its density is constant on the interval $[x_1, x_2]$ and zero elsewhere. Figure 2-8 illustrates the distribution and density of a uniform random variable.

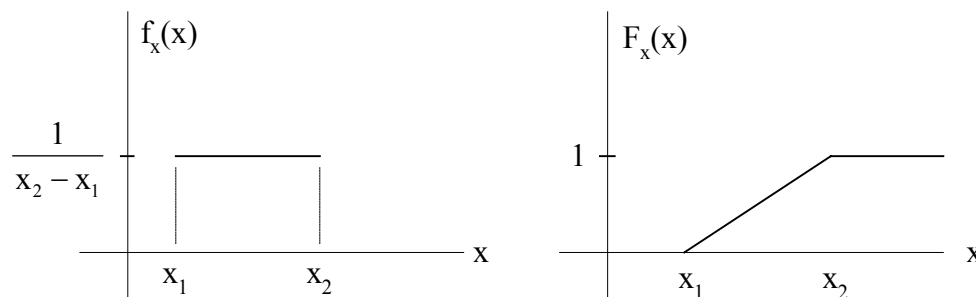


Figure 2-8: Density and distribution functions for a uniform random variable.

Binomial

Random variable X has a *binomial distribution of order n with parameter p* if it takes the integer values $0, 1, \dots, n$ with probabilities

$$\mathbf{P}[X = k] = \binom{n}{k} p^k q^{n-k}, \quad 0 \leq k \leq n. \quad (2-17)$$

Both n and p are known parameters where $p + q = 1$, and

$$\binom{n}{k} \equiv \frac{n!}{k!(n-k)!}. \quad (2-18)$$

We say that binomial random variable X is $B(n,p)$.

The Binomial density function is

$$f_x(x) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \delta(x - k), \quad (2-19)$$

and the Binomial distribution is

$$F_x(x) = \sum_{k=0}^{m_x} \binom{n}{k} p^k q^{n-k}, \quad m_x \leq x < m_x + 1 \leq n \quad (2-20)$$

$$= 1, \quad x \geq n.$$

Note that m_x depends on x .

Poisson

Random variable X is *Poisson* with parameter $a > 0$ if it takes on integer values $0, 1, \dots$

with

$$\mathbf{P}[X = k] = e^{-a} \frac{a^k}{k!}, \quad k = 0, 1, 2, 3, \dots \quad (2-21)$$

The density and distribution of a Poisson random variable are given by

$$f_x(x) = e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!} \delta(x - k) \quad (2-22)$$

$$F_x(x) = e^{-a} \sum_{k=0}^{m_x} \frac{a^k}{k!} \quad m_x \leq x < m_x + 1, \quad m_x = 0, 1, \dots \quad (2-23)$$

Rayleigh

The random variable X is *Rayleigh* distributed with real-valued parameter α , $\alpha > 0$, if it is described by the density

$$\begin{aligned} f_x(x) &= \frac{x}{\alpha^2} \exp\left[-\frac{1}{2}\left(\frac{x}{\alpha}\right)^2\right], \quad x \geq 0 \\ &= 0, \quad x < 0. \end{aligned} \quad (2-24)$$

See Figure 2-9 for a depiction of a Rayleigh density function.

The distribution function for a Rayleigh random variable is

$$\begin{aligned} F_x(x) &= \int_0^x \frac{u}{\alpha^2} e^{-u^2/2\alpha^2} du, \quad x \geq 0 \\ &= 0, \quad x < 0 \end{aligned} \quad (2-25)$$

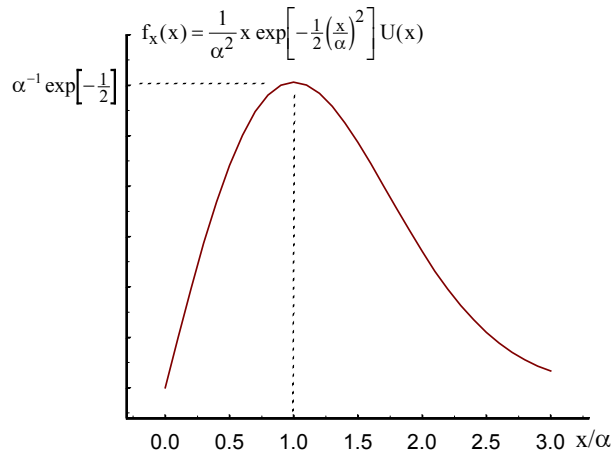


Figure 2-9: Rayleigh density function.

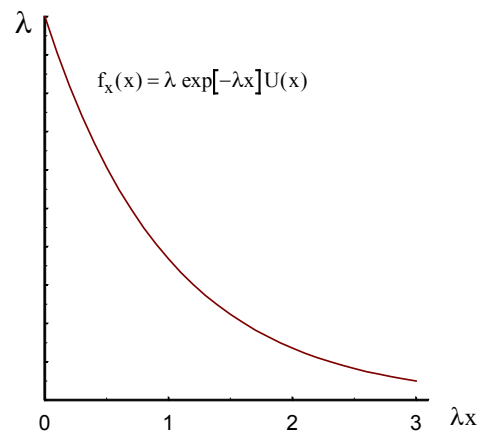


Figure 2-10: Exponential density function.

To evaluate (2-25), use the change of variable $y = u^2/2\alpha^2$, $dy = (u/\alpha^2)du$ to obtain

$$\begin{aligned} F_x(x) &= \int_0^{x^2/2\alpha^2} e^{-y} dy = 1 - e^{-x^2/2\alpha^2}, & x \geq 0 \\ &= 0, & x < 0 \end{aligned} \quad (2-26)$$

as the distribution function for a Rayleigh random variable.

Exponential

The random variable X is *exponentially* distributed with real-valued parameter λ , $\lambda > 0$, if it is described by the density

$$\begin{aligned} f_x(x) &= \lambda \exp[-\lambda x], & x \geq 0 \\ &= 0, & x < 0. \end{aligned} \quad (2-27)$$

See Figure 2-10 for a depiction of an exponential density function.

The distribution function for an exponential random variable is

$$\begin{aligned}
 F_x(x) &= \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x}, & x \geq 0 \\
 &= 0, & x < 0.
 \end{aligned}
 \tag{2-28}$$

Conditional Distribution

Let M denote an event for which $\mathbf{P}[M] \neq 0$. Assuming the occurrence of event M , the *conditional distribution* $F(x | M)$ of random variable X is defined as

$$F(x | M) = \mathbf{P}[X \leq x | M] \equiv \frac{\mathbf{P}[X \leq x, M]}{\mathbf{P}[M]} . \tag{2-29}$$

Note that $F(\infty | M) = 1$ and $F(-\infty | M) = 0$. Furthermore, the conditional distribution has all of the properties of an "ordinary" distribution function. For example,

$$\mathbf{P}[x_1 < X \leq x_2 | M] = F(x_2 | M) - F(x_1 | M) = \frac{\mathbf{P}[x_1 < X \leq x_2, M]}{\mathbf{P}[M]} . \tag{2-30}$$

Conditional Density

The *conditional density* $f(x | M)$ is defined as

$$f(x | M) = \frac{d}{dx} F(x | M) = \lim_{\Delta x \rightarrow 0} \frac{\mathbf{P}[x < X \leq x + \Delta x | M]}{\Delta x} . \tag{2-31}$$

The conditional density has all of the properties of an "ordinary" density function.

Example 2-3: Determine the conditional distribution $F(x | M)$ of random variable $X(f_i) = 10i$ of the fair-die experiment where $M = \{f_2, f_4, f_6\}$ is the event "even" has occurred. First, note that X must take on values in the set $\{10, 20, 30, 40, 50, 60\}$. Hence, if $x \geq 60$, then $[X \leq x]$ is the certain event and $[X \leq x, M] = M$. Because of this,

$$F(x|M) = \frac{P[X \leq x, M]}{P[M]} = \frac{P[M]}{P[M]} = 1, \quad x \geq 60.$$

If $40 \leq x < 60$, then $[X \leq x, M] = [f_2, f_4]$, and

$$F(x|M) = \frac{P[X \leq x, M]}{P[M]} = \frac{P[f_2, f_4]}{P[f_2, f_4, f_6]} = \frac{2/6}{3/6}, \quad 40 \leq x < 60.$$

If $20 \leq x < 40$, then $[X \leq x, M] = [f_2]$, and

$$F(x|M) = \frac{P[X \leq x, M]}{P[M]} = \frac{P[f_2]}{P[f_2, f_4, f_6]} = \frac{1/6}{3/6}, \quad 20 \leq x < 40.$$

Finally, if $x < 20$, then $[X \leq x, M] = [\emptyset]$, and

$$F(x|M) = \frac{P[X \leq x, M]}{P[M]} = \frac{P[\{\emptyset\}]}{P[f_2, f_4, f_6]} = \frac{0}{3/6}, \quad x < 20.$$

Conditional Distribution When Event M is Defined in Terms of X

If M is an event that can be expressed in terms of the random variable X, then $F(x|M)$ can be determined from the "ordinary" distribution $F(x)$. Below, we give several examples.

As a first example, consider $M = [X \leq a]$, and find both $F(x|M) = F(x|X \leq a) = P[X \leq x | X \leq a]$ and $f(x|M)$. Note that

$$F(x|X \leq a) = P[X \leq x | X \leq a] = \frac{P[X \leq x, X \leq a]}{P[X \leq a]}.$$

Hence, if $x \geq a$, we have $[X \leq x, X \leq a] = [X \leq a]$ and

$$F(x|X \leq a) = \frac{\mathbf{P}[X \leq a]}{\mathbf{P}[X \leq a]} = 1, \quad x \geq a.$$

If $x < a$, then $[X \leq x, X \leq a] = [X \leq x]$ and

$$F(x|X \leq a) = \frac{\mathbf{P}[X \leq x]}{\mathbf{P}[X \leq a]} = \frac{F_X(x)}{F_X(a)}, \quad x < a.$$

At $x = a$, $F(x|X \leq a)$ would jump $1 - F_X(a^-)/F_X(a)$ if $F_X(a^-) \neq F_X(a)$. The conditional density is

$$f(x|X \leq a) = \frac{d}{dx} F(x|X \leq a) = \frac{d}{dx} \left[\frac{F_X(x)}{F_X(a)} \right] = \begin{cases} \frac{f_X(x)}{F_X(a)}, & x < a \\ 0, & x \geq a \end{cases} + \left[1 - \frac{F_X(a^-)}{F_X(a)} \right] \delta(x - a).$$

As a second example, consider $M = [b < X \leq a]$ so that

$$F(x|b < X \leq a) = \frac{\mathbf{P}[X \leq x, b < X \leq a]}{\mathbf{P}[b < X \leq a]}.$$

Since

$$\begin{aligned} [X \leq x, b < X \leq a] &= [b < X \leq a], & a \leq x \\ &= [b < X \leq x], & b < x < a \\ &= \{\emptyset\}, & x \leq b, \end{aligned}$$

we have

$$\begin{aligned}
 F(x|b < X \leq a) &= 1, & a \leq x \\
 &= \frac{F_X(x) - F_X(b)}{F_X(a) - F_X(b)}, & b < x < a \\
 &= 0, & x \leq b.
 \end{aligned}$$

$F(x|b < X \leq a)$ is continuous at $x = b$. At $x = a$, $F(x|b < X \leq a)$ jumps $1 - \{F_X(a^-) - F_X(b)\} / \{F_X(a) - F_X(b)\}$, a value that is zero if $F_X(a^-) = F_X(a)$. The corresponding conditional density is

$$\begin{aligned}
 f(x|b < X \leq a) &= \frac{d}{dx} F(x|b < X \leq a) \\
 &= \left\{ \begin{array}{ll} \frac{f_X(x)}{F_X(a) - F_X(b)}, & b < x < a \\ 0, & \text{otherwise} \end{array} \right\} + \left[1 - \frac{F_X(a^-) - F_X(b)}{F_X(a) - F_X(b)} \right] \delta(x - a).
 \end{aligned}$$

Example 2-4: Find $f(x | |X - \eta| \leq \kappa\sigma)$, where X is $N(\eta, \sigma)$. First, note that

$$|X - \eta| \leq \kappa\sigma \Rightarrow \eta - \kappa\sigma \leq X \leq \eta + \kappa\sigma,$$

so that

$$\mathbf{P}[|X - \eta| \leq \kappa\sigma] = \mathbf{P}[\eta - \kappa\sigma \leq X \leq \eta + \kappa\sigma] = \frac{1}{\sqrt{2\pi}\sigma} \int_{\eta - \kappa\sigma}^{\eta + \kappa\sigma} \exp\left[-\frac{1}{2}\left(\frac{x - \eta}{\sigma}\right)^2\right] dx.$$

By a change of variable, this last result becomes

$$\mathbf{P}[|X - \eta| \leq \kappa\sigma] = \frac{1}{\sqrt{2\pi}} \int_{-\kappa}^{\kappa} \exp\left[-\frac{1}{2}u^2\right] du = G(\kappa) - G(-\kappa) = 2G(\kappa) - 1.$$

Hence, by the previous example, we have

$$f(x | |X - \eta| \leq \kappa\sigma) = \frac{\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\eta}{\sigma}\right)^2\right]}{2G(\kappa)-1}, \quad \eta - \kappa\sigma \leq x \leq \eta + \kappa\sigma$$

$$= 0 \quad \text{otherwise.}$$

Total Probability – Continuous Form

Define $B = [X \leq x]$, and let A_1, A_2, \dots, A_n be a *partition* of sample space \mathcal{S} . That is,

$$\bigcup_{i=1}^n A_i = \mathcal{S} \text{ and } A_i \cap A_j = \{\emptyset\} \text{ for } i \neq j. \quad (2-32)$$

From the discrete form of the Theorem of Total Probability discussed in Chapter 1, we have $\mathbf{P}[B] = \mathbf{P}[B | A_1]\mathbf{P}[A_1] + \mathbf{P}[B | A_2]\mathbf{P}[A_2] + \dots + \mathbf{P}[B | A_n]\mathbf{P}[A_n]$. Now, with $B = [X \leq x]$, this becomes

$$\mathbf{P}[X \leq x] = \mathbf{P}[X \leq x | A_1]\mathbf{P}[A_1] + \mathbf{P}[X \leq x | A_2]\mathbf{P}[A_2] + \dots + \mathbf{P}[X \leq x | A_n]\mathbf{P}[A_n]. \quad (2-33)$$

Hence,

$$F_x(x) = F(x|A_1)\mathbf{P}[A_1] + F(x|A_2)\mathbf{P}[A_2] + \dots + F(x|A_n)\mathbf{P}[A_n] \quad (2-34)$$

$$f_x(x) = f(x|A_1)\mathbf{P}[A_1] + f(x|A_2)\mathbf{P}[A_2] + \dots + f(x|A_n)\mathbf{P}[A_n].$$

Several useful formulas can be derived from this result. For example, let A be any event, and let X be any random variable. Then we can write a version of Bayes rule as

$$\begin{aligned}
 \mathbf{P}[A|X \leq x] &= \frac{\mathbf{P}[A, X \leq x]}{\mathbf{P}[X \leq x]} = \frac{\mathbf{P}[X \leq x|A]\mathbf{P}[A]}{\mathbf{P}[X \leq x]} = \frac{F(x|A)\mathbf{P}[A]}{F(x)} \\
 &= \frac{F(x|A)\mathbf{P}[A]}{F(x|A_1)\mathbf{P}[A_1] + F(x|A_2)\mathbf{P}[A_2] + \dots + F(x|A_n)\mathbf{P}[A_n]} .
 \end{aligned}
 \tag{2-35}$$

As a second example, we derive a formula for $\mathbf{P}[A | x_1 < X \leq x_2]$. Now, the conditional distribution $F(x | A)$ has the same properties as an "ordinary" distribution. That is, we can write $\mathbf{P}[x_1 < X \leq x_2 | A] = F(x_2 | A) - F(x_1 | A)$ so that

$$\begin{aligned}
 \mathbf{P}[A | x_1 < X \leq x_2] &= \frac{\mathbf{P}[x_1 < X \leq x_2 | A]}{\mathbf{P}[x_1 < X \leq x_2]} \mathbf{P}[A] \\
 &= \frac{F(x_2 | A) - F(x_1 | A)}{F_x(x_2) - F_x(x_1)} \mathbf{P}[A].
 \end{aligned}
 \tag{2-36}$$

In general, we **cannot** write

$$\mathbf{P}[A | X = x] = \frac{\mathbf{P}[A, X = x]}{\mathbf{P}[X = x]}
 \tag{2-37}$$

since this may result in an indeterminate 0/0 form. Instead, we must write

$$\begin{aligned}
 \mathbf{P}[A | X = x] &= \lim_{\Delta x \rightarrow 0^+} \mathbf{P}[A | x < X \leq x + \Delta x] = \lim_{\Delta x \rightarrow 0^+} \frac{F(x + \Delta x | A) - F(x | A)}{F_x(x + \Delta x) - F_x(x)} \mathbf{P}[A] \\
 &= \lim_{\Delta x \rightarrow 0^+} \frac{[F(x + \Delta x | A) - F(x | A)] / \Delta x}{[F_x(x + \Delta x) - F_x(x)] / \Delta x} \mathbf{P}(A),
 \end{aligned}
 \tag{2-38}$$

which yields

$$\mathbf{P}[A | X = x] = \frac{f(x|A)}{f_x(x)} \mathbf{P}[A]. \quad (2-39)$$

Now, multiply both sides of this last result by f_x and integrate to obtain

$$\int_{-\infty}^{\infty} \mathbf{P}[A | X = x] f_x(x) dx = \mathbf{P}[A] \int_{-\infty}^{\infty} f(x|A) dx. \quad (2-40)$$

But, the area under $f(x | A)$ is unity. Hence, we obtain

$$\mathbf{P}[A] = \int_{-\infty}^{\infty} \mathbf{P}[A | X = x] f_x(x) dx, \quad (2-41)$$

the continuous version of the *Total Probability Theorem*. In Chapter 1, we gave a “finite dimensional” version of this theorem. Compare (2-41) with the result of Theorem 1-1; conceptually, they are similar. $\mathbf{P}[A | X = x]$ is the probability of A given that $X = x$. Equation (2-41) tells us to average this conditional probability over all possible values of X to find $\mathbf{P}[A]$.

Example 2-5

Consider tossing a coin. The sample space is $\mathcal{S} = \{h, t\}$, a “heads” and a “tails”. However, assume that we do not know $\mathbf{P}[h]$ for the coin. So, we model $\mathbf{P}[h]$ as a random variable \tilde{p} . Our goal is to *estimate* $\mathbf{P}[h]$ by repeatedly tossing the coin.

As a random variable, \tilde{p} must map some sample space, call it \mathcal{S}_c , into $[0, 1]$, the range of possible values for the (**probability of heads**) of any coin. To be more definitive, let’s take \mathcal{S}_c as the set of all coins in a given large pot, and note that $\tilde{p} : \mathcal{S}_c \rightarrow [0, 1]$. As assigned by \tilde{p} , each coin in \mathcal{S}_c has a (**probability of heads**) $\in [0, 1]$.

Assume that we can guess (or we know) a density function $f_{\tilde{p}}(p)$ that describes random variable \tilde{p} . Given numbers p_1 and p_2 , the probability of selecting a coin with \tilde{p} (*i.e.*, the probability of heads) between p_1 and p_2 is

$$\mathbf{P}[p_1 < \tilde{p} \leq p_2] = \int_{p_1}^{p_2} f_{\tilde{p}}(p) dp. \quad (2-42)$$

It is reasonable to expect $f_{\tilde{p}}(p)$ to peak at, or near, $p = 1/2$ and drop rapidly as p deviates from $1/2$. After all, most coins are nearly balanced and fair. As we will show by examples that follow, it is possible to use experimental data (*i.e.*, statistics from coin toss trials) to update $f_{\tilde{p}}(p)$. That is, we will condition the density that describes \tilde{p} on experimental data obtained from tossing the selected coin. In this context, we will call $f_{\tilde{p}}(p)$ the *a-priori* density (**before** the coin tosses); the updated version will be called the *a-posteriori* density (**after** the coin tosses).

Our combined, or joint, experiment consists of selecting a coin from \mathcal{S}_c (the first experiment) and then tossing it (the second experiment). For the combined experiment, the relevant product space is

$$\mathcal{S}_c \times \mathcal{S} = [(\rho, h) : \rho \in \mathcal{S}_c] \cup [(\rho, t) : \rho \in \mathcal{S}_c]. \quad (2-43)$$

For the combined experiment, the event “we get a heads” or “heads occurs” is

$$\mathbf{H} = [(\rho, h) : \rho \in \mathcal{S}_c], \quad (2-44)$$

an event in the product, or combined, experiment. However the probability function (measure) is defined on the product space, it is clear that the conditional probability of \mathbf{H} , *given that all coins have* (**Probability of Heads**) = p (*i.e.*, $\tilde{p} = p$), is simply

$$\mathbf{P}[\mathbf{H} \mid \tilde{p} = p] = \mathbf{P}[(\rho, h) : \rho \in \mathcal{S}_c \mid \tilde{p} = p] = p. \quad (2-45)$$

The Theorem of Total Probability can be used to express $\mathbf{P}[\mathbf{H}]$ in terms of known $f_{\tilde{p}}(p)$. Simply use (2-45) and the Theorem of Total Probability (2-41) to write

$$\mathbf{P}[\mathbf{H}] = \int_0^1 \mathbf{P}[\mathbf{H} \mid \tilde{p} = p] f_{\tilde{p}}(p) dp = \int_0^1 p f_{\tilde{p}}(p) dp. \quad (2-46)$$

On the right hand side of (2-46), the integral is called the *ensemble average*, or *expected value*, of the random variable \tilde{p} .

Bayes Theorem - Continuous Form

From (2-39) we get

$$f(x|A) = \frac{\mathbf{P}[A|X=x]}{\mathbf{P}[A]} f_x(x). \quad (2-47)$$

Now, use (2-41) and (2-47) to write

$$f(x|A) = \frac{\mathbf{P}[A|X=x]}{\int_{-\infty}^{\infty} \mathbf{P}[A|X=v] f_x(v) dv} f_x(x), \quad (2-48)$$

a result known as the *continuous form of Bayes Theorem*.

Often, $f_x(x)$ is called the *a-priori* density for random variable X . And, $f(x|A)$ is called the *a-posteriori* density conditioned on the observed event A . In an application, we might “cook up” a density $f_x(x)$ that (crudely) describes a random quantity (*i.e.*, variable) X of interest. To improve our characterization of X , we note the occurrence of a related event A and compute $f(x|A)$ to better characterize X .

The value of x that maximizes $f(x|A)$ is called the *maximum a-posteriori* (MAP) estimate of X . MAP estimation is used in statistical signal processing and many other problems where one must estimate a quantity from observations of related random quantities.

Example 2-6: (MAP estimate of probability of heads in previously discussed coin experiment)

Find the MAP estimate of the probability of heads in the previously discussed coin selection and tossing experiment. First, recall that the probability of heads is modeled as a random variable \tilde{p}

with density $f_{\tilde{p}}(p)$. We call $f_{\tilde{p}}(p)$ the *a-priori* ("before the coin toss experiment") density of \tilde{p} (which we may have to guess). Suppose we toss the coin n times and get k heads. We want to use these experimental results with our *a-priori* density $f_{\tilde{p}}(p)$ to compute the conditional density

$$f(p \mid \underbrace{k \text{ heads, in a specific order, in } n \text{ tosses of a selected coin}}_{\text{Observed Event A}}). \quad (2-49)$$

This density is called *a-posteriori density* ("after the coin toss experiment") of random variable \tilde{p} given the experimentally observed event

$$A = [k \text{ heads, in a specific order, in } n \text{ tosses of a selected coin}]. \quad (2-50)$$

The *a-posteriori* density $f(p \mid A)$ may give us a good idea (better than the *a-priori* density $f_{\tilde{p}}(p)$) of the probability of heads for the randomly selected coin that was tossed. Conceptually, think of $f(p \mid A)$ as a density that results from using experimental data/observation A to "update" $f_{\tilde{p}}(p)$. In fact, given that A occurred, the *a-posteriori* probability that \tilde{p} is between p_1 and p_2 is

$$\int_{p_1}^{p_2} f(p \mid A) dp. \quad (2-51)$$

Finally, the value of p that maximizes $f(p \mid A)$ is the *maximum a-posteriori estimate* (MAP estimate) of \tilde{p} for the selected coin.

To find this *a-posteriori* density, recall that \tilde{p} is defined on the sample space \mathcal{S}_c . The experiment of tossing the randomly selected coin n times is defined on the sample space (*i.e.*, product space) $\mathcal{S}_c \times \mathcal{S}^n$, where $\mathcal{S} = [h, t]$. The elements of $\mathcal{S}_c \times \mathcal{S}^n$ have the form

$$\rho, \underbrace{t \ h \ t \ \cdots \ h}_{n \text{ outcomes}} \quad \text{where } \rho \in \mathcal{S}_c \text{ and } \underbrace{t \ h \ t \ \cdots \ h}_{n \text{ outcomes}} \in \mathcal{S}^n.$$

Now, given that $\tilde{p} = p$, the conditional probability of event A is

$$\underbrace{P[\text{k heads in specific order in n tosses of specific coin}]}_{\text{Observed Event A}} \mid \tilde{p} = p] = p^k (1-p)^{n-k}. \quad (2-52)$$

Substitute this into the continuous form of Bayes rule (2-48) to obtain

$$\begin{aligned} f(p|A) &= f(p \mid \overbrace{\text{k heads, in a specific order, in n tosses of a selected coin}}^{\text{Observed Event A}}) \\ &= \frac{p^k (1-p)^{n-k} f_{\tilde{p}}(p)}{\int_0^1 v^k (1-v)^{n-k} f_{\tilde{p}}(v) dv}, \end{aligned} \quad (2-53)$$

a result known as the *a-posteriori* density of \tilde{p} given A. In (2-53), the quantity v is a dummy variable of integration.

Suppose that the *a-priori* density $f_{\tilde{p}}(p)$ is smooth and slowly changing around $p = k/n$ (indicating a lot of uncertainty in the value of p). Then, for large n , the numerator $p^k (1-p)^{n-k} f_{\tilde{p}}(p)$, and the *a-posteriori* density $f(p|A)$, has a sharp peak at $p = k/n$, indicating little uncertainty in the value of p . When $f(p|A)$ is peaked at $p = k/n$, the MAP estimate of the probability of heads (for the selected coin) is the value $p = k/n$.

Example 2-7: For $\sigma > 0$, we use the *a-priori* density

$$f_{\tilde{p}}(p) = \frac{\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(p-1/2)^2}{2\sigma^2}\right]}{1 - 2G\left(\frac{-1}{2\sigma}\right)}, \quad (2-54)$$

where $G(x)$ is a zero mean, unit variance Gaussian distribution function (verify that there is unit area under $f_{\tilde{p}}(p)$). Also, use numerical integration to evaluate the denominator of (2-53). For σ

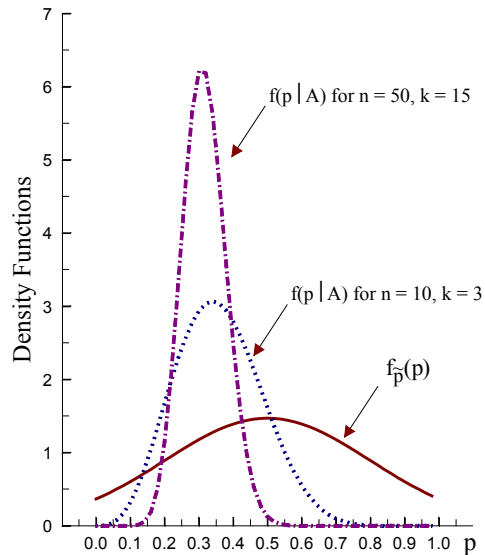


Figure 2-11: Plot of *a-priori* and *a-posteriori* densities.

$\sigma = .3$, $n = 10$ and $k = 3$, *a-posteriori density* $f(p | A)$ was computed and the results are plotted on Figure 2-11. For $\sigma = .3$, $n = 50$ and $k = 15$, the calculation was performed a second time, and the results appear on Figure 2-11. For both, the MAP estimate of \tilde{p} is near $.3$, since this is were the plots of $f(p | A)$ peak.

Expectation

Let X denote a random variable with density $f_x(x)$. The *expected value* of X is defined as

$$E[X] = \int_{-\infty}^{\infty} x f_x(x) dx. \quad (2-55)$$

Also, $E[X]$ is known as the *mean*, or *average value*, of X .

If f_x is symmetrical about some value η , then η is the expected value of X . For example, let X be $N(\eta, \sigma)$ so that

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2} \frac{(x - \eta)^2}{\sigma^2}\right]. \quad (2-56)$$

Now, (2-56) is symmetrical about η , so $E[X] = \eta$.

Suppose that random variable X is discrete and takes on the value of x_i with $\mathbf{P}[X = x_i] = p_i$. Then the expected value of X , as defined by (2-55), reduces to

$$E[X] = \sum_i x_i p_i . \quad (2-57)$$

Example 2-8: Consider a $B(n,p)$ random variable X (that is, X is Binomial with parameters n and p). Using (2-57), we can write

$$E[X] = \sum_{k=0}^n k \mathbf{P}[X = k] = \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} . \quad (2-58)$$

To evaluate this result, first consider the well-know, and extremely useful, Binomial expansion

$$\sum_{k=0}^n \binom{n}{k} x^k = (1+x)^n . \quad (2-59)$$

With respect to x , differentiate (2-59); then, multiply the derivative by x to obtain

$$\sum_{k=0}^n k \binom{n}{k} x^k = xn(1+x)^{n-1} . \quad (2-60)$$

Into (2-60), substitute $x = p/q$, and then multiply both sides by q^n . This results in

$$\begin{aligned} \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} &= npq^{n-1} \left(1 + \frac{p}{q}\right)^{n-1} = np(p+q)^{n-1} \\ &= np . \end{aligned} \quad (2-61)$$

From this and (2-58), we conclude that a $B(n,p)$ random variable has

$$E[X] = np. \quad (2-62)$$

We now provide a second, completely different, evaluation of $E[X]$ for a $B(n,p)$ random variable. Define n new random variables

$$\begin{aligned} X_i &= 1, \text{ if } i^{\text{th}} \text{ trial is a "success", } 1 \leq i \leq n \\ &= 0, \text{ otherwise.} \end{aligned} \quad (2-63)$$

Note that

$$E[X_i] = 1 \cdot p + 0 \cdot (1-p) = p, \quad 1 \leq i \leq n. \quad (2-64)$$

$B(n,p)$ random variable X is the number of successes out of n independent trials; it can be written as

$$X = X_1 + X_2 + \cdots + X_n. \quad (2-65)$$

With the use of (2-64), the expected value of X can be evaluated as

$$\begin{aligned} E[X] &= E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n] \\ &= np, \end{aligned} \quad (2-66)$$

a result equivalent to (2-62).

In what follows, we extend (2-55) to arbitrary functions of random variable X . Let $g(x)$ be any function of x , and let X be any random variable with density $f_x(x)$. In Chapter 4, we will discuss transformations of the form $Y = g(X)$; this defines the new random variable Y in terms of the old random variable X . We will argue that the expected value of Y can be computed as

$$E[Y] = E[g(X)] = \int_{-\infty}^{\infty} g(x)f_x(x)dx . \quad (2-67)$$

This brief “heads-up” note (on what is to come in CH 4) is used next to define certain statistical averages of functions of X .

Variance and Standard Deviation

The *variance* of random variable X is

$$\sigma^2 = \text{Var}[X] = E[(X - \eta)^2] = \int_{-\infty}^{\infty} (x - \eta)^2 f_x(x)dx . \quad (2-68)$$

Almost always, variance is denoted by the symbol σ^2 . The square root of the variance is called the *standard deviation* of the random variable, and it is denoted as σ . Finally, variance is a measure of uncertainty (or dispersion about the mean). The smaller (alternatively, larger) σ^2 is, the more (alternatively, less) likely it is for the random variable to take on values near its mean. Finally, note that (2-68) is an application of the basic result (2-67) with $g = (x - \eta)^2$.

Example 2-9: Let X be $N(\eta, \sigma)$ then

$$\text{VAR}[X] = E[(X - \eta)^2] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - \eta)^2 \exp\left[-\frac{(x - \eta)^2}{2\sigma^2}\right] dx . \quad (2-69)$$

Let $y = (x - \eta)/\sigma$, $dx = \sigma dy$ so that

$$\text{Var}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 y^2 e^{-\frac{1}{2}y^2} dy = \sigma^2, \quad (2-70)$$

a result obtained by looking up the integral in a table of integrals.

Moments

The *nth moment* of random variable X is defined as

$$m_n = E[X^n] = \int_{-\infty}^{\infty} x^n f_x(x) dx. \quad (2-71)$$

The *nth moment about the mean* is defined as

$$\mu_n = E[(X - \eta)^n] = \int_{-\infty}^{\infty} (x - \eta)^n f_x(x) dx. \quad (2-72)$$

Note that (2-71) and (2-72) are basic applications of (2-67).

Variance in Terms of Second Moment and Mean

Note that the variance can be expressed as

$$\sigma^2 = \text{Var}[X] = E[(X - \eta)^2] = E[X^2 - 2\eta X + \eta^2] = E[X^2] - E[2\eta X] + E[\eta^2], \quad (2-73)$$

where the linearity of the operator $E[\cdot]$ has been used. Now, constants “come out front” of expectations, and the expectation of a constant is the constant. Hence, Equation (2-73) leads to

$$\sigma^2 = E[X^2] - E[2\eta X] + E[\eta^2] = m_2 - 2\eta E[X] + \eta^2 = m_2 - \eta^2, \quad (2-74)$$

the second moment minus the square of the mean. In what follows, this formula will be used extensively.

Example 2-10: Let X be $N(0, \sigma)$ and find $E[X^n]$. First, consider the case of **n an odd integer**

$$E[X^n] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x^n \exp\left[-\frac{1}{2} \frac{x^2}{\sigma^2}\right] dx = 0, \quad (2-75)$$

since an integral of an odd function over symmetrical limits is zero. Now, consider the case **n an even integer**. Start with the known tabulated integral

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}} = \alpha^{-1/2} \sqrt{\pi}, \quad \alpha > 0. \quad (2-76)$$

Repeated differentiations with respect to α yields

$$\text{first } d/d\alpha: \quad \int_{-\infty}^{\infty} -x^2 e^{-\alpha x^2} dx = -\frac{1}{2} \alpha^{-3/2} \sqrt{\pi}$$

$$\text{second } d/d\alpha: \quad \int_{-\infty}^{\infty} x^4 e^{-\alpha x^2} dx = \frac{1}{2} \frac{3}{2} \alpha^{-5/2} \sqrt{\pi}$$

⋮

$$k^{\text{th}} \text{ } d/d\alpha: \quad \int_{-\infty}^{\infty} x^{2k} e^{-\alpha x^2} dx = \frac{1}{2} \frac{3}{2} \frac{5}{2} \dots \frac{(2k-1)}{2} \alpha^{-(2k+1)/2} \sqrt{\pi}.$$

Let $n = 2k$ (remember, this is the case n even) and $\alpha = 1/2\sigma^2$ to obtain

$$\begin{aligned} \int_{-\infty}^{\infty} x^n e^{-x^2/2\sigma^2} dx &= \frac{1 \cdot 3 \cdot 5 \cdots (n-1)}{\sqrt{2^n}} \sqrt{\pi 2^{(n+1)} \sigma^{2(n+1)}} \\ &= 1 \cdot 3 \cdot 5 \cdots (n-1) \sqrt{2\pi} \sigma^{n+1}. \end{aligned} \quad (2-77)$$

From this, we conclude that the n^{th} -moment of a zero-mean, Gaussian random variable is

$$m_n \equiv E[X^n] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x^n e^{-x^2/2\sigma^2} dx = 1 \cdot 3 \cdot 5 \cdots (n-1) \sigma^n, \quad n = 2k \quad (n \text{ even})$$

$$= 0, \quad n = 2k-1 \quad (n \text{ odd}). \quad (2-78)$$

Example 2-11 (Rayleigh Distribution): A random variable X is Rayleigh distributed with parameter α if its density function has the form

$$f_X(x) = \frac{x}{\alpha^2} \exp\left(-\frac{x^2}{2\alpha^2}\right), \quad x \geq 0$$

$$= 0, \quad x < 0. \quad (2-79)$$

This random variable has many applications in communication theory; for example, it describes the envelope of a narrowband Gaussian noise process (as described in Chapter 9 of these notes).

The n^{th} moment of a Rayleigh random variable can be computed by writing

$$E[X^n] = \frac{1}{\alpha^2} \int_0^{\infty} x^{n+1} e^{-x^2/2\alpha^2} dx. \quad (2-80)$$

We consider two cases, n even and n odd. For the case n odd, we have $n+1$ even, and (2-80) becomes

$$E[X^n] = \frac{1}{\alpha^2} \frac{1}{2} \int_{-\infty}^{\infty} x^{n+1} e^{-x^2/2\alpha^2} dx = \frac{\sqrt{2\pi}}{2\alpha} \left[\frac{1}{\sqrt{2\pi}\alpha} \int_{-\infty}^{\infty} x^{n+1} e^{-x^2/2\alpha^2} dx \right]. \quad (2-81)$$

On the right-hand-side of (2-81), the bracket contains the $(n+1)^{\text{th}}$ moment of a $N(0, \alpha)$ random variable (compare (2-78) and the right-hand-side of (2-81)). Hence, we can write

$$E[X^n] = \left(\frac{\sqrt{2\pi}}{2\alpha} \right) \cdot 1 \cdot 3 \cdot 5 \cdots n \alpha^{n+1} = 1 \cdot 3 \cdot 5 \cdots n \alpha^n \sqrt{\pi/2}, \quad n = 2k+1 \quad (n \text{ odd}). \quad (2-82)$$

Now, consider the case n even, $n+1$ odd. For this case, (2-80) becomes

$$E[X^n] = \frac{1}{\alpha^2} \int_0^\infty x^{n+1} e^{-x^2/2\alpha^2} dx = \int_0^\infty x^n e^{-x^2/2\alpha^2} \frac{x}{\alpha^2} dx. \quad (2-83)$$

Substitute $y = x^2/2\alpha^2$, so that $dy = (x/\alpha^2)dx$, in (2-83) and obtain

$$E[X^n] = \int_0^\infty (\sqrt{2}\alpha\sqrt{y})^n e^{-y} dy = 2^{n/2} \alpha^n \int_0^\infty y^{n/2} e^{-y} dy = 2^{n/2} \alpha^n \left(\frac{n}{2}\right)!, \quad (2-84)$$

(note that $n/2$ is an integer here) where we have used the Gamma function

$$\Gamma(k+1) = \int_0^\infty y^k e^{-y} dy = k!, \quad (2-85)$$

$k \geq 0$ an integer. Hence, for a Rayleigh random variable X , we have determined that

$$\begin{aligned} E[X^n] &= 1 \cdot 3 \cdot 5 \cdots n \alpha^n \sqrt{\pi/2}, \quad n \text{ odd} \\ &= 2^{n/2} \alpha^n \left(\frac{n}{2}\right)!, \quad n \text{ even.} \end{aligned} \quad (2-86)$$

In particular, Equation (2-86) can be used to obtain the mean and variance

$$E[X] = \sqrt{\frac{\pi}{2}} \alpha \quad (2-87)$$

$$\text{Var}[X] = E[X^2] - (E[X])^2 = 2\alpha^2 - \frac{\pi}{2}\alpha^2 = \left(2 - \frac{\pi}{2}\right)\alpha^2,$$

given that X is a Rayleigh random variable.

Example 2-12: Let X be Poisson with parameter α , so that $\mathbf{P}[X = k] = e^{-\alpha} \frac{\alpha^k}{k!}$, $k \geq 0$, and

$$f(x) = \sum_{k=0}^{\infty} e^{-\alpha} \frac{\alpha^k}{k!} \delta(x - k). \quad (2-88)$$

Show that $E[X] = \alpha$ and $\text{VAR}[X] = \alpha$. To accomplish this, recall that

$$e^{\alpha} = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!}. \quad (2-89)$$

With respect to α , differentiate (2-89) to obtain

$$e^{\alpha} = \sum_{k=0}^{\infty} k \frac{\alpha^{k-1}}{k!} = \frac{1}{\alpha} \sum_{k=1}^{\infty} k \frac{\alpha^k}{k!} \quad (2-90)$$

Multiply both sides of this result by $\alpha e^{-\alpha}$ to obtain

$$\alpha = \sum_{k=1}^{\infty} k \left(e^{-\alpha} \frac{\alpha^k}{k!} \right) = E[X]. \quad (2-91)$$

as claimed. With respect to α , differentiate (2-90) (obtain the second derivative of (2-89)) and obtain

$$e^{\alpha} = \sum_{k=1}^{\infty} k(k-1) \frac{\alpha^{k-2}}{k!} = \frac{1}{\alpha^2} \sum_{k=1}^{\infty} k^2 \frac{\alpha^k}{k!} - \frac{1}{\alpha^2} \sum_{k=1}^{\infty} k \frac{\alpha^k}{k!}.$$

Multiply both sides of this result by $\alpha^2 e^{-\alpha}$ to obtain

$$\alpha^2 = \sum_{k=1}^{\infty} k^2 \left(e^{-\alpha} \frac{\alpha^k}{k!} \right) - \sum_{k=1}^{\infty} k \left(e^{-\alpha} \frac{\alpha^k}{k!} \right) = \sum_{k=1}^{\infty} k^2 \mathbf{P}[X = k] - \sum_{k=1}^{\infty} k \mathbf{P}[X = k]. \quad (2-92)$$

Note that (2-92) is simply $\alpha^2 = E[X^2] - E[X]$. Finally, a Poisson random variable has a variance given by

$$\text{Var}[X] = E[X^2] - (E[X])^2 = E[X^2] - \alpha^2 = E[X] = \alpha, \quad (2-93)$$

as claimed.

Conditional Mean

Let M denote an event. The conditional density $f(x | M)$ can be used to define the *conditional mean*

$$E[X | M] = \int_{-\infty}^{\infty} x f(x | M) dx. \quad (2-94)$$

The conditional mean has many applications, including estimation theory, detection theory, etc.

Example 2-13: Let X be Gaussian with zero mean and variance σ^2 (i.e., X is $N(0, \sigma)$). Let $M = [X > 0]$; find $E[X | M] = E[X | X > 0]$. First, we must find the conditional density $f(x | X > 0)$; from previous work in this chapter, we can write

$$\begin{aligned} F(x | X > 0) &= \frac{\mathbf{P}[X \leq x, X > 0]}{\mathbf{P}[X > 0]} = \frac{\mathbf{P}[X \leq x, X > 0]}{1 - \mathbf{P}[X \leq 0]} = \frac{\mathbf{P}[X \leq x, X > 0]}{1 - F_X(0)} \\ &= \frac{F_X(x) - F_X(0)}{1 - F_X(0)}, \quad x \geq 0 \\ &= 0, \quad x < 0, \end{aligned}$$

so that

$$f(x|X > 0) = 2f_x(x), \quad x \geq 0$$

$$= 0, \quad x < 0.$$

From (2-94) we can write

$$E[X|X > 0] = \frac{2}{\sqrt{2\pi}\sigma} \int_0^{\infty} x \exp\left[-x^2/2\sigma^2\right] dx.$$

Now, set $y = x^2/2\sigma^2$, $dy = xdx/\sigma^2$ to obtain

$$E[X|X > 0] = \frac{2\sigma}{\sqrt{2\pi}} \int_0^{\infty} e^{-y} dy = \sigma\sqrt{\frac{2}{\pi}}.$$

Tchebycheff Inequality

A measure of the concentration of a random variable near its mean is its variance. Consider a random variable X with mean η , variance σ^2 and density $f_X(x)$. The larger σ^2 , the more "spread-out" the density function, and the more probable it is to find values of X "far" from the mean. Let ε denote an arbitrary small positive number. The *Tchebycheff inequality* says that the probability that X is outside $(\eta - \varepsilon, \eta + \varepsilon)$ is negligible if σ/ε is sufficiently small.

Theorem (Tchebycheff's Inequality)

Consider random variable X with mean η and variance σ^2 . For any $\varepsilon > 0$, we have

$$P[|X-\eta| \geq \varepsilon] \leq \frac{\sigma^2}{\varepsilon^2}. \quad (2-95)$$

Proof: Note that

$$\begin{aligned}
 \sigma^2 &= \int_{-\infty}^{\infty} (x - \eta)^2 f_X(x) dx \geq \int_{\{x: |x - \eta| \geq \varepsilon\}} (x - \eta)^2 f_X(x) dx \\
 &\geq \varepsilon^2 \left[\int_{\{x: |x - \eta| \geq \varepsilon\}} f_X(x) dx \right] \\
 &= \varepsilon^2 \mathbf{P}[|X - \eta| \geq \varepsilon].
 \end{aligned}$$

This leads to the Tchebycheff inequality

$$\mathbf{P}[|X - \eta| \geq \varepsilon] \leq \frac{\sigma^2}{\varepsilon^2}. \quad (2-96)$$

The significance of Tchebycheff's inequality is that it holds for any random variable, and it can be used without explicit knowledge of $f(x)$. However, the bound is very "conservative" (or "loose"), so it may not offer much information in some applications. For example, consider Gaussian X . One can show

$$\mathbf{P}[|X - \eta| \geq 3\sigma] = 2 - 2G(3) = .0027, \quad (2-97)$$

where $G(3)$ is obtained from a table containing values of the Gaussian integral. However, the Tchebycheff inequality gives the rather "loose" upper bound of

$$\mathbf{P}[|X - \eta| \geq 3\sigma] \leq 1/9. \quad (2-98)$$

Generalizations of Tchebycheff's Inequality

For a given random variable X , suppose that $f_X(x) = 0$ for $x < 0$. Then, for any $\alpha > 0$, we have

$$\mathbf{P}[X \geq \alpha] \leq \eta/\alpha. \quad (2-99)$$

To show (2-99), note that

$$\eta = E[X] = \int_0^{\infty} x f_X(x) dx \geq \int_{\alpha}^{\infty} x f_X(x) dx \geq \alpha \int_{\alpha}^{\infty} f_X(x) dx = \alpha \mathbf{P}[X \geq \alpha], \quad (2-100)$$

so that $\mathbf{P}[X \geq \alpha] \leq \eta/\alpha$, as claimed.

Corollary: Let X be an arbitrary random variable and α and n an arbitrary real number and positive integer, respectively. The random variable $|X - \alpha|^n$ takes on only nonnegative values.

Hence

$$\mathbf{P}\left[|X - \alpha|^n \geq \epsilon^n\right] \leq \frac{E[|X - \alpha|^n]}{\epsilon^n}, \quad (2-101)$$

which implies

$$\mathbf{P}\left[|X - \alpha| \geq \epsilon\right] \leq \frac{E[|X - \alpha|^n]}{\epsilon^n}. \quad (2-102)$$

The Tchebycheff inequality is a special case with $\alpha = \eta$ and $n = 2$.

Application: System Reliability

Often, systems fail in a random manner. For a particular system, we denote t_f as the time interval from the moment a system is put into operation until it fails; t_f is the *time to failure random variable*. The distribution $F_a(t) = \mathbf{P}[t_f \leq t]$ is the probability the system fails at, or prior to, time t . Implicit here is the assumption that the system is placed into service at $t = 0$. Also, we require that $F_a(t) = 0$ for $t \leq 0$.

The quantity

$$R(t) = 1 - F_a(t) = \mathbf{P}[t_f > t] \quad (2-103)$$

is the *system reliability*. $R(t)$ is the probability the system is functioning at time $t > 0$.

We are interested in simple methods to quantify system reliability. One such measure of system reliability is the *mean time before failure*

$$\text{MTBF} \equiv E[t_f] = \int_0^{\infty} t f_a(t) dt, \quad (2-104)$$

where $f_a = dF_a/dt$ is the density function that describes random variable t_f .

Given that a system is functioning at time t' , $t' \geq 0$, we are interested in the probability that a system fails at, or prior to, time t , where $t > t' \geq 0$. We express this conditional distribution function as

$$F(t|t_f > t') = \frac{\mathbf{P}[t_f \leq t, t_f > t']}{\mathbf{P}[t_f > t']} = \frac{\mathbf{P}[t' < t_f \leq t]}{\mathbf{P}[t_f > t']} = \frac{F_a(t) - F_a(t')}{1 - F_a(t')}, \quad t > t'. \quad (2-105)$$

The conditional density can be obtained by differentiating (2-105) to obtain

$$f(t|t_f > t') = \frac{d}{dt} F(t|t_f > t') = \frac{f_a(t)}{1 - F_a(t')}, \quad t > t'. \quad (2-106)$$

$F(t|t_f > t')$ and $f(t|t_f > t')$ describe t_f conditioned on the event $t_f > t'$. The quantity $f(t|t_f > t')dt$ is, to first-order in dt , the probability that the system fails between t and $t + dt$ given that it was working at t' .

Example 2-14

Suppose that the time to failure random variable t_f is exponentially distributed. That is, suppose that

$$F_a(t) = \begin{cases} 1 - e^{-\kappa t}, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (2-107)$$

$$f_a(t) = \begin{cases} \kappa e^{-\kappa t}, & t \geq 0 \\ 0, & t < 0, \end{cases}$$

for some constant $\kappa > 0$. From (2-106), we see that

$$f(t | t_f > t') = \frac{\kappa e^{-\kappa t}}{e^{-\kappa t'}} = \kappa e^{-\kappa(t-t')} = f_a(t-t'), \quad t > t'. \quad (2-108)$$

That is, if the system is working at time t' , then the probability that it fails between t' and t depends only on the positive difference $t - t'$, **not** on absolute t' . The system does not “wear out” (become more likely to fail) as time progresses!

With (2-106), we define $f(t | t_f > t')$ for $t > t'$. However, the function

$$\beta(t) \equiv \frac{f_a(t)}{1 - F_a(t)}, \quad (2-109)$$

known as the *conditional failure rate* (also known as the *hazard rate*), is very useful. To first-order, $\beta(t)dt$ (when this quantity exists) is the probability that a functioning-at- t system will fail between t and $t + dt$.

Example 2-15 (Continuation of Example 2-14)

Assume the system has f_a and F_a as defined in Example 2-14. Substitute (2-107) into (2-109) to obtain

$$\beta(t) = \frac{\kappa e^{-\kappa t}}{1 - \{1 - e^{-\kappa t}\}} = \kappa. \quad (2-110)$$

That is, the conditional failure rate is the constant κ . As stated in Example 2-14, the system does not “wear out” as time progresses!

If conditional failure rate $\beta(t)$ is a constant κ , we say that the system is a *good as new system*. That is, it does not “wear out” (become more likely to fail) overtime.

Examples 2-14 and 2-15 show that if a system’s time-to-failure random variable t_f is exponentially distributed, then

$$f(t|t_f > t') = f_a(t - t'), \quad t > t' \quad (2-111)$$

and

$$\beta(t) = \text{constant}, \quad (2-112)$$

so the system is a *good as new system*.

The converse is true as well. That is, if $\beta(t)$ is a constant κ for a system, then the system’s time-to-failure random variable t_f is exponentially distributed. To argue this, use (2-109) to write

$$f_a(t) = \kappa[1 - F_a(t)]. \quad (2-113)$$

But (2-113) leads to

$$\frac{dF_a}{dt} = -\kappa F_a(t) + \kappa. \quad (2-114)$$

Since $F_a(t) = 0$ for $t \leq 0$, we must have

$$F_a(t) = \begin{cases} 1 - e^{-\kappa t}, & t \geq 0 \\ 0, & t < 0, \end{cases}$$

so t_f is exponentially distributed. For $\beta(t)$ to be equal to a constant κ , failures must be truly random in nature. A constant β requires that there be no time epochs where failure is more likely (no Year 2000 – type problems!).

Random variable t_f is said to exhibit the *Markov* property, or it is said to be *memoryless*, if its conditional density obeys (2-111). We have established the following theorem.

Theorem

The following are equivalent

- 1) A system is a *good-as-new system*
- 2) $\beta = \text{constant}$
- 3) $f(t \mid t_f > t') = f_a(t - t')$, $t > t'$
- 4) t_f is exponentially distributed.

The previous theorem, and the Markov property, is stated in the context of system reliability. However, both have far reaching consequences in other areas that have nothing to do with system reliability. Basically, in problems dealing with random arrival times, the time between two successive arrivals is exponentially distributed if

- 1) the arrivals are independent of each other, and
- 2) the arrival time t *after* any specified fixed time t' is described by a density function that depends only on the difference $t - t'$ (the arrival time random variable obeys the Markov property).

In Chapter 9, we will study *shot noise* (caused by the random arrival of electrons at a semiconductor junction, vacuum tube anode, etc), an application of the above theorem and the Markov property.